Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases

IMMUNOLOGY

Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases

Lucija Klarić[1,2*†], Yakov A. Tsepilov[3,4†], Chloe M. Stanton[2†], Massimo Mangino[5,6], Timo Tõnis Sikka[7,8], Tõnu Esko[7,9,10], Eugene Pakhomov[3,4], Perttu Salo[11], Joris Deelen[12,13], Stuart J. McGurnaghan[2], Toma Keser[14], Frano Vučković[1], Ivo Ugrina[1,15], Jasminka Krištić[1], Ivan Gudelj[1], Jerko Štambuk[1], Rosina Plomp[16], Maja Pučić-Baković[1], Tamara Pavić[14], Marija Vilaj[1], Irena Trbojević-Akmačić[1], Camilla Drake[2], Paula Dobrinić[17], Jelena Mlinarec[17], Barbara Jelušić[17], Anne Richmond[2], Maria Timofeeva[18], Alexander K. Grishchenko[3,4], Julia Dmitrieva[19], Mairead L. Bermingham[2], Sodbo Zh. Sharapov[3], Susan M. Farrington[18], Evropi Theodoratou[20,21], Hae-Won Uh[16,22], Marian Beekman[12], Eline P. Slagboom[12], Edouard Louis[23], Michel Georges[19], Manfred Wuhrer[16], Helen M. Colhoun[2,24], Malcolm G. Dunlop[18], Markus Perola[11], Krista Fischer[7], Ozren Polasek[25,26,27], Harry Campbell[20], Igor Rudan[20], James F. Wilson[2,20], Vlatka Zoldoš[17], Veronique Vitart[2], Tim Spector[5], Yurii S. Aulchenko[3,28,29‡], Gordan Lauc[1,14‡], Caroline Hayward[2,30‡]

Effector functions of immunoglobulin G (IgG) are regulated by the composition of a glycan moiety, thus affecting activity of the immune system. Aberrant glycosylation of IgG has been observed in many diseases, but little is understood about the underlying mechanisms. We performed a genome-wide association study of IgG N-glycosylation (*N* = 8090) and, using a data-driven network approach, suggested how associated loci form a functional network. We confirmed in vitro that knockdown of *IKZF1* decreases the expression of fucosyltransferase FUT8, resulting in increased levels of fucosylated glycans, and suggest that RUNX1 and RUNX3, together with SMARCB1, regulate expression of glycosyltransferase MGAT3. We also show that variants affecting the expression of genes involved in the regulation of glycoenzymes colocalize with variants affecting risk for inflammatory diseases. This study provides new evidence that variation in key transcription factors coupled with regulatory variation in glycogenes modifies IgG glycosylation and has influence on inflammatory diseases.

INTRODUCTION

Glycosylation, a series of reactions that creates complex carbohydrate structures (glycans) attached to a polypeptide backbone, is among the most common and complex posttranslational protein modifications. Highly regulated attachment of different glycans to the same glycosylation site (alternative glycosylation) can greatly contribute to variability in glycoprotein structure and influence function in a way that is analogous to changes in protein sequence (*1*). Immunoglobulin G (IgG) is a simple glycoprotein with only one conserved N-glycosylation site per heavy chain and glycans that have no more than two antennae, but even this simple glycosylation has the capacity to generate hundreds of different glycoforms of IgG. It is well established that alternative glycosylation of IgG can act as a molecular switch between different immune response outcomes and thus significantly affect the function of the immune system (*2*). Aberrant protein glycosylation has been observed in many physiological states,

[1]Genos Glycoscience Research Laboratory, Zagreb, Croatia. [2]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [3]Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science, 630090 Novosibirsk, Russia. [4]Novosibirsk State University, 630090 Novosibirsk, Russia. [5]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [6]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK. [7]Estonian Genome Center, University of Tartu, Tartu, Estonia. [8]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [9]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. [10]Division of Endocrinology, Boston Children's Hospital, Cambridge, MA, USA. [11]Genomics and Biomarkers Unit, Department of Health, National Institute for Health and Welfare (THL), Helsinki, Finland. [12]Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Centre, Leiden, Netherlands. [13]Max Planck Institute for Biology of Ageing, Cologne, Germany. [14]Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia. [15]University of Split, Faculty of Science, Split, Croatia. [16]Leiden University Medical Centre, Leiden, Netherlands. [17]Division of Molecular Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia. [18]Colon Cancer Genetics Group, Cancer Research UK Edinburgh

from aging- and age-related diseases to autoimmune diseases and cancer (*3*).

Glycans are synthesized in the endoplasmic reticulum and the Golgi apparatus by a complex interplay of glycosyltransferases—enzymes that add monosaccharides to a growing glycan chain—and various other enzymes and transporters involved in synthesis and delivery of donors and substrates needed for chemical reactions (*4*). However, our understanding of the mechanisms by which alternative glycosylation is achieved and regulated is very limited.

N-linked glycans attached to a protein can be released, and their abundance can be quantified using a range of analytical methods. These measurements represent a set of quantitative traits characterizing the glycome of the studied protein. As with any quantitative trait, the genetic determinants of the glycome can be studied by associating levels of glycans to polymorphic sequence variants measured in large cohorts of individuals [genome-wide association studies (GWAS)]. The first GWAS of the glycome identified associated variants in or near 7 glycosyltransferase genes and in 15 additional loci with no apparent role in protein glycosylation (*5*, *6*). The list was recently extended by six additional loci (*7*, *8*). However, there is still limited understanding of how these genes are functionally related or what their role is in the genetic regulation of IgG N-glycosylation in health and disease.

To address these questions, we performed the largest GWAS of the total IgG *N*-glycome to date, on 8090 samples from individuals of European ancestry, and more than doubled the number of associated loci. To prioritize genes with a plausible role in IgG glycosylation, we assessed whether variants were located in the coding regions of genes, showed pleiotropy with gene expression in biologically relevant cell types, and assessed enrichment in gene sets from different biological pathways. We explored how these genes are connected in a functional network and confirmed some of the network connections with in vivo functional follow-up. Last, we investigated pleiotropy with other complex traits and diseases by interrogating the overlap of glycosylation associations with susceptibility loci for other phenotypes. Where available, we also assessed whether IgG N-glycosylation, diseases, and gene expression are likely to be controlled by the same underlying causal variant using summary-level Mendelian randomization (SMR). This strategy not only resulted in the discovery of new candidate genes but also suggested how some of these genes could regulate glycosylation enzymes and how they could influence the aberrant glycosylation observed in diseases with an inflammatory signature.

RESULTS

Discovery and replication meta-analyses

We performed a discovery IgG N-glycosylation GWAS on four cohorts of European descent ($N = 8090$). Associations of 77 ultraperformance liquid chromatography (UPLC) IgG *N*-glycan traits with the suggestive association near *SLC38A10* from Lauc *et al.*), and 1 confirmed those from Wahl *et al.* (*8*) (Fig. 1 and table S1), while 14 loci were not previously associated with IgG glycosylation. Individual associations and Manhattan plots can be found in the online resource available at https://shiny.igmm.ed.ac.uk/igg_glycans_gwas/. For 19 of 27 significant loci (9 of 14 previously unassociated), the same single-nucleotide polymorphism (SNP)–glycan association was replicated in a meta-analysis of four independent European cohorts ($N = 2388$) with $P \leq 0.05/27 \approx 1.9 \times 10^{-3}$. None of the genes in novel loci have a known role in glycosylation. For all loci, the direction of effect estimates obtained for the UPLC data (effect either increasing or decreasing with the same allele) was the same in the discovery and the replication parts of the study (Table 1).

To further validate our findings, we analyzed all significantly associated SNPs in a cohort where glycans were measured with liquid chromatography coupled to electrospray mass spectrometry (LCMS; $N = 1842$). From the 27 genome-wide significant SNPs, 17 were associated with at least one of the LCMS-measured IgG glycopeptide levels at Bonferroni-corrected $P \leq 3.7 \times 10^{-5}$. Of these 17, 14 loci were also replicated in the UPLC replication study (table S2).

We performed an approximate conditional and joint analysis using GCTA-COJO (genome-wide complex trait conditional and joint analysis) software (*9*) to investigate secondary associations in the 27 loci. We found evidence of multiple SNPs independently contributing to glycan level variation for four loci. Three of these loci span glycosyltransferase genes, coding for enzymes directly involved in biosynthesis of glycans (*4*) (table S3). The greatest number of independently associated SNPs (six) was observed for the fucosyltransferase locus, *FUT8*, followed by three SNPs in the galactosyltransferase locus, *B4GALT1*, and two SNPs in the sialyltransferase gene, *ST6GAL1*. Two SNPs in the fourth locus, spanning the human leukocyte antigen (HLA) region on chromosome 6, are likely to be due to the complexity of this region, where multiple antigen-presenting genes are in close proximity and in high linkage disequilibrium (LD).

Using only the marginal effects of top SNPs, we were able to explain up to 20% of the genetic variance (for a single glycan trait). Uncovering additional variants in the conditional analysis resulted in a maximum of 22% variance explained. These maxima were reached for IGP29, the percentage of monosialylation of fucosylated digalactosylated structures without bisecting *N*-acetylglucosamine, GlcNAc (table S3).

Prioritizing genes associated with IgG N-glycosylation

We applied the following strategy to prioritize the most likely functional genes. The Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) framework (*10*) was used to perform a gene set and tissue/cell type enrichment analysis and to provide additional evidence for gene prioritization. Next, for all candidate genes in every locus, we explored evidence for genes containing SNPs affecting amino acid sequence using the Variant Effect Predictor

Table 1. Loci associated with IgG glycosylation. Each locus is represented by the SNP with the strongest association in the region. Locus, coded by "chromosome: locus start–locus end" (GRCh37); No. of SNPs, maximum number of SNPs independently contributing to trait variation; SNP, variant with the strongest association in the locus; position, position of the SNP on the NCBI (National Center for Biotechnology Information) build 37; EA, allele for which effect estimate is reported; OA, other allele; EAF, frequency of the effect allele; R^2, sample size weighted average of imputation quality for the SNP with the strongest association in the locus; glycan, glycan associated with the reported SNP; No. of glycans, number of other glycans suggestively or significantly associated with variants in given locus; β, effect estimate for the SNP and glycan with the strongest association in the locus in the discovery; *P*, *P* value for the discovery effect estimate; β UPLC repl, effect estimate for the SNP and glycan with the strongest association in the locus in replication; *P* UPLC repl, *P* value for the effect estimate in replication. The loci replicated in UPLC replication at $P \leq 0.0019$ are in bold. The loci from Lauc *et al.* (*6*), Shen *et al.* (*7*), and Wahl *et al.* (*8*) are in italics.

