OXFORD

ASSOCIATION STUDIES ARTICLE

Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function

Alexandra S. Shadrina[1,*], Alexander S. Zlobin[1], Olga O. Zaytseva[2],
Lucija Klarić[2,3], Sodbo Z. Sharapov[1], Eugene D. Pakhomov[1], Marcus Perola[4],
Tonu Esko[5], Caroline Hayward[3], James F. Wilson[3,6], Gordan Lauc[2,†],
Yurii S. Aulchenko[1,7,†] and Yakov A. Tsepilov[1,8,†]

[1]Laboratory of Glycogenomics, Institute of Cytology and Genetics, Novosibirsk 630090, Russia, [2]Genos
Glycoscience Research Laboratory, Zagreb 10000, Croatia, [3]MRC Human Genetics Unit, MRC Institute of
Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK, [4]Genomics and
Biomarkers Unit, Department of Health, National Institute for Health and Welfare (THL), Helsinki, Finland,
[5]Estonian Genome Center, University of Tartu, Tartu, Estonia, [6]Centre for Global Health Research, Usher
Institute, University of Edinburgh, Edinburgh EH8 9AG, Scotland, [7]PolyOmica, 's-Hertogenbosch 5237 PA, The
Netherlands and [8]Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University,
Novosibirsk 630090, Russia

*To whom correspondence should be addressed at: Laboratory of Glycogenomics, Institute of Cytology and Genetics, 10 Lavrentiev Avenue, Novosibirsk
630090, Russia. Tel: +7 3833634963*1228; Fax: +7 3833331278; Email: weiner.alexserg@gmail.com

Abstract

The N-glycosylation of immunoglobulin G (IgG) affects its structure and function. It has been demonstrated that IgG
N-glycosylation patterns are inherited as complex quantitative traits. Genome-wide association studies identified loci
harboring genes encoding enzymes directly involved in protein glycosylation as well as loci likely to be involved in
regulation of glycosylation biochemical pathways. Many of these loci could be linked to immune functions and risk of
inflammatory and autoimmune diseases. The aim of the present study was to discover and replicate new loci associated
with IgG N-glycosylation and to investigate possible pleiotropic effects of these loci onto immune function and the risk of
inflammatory and autoimmune diseases. We conducted a multivariate genome-wide association analysis of 23 IgG
N-glycosylation traits measured in 8090 individuals of European ancestry. The discovery stage was followed up by
replication in 3147 people and *in silico* functional analysis. Our study increased the total number of replicated loci from 22 to
29. For the discovered loci, we suggest a number of genes potentially involved in the control of IgG N-glycosylation. Among
the new loci, two (near *RNF168* and *TNFRSF13B*) were previously implicated in rare immune deficiencies and were associated
with levels of circulating immunoglobulins. For one new locus (near *AP5B1/OVOL1*), we demonstrated a potential pleiotropic
effect on the risk of asthma. Our findings underline an important link between IgG N-glycosylation and immune function
and provide new clues to understanding their interplay.

Introduction

Immunoglobulin G (IgG) is the most abundant immunoglobulin isotype and the most prevalent glycoprotein in human plasma (1,2). IgG has a broad spectrum of activity and participates in multiple immune reactions including phagocytosis, complement activation and antibody-dependent cell-mediated cytotoxicity (1,3,4). All IgG molecules are N-glycosylated at a highly conserved asparagine 297 residue (also referred to as CH2-84.4) located in the Fc region, which interacts with effector cells of the immune system and some proteins of the complement system (5). Additionally, glycosylation of variable regions of light or heavy chains can occur in approximately 20% of IgG molecules (6). The attachment of different glycans generates a wide range of IgG glycoforms. The spectrum of glycans linked to IgG molecules is called the IgG N-glycome.

N-glycosylation of IgG contributes to the maintenance of its structure, stability and solubility. No less importantly, N-glycosylation drives IgG's effector functions, thus playing an important regulatory role in the immune system (4,5,7). For example, addition of fucose to the IgG glycan core structure decreases the affinity of IgG to FcγRIIIA and FcγRIIIB receptors and diminishes antibody-dependent cell-mediated cytotoxicity (ADCC). On the other hand, increased levels of IgG sialylation and terminal galactosylation are generally considered to have an anti-inflammatory effect (reviewed in (4)). IgG N-glycome composition is affected in various physiological and pathological states and is associated with a number of diseases, such as inflammatory, infectious, autoimmune and alloimmune diseases, as well as different types of cancer (4). IgG glycosylation patterns are also considered as biomarkers of ageing (8).

Although biochemical reactions underlying protein glycosylation are well studied, the biological networks regulating these reactions remain poorly understood. Elucidating the genetic factors influencing IgG N-glycome variation helps to uncover molecules and pathways involved in establishing IgG glycosylation patterns and facilitates research linking IgG N-glycome and human complex diseases. To date, four genome-wide association studies (GWASes) of the total plasma IgG N-glycome have been published (9–12). The largest study was conducted by Klarić *et al*. (12), which identified and replicated 19 loci and brought the number of loci robustly associated with the IgG N-glycome composition to 22. This work suggested how these loci form a functional network and proposed new regulators of IgG glycosylation.

In addition to the loci that contain glycosyltransferase genes (*MGAT3, ST6GAL1, B4GALT1, FUT6/FUT3, FUT8*), previous analyses identified loci that contain genes encoding transcription factors known to play central roles in lymphocyte maturation and differentiation: RUNX1, RUNX3, IKZF1 and IKZF3 (9,11,12). For several IgG N-glycosylation loci, colocalization analysis suggested pleiotropic effects onto autoimmune and inflammatory conditions (12). Notably, for the most highly replicated and significant asthma locus at 17q12-21 (near *ORMDL3-GSDMB-IKZF3* genes) (13), the asthma risk alleles were associated with the decreased levels of afucosylated, agalactosylated and monogalactosylated glycans and increased levels of fucosylated glycans (12). This locus was also associated with the risk of autoimmune and inflammatory diseases such as rheumatoid arthritis, Crohn's disease, ulcerative colitis and primary biliary cirrhosis, and the autoimmune disease risk allele was also associated with increased levels of afucosylated agalactosylated and monogalactosylated glycans and decreased levels of fucosylated glycans (12). Consistent with these observations, it is known

that although the same single nucleotide polymorphisms (SNPs) are associated with both asthma and autoimmune disease, the risk alleles associated with autoimmune disease are opposite of those associated with asthma risk (13). Similar to observations in the *ORMDL3-GSDMB-IKZF3* locus, at another locus near the *IRF1-SLC22A4* genes, the Crohn's disease risk allele was demonstrated to be associated with the increased levels of afucosylated agalactosylated glycans and decreased levels of fucosylation (12).

The above observations, in general, suggest that a more proinflammatory and pro-ADCC glycan profile is associated with increased risk of autoimmune and inflammatory diseases. However, the evidence is still scattered, and identification of additional IgG N-glycosylation loci that overlap with the loci affecting the disease risk may help in understanding the relationship between glycosylation and immune function. Application of multivariate methods in genetic association studies of correlated phenotypes can increase the statistical power to discover novel loci (10,14,15) as well as facilitate the interpretation of the results (16). The efficacy of a multivariate approach in investigating the genetic architecture of IgG N-glycosylation traits has recently been demonstrated by Shen *et al*. (10). At the time when only nine loci were shown to be associated with IgG N-glycosylation (9), of which two were replicated, the study employing multivariate methodology identified and replicated five novel loci despite using a smaller discovery sample size.

In the present study, we aimed to provide further insights into the genetics underpinning IgG N-glycosylation by applying a multivariate approach to the largest currently available dataset linking genotypes with IgG glycan measurements [$N = 8090$; Klarić *et al*.'s study (12)] and replicating our results using data from 3147 independent samples.

Results

Phenotype grouping

Following the study by Shen *et al*. (10), we grouped 23 directly measured IgG N-glycan traits (described in Supplementary Material, Table S1) into nine groups based on the general structural and chemical properties of glycans, including one group that combined all 23 IgG N-glycosylation traits. The description of these groups is provided in Figure 1.