Locus	Candidate genes	No. of SNPs	SNP	EA	OA	EAF	R^2	Glycan	No. of glycans	β	*P*	β UPLC repl	*P* UPLC repl
Genome-wide significant													
1:25226001-25345011	***RUNX3***	***1***	***rs10903118***	***T***	***C***	***0.484***	***0.886***	***IGP74***	***13***	***−0.121***	$\boldsymbol{5.14\times10^{-13}}$	***−0.097***	$\boldsymbol{7.93\times10^{-4}}$
3:186712711-186738421	***ST6GAL1***	***2***	***rs7621161***	***A***	***C***	***0.284***	***0.937***	***IGP29***	***27***	***−0.653***	$\boldsymbol{4.65\times10^{-276}}$	***−0.682***	$\boldsymbol{1.55\times10^{-111}}$
5:95211647-95347786	*ELL2*	*1*	*rs7700895*	*A*	*T*	*0.215*	*0.997*	*IGP35*	*4*	*0.155*	1.20×10^{-14}	*0.097*	4.61×10^{-3}
5:131026218-131833599	***IRF1-SLC22A4***	**1**	**rs11748193**	**A**	**T**	**0.399**	**0.987**	**IGP2**	**2**	**0.110**	$\mathbf{4.31\times10^{-10}}$	**0.116**	$\mathbf{1.31\times10^{-4}}$
6:29299390-33883424	*HLA* region	*2*	*rs3099844*	*A*	*C*	*0.136*	*0.995*	*IGP15*	*13*	*−0.227*	1.12×10^{-13}	*−0.083*	2.28×10^{-1}
6:139617590-139636003	***TXLNB***	**1**	**rs9385856**	**T**	**C**	**0.574**	**0.977**	**IGP70**	**24**	**0.151**	$\mathbf{5.05\times10^{-19}}$	**0.111**	$\mathbf{1.54\times10^{-4}}$
6:143150223-143203591	***HIVEP2***	**1**	**rs7758383**	**A**	**G**	**0.521**	**0.952**	**IGP13**	**3**	**0.125**	$\mathbf{9.61\times10^{-14}}$	**0.225**	$\mathbf{1.12\times10^{-14}}$
7:6520676-6537913	*DAGLB*	1	rs6964421	T	C	0.328	0.957	IGP14	7	0.117	5.31×10^{-11}	0.000	9.95×10^{-1}
7:50325717-50361683	***IKZF1***	***1***	***rs6421315***	***C***	***G***	***0.383***	***0.938***	***IGP62***	***37***	***0.192***	$\boldsymbol{4.70\times10^{-27}}$	***0.133***	$\boldsymbol{8.97\times10^{-6}}$
7:150906453-150969535	***ABCF2***	***1***	***rs7812088***	***A***	***G***	***0.119***	***0.981***	***IGP2***	***16***	***−0.250***	$\boldsymbol{2.06\times10^{-22}}$	***−0.269***	$\boldsymbol{3.83\times10^{-9}}$
8:103538266-103550211	*ODF1*	1	rs10096810	A	G	0.627	1.000	IGP77	4	−0.110	9.52×10^{-11}	−0.074	1.37×10^{-2}
9:33041761-33186080	***B4GALT1***	***3***	***rs10813951***	***A***	***G***	***0.732***	***0.967***	***IGP17***	***30***	***0.230***	$\boldsymbol{8.84\times10^{-34}}$	***0.156***	$\boldsymbol{4.57\times10^{-6}}$
9:33205136-33375592	***SPINK4***	**1**	**rs12341905**	**A**	**G**	**0.896**	**0.964**	**IGP53**	**4**	**0.168**	$\mathbf{1.46\times10^{-09}}$	**0.199**	$\mathbf{4.99\times10^{-5}}$
11:114323627-114450529	***NXPE1-NXPE1***	**1**	**rs481080**	**A**	**G**	**0.475**	**0.999**	**IGP29**	**3**	**−0.140**	$\mathbf{1.05\times10^{-16}}$	**−0.099**	$\mathbf{5.28\times10^{-4}}$
14:65472891-66284991	***FUT8***	***6***	***rs11847263***	***T***	***G***	***0.647***	***0.985***	***IGP42***	***16***	***−0.283***	$\boldsymbol{1.13\times10^{-58}}$	***−0.252***	$\boldsymbol{5.93\times10^{-17}}$
14:105966019-106002352	***TMEM121***	***1***	***rs4074453***	***T***	***C***	***0.758***	***0.996***	***IGP48***	***6***	***−0.218***	$\boldsymbol{3.82\times10^{-29}}$	***−0.299***	$\boldsymbol{1.25\times10^{-13}}$
16:23397113-23613191	*GGA2-COG7*	1	rs250555	T	C	0.860	0.961	IGP26	4	−0.155	6.76×10^{-10}	−0.044	2.24×10^{-1}
17:37903731-38112190	***ORMDL3-GSDMB-IKZF3-ZPBP2***	**1**	**rs7216389**	**T**	**C**	**0.487**	**0.999**	**IGP59**	**11**	**0.137**	$\mathbf{1.17\times10^{-15}}$	**0.094**	$\mathbf{1.53\times10^{-3}}$
17:43463492-	***CRHR1-SPPL2C-***												

Locus	Candidate genes	No. of SNPs	SNP	EA	OA	EAF	R^2	Glycan	No. of glycans	β	*P*	β UPLC repl	*P* UPLC repl
19:5822316-5845974	*FUT6*	*1*	*rs874232*	*T*	*C*	*0.565*	*0.971*	*IGP12*	*5*	*0.119*	7.85×10^{-13}	*0.070*	2.08×10^{-2}
19:19260586-19296217	*RFXANK*	1	rs7257072	T	C	0.515	0.997	IGP9	12	−0.122	1.59×10^{-13}	−0.068	1.97×10^{-2}
20:17818141-17833534	*MGME1*	1	rs2745851	A	G	0.376	0.982	IGP38	10	−0.125	4.61×10^{-13}	−0.094	8.82×10^{-3}
21:36546756-36665202	***RUNX1***	**1**	**rs7281587**	**A**	**G**	**0.247**	**0.998**	**IGP45**	**19**	**−0.144**	$\mathbf{1.13 \times 10^{-13}}$	**−0.171**	$\mathbf{7.51 \times 10^{-7}}$
22:24093789-24182500	***SMARCB1-DELR3-CHCHD10-VPREB3***	***1***	***rs17630758***	***A***	***G***	***0.152***	***0.991***	***IGP66***	***25***	***−0.308***	$\boldsymbol{1.20 \times 10^{-41}}$	***−0.386***	$\boldsymbol{4.94 \times 10^{-23}}$
22:39774448-39860868	***MGAT3***	***1***	***rs5750830***	***A***	***C***	***0.740***	***0.965***	***IGP40***	***28***	***−0.344***	$\boldsymbol{7.74 \times 10^{-69}}$	***−0.363***	$\boldsymbol{5.26 \times 10^{-29}}$
Suggestive													
1:246854862-246963137	LINC01341	1	rs3795464	A	G	0.383	0.984	IGP73	2	−0.099	1.01×10^{-08}		
2:100636757-100805273	AC092667.1	1	rs2309748	A	T	0.647	0.993	IGP34	2	−0.105	3.38×10^{-09}		
17:16820099-16875636	TBC1D27-TNFRSF13B	1	rs4561508	T	C	0.105	0.975	IGP9	2	0.163	5.70×10^{-09}		
17:56398006-56410041	MIR142	1	rs2526378	A	G	0.549	0.935	IGP31	1	−0.100	4.44×10^{-09}		
19:1614910-1658699	TCF3	1	rs4807942	T	C	0.878	0.661	IGP68	1	0.154	1.54×10^{-08}		
20:52170177-52212273	ZNF217	1	rs1555926	T	C	0.789	0.992	IGP11	3	−0.118	6.96×10^{-09}		

traits or due to the variant being in LD with two independent causal variants, each exhibiting independent effects on each trait. More specifically, a pleiotropic scenario would assume that the same unobserved variant is responsible for the association with both gene expression and glycan levels, while an LD scenario would assume that coassociation is due to two or more underlying variants that are in LD and independently contribute to either gene expression or glycan levels.

The DEPICT gene prioritization tool provided evidence of prioritization for 18 genes in 14 loci at false discovery rate (FDR) <0.05 and another 3 genes at FDR <0.2 (Supplementary Note, Appendix Table 9). In four genes, associated variants were predicted to result in a potentially deleterious amino acid change, and in five genes, probably benign amino acid changes. SMR/HEIDI indicated that IgG N-glycosylation–associated variants in 9 loci had pleiotropic effects on expression of 13 genes in different tissues, including B cells (4 genes in 3 loci), the cell lineage responsible for IgG synthesis. For processes: glycosylation (23 gene sets), immune system processes (22 gene sets), and transcription (7 gene sets) (table S4). DEPICT tests for enrichment in preconstructed gene sets that are based on information from protein-protein interactions (PPIs), molecular and biochemical pathways, gene coexpression, and gene sets based on mouse gene knockout studies. No gene sets were enriched at FDR <0.05, but 397 gene sets were enriched at FDR <0.2. These gene sets are connected to various processes involved in immunity, B cell life cycle, and antibody production and quantity (table S5).

Functional network of loci associated with IgG N-glycosylation

Next, we investigated how genes in associated loci are functionally connected with each other. We assume that two loci with a similar glycome-wide effect (effect of the top SNP in the locus on all glycan traits) are likely to have a similar role in the regulation of glycosylation and may be a part of the same biological pathway. Our assumption

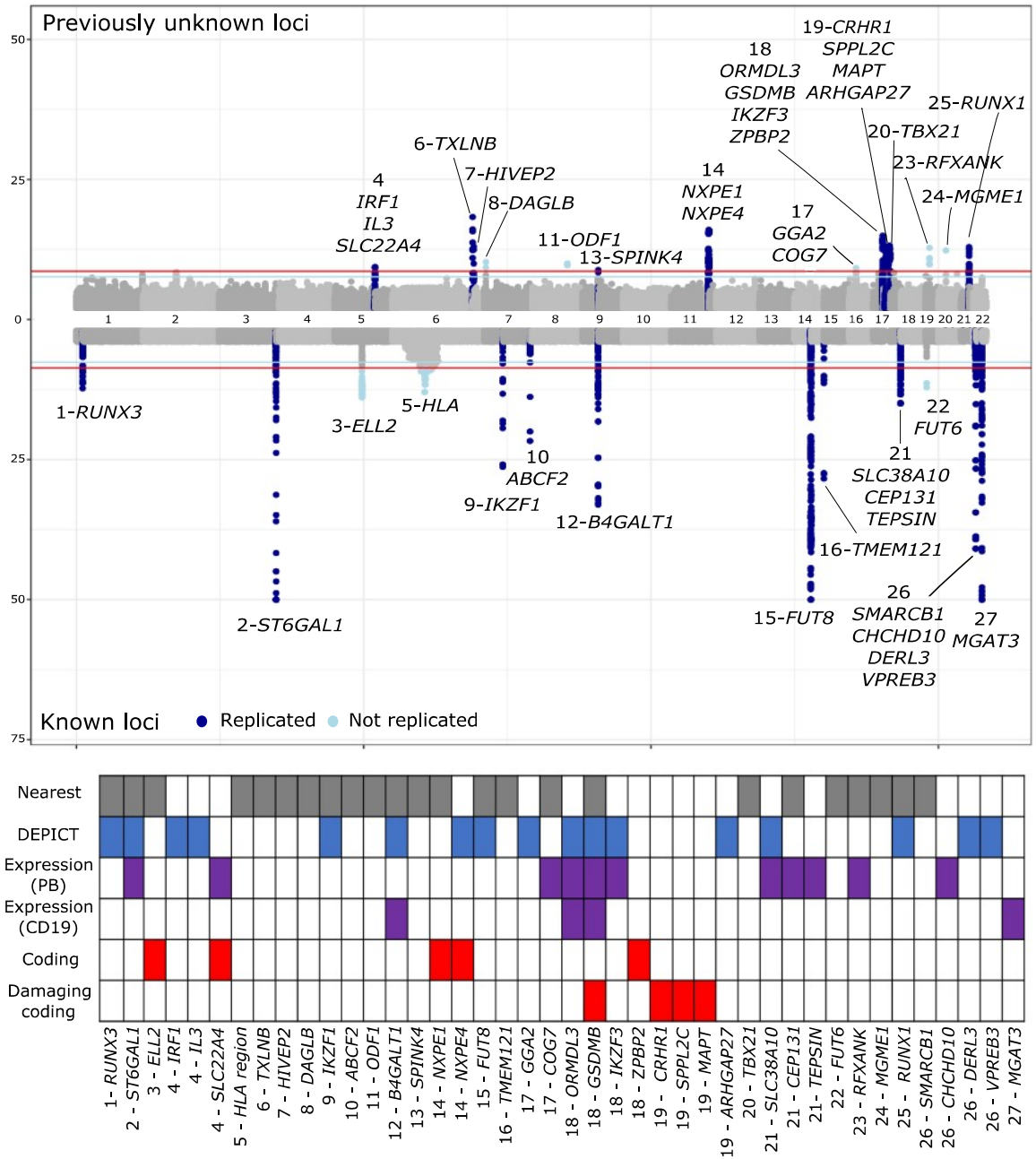


Fig. 1. Gene prioritization in loci associated with IgG N-glycosylation. The Manhattan plot was created by taking the lowest *P* value at every genomic position from all 77 GWAS. For simplicity, the plot was trimmed at the equivalent of $P = 10^{-50}$. The lowest observed *P* value in this analysis was 4.65×10^{-276} at *ST6GAL1*. Known loci, loci detected in previous IgG N-glycosylation GWAS; replicated, UPLC replication GWAS. Bottom: Summary of support for prioritization of every gene in the Manhattan plot. DEPICT, genes from enriched gene sets; expression, genes whose expression is pleiotropic with IgG N-glycosylation; CD19, B cells; PB, peripheral blood; coding, genes for which IgG N-glycosylation–associated SNP results in a changed amino acid sequence.

Given that the direction of the GWAS estimates depends on the allele used as a reference, the sign of correlation is not informative in this analysis. Therefore, we report absolute values of correlation coefficients. Although nodes in this network represent effects of top SNPs, we marked nodes by using the names of candidate genes in each locus, and we used these names to denote loci throughout the text. Below, we focus our discussion only on clusters that contain glycosyltransferases (*MGAT3*, *FUT8*, *B4GALT1*, and *ST6GAL1*), because these enzymes catalyze the transfer of monosaccharides to a growing glycan chain and therefore have a known and well-established role in the biosynthesis of glycans. As outlined above, we hypothesize that any locus with a similar glycome-wide effect to a glycosyltransferase locus is likely to either regulate the expression levels of the enzyme or modulate its activity through other indirect effects.

To validate the functional network, we applied both computational and experimental approaches. We first pruned the correlation network for significant correlations [$P \leq 0.05/(27 \times 26)/2 \approx 1.4 \times 10^{-4}$] and performed permutation analysis with 100,000 random SNPs. We also compared our network with the STRING human PPI database (*16*).

The strongest overall correlation, a Spearman's ρ of 0.97, was observed between IgG–*N*-glycome–wide effects of the top SNP in the *MGAT3* locus and the top SNP in the *SMARCB1-DELR3-CHCHD10-VPREB3* locus (Fig. 2A). None of the randomly selected SNPs exhibited such a strong correlation with these two loci in the permutation analysis (Table 2 and table S6). Recent work by Sharapov *et al.* (*17*) on genetics of the total plasma proteome N-glycosylation, of which IgG is a major part, suggested that the likely candidate gene is *DERL3*. *DERL3* encodes a protein involved in endoplasmic reticulum–associated degradation for misfolded luminal glycoproteins. Data provided in this study shift the balance of evidence toward the *SMARCB1*, as it is unlikely that modulation of glycoprotein degradation would result in effects similar to those observed when perturbing MGAT3.

The second strongest association, a Spearman's ρ of 0.94, was between lead SNPs in TFs *RUNX3* and *RUNX1* loci. This link was validated in the permutation analysis and observed in the STRING PPI network. *RUNX3* and *RUNX1* loci were also strongly correlated with both the *MGAT3* and *SMARCB1-DELR3-CHCHD10-VPREB3* loci (Table 2). A proxy for the top SNP in the *MGAT3* locus, SNP rs8137426 (LD $R^2 = 1$ with rs5750830), is located in a region bound by TF RUNX3, although not directly within its binding motif (fig. S2A). In our SMR/HEIDI analysis, we found evidence of the same association having a pleiotropic effect on both *MGAT3* expression in $CD19^+$ B cells and IGP40, a glycan trait related with bisecting GlcNAc (https://shiny.igmm.ed.ac.uk/igg_glycans_gwas/). This suggests that the associated SNP could influence binding of TFs (RUNX1 or RUNX3), which, together with the chromatin remodeling protein SMARCB1, regulate expression of *MGAT3*, resulting in an increased incidence of bisecting GlcNAc in all fucosylated

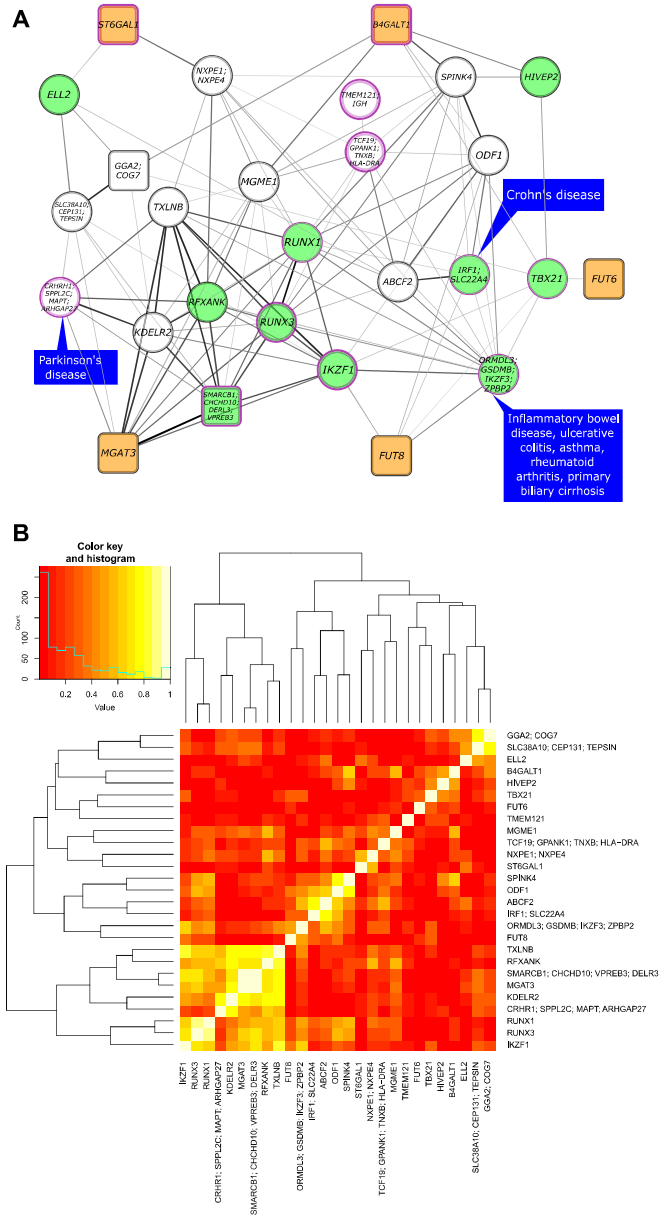


Fig. 2. Functional network of loci associated with IgG N-glycosylation. Correlation estimates are computed on the basis of squared pairwise Spearman's correlation of SNP effects. The loci are denoted with names of genes that were prioritized in regions tagged by the given lead SNP. (**A**) Functional network of loci associated with IgG N-glycosylation. In this network, each node represents a lead SNP in the locus, and each edge represents the squared correlation of glycome-wide effects of the two nodes. Only significant correlations after multiple testing correction ($P \leq 1.4 \times 10^{-4}$) are shown. The thickness and intensity of edges depend on variation in one locus explained by the effect estimates in the second locus. Round-edged rectangular nodes denote genes

Table 2. In silico evidence for edges from functional network of IgG N-glycosylation loci. Permutation analysis was used to assess the distribution of glycome-wide SNP effect pairwise correlations and the STRING PPI database (*16*) to search for biological validation of observed links. All reported correlations are statistically significant after Bonferroni correction. Corr, Spearman's correlation coefficient of *z* scores of SNP1 and *z* scores of SNP2; quant, quantile of distribution at which the given ρ2 was observed in permutation analysis; STRING score, probability that the link exists obtained by combining evidence from different sources. SNPs that were replicated in UPLC replication are in bold.