Discovery and replication of loci associated with IgG N-glycosylation trait groups

Multivariate analysis in the discovery cohort of 8090 samples revealed 32 loci associated with at least one group of IgG N-glycosylation traits at a genome-wide significant level of $P < 5.6 \times 10^{-9}$ (Table 1, quantile–quantile plots for all trait groups are given in Supplementary Material, Fig. S1 and the matrix of correlations between 23 IgG N-glycosylation traits in the discovery cohort is presented in Supplementary Material, Fig. S2a). These loci include the major histocompatibility complex (MHC) region and two loci (rs11847263 near *FUT8* and rs10445337 near *MAPT*), which map in regions of extended linkage disequilibrium (LD) and therefore do not meet the locus definition '±500 kb around the lead SNP' (regional association plots for these loci can be found in Supplementary Material, Fig. S3). Twenty-six loci were associated with ≥2 trait groups. The median number of trait groups associated with the identified loci was 6 (range 1–9). The most significantly associated glycan trait groups ('top trait groups') are presented in Table 1. All

Directly measured IgG N-glycan traits

IGP1	FA1	IGP13	FA2G2
IGP2	A2	IGP14	FA2BG2
IGP3	FA2	IGP15	FA2G1S1
IGP4	M5	IGP16	A2G2S1
IGP5	FA2B	IGP17	FA2G2S1
IGP6	A2G1	IGP18	FA2BG2S1
IGP7	FA2[6]G1	IGP19	FA2FG2S1
IGP8	FA2[3]G1	IGP20	A2G2S2
IGP9	FA2[6]BG1	IGP21	A2BG2S2
IGP10	FA2[3]BG1	IGP22	FA2G2S2
IGP11	A2G2	IGP23	FA2BG2S2
IGP12	A2BG2		

Trait grouping for multivariate analysis

N-glycosylation: IGP1, IGP2, IGP3, IGP4, IGP5, IGP6, IGP7, IGP8, IGP9, IGP10, IGP11, IGP12, IGP13, IGP14, IGP15, IGP16, IGP17, IGP18, IGP19, IGP20, IGP21, IGP22, IGP23

Monogalactosylation: IGP6, IGP7, IGP8, IGP9, IGP10, IGP15

Digalactosylation: IGP11, IGP12, IGP13, IGP14, IGP16, IGP17, IGP18, IGP20, IGP21, IGP22, IGP23

Galactosylation: IGP6, IGP7, IGP8, IGP9, IGP10, IGP11, IGP12, IGP13, IGP14, IGP15, IGP16, IGP17, IGP18, IGP20, IGP21, IGP22, IGP23

Monosialylation: IGP15, IGP16, IGP17, IGP18

Disialylation: IGP20, IGP21, IGP22, IGP23

Sialylation: IGP15, IGP16, IGP17, IGP18, IGP20, IGP21, IGP22, IGP23

Fucosylation: IGP1, IGP3, IGP5, IGP7, IGP8, IGP9, IGP10, IGP13, IGP14, IGP15, IGP17, IGP18, IGP22, IGP23

BisectingGlcNAc: IGP5, IGP9, IGP10, IGP12, IGP14, IGP18, IGP23

● Mannose ■ N-Acetylglucosamine ○ Galactose ▲ Fucose ◆ N-Acetylneuraminic acid
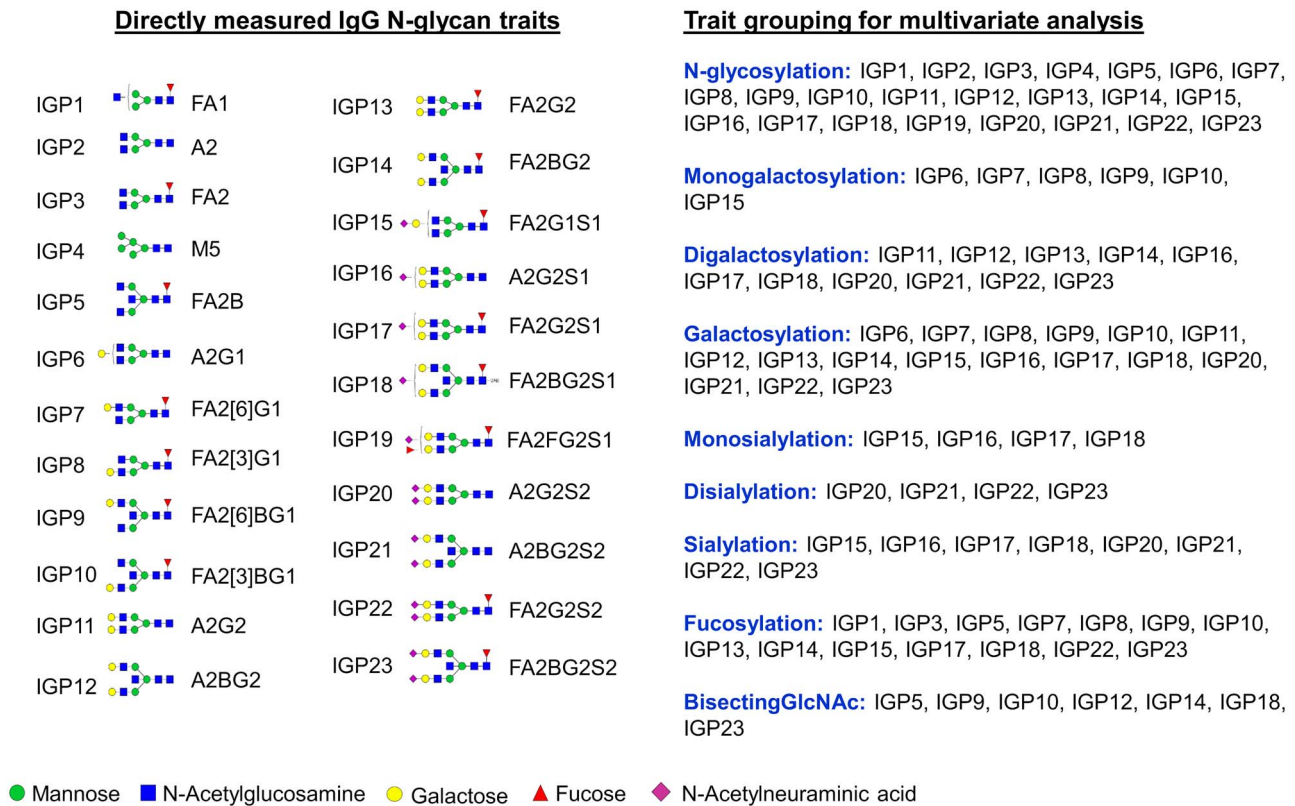
Figure 1. Directly measured IgG N-glycan traits and trait grouping for multivariate analysis. IGP1-23 traits (Supplementary Material, Table S1) correspond to UPLC peaks and approximate the percentage of a certain glycan (the most abundant glycan in the peak) in total IgG glycans (18, 19). Sugar residues are graphically represented according to *Essentials of Glycobiology, 2nd edition* (20). Glycan names are given according to the Oxford nomenclature (described in (21)). GlcNAc, N-Acetylglucosamine.

trait groups revealed in the multivariate analysis are listed in Supplementary Material, Table S2.

The six loci we discovered were not reported as genome-wide significant in any previous IgG N-glycome GWAS and were therefore considered novel; although the locus rs4561508 near *TNFRSF13B* reached a suggestive level of significance in the most recent GWAS by Klarić *et al.* (12). Their regional association plots are presented in Supplementary Material, Figure S4. Twenty six other loci overlap with those identified previously (which means that the lead SNPs corresponding to these 26 loci reside within the genome-wide significant loci reported in the studies (9–12)). Of those, 21 were replicated in at least one previous GWAS (9–12), and 5 were first discovered in Klarić *et al.*'s GWAS (12), but their associations were not replicated in the UPLC-measured replication cohort of that study.

Replication, which is a reproduction of statistically significant associations in an independent set of samples, is an important step in GWAS. Replication reduces the likelihood that the observed genotype–phenotype association is a chance finding or an analysis artifact (17) and is, therefore, essential before undertaking complex and expensive experimental follow-up studies. In our replication analysis, we included six novel loci discovered at the genome-wide significance level in this study and the five loci identified, but not replicated previously. Our replication analysis used 3147 samples from three cohorts. From 11 loci analyzed, 7 [5 novel and 2 discovered by Klarić *et al.* (12)] passed our criteria for replication: (1) were associated at $P < 0.0045$ with the discovery top trait group in the MANOVA replication test and (2) were associated at $P < 0.0045$ with the linear combinations of glycan traits (constructed in the replication cohort using the

coefficients estimated in the discovery analysis), with the beta sign for the effect allele (EA) consistent with that observed in the discovery analysis (Table 2). Details are provided in Supplementary Material, Table S2, Supplementary Material, Figure S2b (correlations between the traits in the replication cohort) and Supplementary Material, Table S3 (partial regression coefficients for individual IgG N-glycosylation traits estimated for the top trait groups in the discovery and replication cohorts). Thus, our study identified and replicated seven additional loci associated with IgG N-glycome, which constitutes nearly one-third of the number of loci that are already known to affect IgG N-glycosylation ($n = 22$). For these loci, we performed an *in silico* functional analysis (see below).

Functional annotation of replicated loci

Variant Effect Predictor analysis. The results of the Ensembl Variant Effect Predictor (VEP) analysis (22) are shown in Supplementary Material, Table S4. The analysis identified two missense SNPs: rs34562254 in the *TNFRSF13B* (TNF receptor superfamily member 13B) gene and rs3796129 in the *RNF168* (ring finger protein 168) gene. These polymorphisms are in high LD with the lead SNPs rs4561508 and rs12635457 associated with the trait group 'N-glycosylation' in our multivariate analysis. The missense variant rs34562254 A is correlated with the allele rs4561508 T ($r^2 = 0.89$ in European ancestry populations according to LDlink (23), https://analysistools.nci.nih.gov/LDlink/), and the missense allele rs3796129 T is correlated with the allele rs12635457 A ($r^2 = 0.94$ in European ancestry populations). SNP

Table 1. Loci associated with IgG N-glycosylation trait groups in the multivariate analysis (discovery cohort)

Lead SNP	Chr: position[a]	EA/RA[b]	Top[c] group of IgG N-glycosylation traits	β (SE)	P-value	EAF[d]	N[e]	Number of traits groups[f]	Nearest gene(s)[g]
Novel loci identified in a multivariate analysis									
rs12049042	1:246288812	C/T	Galactosylation	0.02 (0.003)	1.20×10^{-09}	0.01	4417	1	SMYD3
rs11895615	2:26113120	C/T	BisectingGlcNAc	0.01 (0.001)	5.69×10^{-10}	0.70	8024	1	ASXL2
rs1372288	3:142901537	C/T	N-glycosylation	0.01 (0.002)	8.73×10^{-11}	0.23	8090	2	CHST2, SLC9A9
rs12635457	3:196203979	A/G	N-glycosylation	0.02 (0.002)	1.61×10^{-13}	0.64	8016	1	RNF168
rs479844	11:65551957	A/G	N-glycosylation	0.02 (0.002)	1.97×10^{-13}	0.42	8089	3	OVOL1
rs4561508	17:16848750	C/T	N-glycosylation	0.01 (0.002)	1.38×10^{-10}	0.90	8002	2	TNFRSF13B
Loci identified but not replicated in Klarić *et al.* (12)									
rs6964421	7:6520676	C/T	Fucosylation	0.01 (0.002)	2.04×10^{-12}	0.67	8090	5	KDELR2
rs10096810	8:103545436	A/G	BisectingGlcNAc	0.01 (0.001)	4.69×10^{-09}	0.63	8090	1	ODF1
rs540719	16:23456578	C/T	Fucosylation	0.01 (0.002)	8.20×10^{-11}	0.19	8087	2	COG7
rs7257072	19:19267990	C/T	Fucosylation	0.01 (0.002)	5.83×10^{-14}	0.49	8090	7	MEF2B, BORCS8-MEF2B
rs2745851	20:17829280	A/G	Sialylation	0.02 (0.002)	4.19×10^{-20}	0.38	8087	9	MGME1, BANF2, SNX5
Loci identified and replicated in at least one previous study (9–12)									
rs311438	1:25320236	A/G	Digalactosylation	0.01 (0.001)	1.66×10^{-09}	0.09	8070	1	MIR4425, RUNX3
rs6764279	3:186723975	A/C	Galactosylation	0.22 (0.006)	3.51×10^{-302}	0.28	8086	9	ST6GAL1
rs3777184	5:95262044	C/G	N-glycosylation	0.03 (0.002)	1.22×10^{-27}	0.26	8084	8	ELL2
rs4705950	5:131793286	C/T	N-glycosylation	0.01 (0.002)	1.99×10^{-09}	0.58	8010	1	IRF1-AS1, IRF1, SLC22A5
rs3099844[h]	6:31448976	A/C	Fucosylation	0.03 (0.003)	2.81×10^{-20}	0.14	8090	8	MICB-DT, MICB, HCG26
rs9385856	6:139625074	C/T	Fucosylation	0.02 (0.002)	1.16×10^{-17}	0.43	8089	8	TXLNB
rs7758383	6:143169723	A/G	Fucosylation	0.02 (0.002)	4.86×10^{-28}	0.52	8020	6	HIVEP2
rs7789913	7:50352695	C/T	Galactosylation	0.03 (0.002)	3.63×10^{-36}	0.61	8089	8	IKZF1
rs7812088	7:150919829	A/G	N-glycosylation	0.03 (0.003)	6.32×10^{-35}	0.12	8089	8	ABCF2
rs10124479	9:33136233	C/T	Fucosylation	0.06 (0.003)	7.52×10^{-76}	0.32	8070	9	B4GALT1
rs481080	11:114442265	A/G	N-glycosylation	0.02 (0.002)	5.43×10^{-14}	0.48	8090	5	NXPE4, NXPE2
rs11847263[i]	14:65775695	G/T	N-glycosylation	0.06 (0.003)	6.92×10^{-71}	0.35	8044	9	MIR4708, FUT8
rs4074453	14:105998544	C/T	Fucosylation	0.04 (0.003)	1.55×10^{-51}	0.24	8089	7	TMEM121
rs4795400	17:38067020	C/T	Monogalactosylation	0.01 (0.001)	3.41×10^{-13}	0.53	7978	3	IKZF3, GSDMB, ORMDL3, LRRC3C, ZPBP2
rs10445337[i]	17:44067400	C/T	N-glycosylation	0.02 (0.002)	1.62×10^{-22}	0.21	8076	6	MAPT
rs11650354	17:45822092	C/T	Galactosylation	0.01 (0.002)	3.34×10^{-10}	0.83	8090	2	TBX21
rs9319617	17:79192446	C/T	Fucosylation	0.02 (0.002)	5.95×10^{-19}	0.50	7750	6	CEP131
rs12986368	19:5830922	C/T	N-glycosylation	0.02 (0.002)	3.45×10^{-21}	0.63	7717	4	FUT6
rs7281587	21:36565278	A/G	N-glycosylation	0.02 (0.002)	5.59×10^{-13}	0.25	8090	5	RUNX1, RUNX1-IT1
rs17630758	22:24136542	A/G	N-glycosylation	0.03 (0.003)	3.31×10^{-41}	0.15	8041	8	SMARCB1
rs5750830	22:39840828	A/C	N-glycosylation	0.08 (0.004)	8.11×10^{-101}	0.74	7867	8	TAB1, MGAT3

β (SE), *P*-values and *N* are given for associations with top group of IgG N-glycosylation traits. GlcNAc, *N*-Acetylglucosamine. Associated loci were defined as regions within ±500 kb around the lead SNPs.
[a]Chromosome: position on chromosome according to GRCh37.p13 assembly.
[b]Effect allele/reference allele.
[c]Trait group associated with SNP at the highest level of statistical significance.
[d]Effect allele frequency (discovery cohort).
[e]Number of subjects in a multivariate analysis.
[f]Total number of traits groups associated with the locus.
[g]Nearest gene(s) according to the dbSNP (https://www.ncbi.nlm.nih.gov/snp/).
[h]MHC locus (chromosome 6, 25–35 Mb).
[i]For these loci, regional association plots were manually inspected, and neighboring loci were merged (three loci merged into one for the locus marked by rs10445337; two loci merged into one for the locus marked by rs11847263; Supplementary Material, Figure S3).

rs3796129 is triallelic (G >C, T; Pro401Arg or Pro401Gln) with the allele C being extremely rare, and LD between rs3796129 C and rs12635457 alleles is unknown. The rs34562254 (Pro205Leu) was classified as possibly damaging by the PolyPhen algorithm and as tolerated by the SIFT tool. The rs3796129 was classified as benign/tolerated by both algorithms. Based on these data, we only prioritized the gene *TNFRSF13B* bearing the polymorphism rs34562254.

Colocalization analysis. Colocalization analysis was performed for five out of seven loci identified and replicated in our study (rs11895615, rs479844, rs4561508, rs7257072 and rs2745851). The

loci rs1372288 and rs12635457 were not included in the analysis since they did not reach a genome-wide significance in the meta-analysis of univariate GWAS results for the 23 IgG N-glycosylation traits obtained in the discovery and replication cohorts. The results of the SMR/HEIDI (Summary data-based Mendelian Randomization analysis/Heterogeneity in Dependent Instruments test) colocalization analysis (24) with expression quantitative trait loci (eQTLs) are presented in Supplementary Material, Table S5a. The analysis suggested two genes whose expression levels may be associated with the same functional polymorphisms that affect IgG N-glycosylation: *OVOL1* (ovo-like transcriptional repressor 1) and *EFEMP2* (EGF containing fibulin extracellular matrix protein 2). The strongest result (nominal

Table 2. Replication of loci associated with IgG N-glycosylation trait groups

Lead SNP	Chr: position[a]	EA/RA[b]	Top[c] group of IgG N-glycosylation traits	P_{MANOVA}[d]	$\beta_{discover}$ (SE)[e]	$\beta;lin.com$ (SE)[f]	$P_{lin.com}$[f]	N[g]	Nearest gene(s)[h]	Genes prioritized by *in silico* methods[i] or suggested based on literature data[j]
Novel loci identified in a multivariate analysis										
rs12049042	1:246288812	C/T	*Galactosylation*	0.45	0.02 (0.003)	−0.13 (0.08)	0.11	2572	*SMYD3*	*Not tested*
rs11895615	2:26113120	C/T	BisectingGlcNAc	3.01×10^{-08}	0.01 (0.001)	0.16 (0.02)	1.24×10^{-11}	3147	ASXL2	ASXL2 (D)
rs1372288	3:142901537	C/T	N-glycosylation	2.06×10^{-13}	0.01 (0.002)	0.17 (0.03)	5.06×10^{-10}	3147	CHST2, SLC9A9	SLC9A9 (L)
rs12635457	3:196203979	A/G	N-glycosylation	1.13×10^{-04}	0.02 (0.002)	0.08 (0.02)	4.77×10^{-04}	3147	RNF168	RNF168 (L)
rs479844	11:65551957	A/G	N-glycosylation	2.69×10^{-14}	0.02 (0.002)	0.20 (0.02)	5.62×10^{-18}	3147	OVOL1	AP5B1 (D), OVOL1 (S), EFEMP2 (S[*])
rs4561508	17:16848750	C/T	N-glycosylation	3.27×10^{-04}	0.01 (0.002)	0.13 (0.04)	2.91×10^{-04}	3147	TNFRSF13B	TNFRSF13B (V, D)
Loci identified but not replicated in Klarić et al. (12)										
rs6964421	7:6520676	C/T	*Fucosylation*	7.20×10^{-03}	0.01 (0.002)	0.10 (0.02)	5.35×10^{-05}	3147	*KDELR2*	*Not tested*
rs10096810	8:103545436	A/G	*BisectingGlcNAc*	0.09	0.01 (0.001)	0.07 (0.02)	2.75×10^{-03}	3147	*ODF1*	*Not tested*
rs540719	16:23456578	C/T	*Fucosylation*	0.12	0.01 (0.002)	0.03 (0.03)	0.27	3147	*COG7*	*Not tested*
rs7257072	19:19267990	C/T	*Fucosylation*	5.61×10^{-05}	0.01 (0.002)	0.13 (0.02)	8.19×10^{-08}	3147	*MEF2B, BORCS8-MEF2B*	
rs2745851	20:17829280	A/G	*Sialylation*	1.65×10^{-04}	0.02 (0.002)	0.20 (0.04)	7.91×10^{-06}	1127	*MGME1, BANF2, SNX5*	