SNP1	SNP2	Locus 1	Locus 2	Corr	SNP 1 quant	SNP 2 quant	STRING gene 1	STRING gene 2	STRING score
Glycosyltransferase loci									
rs5750830	**rs17630758**	*MGAT3*	*SMARCB1-DERL3-CHCHD10-VPREB3*	0.967	0	0			
rs10813951	**rs12341905**	*B4GALT1*	*SPINK4*	0.805	3.54×10^{-3}	7.72×10^{-3}			
rs10813951	**rs7758383**	*B4GALT1*	*HIVEP2*	0.711	1.87×10^{-2}	3.83×10^{-3}			
rs7621161	rs481080	*ST6GAL1*	*NXPE1; NXPE4*	0.773	6.00×10^{-5}	3.01×10^{-3}			
rs7621161	**rs7700895**	*ST6GAL1*	*ELL2*	0.581	8.32×10^{-3}	2.45×10^{-2}			
rs11847263	**rs7216389**	*FUT8*	*ORMDL3-GSDMB-IKZF3-ZPBP2*	0.709	8.42×10^{-3}	2.59×10^{-2}			
rs11847263	**rs6421315**	*FUT8*	*IKZF1*	0.624	2.62×10^{-2}	9.69×10^{-2}			
Other loci									
rs10903118	**rs7281587**	*RUNX3*	*RUNX1*	0.891	2.00×10^{-5}	3.00×10^{-5}	*RUNX3*	*RUNX1*	0.9
rs7281587	**rs17630758**	*RUNX1*	*SMARCB1-DERL3-CHCHD10-VPREB3*	0.597	3.25×10^{-2}	1.59×10^{-2}	*RUNX1*	*SMARCB1*	0.4
rs10903118	**rs17630758**	*RUNX3*	*SMARCB1-DERL3-CHCHD10-VPREB3*	0.694	1.15×10^{-2}	3.33×10^{-3}			
rs7216389	**rs6421315**	*ORMDL3-GSDMB-IKZF3-ZPBP2*	*IKZF1*	0.609	9.57×10^{-3}	2.49×10^{-2}	*IKZF3*	*IKZF1*	0.42
rs7216389	**rs6421315**	*ORMDL3-GSDMB-IKZF3-ZPBP2*	*IKZF1*	0.609	9.57×10^{-3}	2.49×10^{-2}	*ZPBP2*	*IKZF1*	0.48
rs7281587	**rs6421315**	*RUNX1*	*IKZF1*	0.626	2.56×10^{-2}	2.15×10^{-2}	*IKZF1*	*RUNX1*	0.6
rs11748193	rs3099844	*IRF1-SLC22A4*	*TCF19-GPANK1-TNXB-HLA-DRA*	0.231	8.82×10^{-2}	9.86×10^{-2}	*HLA-DRA*	*IRF1*	0.93
rs7281587	**rs11748193**	*RUNX1*	*IRF1-SLC22A4*	0.283	1.63×10^{-1}	6.02×10^{-2}	*RUNX1*	*SLC22A4*	0.7
rs7257072	rs3099844	*RFXANK*	*TCF19-GPANK1-TNXB-HLA-DRA*	0.23	1.77×10^{-1}	9.85×10^{-2}	*HLA-DRA*	*RFXANK*	0.62

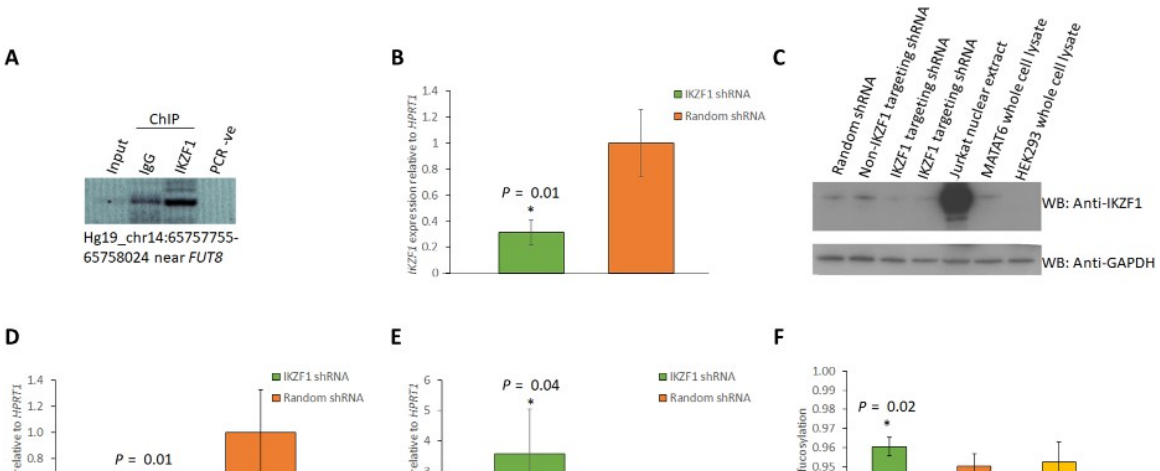
the sialyltransferase *ST6GAL1* locus (ρ = 0.77 and 0.58). Full results of the permutation analysis and STRING PPI are avail-

glycosyltransferase gene (and/or for the factors regulating this gene, i.e., RUNX1 is repressed by RUNX3) (Fig. 2B).

strongly affect binding of the TF. Alternatively, the allele can also completely disrupt or create the binding motif, resulting in a loss of a known or introduction of a new TF-binding site. In four loci spanning the glycosyltransferase genes, the associated SNPs resulted in either introduction or disruption of binding sites for TFs RUNX1, RUNX3, IKZF1, SMARCB1, TBX21, and IRF1. We observed the same disruption/introduction of TF-binding sites for all these TFs too and therefore confirmed some well-known regulation feedback loops between TFs RUNX1 and RUNX3 and IKZF1 and IKZF3 (table S8) (*19*, *20*). To assess whether these TF-binding site alterations are specific for associated SNPs, we compared the frequency of TF-binding alterations of associated and nonassociated SNPs. In seven glycosylation loci, the associated SNPs were at least two times more likely to affect binding of TF than nonassociated SNPs from the same region. The highest difference in the number of SNPs affecting the TF-binding site was observed for *FUT8*, *MGAT3*, and loci coding for TFs that are members of their cluster. Associated SNPs from the *FUT8* and *MGAT3* loci have a significant effect on binding of TF from our network, while nonassociated SNPs from their proximity have no effect on binding (table S9). In summary, we not only found TFs associated with IgG glycosylation but also showed that glycosylation-associated SNPs can potentially alter their binding in other glycosylation-associated loci. These results not only reinforce some of the links suggested in our network but also confirm some known feedback loops between different TFs involved in B cell differentiation and biology (e.g., interactions of IKZF1 and IKZF3 and RUNX1 and RUNX3) (*19*, *20*).

Experimental validation of the functional links between *IKZF1* and *FUT8*

Our network-based approach allowed us to show that variants in or near *IKZF1* and *IKZF3* share glycome-wide similarities with the top SNP in the *FUT8* locus. We hypothesized that the TFs IKZF1 and IKZF3 may regulate the expression of *FUT8*. We further observed notable pleiotropy between variants in the *IKZF3-ORMDL3-GSDMB-ZPBP2* locus with inflammatory conditions including inflammatory bowel disease (IBD), ulcerative colitis (UC), and rheumatoid arthritis (RA) (Fig. 2). We therefore aimed to disrupt this node of our network in our cellular model. Microarray and RNA sequencing data from the lymphoblastoid cell line (LCL) used in this study showed that *IKZF1* was more highly expressed than *IKZF3*. Because these genes are known to regulate one another, we decided to target *IKZF1*. To investigate the potential link between *IKZF1* and *FUT8*, we identified IKZF1-binding sites on chromosome 14 near *FUT8* in the LCL GM12878 Encode chromatin immunoprecipitation sequencing (ChIP-seq) data. We performed ChIP followed by polymerase chain reaction (PCR) to confirm IKZF1 binding at one of these sites in an IgG-secreting human B cell–derived LCL, MATAT6 (Fig. 3A). We then investigated the role of IKZF1 in the regulation of *FUT8* by conducting stable short hairpin RNA (shRNA)–mediated knockdown of *IKZF1* using shRNA in MATAT6 cells. This resulted in depletion of the *IKZF1* transcript and protein (Fig. 3, B and C). Knockdown of *IKZF1* was accompanied by a significant down-regulation of *IKZF3* (Fig. 3D) and a 3.5-fold up-regulation in the expression of *FUT8* (*t* test, $P = 0.04$; Fig. 3E), corroborating a link between *IKZF1* and *FUT8* and highlighting the dynamic relationships between TFs in B cells. Next, we looked at changes in the glycosylation of secreted IgG in cells with depleted IKZF1 at two different time points and observed a small but significant ($P = 0.02$; table S10) increase in fucosylation (Fig. 3F) resulting from *IKZF1* knockdown. While this change is small, the low SE of triplicate measurements and independent replication of the experiment (second time point) demonstrate its robustness (table S10). In addition, one must consider that most of the glycans are fucosylated and that further fucosylation might be limited by specificities of glycosyltransferases involved in the process—the addition of bisecting GlcNAc to a growing glycan chain by MGAT3



inhibits fucosylation of these glycans (*21*). This change also results in a 20% decrease (from 5 to 4%) in afucosylated glycans [glycoforms activating antibody-dependent cellular cytotoxicity (ADCC)]. These data are consistent with decreased expression of *IKZF1* resulting in increased *FUT8* expression and, thus, increased IgG fucosylation. Hence, although there is no in silico evidence that the lead SNP in the *FUT8* locus overlaps with a predicted IKZF1-binding motif, and there is no evidence for the associated SNP being a *FUT8* eQTL, we show that SNPs that are in LD with the lead *FUT8* SNP can alter binding of *IKZF1* (table S8). Furthermore, this TF can bind to regulatory regions of *FUT8* and may act cooperatively with additional factors to modify its expression. To explore this further, we used publicly available Hi-C profiling performed at 1-kb resolution in the LCL GM12878. These data indicate that SNPs in the *FUT8* locus lie in the same chromosomal topologically associating domain as the transcription start site of *FUT8* (fig. S2B). CCCTC-binding factor (CTCF) plays a crucial role in maintaining the three-dimensional structure of the genome, and disruption of CTCF binding by SNPs in LD with rs4400971 may modify the interactions formed between enhancer and promoter regions of *FUT8* to affect expression of the glycosyltransferase enzyme (for details, see Supplementary Note).

Pleiotropy with complex traits and diseases

Because aberrant glycosylation patterns have been observed in various diseases (*22*–*24*), we explored pleiotropy between the glycosylation loci and other traits. We first searched for whether suggestive or significant glycosylation-associated SNPs and their proxies ($P_{GWASglyco} \leq 2.4 \times 10^{-8}$; LD $R^2 \geq 0.8$) were previously reported as associated ($P_{GWAStraits} \leq 5 \times 10^{-8}$) with other traits using Ensembl's BioMart tool (*25*). Overall, 219 SNPs from 16 glycosylation loci were reported to be associated with 83 other traits in various databases, 47 of which map outside the HLA region (chr6:29570005-33377657, GRCh37) (https://shiny.igmm.ed.ac.uk/igg_glycans_gwas/).

To expand on the findings from Lauc *et al.* (*6*), where they report that some variants in LD with glycosylation variants are pleiotropic with diseases, we tested whether these coassociations may be due to the same underlying causal variant (pleiotropic) or are suggestive of LD between distinct causal variants each controlling different traits. We performed the SMR/HEIDI test on traits for which the full summary-level GWAS data were available. We excluded the HLA region from our analysis because of its complex LD structure, which is incompatible with this method's sensitivity to LD. In total, we analyzed 10 traits (see details in Supplementary Note) and 55 glycan-trait combinations.

In three glycosylation loci, we found evidence for pleiotropy ($P_{SMR} \leq 9.1 \times 10^{-4}$, $P_{HEIDI} \geq 0.05$) between at least one IgG glycan and eight other diseases or traits [Crohn's disease (CD), IBD, UC, RA, primary biliary cirrhosis (PBC), asthma, high-density lipopro-

DISCUSSION

A recent study of the IgG glycome in 95 mouse strains from the collaborative cross cohort revealed that differences in glycosylation make IgG structure quite diverse between different strains (*26*). Despite the absence of a direct genetic template for individual glycans, these differences are heritable, indicating that glycoprotein structure is being inherited as a complex trait. The main enzymes involved in protein N-glycosylation are well characterized (*4*), but the genetic regulation of the process of glycosylation is poorly understood. To address this, we performed the largest GWAS of the IgG glycome to date, identifying 27 loci and therefore doubling the number of previously associated loci. While 10 of these loci contain known glycosylation enzymes and other previously associated genes, 14 novel loci contain genes with no known role in glycosylation. We were able to explain as much as 22% of variance in glycan levels (for the degree of monosialylation of fucosylated digalactosylated structures without bisecting GlcNAc, IGP29). Compared to previous studies, we do not just report the gene closest to the strongest association signal but have performed detailed analyses to seek the most biologically relevant annotation of our associated loci. Our candidate genes either have associated variants located in the coding regions of genes, show pleiotropy with gene expression in biologically relevant cell types, or are enriched in gene sets from appropriate biological pathways. We also explored how these genes are connected in a functional network, proposing mechanisms of the regulation of known glycosylation enzymes, and confirmed some of the network connections with in vitro functional follow-up.

As previously observed for GWAS of other complex traits, SNPs associated with IgG glycosylation predominantly map to noncoding regions of the genome, suggesting that regulatory variation at distal enhancers influences interindividual differences in glycan profiles. Several lines of evidence support this: the association of variants in or near genes coding for TFs, a high incidence of associated variants that are predicted to affect binding of those TFs, and pleiotropy between variants regulating the expression of the glycosyltransferases in B cells and IgG N-glycosylation. In addition, missense variants were observed only in genes without a previously known role in glycosylation and not in any glycosyltransferase genes where damaging mutations are known to result in severe disease (*27*). IgG is synthesized and secreted by terminally differentiated B cells called plasma cells. Gene set enrichment analysis of our GWAS loci identified enrichment not only for GO terms associated with glycosylation and processes involved in the differentiation and maturation of B cells but also for GO terms associated with TF activity. This suggests that IgG N-glycosylation is more likely to be regulated at the level of expression of the main enzymes, their intracellular localization, or substrate availability than by impaired protein function.

We also capitalized on the omics setting of this study to create a data-driven network of the associated loci, suggesting mechanistic hypotheses on how new loci may influence known glycosyltransferases. We then used in vitro studies to validate selected findings. Within

Table 3. Summary of the HEIDI testing for pleiotropy versus LD between variants coassociated with IgG glycan levels, complex traits, and gene expression. All traits have significant ($P_{SMR} \leq 9.1 \times 10^{-4}$) coassociation. In these loci where the associated variant was the same in IgG glycosylation, complex traits, and gene expression, we report only glycans pleiotropic with both expression and complex traits. The HEIDI test distinguishes between pleiotropy (shared causal variant, HEIDI $P \geq 0.05$) or causal variants in LD (HEIDI $P < 0.05$). Glycan trait descriptions: G, galactose; F, fucose; N, bisecting GlcNAc; S, sialic acid; gene expression: CD4 and CD8, helper and cytotoxic T cells; CD19, B cells; PB, peripheral blood.

Locus	IGP	Glycan description	Complex traits	SMR direction	Gene expression	SMR direction
Shared causal variant (pleiotropy)						
IRF1-SLC22A4	IGP2 and IGP42	G0	CD	+	*SLC22A4* (PB)	–
ORMDL3-GSDMB-IKZF3-ZPBP2	IGP2, IGP6, IGP42, and IGP46	G0 and G1	Asthma, HDL, PBC, IBD, RA, and UC	+	*ORMDL3* (CD4, CD8, CD19, and PB)	–
					GSDMB (CD19 and PB)	–
					IKZF3 (PB)	+
	IGP58, IGP59, IGP60, and IGP61	% Fucosylation in total and in agalactosylated/ monogalactosylated/ digalactosylated glycans	Asthma, PBC, HDL, RA, and UC	–	*ORMDL3* (CD4, CD8, CD19, and PB)	+
					GSDMB (CD19 and PB)	+
					IKZF3 (PB)	–
CRHR1-SPPL2C-MAPT-ARHGAP27	IG14, IGP49, and IGP54	G1FN, G2FN, % fucosylation of G1 glycans	PD	–		
Distinct causal variants in LD						
RUNX3	IGP65, IGP69, IGP74, IGP75, and IGP76	Ratio of fucosylated digalactosylated structures	Ankylosing spondylitis			
IRF1-SLC22A4	IGP2 and IGP42	G0	Height and IBD			

We identified seven loci as containing associated SNPs for which one allele disrupts a binding motif for at least one TF in our network, two or more times more frequently than nonassociated SNPs from the same region. This is consistent with the identification of functionally important variants (or in LD with) in cis-regulatory elements acting to alter gene expression. For example, in our network, we suggest that a number of TFs and chromatin remodeling proteins regulate expression of *MGAT3*, the enzyme that transfers a GlcNAc to the mannose core of N-linked oligosaccharides to produce bisecting GlcNAc glycan structures. We have provided in silico evidence that associated SNPs from the *MGAT3* region disrupt binding sites for these TFs and are pleiotropic with *MGAT3* expression in $CD19^+$ B cells. In our network, we also observed a link between *IKZF1* and *FUT8*. We provide direct evidence that IKZF1 transcriptionally regulates *FUT8*, showing, first, that IKZF1 binds to regulatory regions of *FUT8* and, second, that *IKZF1* knockdown results in increased *FUT8* expression and increased IgG fucosylation. We also

Using glycosyltransferases as a positive control for gene prioritization, it is apparent that none of the methods used are sufficient in isolation, highlighting the complementarity and limitations of different approaches. In current databases and pathways, there is no information on the regulation of glycosyltransferases, so complementing GWAS with data-driven network analysis, as performed here, can provide useful insights into potential mechanisms regulating genes of interest.

Aberrant glycosylation profiles were in the past observed in many diseases. An increase in agalactosylated structures is related to the proinflammatory effects of IgG (*31*). The absence of the core fucose is associated with increased ADCC (*2*, *32*). Previous studies reported that some IgG glycosylation–related SNPs were also associated with a range of autoimmune diseases and hematological cancers (*6*, *7*), but the mechanisms are still mostly unknown. We explored pleiotropy with complex traits and diseases and showed that diseases having the same associated SNPs are enriched in immune system disorders.

agalactosylated glycan traits was observed in epidemiological studies of patients with IBD, CD, UC (*33*, *34*), and RA (*35*), and decreased fucosylation in UC, but not in CD (*34*).

In the same locus on chromosome 17, expression of *GSDMB* and *ORMDL3* in peripheral blood and B and T cells—and of *IKZF3* in peripheral blood—followed the opposite pattern of association with IgG glycans to that for the disease risks. While these genes are not glycosyltransferases and therefore do not have a direct influence on glycosylation levels, we showed a direct link between the *GSDMB-ORMDL3-IKZF3-ZPBP2* locus and *FUT8* and indirect links with *B4GALT1*, potentially explaining how pleiotropy with genes previously unrelated to glycosylation could influence glycosylation and suggesting why specific glycans are altered in these diseases.

One drawback of this study is the use of a relatively old HapMap2 imputation panel. However, at the onset of analyses, for many complex traits and expression studies, summary association statistics were only available for older panels. Still, using a newer imputation panel or even whole-genome sequencing data would increase the power of the study to detect rarer variants not well tagged in the HapMap2 imputation.

B cell development requires a regulatory network driven by coordinated activity of TFs, the end point being antibody production by plasma cells in response to antigen stimulus. The TFs IKZF1, IKZF3, RUNX1, and RUNX3 are lineage specific and have key roles in controlling the appropriate gene expression pattern at given time points, acting to both activate and repress gene expression and to alter the epigenetic landscape at functionally important genes. Furthermore, their relative expression levels are dynamically and reciprocally regulated (*19*). A limitation of our network is that it is one dimensional; that is, it cannot delineate hierarchal relationships nor indicate at which stage during B cell development these interactions occur. Functional characterization of individual SNPs and of these putative enhancers to establish causality is beyond the scope of this study and remains to be explored.

In conclusion, we have shown that GWAS together with gene prioritization and data-driven network analysis is a powerful strategy for identifying biologically meaningful regulatory mechanisms underlying a complex biological process as is IgG glycosylation. The complexity of our functional network appears to reflect both the complexity of glycans themselves and the complex differentiation process that B cells undergo to reach the terminally differentiated state of IgG-producing plasma cells. However, it is clear that variation in key TFs coupled with regulatory variation in glycosylation enzyme genes drives changes in IgG glycosylation, modifying IgG function and influencing health and disease.

METHODS

All studies were approved by local research ethics committees,

using affinity chromatography binding to protein G plates, followed by release and labeling of glycans with 2-AB (2-aminobenzamide) fluorescent dye. Glycans were then separated and quantified by hydrophilic interaction UPLC, resulting in 24 chromatographic peaks. Most of the peaks contain a single glycan structure, while some contain more than one, but with at least 63% of the peak being contributed by the most abundant glycan (*36*).