The loci which were not replicated are shown in italics. More details are provided in Supplementary Material, Table S2. GlcNAc, N-Acetylglucosamine.
[a]Chromosome: position on chromosome according to GRCh37.p13 assembly.
[b]Effect allele/reference allele.
[c]Trait group associated with SNP at the highest level of statistical significance in the discovery cohort.
[d]P-value for the association with the top trait group in the replication cohort (MANOVA test).
[e]Beta (SE) for the association with IgG N-glycosylation trait groups in the multivariate analysis in the discovery cohort.
[f]Beta (SE) and P-value for the association with the linear combination of glycan traits in the replication cohort (constructed using the coefficients estimated in the discovery cohort; each coefficient corresponds to the trait in the discovery top trait group).
[g]Number of subjects in the replication analysis.
[h]Nearest gene(s) according to the dbSNP (https://www.ncbi.nlm.nih.gov/snp/).
[i]Genes prioritized in the VEP analysis (V; Supplementary Material, Table S4), SMR/HEIDI analysis (S, likely shared, $P_{HEIDI} \geq 0.05$; S[*], possibly shared, $0.05 > P_{HEIDI} \geq 0.001$; Supplementary Material, Table S5a) and DEPICT analysis (D; Supplementary Material, Table S6a).
[j]Genes that can be considered potentially causal based on published data on their function (L).

$P_{SMR} = 6 \times 10^{-8}$, Bonferroni-corrected $P_{SMR} = 5 \times 10^{-5}$, $P_{HEIDI} = 0.4$) was obtained for the locus marked by rs479844 where increased expression of *OVOL1* in Epstein–Barr virus (EBV)-transformed lymphocytes was found to be associated with the same allele as associated with the increased percentage of FA2FG2S1 glycan in total IgG glycans (IGP19 UPLC peak). For the same locus, we also observed colocalization with the expression of the *EFEMP2* gene in peripheral blood; however, the SMR test significance was weaker (Bonferroni-corrected $P_{SMR} = 0.003$), and the sharing was characterized as 'possible' ($P_{HEIDI} = 0.03$). Based on these observations, we consider *OVOL1* and *EFEMP2* as possible candidate genes, which may influence IgG N-glycosylation or its regulation.

We also used colocalization analysis to check for overlap between the IgG N-glycosylation loci and the loci affecting the risk of autoimmune and inflammatory diseases (asthma, rheumatoid arthritis, primary biliary cirrhosis, ankylosing spondylitis and inflammatory bowel disease, including Crohn's disease and ulcerative colitis). For the locus marked by rs479844 (near *OVOL1*), we detected colocalization with self-reported asthma (the UK Biobank phenotype). The same allele was found to be associated with the increased percentage of FA2FG2S1 glycan in total IgG glycans (IGP19 UPLC peak) and decreased asthma risk (Supplementary Material, Table S5b). For other diseases and loci analyzed, colocalization was not established.

Data-driven Expression Prioritized Integration for Complex Traits analysis. The genes prioritized by the Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) (25) analysis are listed in Supplementary Material, Table S6a. Statistically significant results (FDR < 0.05) were only observed when the P-value threshold for input SNPs was set at 5.0×10^{-8}. DEPICT prioritized three genes in three loci replicated in our study: *ASXL2* (ASXL transcriptional regulator 2; locus marked by rs11895615), *AP5B1* (adaptor related protein complex 5 subunit beta 1; locus marked by rs479844) and *TNFRSF13B* (locus marked by rs4561508).

DEPICT gene set enrichment analysis provided results with FDR < 0.05 only when we included all SNPs associated with IgG N-glycosylation at $P < 1.0 \times 10^{-5}$. The only enriched category identified was 'decreased mature B cell number' (Supplementary Material, Table S6b). DEPICT tissue/cell type set enrichment analysis yielded results with FDR < 0.05 only when the P-value threshold for input SNPs was set at 5.0×10^{-8}. Nineteen tissues/cell types were revealed, and almost all of them were related to the hemic and immune systems: blood, palatine tonsil, lymphocytes, leukocytes, antibody producing cells, plasma cells, natural killer cells, etc. (Supplementary Material, Table S6c).

Summary of in silico gene prioritization. *In silico* functional analysis allowed us to suggest a total of five potentially causal

genes for the three loci marked by rs11895615, rs479844 and rs4561508 (Table 2). For the locus rs11895615, a single *ASXL2* gene (nearest to the lead SNP) was proposed by the DEPICT tool. For the loci rs479844 and rs4561508, three genes and one gene were proposed, respectively, of which the gene *TNFRSF13B* (for rs4561508) was prioritized using two approaches (eQTL colocalization analysis and VEP analysis), *AP5B1* (for rs479844) was proposed only by the DEPICT tool and *OVOL1* and *EFEMP2* (for rs479844) were suggested only by the eQTL colocalization analysis. Colocalization with eQTL, on the one hand, can provide evidence for expression-mediated effects on the analyzed trait and, on the other hand, is not able to distinguish causation from pleiotropy and in some cases indicates a group of genes whose expression may be affected by the same regulatory locus, or even a group of functional SNPs in very tight LD. We therefore avoid making definite conclusions on the best candidate genes for the locus rs479844.

Discussion

Identification of new loci associated with IgG N-glycosylation can provide clues towards understanding the mechanisms of *in vivo* regulation of this important biological process intimately connected to immune system function. These leads and hypotheses should be followed up in subsequent targeted functional experiments. However, before undertaking lengthy and expensive experimental work, much effort should be invested in excluding the possibility of false-positive associations by performing replication in independent samples as well as conducting *in silico* functional annotation.

In this study, we performed a multivariate genome-wide association analysis of 23 plasma IgG N-glycosylation traits having combined them into nine groups based on general structural characteristics of IgG-attached glycans (see Fig. 1; one group included all 23 traits). A multivariate approach can increase the statistical power compared to standard univariate GWAS (10,14,15), and here we demonstrate this by revealing six novel genome-wide significant loci, which have not been identified in the study of Klarić *et al*. (12), providing us with the GWAS summary statistics for individual glycan traits. Of note, 26 out of 27 loci revealed in that study (12) (all except rs12341905 near *SPINK4*) also reached genome-wide significance in our work (Table 1, Supplementary Material, Table S2). We replicated five out of six of the novel loci as well as an additional two loci that have not been replicated previously. Thus, our study increases the number of established IgG N-glycosylation loci from 22 to 29.

It may be interesting to discuss why in our analysis the association with rs12341905 near *SPINK4*—a locus not only detected but also replicated by Klarić *et al*. (12)—was not genome-wide significant. First, we should mention that while a multivariate analysis increases power for identification of genetic associations with glycan traits, this is a statistical relationship that holds asymptotically but is not guaranteed to hold for every specific test. This is especially true for scenarios under which the gain of power by the multivariate analysis is minimal (10,16). Note also that in the study of Klarić *et al*. (12) the association with rs12341905 was the weakest among 27 others and had barely passed the genome-wide significance threshold. Therefore, it may be not too surprising that in our study that used another analysis method this SNP has just missed the study-specific genome-wide significance threshold: the P-value for the trait 'monosialylation' was 1.46×10^{-8}, which is greater than 5.6×10^{-9}. Still, we should emphasize that, given replication of

this locus by Klarić *et al*. (12), we consider this locus as established, and we attribute the fact that we could not detect this locus at genome-wide significance using multivariate analysis to the stochastic variation in power of the methods.

For the seven newly established loci, we have performed an extensive *in silico* follow-up analysis. None of these loci contain genes encoding glycosyltransferases. For five of the seven replicated loci, we suggested seven candidate genes based on the analysis of literature (two genes) and/or through use of several bioinformatics tools and methods (five genes) (Table 2), with the *TNFRSF13B* gene (locus rs4561508) being prioritized using two independent approaches.

Among the genes prioritized as possible candidates, several could be linked to N-glycosylation and/or immune function based on current knowledge.

In the locus tagged by rs1372288, the top associated SNP is in LD with rs4839604 ($r^2 = 0.79$ in European ancestry populations according to LDlink) associated with plasma protein N-glycosylation (26,27). IgG is among the most abundant plasma glycoprotein (1,2), and we assume that changes in its glycosylation patterns may be responsible for the effects observed for the total plasma protein N-glycome. These previous studies (26,27) did not apply *in silico* methods to find potentially causal genes but prioritized the gene *SLC9A9* (carrier family 9 member A9) based on its possible involvement in the regulation of endosomal pH (28).

The *TNFRSF13B* gene, which was suggested by both the VEP and DEPICT tools, encodes a lymphocyte-specific member of the tumor necrosis factor receptor superfamily known as the TACI protein (the transmembrane activator and calcium modulator and cyclophilin ligand interactor). Signaling through this protein activates several transcription factors, including NFAT, AP-1 and NF-kappa-B (29). TACI plays a crucial role in humoral immunity acting in synergy with other signaling networks to promote B cell differentiation into plasma cells and immunoglobulin production (30,31). Coding variants in *TNFRSF13B* are associated with common variable immunodeficiency and selective IgA deficiency (32), rare *TNFRSF13B* mutations can increase the risk of asthma symptoms (33) and several common variants in this gene were reported to be associated with circulating IgG levels in GWASes conducted by Liao *et al*. (34) and Jonsson *et al*. (35). One of the lead SNPs in the study of Liao *et al*., rs4792800, is in LD with our lead N-glycosylation-associated SNP rs4561508 ($r^2 = 0.95$ in European ancestry populations according to LDlink). These findings indicate a possible pleiotropic action of *TNFRSF13B* with respect to N-glycosylation of IgG and immune function.

The gene *RNF168* (ring finger protein 168) bearing rs12635457 in its intron encodes an E3 ubiquitin ligase protein involved in DNA double-strand break repair and immunoglobulin class switch recombination (36). Mutations in this gene lead to a RIDDLE syndrome (an immunodeficiency and radiosensitivity disorder), the features of which include decreased immunoglobulin production (37). An association between SNPs at/near *RNF168* and IgM levels was previously reported (35). As far as we are aware, there is no reported evidence for *RNF168* involvement in glycosylation processes, but if so, its pleiotropic action on immune-related and glycosylation traits may be assumed.

Among IgG N-glycosylation loci established in this work, the locus tagged by rs479844 near *OVOL1* is the most interesting in our opinion. We demonstrated the potential pleiotropic effect of this locus on asthma (Supplementary Material, Table S5b); the same locus was identified in GWAS for the combined eczema plus asthma phenotype (the atopic march) (38), combined phenotypes of asthma/hay fever/eczema (39) and atopic dermatitis

(40,41). This locus was also found to have pleiotropic effects on dermatological and allergic diseases (42) and to be differentially methylated in childhood asthma (43). Three genes, *AP5B1, OVOL1* and *EFEMP2*, were proposed for the rs479844 locus by our bioinformatic analysis, with colocalization results for *OVOL1* being statistically strong and *AP5B1* being prioritized by the DEPICT gene prioritization tool. AP5B1 (also known as DKFZp761E198) is a subunit of an adaptor protein complex AP-5 (44) likely involved in facilitating late endosome-to-Golgi retrieval—a pathway contributing to a normal lysosomal function (45). *OVOL1* genes encode a putative zinc finger containing transcription factor, which is involved in the development and differentiation of epithelial tissues and germ cells (46–48) and plays a role in a skin barrier function (49). Finally, *EFEMP2* encodes the EGF-containing fibulin-like extracellular matrix protein 2 (also known as fibulin-4). Rare mutations in this gene lead to cutis laxa type IB, an autosomal recessive disorder characterized by the presence of severe systemic connective tissue abnormalities [phenotype number in the Mendelian Inheritance in Man (MIM, https://www.omim.org/) database: 614437] (50,51). A potential link between the genes *AP5B1, OVOL1, EFEMP2* and IgG N-glycosylation is not apparent. Therefore, we conclude that further research is needed to define the causal gene in the rs479844 locus.

Besides searching for causal genes and colocalization analysis, the functional annotation in our study involved enrichment analysis and assessment of colocalization with a set of inflammatory and autoimmune diseases. We observed an enrichment of tissue/cell/gene set categories related to the hemic and immune systems (Supplementary Material, Table S6b, S6c).

An intriguing pattern that emerges from this and previous studies is the overlap between loci that regulate IgG N-glycosylation and immune function. In previous works (9–12), a total of 22 IgG N-glycome-associated loci have been identified and replicated. Of these, five contained genes encoding N-glycosyltransferases. Of the other 17, 11 loci contained candidate genes implicated in the immune response. These loci, genes and proteins are as follows: the human leukocyte antigen (HLA) region, which contains multiple immune-related genes; HIVEP2 involved in cellular immunity (52,53); transcription factors RUNX1, RUNX3, IKZF1 and IKZF3 that play a central role in B-cell maturation and differentiation; VPREB3, which likely participates in the transport and assembly of the pre-B cell receptor (54); a transcription factor TBX21, mouse homolog of which regulates the expression of IFN*γ* and controls Th1 lineage commitment (55); a transcription elongation factor ELL2, which directs immunoglobulin secretion in plasma cells (56); a locus near the *IRF1* (interferon regulatory factor 1) gene and a locus near *IGH* that codes for immunoglobulin heavy chains. Not surprisingly, the enrichment analysis highlighted the gene sets that are known to be connected to various processes involved in immunity, such as B-cell life cycle and antibody production (12). Moreover, for two loci, pleiotropic effects were demonstrated on the pro-inflammatory pattern of IgG N-glycosylation and the risk of inflammatory and autoimmune diseases such as Crohn's (locus near *IRF1-SLC22A4*), ulcerative colitis, rheumatoid arthritis and primary biliary cirrhosis (locus near *ORMDL3-GSDMB-IKZF3*). Interestingly, the latter locus also exhibited pleiotropy with the risk of asthma, although the asthma risk allele was 'protective' for autoimmune disease. This is a well-known phenomenon for the 17q12-21 locus (13).

Among five novel loci detected and replicated in the present study, three can be connected to the immune system's function. Given that this work has detected a second locus that is likely to be shared between IgG N-glycosylation and asthma, it may be tempting to speculate about biological mechanisms linking these traits. However, we believe that it is too early to draw strong conclusions, and more data should be accumulated by more powerful GWAS. Several observations suggest that the mechanism linking asthma and IgG N-glycome via the loci near *ORMDL3-GSDMB-IKZF3* and *OVOL1* are different and support a cautious approach to interpretation of the results. First, different IgG N-glycan traits are associated with these loci. Second, the mechanisms by which these loci alter asthma risk are not likely to be the same. For example, the *ORMDL3-GSDMB-IKZF3* locus was found to be more strongly associated with asthma without hay fever/eczema than to hay fever/eczema without asthma; the reverse is true for the locus near *OVOL1* (39), which is also strongly implicated in atopic dermatitis (40,41).

Our findings are consistent with the conclusion drawn in the previous studies (9,10,12) that the loci and genes associated with IgG N-glycosylation are also related to immune function and immune-related conditions. In general, our results provide additional information characterizing the interplay between glycosylation and immune function.

In conclusion, our study establishes seven new loci associated with plasma IgG N-glycosylation and suggests a number of genes potentially involved in this process. Some loci and genes (*TNFRSF13B*, *RNF168* and *OVOL1/AP5B1*) are likely to have pleiotropic effects on the traits related to the immune system and inflammatory diseases. For the locus near *OVOL1*, we demonstrate colocalization with variants affecting the risk of asthma, and the genes *TNFRSF13B* and *RNF168* have previously been implicated in immunity disorders and immunoglobulin production. Future studies should focus on increasing statistical power to provide better resolution for the *in silico* studies and on providing mechanistic insights into the observed associations by targeted functional follow-up. Besides this, it will be fruitful to investigate causative relationships between alteration of IgG N-glycosylation and immune-related conditions.

Materials and Methods

Study cohorts

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

Discovery cohort. In the discovery stage of our analysis, we used GWAS summary statistics for plasma IgG N-glycosylation traits obtained in the study by Klarić *et al*. (12) (discovery dataset, European ancestry individuals, *N* = 8090; GWAS summary statistics can be downloaded from (57)). In that study, GWAS was conducted for a total of 77 quantitative IgG N-glycosylation traits—23 traits directly measured by ultra-performance liquid chromatography and 54 derived traits combining directly measured glycans in a biologically meaningful way. The full list of traits is given in Supplementary Material, Table S1. Description of the analyzed cohorts and details on IgG isolation from plasma samples, IgG N-glycome quantification, genotyping, glycan and genetic data pre-processing, quality control, imputation, genome-wide association analyses and meta-analysis are provided in Klarić *et al*. (12).

Replication cohort. Replication was performed using GWAS data obtained for a total of 3147 samples from four cohorts of European descent: EGCUT ($N = 575$), FINRISK ($N = 552$), CROATIA-Korcula2 ($N = 941$) and VIKING ($N = 1079$). Description of the EGCUT and FINRISK cohorts as well as details on phenotyping- and genotyping-related procedures can be found in the study of Klarić *et al.* (12).

The Viking Health Study—Shetland (VIKING) is a cross-sectional volunteer study from the isolated archipelago of Shetland in Northern Scotland, with high kinship and considerable genetic drift (58).