Phenotype preprocessing

To reduce experimental variation in glycan measurements, before genetic studies, raw glycan data were total area normalized and batch corrected using the "ComBat" function of "sva" (*37*) R package centrally by the phenotype provider (Genos Ltd. for UPLC cohorts and Leiden University Medical Center for the Leiden Longevity Study). More detailed information on glycan preprocessing can be found in the Supplementary Note.

Genetic association studies

Discovery

Discovery genome-wide association studies (GWAS) were performed in four cohorts of European descent, CROATIA-Korcula ($N = 849$), CROATIA-Vis ($N = 802$), ORCADES (Orkney Complex Disease Study) ($N = 1960$), and TwinsUK (N = 4479), with a combined sample size of 8090. Each glycan trait was first rank transformed to a normal distribution and then corrected for age, sex, cohort-specific covariates, and cryptic relatedness using linear mixed models and a kinship matrix estimated from genotyped data (as appropriate in each specific cohort). Within each cohort, a genome-wide association scan was performed based on the HapMap2 (release 22) imputed genetic data (or corresponding subset from a newer reference panel), assuming an additive linear model of association. Details of individual-level GWAS and parameters specific for each cohort can be seen in the Supplementary Note.

Cohort-level GWAS were examined for inconsistencies and informative data using the EasyQC software package (*38*), and results were pooled and analyzed with METAL (*39*) using a fixed-effect inverse-variance meta-analysis method. Before meta-analysis, each GWAS was corrected for genomic control inflation factor. The genomic control inflation factor varied from 0.95 to 1.05, suggesting little residual influence of population stratification. Additional genomic control was performed on the aggregated meta-analysis results. As suggested by Winkler *et al.* (*38*), GWAS cleaning and meta-analysis were performed by two independent analysts, and the obtained effect size estimates and *P* values of the two meta-analyses displayed perfect concordance.

Multiple testing was controlled for by considering an association as genome-wide significant if the *P* value of association was $\leq 2.4 \times 10^{-9}$, as suggested by Li and Ji (*40*) and applied in Lauc *et al.* (*6*).

A locus was defined using the DEPICT tool (*10*) as a region spanned by SNPs in LD from the lead variant in the region. In cases

individual replication study and pooled using fixed-effect inverse-variance meta-analysis. The top SNP was defined as the SNP with the lowest P value in a locus. Replication significance level was set to $P \leq 1.9 \times 10^{-3}$ (0.05 divided by the number of genome-wide significant loci, 27).

Additional validation was performed on glycosylation data measured with LCMS in the Leiden Longevity Study ($N = 1842$). Given that UPLC and LCMS measure similar but slightly different glycan traits (see "Validation genome-wide association studies" in the Supplementary Note), it was not possible to replicate the same discovery SNP-glycan pairs. Instead, we looked at associations of the top SNP from every locus with any LCMS glycan, with an aim of replicating a locus rather than SNP-glycan association. Therefore, we set the significance level to $P \leq 3.7 \times 10^{-5}$ [0.05/50/27—Bonferroni correction for 50 directly measured LCMS glycan traits, as used in Wahl *et al.* (*8*), and 27 loci].

GWAS of gene expression

Peripheral blood eQTL summary-level genome-wide statistics were downloaded from the SMR Results Database (*41*) in August 2016. Briefly, this study meta-analyzed gene expression quantified with Illumina arrays in 5311 samples and HapMap2 reference panel imputed genotypes (*14*).

We also used eQTL summary-level genome-wide statistics for six circulating immune cell types [$CD4^+$ T lymphocytes, $CD8^+$ T lymphocytes, $CD19^+$ B lymphocytes, $CD14^+$ monocytes, $CD15^+$ granulocytes, and platelets from the CEDAR dataset; (*13*)].

Conditional analyses and variance explained

To test for the association of secondary SNPs while conditioning on the top SNP in the region, we performed conditional analysis on summary statistics from the discovery meta-analysis. Briefly, this method performs forward stepwise selection on summary-level data, where, based on an LD estimated from a reference dataset in each iteration, the SNP with the strongest association in the region is added in the regression model until no additional SNPs reach genome-wide significance. GCTA stepwise variable selection (*9*) was performed using more than 6000 unrelated individuals from the Generation Scotland (*42*) as an independent reference sample and restricting collinearity to 0.9, with $P \leq 2.4 \times 10^{-9}$ as a genome-wide significance level. Reported joint P value was corrected with genomic control inflation factor λ.

For every independent SNP_i, we calculated the proportion of explained phenotypic variance as

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

where β_i is the effect estimate of SNP_i in univariate meta-analysis. and p_i and q_i are the minor and major allele frequencies of SNP_i calculated in the Generation Scotland cohort, respectively. For each glycan that has at least one significantly associated SNP, we calculated total univariate explained variance as a sum of the pro-

where β_i^u is the effect estimate of SNP_i in the univariate analysis, and β_i^J is the joint effect estimate of the same SNP in the joint analysis.

Prioritizing genes associated with IgG N-glycosylation

For the genome-wide suggestive and significant loci, we explored potential causative genes at each locus using a strategy that combined publicly available datasets and prioritization tools and eQTLs from peripheral blood– and immune cell type–specific datasets.

Coding variation

To obtain the putative functional effect of associated variants, significantly or suggestively associated SNPs were annotated with VEP (*11*) in May 2016. Because of alternative splicing, each gene can have more than one transcript, and, consequently, depending on the transcript, one SNP can have a different effect on protein function.

SMR/HEIDI analysis for pleiotropy with gene expression

To test for potential pleiotropy between gene expression and IgG glycosylation, we performed SMR with HEIDI analysis, developed by Zhu *et al.* (*12*). SMR analysis provides evidence for pleiotropy but cannot distinguish if the associations are driven by the same or highly correlated but distinct causal variants. The subsequent HEIDI test allowed us to distinguish pleiotropy from LD. The test was performed on a publicly available peripheral blood dataset from Westra *et al* (*14*) and immune cell–specific gene expression from the CEDAR dataset (*13*). Among others, this dataset contains expression in B lymphocytes (CD19), helper T lymphocytes (CD4), cytotoxic T cells (CD8), macrophages (CD14), neutrophils (CD15), and platelets (PLA).

We set a threshold for the SMR test at $P_{SMR} \leq 1.9 \times 10^{-5}$, corresponding to a Bonferroni correction for 2622 tests, number of regions where genome-wide significant top regional IgG glycan SNP (or its proxy) was also available in any of the gene expression associations. All regions with significant P_{SMR} and $P_{HEIDI} \geq 0.05$ were considered to exhibit concordance in regional association patterns and therefore showed evidence of sharing the underlying unobserved causal variant. For these regions, we can suggest that they are pleiotropic. Details of the algorithm can be seen in the Supplementary Note.

Prioritization using DEPICT

DEPICT (*10*) was run on the merged list of independent SNPs obtained from the GCTA-COJO analysis to identify gene sets enriched for genes near associated variants. Suggestive independent SNPs ($P \leq 5 \times 10^{-8}$) were submitted to DEPICT, release 194 (*10*). The list of independent SNPs was created by merging glycan-wise GCTA-COJO results, resulting in 113 SNPs. Given that different glycans can have a different lead SNP (but in high LD), this list was additionally pruned by applying PLINK clumping (*43*), where all SNPs within 500 kb and LD $R^2 > 0.1$ of the SNP with the strongest association were assigned to the same clump. LD was estimated using 1000 Genomes Project Phase 1 CEU [Utah Residents (CEPH) with Northern and Western European Ancestry], GBR (British in England and Scotland), and TSI (Toscani in Italy) data.

defined using the keywords "transcription" or "expression." DEPICT pathways and tissue enrichments analyses were performed as described above.

Functional network of genes associated with IgG N-glycosylation

To suggest a functional relationship of loci associated with IgG glycosylation, we performed pairwise association analyses of glycome-wide effects of lead SNPs. Glycome-wide effect of the SNP was defined as the vector of *z* scores [*z* score = β/*se*(β)] of this SNP on each of glycans.

Before analysis, we removed 15 derived traits (IGP41 to IGP54) that represent directly measured glycans (IGP1 to IGP15) that were normalized with the surface area of neutrally charged glycans (see the Supplementary Note for details).

We then constructed glycome-wide effects for all 27 lead SNPs from the meta-analysis results for 62 glycans, resulting in 27 vectors of 62 *z* scores. For every pair of SNPs, we computed the Spearman's correlation coefficient and corresponding *P* value between 27 glycome-wide effects.

The network of significant Spearman's correlations [*P* value corrected for $(27 \times 26)/2 = 351$ tests, $P \leq 1.4 \times 10^{-4}$] was visualized using Cytoscape (*44*), where each node represents one lead SNP annotated with the gene prioritized in the region of that SNP and width and color intensity of edges present squared Spearman's correlation coefficient of the two nodes. We next performed clustering analysis of the full correlation matrix by applying hierarchical clustering with Euclidean distance and complete linkage. To validate the network, we performed permutation analysis and compared the network with STRING PPI networks (*16*).

The permutation analysis was performed by estimating the correlation of glycome-wide effects of every top SNP and glycome-wide effects of 100,000 random SNPs. To ensure that no glycosylation-associated SNPs were included in the permutation analysis, the following procedure was used. From the list of all HapMap release 22 SNPs (2,574,585 SNPs), all variants from the associated loci were removed, resulting in 2,510,568 remaining SNPs. An additional 6641 SNPs were removed because they had $P_{GWAS} \leq 5 \times 10^{-5}$ in at least one glycosylation GWAS. The final list contained 2,503,927 SNPs. From this list, 100,000 SNPs were randomly selected for the permutation analysis. A Spearman's correlation between glycome-wide effects of each of the 27 top SNPs with glycome-wide effects of 100,000 random SNPs was computed. These correlations were then compared with correlations used to construct the network.

Additional validation was performed by comparing the network obtained by submitting IgG N-glycosylation candidate genes to the STRING database of PPIs (version 10.5, accessed in September 2017) (*16*).

ChIP-seq data (.narrowPeak files) for the TFs were downloaded from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/ for the GM12878 LCL. The peak-motifs tool (*18*) in the Regulatory

site disruptions of glycosylation SNPs and other SNPs in the region, we performed the same analysis using all significantly and suggestively associated glycosylation SNPs (associated SNPs) and all SNPs within 50 kb of every glycosylation locus whose association *P* value was $\leq 5 \times 10^{-4}$ (nonassociated SNPs) and compared frequency of SNPs that are predicted to significantly (Bonferroni-corrected $P \leq 1.8 \times 10^{-8}$) disrupt binding for the given TF.

Hi-C data derived from GM12878, digested with Mbo I, were interrogated using the online tool available at http://higlass.io/app/ (*45*). More details can be found in the Supplementary Note.