The CROATIA-Korcula2 study is a population-based study that sampled participants, aged 18–92 years, from the Adriatic island of Korčula in 2012 as an extension of the 10 001 Dalmatians Biobank (59). All subjects visited the clinical research center in the region, where they were examined in-person and where fasting blood was drawn and stored for future analyses. Many biochemical and physiological measurements were performed, and questionnaires of medical history as well as lifestyle and environmental exposures were collected.

GWAS results for replication cohorts were combined by the inverse-variance-weighted meta-analysis (fixed-effects model). Analysis was performed using the GWAS-MAP platform (60). SNPs with MAF < 1% were excluded. Before meta-analysis, a genomic control (61) was applied to the summary statistics of each study.

Phenotypic correlations

Phenotypic correlations between 23 IgG N-glycosylation traits were estimated as correlations between the vectors of Z-statistics obtained in GWAS for each trait (method described by Stephens (15)). Pearson correlations were computed using the R statistical software (62) version 3.4.4.

Multivariate analysis

Multivariate analysis was performed using the summary statistics-based method described by Ning *et al.* (63). This method enables multivariate analysis of variance (MANOVA) using the results of multiple GWAS for individual traits. Calculations were performed using the MultiABEL package version 1.1–9 (https://cran.r-project.org/src/contrib/Archive/MultiABEL/, tutorial: https://github.com/xiashen/MultiABEL/; 'MultiSummary' function) for R software environment (64).

Following Shen *et al.* (10), the statistical significance threshold for multivariate analysis was set at P-value $< 5.6 \times 10^{-9}$ ($5.0 \times 10^{-8}/9$, where 9 is the number of glycan groups).

Associated loci were defined as regions within ±500 kb around the SNPs most significantly associated with IgG N-glycosylation trait groups (lead SNPs). SNPs located between 25 and 35 Mb on the chromosome 6 were merged into a single MHC locus. Additionally, for neighboring loci, regional association plots were manually inspected, and the loci were merged if they fell into one LD block.

Replication

Replication was performed according to a protocol similar to that described in Shen *et al.* (10). First, summary statistics obtained in the meta-analysis of replication cohorts was used to get MANOVA results for each lead SNP and each 'top' trait group (the group of IgG N-glycosylation traits most significantly associated

with the lead SNP in the discovery cohort). Analysis was performed using the MultiABEL R package version 1.1-9 ('MultiSummary' function). Association was considered replicated if the association between the lead SNP and the corresponding trait group in the replication MANOVA test passed the Bonferroni-corrected threshold of $P < 0.0045$ (0.05/11, where 11 is the number of loci in the replication analysis; the first criterion of replication). Next, for each lead SNP-top trait group pair, we extracted the coefficients of linear combinations of individual traits in the discovery cohort and used them to construct the corresponding trait group in the replication cohort. Associations between the lead SNPs and the constructed linear combinations were tested as described in Tsepilov *et al.* (64) and McDaid *et al.* (65). In brief, the effects and corresponding standard errors (SEs) were calculated as follows:

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

$$SE_{new} = \frac{\sqrt{\left(SE_1^2 + \frac{\beta_1^2}{N}\right)*sum(\Sigma_{ph}*(a \otimes a)) - \frac{(B*a)^2}{N}}}{\sqrt{sum(\Sigma_{ph}*(a \otimes a))}}, \text{respectively};$$

where B—matrix, where each column β_i—effect of each SNP in GWAS$_i$; a—vector of linear combination coefficients estimated in the discovery cohort; Σ_{ph}—matrix of correlations between 23 IgG N-glycosylation traits in the replication cohort; $\otimes$—outer product; SE$_1$—standard error of each SNP's effect in the first GWAS$_1$; N—minimal N out of all GWASes for each SNP. If the association of the EA with the constructed linear combinations had the same direction of effect as in the discovery cohort and passed the Bonferroni-corrected threshold of $P < 0.0045$, the locus was considered replicated (the second criterion of replication).

In silico functional annotation

Prediction of SNP effects. We used the Ensembl VEP (22) to analyze the potential effects of SNPs and indels in high LD ($r^2 > 0.8$) with the replicated polymorphisms. LD was calculated using PLINK 1.9 (66) (–*show-tags* option) and genotype data for 503 European ancestry individuals (1000 Genomes phase 3 version 5 data). The analysis was performed using software available online (https://www.ensembl.org/info/docs/tools/vep/index.html).

Colocalization analysis. In order to find potentially causal genes involved in IgG N-glycosylation and to investigate pleiotropic effects on autoimmune and inflammatory conditions, we analyzed the colocalization between the loci identified and replicated in our multivariate analysis, eQTLs from selected tissues and loci associated with asthma, rheumatoid arthritis, primary biliary cirrhosis, ankylosing spondylitis, inflammatory bowel disease, Crohn's disease and ulcerative colitis [these traits were selected in accordance with the similar analysis performed in the study of Klarić *et al.* (12)]. Colocalization was examined using SMR analysis followed by the HEIDI test (24). This approach allows to identify the genes whose expression level may be affected by the same causal SNP that is associated with the analyzed trait as well as to reveal pleiotropic loci influencing multiple traits or diseases. The SMR test indicates whether two traits (e.g. IgG N-glycosylation and gene expression level in a certain tissue or the disease) are associated with the same locus, and HEIDI test specifies whether both traits are affected by the same underlying functional SNP.

Results of multivariate GWAS cannot be directly used for a colocalization analysis. Therefore, we have meta-analyzed univariate GWAS results for the 23 IgG N-glycosylation traits obtained in discovery and replication cohorts using the inverse-variance-weighted meta-analysis method (fixed-effects model). For each locus included in the colocalization analysis, we selected the 'primary' IgG N-glycosylation GWAS from the resulting meta-analyses. The selection criterion was the lowest *P*-value for the association between the lead SNP marking the locus and IgG N-glycosylation trait for which the meta-analysis was conducted. Loci where associations did not reach *P*-value $<5.0 \times 10^{-8}$ (rs1372288 and rs12635457) were excluded from the analysis.

Summary statistics for gene expression levels in tissues/cell types were obtained from the Blood eQTL study (67) (http://cnsgenomics.com/software/smr/#eQTLsummarydata), the CEDAR project (68) (http://cedar-web.giga.ulg.ac.be/) and the GTEx project version 7 (69) (https://gtexportal.org). In total, we used data for eight tissues/cell types: peripheral blood (from Blood eQTL study); six circulating immune cell types—CD4+ T lymphocytes, CD8+ T lymphocytes, CD19+ B lymphocytes, CD14+ monocytes, CD15+ granulocytes and platelets (from CEDAR); and the EBV-transformed lymphocytes (from GTEx).

Summary statistics for autoimmune/inflammatory diseases were obtained from the following resources: for asthma (self-reported), we used the UK Biobank (70) GWAS summary statistics provided by the Neale Lab (http://www.nealelab.is/; 39 049 case and 298 110 control individuals; data were downloaded in 2017); for rheumatoid arthritis, GWAS summary statistics from Okada *et al.*'s study (71); for primary biliary cirrhosis, from Cordell *et al.*'s study (72); for ankylosing spondylitis, from Cortes *et al.*'s study (73) and for inflammatory bowel disease and its subtypes, from Liu *et al.*'s study (74).

SMR/HEIDI analysis was performed according to the protocol described by Zhu *et al.* (24). We used sets of SNPs having the following properties: (1) being located within ±250 kb from the lead SNPs replicated in the present study; (2) being present in both the primary GWAS and eQTL data/GWAS for the disease; (3) having MAF ≥ 0.03 in both datasets; (4) having squared Z-test value ≥ 10 in the primary GWAS. Those SNPs that met criteria (1), (2), (3) and (4) had the lowest *P*-value in the primary GWAS and were in high LD ($r^2 > 0.8$) with the lead SNPs were used as instrumental variables to elucidate the relationship between gene expression/disease and IgG N-glycosylation (we define them as 'top SNPs'). It should be noted that SMR/HEIDI analysis does not identify a causative SNP affecting both traits. It can be either the top SNP or any other SNP in strong LD.

After defining the set of eligible SNPs for each locus, we made the 'target' and 'rejected' SNP sets and added the top SNP to the 'target' set. Then, we performed the following iterative procedure of SNP filtration: if the SNP from the eligible SNP set with the lowest P_{SMR} had $r^2 > 0.9$ with any SNP in the 'target' SNP set, it was added to the 'rejected' set; otherwise, it was added to the 'target' set. The procedure was repeated until eligible SNP set was exhausted, or the 'target' set had 20 SNPs. If we were unable to select three or more SNPs, the HEIDI test was not conducted. HEIDI statistics was calculated as $T_{HEIDI} = \sum_{i}^{m} z_{d(i)}^2$, where *m* is the number of SNPs selected for analysis,

$$z_{d(i)} = d_i \big/ SE_{(d_i)}$$

$$d_i = \beta_{SMR_i} - \beta_{SMR \, (top \, SNP)}$$

All SNP effects and SEs were standardized. Analysis was performed using the GWAS-MAP platform (60).

In the analysis of pleiotropy, we explored colocalization of regional associations of IgG N-glycosylation and selected diseases and gene expression traits in selected tissues/cell types. Similar to the previous works (12,75), the results of the SMR test were considered statistically significant if $P_{SMR} < 6.2 \times 10^{-5}$ (0.05/801, where 801 is a total number of tests corresponding to analyzed loci and gene expression/disease traits). Inference of whether a functional variant may be shared between the IgG N-glycan trait and gene expression/disease was made based on the HEIDI test: $P_{HEIDI} \geq 0.05$ (likely shared), $0.05 > P_{HEIDI} \geq 0.001$ (possibly shared) and $P_{HEIDI} < 0.001$ (sharing is unlikely).

DEPICT analysis. Gene prioritization and gene set and tissue/cell type enrichment analyses were performed using the DEPICT (25) software version 1.1, release 194 with default parameters (https://data.broadinstitute.org/mpg/depict/). Analysis was conducted for SNPs associated with any IgG N-glycosylation trait group at $P < 5.0 \times 10^{-8}$ and $P < 1.0 \times 10^{-5}$. The MHC region was excluded. The significance threshold for DEPICT analysis was set at FDR < 0.05.

Supplementary Material

Supplementary Material is available at *HMG* online.

Data and code availability

All relevant data are within the manuscript and its Supporting Information files. All computer code used in this research is available at https://github.com/Defrag1236/MV_IGG.

Acknowledgements

Funding

References

1. Vidarsson, G., Dekkers, G. and Rispens, T. (2014) IgG subclasses and allotypes: from structure to effector functions. *Front. Immunol.*, **5**, 520.

2. Clerc, F., Reiding, K.R., Jansen, B.C., Kammeijer, G.S.M., Bondt, A. and Wuhrer, M. (2016) Human plasma protein N-glycosylation. *Glycoconj. J.*, **33**, 309–343.

3. Ward, E.S. and Ghetie, V. (1995) The effector functions of immunoglobulins: implications for therapy. *Ther. Immunol.*, **2**, 77–94.

4. Gudelj, I., Lauc, G. and Pezer, M. (2018) Immunoglobulin G glycosylation in aging and diseases. *Cell. Immunol.*, **333**, 65–79.

5. Maverakis, E., Kim, K., Shimoda, M., Gershwin, M.E., Patel, F., Wilken, R., Raychaudhuri, S., Ruhaak, L.R. and Lebrilla, C.B. (2015) Glycans in the immune system and the altered glycan theory of autoimmunity: a critical review. *J. Autoimmun.*, **57**, 1–13.

6. Holland, M., Yagi, H., Takahashi, N., Kato, K., Savage, C.O.S., Goodall, D.M. and Jefferis, R. (2006) Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. *Biochim. Biophys. Acta*, **1760**, 669–677.

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

8. Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K. *et al.* (2014) Glycans are a novel biomarker of chronological and biological ages. *J. Gerontol.*, **69**, 779–789.

9. Lauc, G., Huffman, J.E., Pučić, M., Zgaga, L., Adamczyk, B., Mužinić, A., Novokmet, M., Polašek, O., Gornik, O., Krištić, J. *et al.* (2013) Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLoS Genet.*, **9**, e1003225.

10. Shen, X., Klarić, L., Sharapov, S., Mangino, M., Ning, Z., Wu, D., Trbojević-Akmačić, I., Pučić-Baković, M., Rudan, I., Polašek, O. *et al.* (2017) Multivariate discovery and replication of five novel loci associated with immunoglobulin G N-glycosylation. *Nat. Commun.*, **8**, 447.

11. Wahl, A., van den Akker, E., Klaric, L., Štambuk, J., Benedetti, E., Plomp, R., Razdorov, G., Trbojević-Akmačić, I., Deelen, J., van Heemst, D. *et al.* (2018) Genome-wide association study on immunoglobulin G glycosylation patterns. *Front. Immunol.*, **9**, 277.

12. Klarić, L., Tsepilov, Y.A., Stanton, C.M., Mangino, M., Sikka, T.T., Esko, T., Pakhomov, E., Salo, P., Deelen, J., McGurnaghan, S.J. *et al.* (2020) Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.*, **6**, eaax0301.

13. Stein, M.M., Thompson, E.E., Schoettler, N., Helling, B.A., Magnaye, K.M., Stanhope, C., Igartua, C., Morin, A., Washington, C., Nicolae, D. *et al.* (2018) A decade of research on the 17q12-21 asthma locus: piecing together the puzzle. *J. Allergy Clin. Immunol.*, **142**, 749–764.e3.

14. Inouye, M., Ripatti, S., Kettunen, J., Lyytikäinen, L.P., Oksala, N., Laurila, P.P., Kangas, A.J., Soininen, P., Savolainen, M.J., Viikari, J. *et al.* (2012) Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.*, **8**, e1002907.

15. Stephens, M. (2013) A unified framework for association analysis with multiple related phenotypes. *PLoS One*, **8**, e65245.

16. Tsepilov, Y.A., Sharapov, S.Z., Zaytseva, O.O., Krumsek, J., Prehn, C., Adamski, J., Kastenmüller, G., Wang-Sattler, R., Strauch, K., Gieger, C. *et al.* (2018) A network-based conditional genetic association analysis of the human metabolome. *Gigascience*, **7**, giy137.

17. Kraft, P., Zeggini, E. and Ioannidis, J.P.A. (2009) Replication in genome-wide association studies. *Stat. Sci.*, **24**, 561–573.

18. Pučić, M., Knežević, A., Vidič, J., Adamczyk, B., Novokmet, M., Polašek, O., Gornik, O., Šupraha-Goreta, S., Wormald, M.R., Redžic, I. *et al.* (2011) High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol. Cell. Proteomics*, **10**, M111.010090.

19. Russell, A.C., Šimurina, M., Garcia, M.T., Novokmet, M., Wang, Y., Rudan, I., Campbell, H., Lauc, G., Thomas, M.G. and Wang, W. (2017) The N-glycosylation of immunoglobulin G as a novel biomarker of Parkinson's disease. *Glycobiology*, **27**, 501–510.

20. Varki, A. and Sharon, N. (2009) Historical background and overview. In Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W. and Etzler, M.E. (eds), *Essentials of Glycobiology*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

21. Doherty, M., Theodoratou, E., Walsh, I., Adamczyk, B., Stöckmann, H., Agakov, F., Timofeeva, M., Trbojević-Akmačić, I., Vučković, F., Duffy, F. *et al.* (2018) Plasma N-glycans in colorectal cancer risk. *Sci. Rep.*, **8**, 8655.

22. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P. and Cunningham, F. (2016) The Ensembl Variant Effect Predictor. *Genome Biol.*, **17**, 122.

23. Machiela, M.J. and Chanock, S.J. (2015) LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*, **31**, 3555–3557.

24. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M. *et al.* (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.*, **48**, 481–487.

25. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T. *et al.* (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.*, **6**, 5890.

26. Huffman, J.E., Knežević, A., Vitart, V., Kattla, J., Adamczyk, B., Novokmet, M., Igl, W., Pučić, M., Zgaga, L., Johannson, Å. *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.

27. Sharapov, S.Z., Shadrina, A.S., Tsepilov, Y.A., Elgaeva, E.E., Tiys, E.S., Feoktistova, S.G., Zaytseva, O.O., Vuckovic, F., Cuadrat, R., Jäger, S. *et al.* (2021) Replication of fifteen loci involved in human plasma protein N-glycosylation in 4,802 samples from four cohorts. *Glycobiology.*, **31**, 82–88.

28. Roxrud, I., Raiborg, C., Gilfillan, G.D., Strømme, P. and Stenmark, H. (2009) Dual degradation mechanisms ensure disposal of NHE6 mutant protein associated with neurological disease. *Exp. Cell Res.*, **315**, 3014–3027.

29. Von Bülow, G.U. and Bram, R.J. (1997) NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. *Science*, **278**, 138–141.

30. Ozcan, E., Garibyan, L., Lee, J.J.Y., Bram, R.J., Lam, K.P. and Geha, R.S. (2009) Transmembrane activator, calcium modulator, and cyclophilin ligand interactor drives plasma cell differentiation in LPS-activated B cells. *J. Allergy Clin. Immunol.*, **123**, 1277–1286.e5.

31. Castigli, E., Wilson, S.A., Elkhal, A., Ozcan, E., Garibyan, L. and Geha, R.S. (2007) Transmembrane activator and calcium modulator and cyclophilin ligand interactor enhances CD40-driven plasma cell differentiation. *J. Allergy Clin. Immunol.*, **120**, 885–891.

32. Castigli, E., Wilson, S., Garibyan, L., Rachid, R., Bonilla, F., Schneider, L., Morra, M., Curran, J. and Geha, R. (2007) Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat. Genet.*, **39**, 430–431.

33. Janzi, M., Melén, E., Kull, I., Wickman, M. and Hammarström, L. (2012) Rare mutations in TNFRSF13B increase the risk of asthma symptoms in Swedish children. *Genes Immun.*, **13**, 59–65.

34. Liao, M., Ye, F., Zhang, B., Huang, L., Xiao, Q., Qin, M., Mo, L., Tan, A., Gao, Y., Lu, Z. *et al.* (2012) Genome-wide association study identifies common variants at TNFRSF13B associated with IgG level in a healthy Chinese male population. *Genes Immun.*, **13**, 509–513.