In vitro validation of the functional links between *IKZF1* and *FUT8*

ShRNA cloning, cell culture, transfections, and shRNA knockdown

Four shRNAs targeting the coding region or the 3′ untranslated region of all *IKZF1* isoforms were designed using Block-iT RNAi Designer (https://rnaidesigner.thermofisher.com/rnaiexpress/). Single-stranded oligonucleotides for IKZF1 shRNA and a sequence with no corresponding target in the human genome were annealed and cloned into pENTR/H1/TO vector and used for electroporation of an IgG1-secreting human LCL, MATAT6. RNA was extracted to assess gene knockdown efficiency for each shRNA tested by quantitative PCR (qPCR), and the IKZF1 shRNA resulting in the most significant knockdown was used for subsequent experiments. More details can be found in the Supplementary Note.

RNA, complementary DNA, and real-time PCR

RNA was extracted from MATAT6 cells and stable shRNA lines, followed by on-column DNase (deoxyribonuclease) digestion and synthesis of complementary DNA and qPCR. qPCR gene expression assays were performed for *IKZF1*, *FUT8*, *IKZF3*, and *HPRT1*. Samples were run in triplicate. Relative gene expression level was determined using the comparative C_t (cycle threshold) method. Statistically significant differences in gene expression were determined using the paired *t* test. More details can be found in the Supplementary Note.

Western blot analysis

Whole-cell lysates were prepared by lysing cells in radioimmunoprecipitation assay buffer on ice. Jurkat nuclear extract and human embryonic kidney 293T lysates were used as positive and negative controls for IKZF1. Ten micrograms of total protein was reduced and denatured before separation on 4 to 12% Bis-Tris NuPAGE gel. Membranes were probed using rabbit polyclonal anti-Ikaros and mouse anti–α-tubulin antibodies. Secondary antibodies were horseradish peroxidase–conjugated anti-rabbit IgG or anti-mouse IgG. Antibody detection was achieved using the enhanced chemiluminescence detection system. More details can be found in the Supplementary Note.

Glycan profiling of secreted IgG

Five million cells from stable bulk cultures of shRNA-expressing MATAT6 cells were washed in phosphate-buffered saline and then resuspended in (serum-free) Opti-MEM. After 72 hours, conditioned media were collected by centrifugation and immediately frozen at

described in the Supplementary Note. ChIP input was incubated with either anti-IKZF1 or goat IgG isotype control antibody overnight at 4°C with rotation. Five micrograms of biotinylated anti-goat IgG was added for a further 2 hours before the addition of streptavidin-agarose beads. Agarose beads were collected by centrifugation. After the final wash, Chelating resin solution was added to the beads, and the samples were boiled for 10 min. ChIP-PCR was performed using primers flanking a binding site upstream of *FUT8* identified by ChIP-seq analysis in the GM12878 cell line to confirm that IKZF1-DNA complexes also occur in MATAT6 cells. PCR products were analyzed on 2% (w/v) agarose gel by electrophoresis in tris-borate EDTA buffer.

SMR/HEIDI analysis for pleiotropy with complex traits

To assess regional association concordance between IgG glycosylation and other complex traits (SMR/HEIDI test), we were able to download summary-level statistics including signed regression coefficient estimates and SE of these estimates for 10 traits (see the Supplementary Note for details). For each unique locus and glycan-trait combination for the test, we used only the SNP with the lowest *P* value in glycan GWAS. For all the available traits, the same procedure as outlined in the "SMR/HEIDI analysis for pleiotropy with gene expression" was applied.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/8/eaax0301/DC1

Supplementary Note

Appendix Table 1. Participating studies.

Appendix Table 2. Genotyping overview.

Appendix Table 3. Overview of imputation software and reference panels.

Appendix Table 4. Description of structures for UPLC IgG glycans.

Appendix Table 5. Details of genome-wide association analyses.

Appendix Table 6. Individual GWAS file-level quality control.

Appendix Table 7. Description of structures for LCMS glycans.

Appendix Table 8. Summary-level statistics downloaded for SMR or HEIDI test.

Appendix Table 9. DEPICT gene prioritization.

Appendix Table 10. SNPs associated with IgG glycosylation with nonsynonymous amino acid change.

Appendix Table 11. Strongest eQTL for each probe in each cell type in the CEDAR dataset.

Appendix Figure 1. Phenotypic correlation of UPLC IgG *N*-glycans.

Fig. S1. Glycome-wide effect estimates of genome-wide significant loci.

Fig. S2A. RUNX3 binding in the proximity of MGAT3 rs8137426, SNP strongly associated with IgG N-glycosylation is the region bound by RUNX3, in the proximity of MGAT3.

Fig. S2B. Top SNPs in *FUT8* locus lie in the same chromosomal topological associating domain as the transcription start site of FUT8.

Table S1. Comparison of *P* values from current meta-analysis (*P*-new) and Lauc *et al.* (*P*-old), Shen *et al.* (*P*-multivariate), and Wahl *et al.* (*P*-LCMS).

Table S2. Results of replication and validation analysis.

Table S3. Phenotypic variance explained by significantly associated SNPs ($P \leq 2.4 \times 10^{-9}$).

Table S4. FUMA GO gene set enrichment analysis.

Table S5. DEPICT analysis of gene set enrichment.

Table S6. Correlation of glycome-wide effects of top genome-wide significant SNPs associated with IgG glycosylation.

Table S7. STRING PPI analysis of genes prioritized in loci associated with IgG N-glycosylation.

Table S8. TF motif alterations by glycosylation-associated SNPs.

Table S9. Effect of IgG glycosylation–associated SNPs on TF motif–binding disruption compared with nonassociated SNPs from the region.

3. I. Gudelj, G. Lauc, M. Pezer, Immunoglobulin G glycosylation in aging and diseases. *Cell. Immunol.* **333**, 65–79 (2018).
4. N. Taniguchi, K. Honke, M. Fukuda, *Handbook of Glycosyltransferases and Related Genes* (Springer Science & Business Media, 2011).
5. J. E. Huffman, A. Knežević, V. Vitart, J. Kattla, B. Adamczyk, M. Novokmet, W. Igl, M. Pučić, L. Zgaga, Å. Johannson, I. Redžić, O. Gornik, T. Zemunik, O. Polašek, I. Kolčić, M. Pehlić, C. A. M. Koeleman, S. Campbell, S. H. Wild, N. D. Hastie, H. Campbell, U. Gyllensten, M. Wuhrer, J. F. Wilson, C. Hayward, I. Rudan, P. M. Rudd, A. F. Wright, G. Lauc, 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 (2011).
6. G. Lauc, J. E. Huffman, M. Pučić, L. Zgaga, B. Adamczyk, A. Mužinić, M. Novokmet, O. Polašek, O. Gornik, J. Krištić, T. Keser, V. Vitart, B. Scheijen, H.-W. Uh, M. Molokhia, A. L. Patrick, P. McKeigue, I. Kolčić, I. K. Lukić, O. Swann, F. N. van Leeuwen, L. R. Ruhaak, J. J. Houwing-Duistermaat, P. E. Slagboom, M. Beekman, A. J. M. de Craen, A. M. Deelder, Q. Zeng, W. Wang, N. D. Hastie, U. Gyllensten, J. F. Wilson, M. Wuhrer, A. F. Wright, P. M. Rudd, C. Hayward, Y. Aulchenko, H. Campbell, I. Rudan, Loci associated with *N*-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. *PLOS Genet.* **9**, e1003225 (2013).
7. X. Shen, L. Klarić, S. Sharapov, M. Mangino, Z. Ning, D. Wu, I. Trbojević-Akmačić, M. Pučić-Baković, I. Rudan, O. Polašek, C. Hayward, T. D. Spector, J. F. Wilson, G. Lauc, Y. S. Aulchenko, Multivariate discovery and replication of five novel loci associated with Immunoglobulin G *N*-glycosylation. *Nat. Commun.* **8**, 447 (2017).
8. A. Wahl, E. van den Akker, L. Klaric, J. Štambuk, E. Benedetti, R. Plomp, G. Razdorov, I. Trbojević-Akmačić, J. Deelen, D. van Heemst, P. E. Slagboom, F. Vučković, H. Grallert, J. Krumsiek, K. Strauch, A. Peters, T. Meitinger, C. Hayward, M. Wuhrer, M. Beekman, G. Lauc, C. Gieger, Genome-wide association study on immunoglobulin G glycosylation patterns. *Front. Immunol.* **9**, 277 (2018).
9. J. Yang, T. Ferreira, A. P. Morris, S. E. Medland; Genetic Investigation of ANthropometric Traits (GIANT) Consortium; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, P. A. F. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1-3 (2012).
10. T. H. Pers, J. M. Karjalainen, Y. Chan, H.-J. Westra, A. R. Wood, J. Yang, J. C. Lui, S. Vedantam, S. Gustafsson, T. Esko, T. Frayling, E. K. Speliotes; Genetic Investigation of ANthropometric Traits (GIANT) Consortium, M. Boehnke, S. Raychaudhuri, R. S. N. Fehrmann, J. N. Hirschhorn, L. Franke, Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
11. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F. Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
12. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
13. Y. Momozawa, J. Dmitrieva, E. Théâtre, V. Deffontaine, S. Rahmouni, B. Charloteaux, F. Crins, E. Docampo, M. Elansary, A.-S. Gori, C. Lecut, R. Mariman, M. Mni, C. Oury, I. Altukhov, D. Alexeev, Y. Aulchenko, L. Amininejad, G. Bouma, F. Hoentjen, M. Löwenberg, B. Oldenburg, M. J. Pierik, A. E. vander Meulen-de Jong, C. Janneke van der Woude, M. C. Visschedijk; The International IBD Genetics Consortium, M. Lathrop, J.-P. Hugot, R. K. Weersma, M. De Vos, D. Franchimont, S. Vermeire, M. Kubo, E. Louis, M. Georges, IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
14. H.-J. Westra, M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. C. 't Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Völker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. J. van Meurs, L. Franke, Systematic