35. Jonsson, S., Sveinbjornsson, G., De Lapuente Portilla, A.L., Swaminathan, B., Plomp, R., Dekkers, G., Ajore, R., Ali, M., Bentlage, A.E.H., Elmér, E. *et al.* (2017) Identification of sequence variants influencing immunoglobulin levels. *Nat. Genet.*, **49**, 1182–1191.

36. Ramachandran, S., Chahwan, R., Nepal, R.M., Frieder, D., Panier, S., Roa, S., Zaheen, A., Durocher, D., Scharff, M.D. and Martin, A. (2010) The RNF8/RNF168 ubiquitin ligase cascade facilitates class switch recombination. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 809–814.

37. Stewart, G.S., Panier, S., Townsend, K., Al-Hakim, A.K., Kolas, N.K., Miller, E.S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M. *et al.* (2009) The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell*, **136**, 420–434.

38. Marenholz, I., Esparza-Gordillo, J., Rüschendorf, F., Bauerfeind, A., Strachan, D.P., Spycher, B.D., Baurecht, H., Margaritte-Jeannin, P., Sääf, A., Kerkhof, M. *et al.* (2015) Meta-analysis identifies seven susceptibility loci involved in the atopic march. *Nat. Commun.*, **6**, 8804.

39. Johansson, Å., Rask-Andersen, M., Karlsson, T. and Ek, W.E. (2019) Genome-wide association analysis of 350 000 Caucasians from the UK Biobank identifies novel loci for asthma, hay fever and eczema. *Hum. Mol. Genet.*, **28**, 4022–4041.

40. Paternoster, L., Standl, M., Chen, C.M., Ramasamy, A., Bøpnnelykke, K., Duijts, L., Ferreira, M.A., Alves, A.C., Thyssen, J.P., Albrecht, E. *et al.* (2012) Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. *Nat. Genet.*, **44**, 187–192.

41. Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., Curtin, J.A., Bønnelykke, K., Tian, C., Takahashi, A. *et al.* (2015) Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nat. Genet.*, **47**, 1449–1456.

42. Christou, M.A., Ntritsos, G., Markozannes, G., Koskeridis, F., Nikas, S.N., Karasik, D., Kiel, D.P., Evangelou, E. and Ntzani, E.E. (2020) A genome-wide scan for pleiotropy between bone mineral density and nonbone phenotypes. *Bone Res.*, **8**, 1–9.

43. Reese, S.E., Xu, C.J., den Dekker, H.T., Lee, M.K., Sikdar, S., Ruiz-Arenas, C., Merid, S.K., Rezwan, F.I., Page, C.M., Ullemar, V. *et al.* (2019) Epigenome-wide meta-analysis of DNA methylation and childhood asthma. *J. Allergy Clin. Immunol.*, **143**, 2062–2074.

44. Hirst, J., D. Barlow, L., Francisco, G.C., Sahlender, D.A., Seaman, M.N.J., Dacks, J.B. and Robinson, M.S. (2011) The fifth adaptor protein complex. *PLoS Biol.*, **9**, e1001170.

45. Hirst, J., Itzhak, D.N., Antrobus, R., Borner, G.H.H. and Robinson, M.S. (2018) Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. *PLoS Biol.*, **16**, e2004411.

46. Nair, M., Teng, A., Bilanchone, V., Agrawal, A., Li, B. and Dai, X. (2006) Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. *J. Cell Biol.*, **173**, 253–264.

47. Li, B., Nair, M., Mackay, D.R., Bilanchone, V., Hu, M., Fallahi, M., Song, H., Dai, Q., Cohen, P.E. and Dai, X. (2005) Ovol1 regulates meiotic pachytene progression during spermatogenesis by repressing Id2 expression. *Development*, **132**, 1463–1473.

48. Dai, X., Schonbaum, C., Degenstein, L., Bai, W., Mahowald, A. and Fuchs, E. (1998) The ovo gene required for cuticle formation and oogenesis in flies is involved in hair formation and spermatogenesis in mice. *Genes Dev.*, **12**, 3452–3463.

49. Tsuji, G., Ito, T., Chiba, T., Mitoma, C., Nakahara, T., Uchi, H. and Furue, M. (2018) The role of the OVOL1–OVOL2 axis in normal and diseased human skin. *J. Dermatol. Sci.*, **90**, 227–231.

50. Hucthagowder, V., Sausgruber, N., Kim, K.H., Angle, B., Marmorstein, L.Y. and Urban, Z. (2006) Fibulin-4: a novel gene for an autosomal recessive cutis laxa syndrome. *Am. J. Hum. Genet.*, **78**, 1075–1080.

51. Hoyer, J., Kraus, C., Hammersen, G., Geppert, J.P. and Rauch, A. (2009) Lethal cutis laxa with contractural arachnodactyly, overgrowth and soft tissue bleeding due to a novel homozygous fibulin-4 gene mutation. *Clin. Genet.*, **76**, 276–281.

52. Nakayama, T. and Kimura, M.Y. (2010) Memory Th1/Th2 cell generation controlled by Schnurri-2. *Adv. Exp. Med. Biol.*, **684**, 1–10.

53. Staton, T.L., Lazarevic, V., Jones, D.C., Lanser, A.J., Takagi, T., Ishii, S. and Glimcher, L.H. (2011) Dampening of death pathways by schnurri-2 is essential for T-cell development. *Nature*, **472**, 105–110.

54. Rosnet, O., Blanco-Betancourt, C., Grivel, K., Richter, K. and Schiff, C. (2004) Binding of free immunoglobulin light chains to VpreB3 inhibits their maturation and secretion in chicken B cells. *J. Biol. Chem.*, **279**, 10228–10236.

55. Szabo, S.J., Kim, S.T., Costa, G.L., Zhang, X., Fathman, C.G. and Glimcher, L.H. (2000) A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell*, **100**, 655–669.

56. Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L. and Mil-carek, C. (2009) Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat. Immunol.*, **10**, 1102–1109.

57. Klarić, L., Tsepilov, Y.A., Aulchenko, Y.S., Lauc, G. and Hayward, C. (2019) GWAS summary statistics for UPLC IgG N-glycosylation traits [dataset]. MRC Human Genetics Unit. Institute of Genetics and Molecular Medicine. University of Edinburgh. doi:10.7488/ds/2481.

58. Kerr, S.M., Klaric, L., Halachev, M., Hayward, C., Boutin, T.S., Meynert, A.M., Semple, C.A., Tuiskula, A.M., Swan, H., Santoyo-Lopez, J. *et al.* (2019) An actionable KCNH2 long QT syndrome variant detected by sequence and haplotype analysis in a population research cohort. *Sci. Rep.*, **9**, 10964.

59. Rudan, I., Marušić, A., Janković, S., Rotim, K., Boban, M., Lauc, G., Grković, I., Dogaš, Z., Zemunik, T., Vatavuk, Z. *et al.* (2009) 10 001 Dalmatians:' Croatia launches its National Biobank. *Croat. Med. J.*, **50**, 4–6.

60. Shashkova, T.I., Gorev, D.D., Pakhomov, E.D., Shadrina, A.S., Sharapov, S.Z., Tsepilov, Y.A., Karssen, L.C. and Aulchenko, Y.S. (2020) The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits. *Vavilovskii Zhurnal Genet. Selektsii*, **24**, 876–884.

61. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.

62. Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299.

63. Ning, Z., Tsepilov, Y.A., Sharapov, S.Z., Grishenko, A.K., Feng, X., Shirali, M., Joshi, P.K., Wilson, J.F., Pawitan, Y., Haley, C.S. *et al.* (2015) Beyond power: multivariate discovery, replication, and interpretation of pleiotropic loci using summary association statistics. *bioRxiv*. doi: 10.1101/022269.

64. Tsepilov, Y.A., Freidin, M.B., Shadrina, A.S., Sharapov, S.Z., Elgaeva, E.E., Zundert, J., Karssen, L.C., Suri, P., Williams, F.M.K. and Aulchenko, Y.S. (2020) Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun. Biol.*, **3**, 329.

65. McDaid, A.F., Joshi, P.K., Porcu, E., Komljenovic, A., Li, H., Sorrentino, V., Litovchenko, M., Bevers, R.P.J., Ruëger, S., Reymond, A. *et al.* (2017) Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nat. Commun.*, **8**, 1–11.

66. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. and Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, **4**, 7.

67. Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E. *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–1243.

68. Momozawa, Y., Dmitrieva, J., Théâtre, E., Deffontaine, V., Rahmouni, S., Charloteaux, B., Crins, F., Docampo, E., Elansary, M., Gori, A.-S. *et al.* (2018) IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.*, **9**, 2427.

69. GTEx, Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.

70. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. *et al.* (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, **12**, e1001779.

71. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S. *et al.* (2013) Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, **506**, 376–381.

72. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C. *et al.* (2015) International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.*, **6**, 8019.

73. Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S. *et al.* (2013) Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.*, **45**, 730–738.

74. Liu, J.Z., Van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T. *et al.* (2015) Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.*, **47**, 979–986.

75. Sharapov, S.Z., Tsepilov, Y.A., Klaric, L., Mangino, M., Thareja, G., Shadrina, A.S., Simurina, M., Dagostino, C., Dmitrieva, J., Vilaj, M. *et al.* (2019) Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum. Mol. Genet.*, **28**, 2062–2077.