J. Kristic, J. Simunovic, A. Momcilovic, H. Campbell, M. Doherty, M. G. Dunlop, S. M. Farrington, M. Pucic-Bakovic, C. Gieger, M. Allegri, E. Louis, M. Georges, K. Suhre, T. Spector, F. M. K. Williams, G. Lauc, Y. S. Aulchenko, Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum. Mol. Genet.* **28**, 2062–2077 (2019).
18. M. Thomas-Chollier, C. Herrmann, M. Defrance, O. Sand, D. Thieffry, J. van Helden, RSAT peak-motifs: Motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res.* **40**, e31 (2012).
19. L. C. Spender, H. J. Whiteman, C. E. Karstegl, P. J. Farrell, Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. *Oncogene* **24**, 1873–1881 (2005).
20. B. Morgan, L. Sun, N. Avitahl, K. Andrikopoulos, T. Ikeda, E. Gonzales, P. Wu, S. Neben, K. Georgopoulos, Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004–2013 (1997).
21. I. Brockhausen, H. Schachter, in *Glycosyltransferases Involved in N- and O-Glycan biosynthesis*, H. Gabius, S. Gabius, Eds. (Chapman & Hall, 1997), pp. 79–113.
22. Y. Miura, T. Endo, Glycomics and glycoproteomics focused on aging and age-related diseases—Glycans as a potential biomarker for physiological alterations. *Biochim. Biophys. Acta* **1860**, 1608–1614 (2016).
23. E. Maverakis, K. Kim, M. Shimoda, M. E. Gershwin, F. Patel, R. Wilken, S. Raychaudhuri, L. R. Ruhaak, C. B. Lebrilla, Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review. *J. Autoimmun.* **57**, 1–13 (2015).
24. N. Taniguchi, Y. Kizuka, Glycans and cancer: Role of *N*-glycans in cancer biomarker, progression and metastasis, and therapeutics. *Adv. Cancer Res.* **126**, 11–51 (2015).
25. S. Durinck, P. T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
26. J. Krištić, O. O. Zaytseva, R. Ram, Q. Nguyen, M. Novokmet, F. Vučković, M. Vilaj, I. Trbojević-Akmačić, M. Pezer, K. M. Davern, G. Morahan, G. Lauc, Profiling and genetic control of the murine immunoglobulin G glycome. *Nat. Chem. Biol.* **14**, 516–524 (2018).
27. H. H. Freeze, Genetic defects in the human glycome. *Nat. Rev. Genet.* **7**, 537–551 (2006).
28. B. Niebuhr, N. Kriebitzsch, M. Fischer, K. Behrens, T. Günther, M. Alawi, U. Bergholz, U. Müller, S. Roscher, M. Ziegler, F. Buchholz, A. Grundhoff, C. Stocking, Runx1 is essential at two stages of early murine B-cell development. *Blood* **122**, 413–423 (2013).
29. M. Sellars, B. Reina-San-Martin, P. Kastner, S. Chan, Ikaros controls isotype selection during immunoglobulin class switch recombination. *J. Exp. Med.* **206**, 1073–1087 (2009).
30. J. H. Wang, N. Avitahl, A. Cariappa, C. Friedrich, T. Ikeda, A. Renold, K. Andrikopoulos, L. Liang, S. Pillai, B. A. Morgan, K. Georgopoulos, Aiolos regulates B cell activation and maturation to effector state. *Immunity* **9**, 543–553 (1998).
31. C. M. Karsten, M. K. Pandey, J. Figge, R. Kilchenstein, P. R. Taylor, M. Rosas, J. U. McDonald, S. J. Orr, M. Berger, D. Petzold, V. Blanchard, A. Winkler, C. Hess, D. M. Reid, I. V. Majoul, R. T. Strait, N. L. Harris, G. Köhl, E. Wex, R. Ludwig, D. Zillikens, F. Nimmerjahn, F. D. Finkelman, G. D. Brown, M. Ehlers, J. Köhl, Anti-inflammatory activity of IgG1 mediated by Fc galactosylation and association of FcγRIIB and dectin-1. *Nat. Med.* **18**, 1401–1406 (2012).
32. R. L. Shields, J. Lai, R. Keck, L. Y. O'Connell, K. Hong, Y. G. Meng, S. H. A. Weikert, L. G. Presta, Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human FcγRIII and antibody-dependent cellular toxicity. *J. Biol. Chem.* **277**, 26733–26740 (2002).
33. S. Shinzaki, H. Iijima, T. Nakagawa, S. Egawa, S. Nakajima, S. Ishii, T. Irie, Y. Kakiuchi, T. Nishida, M. Yasumaru, T. Kanto, M. Tsujii, S. Tsuji, T. Mizushima, H. Yoshihara, A. Kondo, E. Miyoshi, N. Hayashi, IgG oligosaccharide alterations are a novel diagnostic marker for disease activity and the clinical course of inflammatory bowel disease. *Am. J. Gastroenterol.* **103**, 1173–1181 (2008).
34. M. Šimurina, N. de Haan, F. Vučković, N. A. Kennedy, J. Štambuk, D. Falck, I. Trbojević-Akmačić, F. Clerc, G. Razdorov, A. Khon, A. Latiano, R. D'Incà, S. Danese, S. Targan, C. Landers, M. Dubinsky; Inflammatory Bowel Disease Biomarkers Consortium, D. P. B. McGovern, V. Annese, M. Wuhrer, G. Lauc, Glycosylation of immunoglobulin G associates with clinical features of inflammatory bowel diseases. *Gastroenterology* **154**, 1320–1333.e10 (2018).
35. R. B. Parekh, R. A. Dwek, B. J. Sutton, D. L. Fernandes, A. Leung, D. Stanworth, T. W. Rademacher, T. Mizuochi, T. Taniguchi, K. Matsuta, F. Takeuchi, Y. Nagano, T. Miyamoto, A. Kobata, Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452–457 (1985).
Genetic Investigation of Anthropometric Traits (GIANT) Consortium, Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
39. C. J. Willer, Y. Li, G. R. Abecasis, METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
40. J. Li, L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95**, 221–227 (2005).
41. J. M. Whiteheads Pavlides, Z. Zhu, J. Gratten, A. F. McRae, N. R. Wray, J. Yang, Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Med.* **8**, 84 (2016).
42. B. H. Smith, A. Campbell, P. Linksted, B. Fitzpatrick, C. Jackson, S. M. Kerr, I. J. Deary, D. J. Macintyre, H. Campbell, M. McGilchrist, L. J. Hocking, L. Wisely, I. Ford, R. S. Lindsay, R. Morton, C. N. A. Palmer, A. F. Dominiczak, D. J. Porteous, A. D. Morris, Cohort profile: Generation Scotland: Scottish family health study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
43. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
44. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
45. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
46. J. E. Huffman, M. Pučić-Baković, L. Klarić, R. Hennig, M. H. J. Selman, F. Vučković, M. Novokmet, J. Krištić, M. Borowiak, T. Muth, O. Polašek, G. Razdorov, O. Gornik, R. Plomp, E. Theodoratou, A. F. Wright, I. Rudan, C. Hayward, H. Campbell, A. M. Deelder, U. Reichl, Y. S. Aulchenko, E. Rapp, M. Wuhrer, G. Lauc, Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. *Mol. Cell. Proteomics* **13**, 1598–1610 (2014).
47. J. Krištić, F. Vučković, C. Menni, L. Klarić, T. Keser, I. Beceheli, M. Pučić-Baković, M. Novokmet, M. Mangino, K. Thaqi, P. Rudan, N. Novokmet, J. Sarac, S. Missoni, I. Kolčić, O. Polašek, I. Rudan, H. Campbell, C. Hayward, Y. Aulchenko, A. Valdes, J. F. Wilson, O. Gornik, D. Primorac, V. Zoldoš, T. Spector, G. Lauc, Glycans are a novel biomarker of chronological and biological ages. *J. Gerontol. A Biol. Sci. Med. Sci.* **69**, 779–789 (2014).
48. C. Menni, T. Keser, M. Mangino, J. T. Bell, I. Erte, I. Akmačić, F. Vučković, M. Pučić Baković, O. Gornik, M. I. McCarthy, V. Zoldoš, T. D. Spector, G. Lauc, A. M. Valdes, Glycosylation of immunoglobulin G: Role of genetic and epigenetic influences. *PLOS ONE* **8**, e82558 (2013).
49. I. Trbojević Akmačić, I. Ugrina, G. Lauc, Comparative analysis and validation of different steps in glycomics studies, in *Methods in Enzymology, Volume 586: Chapter Three*, A. K. Shukla, Eds. (Elsevier, Netherlands, 2017).
50. A. Moayyeri, C. J. Hammond, D. J. Hart, T. D. Spector, The UK Adult Twin Registry (TwinsUK Resource). *Twin Res. Hum. Genet.* **16**, 144–149 (2012).
51. T. D. Spector, F. M. K. Williams, The UK adult twin registry (TwinsUK). *Twin Res. Hum. Genet.* **9**, 899–906 (2006).
52. R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright, H. Campbell, J. F. Wilson, Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).
53. I. Rudan, A. Marušić, S. Janković, K. Rotim, M. Boban, G. Lauc, I. Grković, Z. Đogaš, T. Zemunik, Z. Vatavuk, G. Benčić, D. Rudan, R. Mulić, V. Krželj, J. Terzić, D. Stojanović, D. Puntarić, E. Bilić, D. Ropac, A. Vorko-Jović, A. Znaor, R. Stevanović, Z. Biloglav, O. Polašek, "10 001 Dalmatians:" Croatia launches its national biobank. *Croat. Med. J.* **50**, 4–6 (2009).
54. V. Vitart, I. Rudan, C. Hayward, N. K. Gray, J. Floyd, C. N. Palmer, S. A. Knott, I. Kolcic, O. Polasek, J. Graessler, J. F. Wilson, A. Marinaki, P. L. Riches, X. Shu, B. Janicijevic, N. Smolej-Narancic, B. Gorgoni, J. Morgan, S. Campbell, Z. Biloglav, L. Barac-Lauc, M. Pericic, I. M. Klaric, L. Zgaga, T. Skaric-Juric, S. H. Wild, W. A. Richardson, P. Hohenstein,

57. COGENT Study, R. S. Houlston, E. Webb, P. Broderick, A. M. Pittman, M. C. Di Bernardo, S. Lubbe, I. Chandler, J. Vijayakrishnan, K. Sullivan, S. Penegar; Colorectal Cancer Association Study Consortium, L. Carvajal-Carmona, K. Howarth, E. Jaeger, S. L. Spain, A. Walther, E. Barclay, L. Martin, M. Gorman, E. Domingo, A. S. Teixeira; CoRGI Consortium, D. Kerr, J.-B. Cazier, I. Niittymäki, S. Tuupanen, A. Karhu, L. A. Aaltonen, I. P. M. Tomlinson, S. M. Farrington, A. Tenesa, J. G. D. Prendergast, R. A. Barnetson, R. Cetnarskyj, M. E. Porteous, P. D. P. Pharoah, T. Koessler, J. Hampe, S. Buch, C. Schafmayer, J. Tepel, S. Schreiber, H. Völzke, J. Chang-Claude, M. Hoffmeister, H. Brenner, B. W. Zanke, A. Montpetit, T. J. Hudson, S. Gallinger, H. Campbell, M. G. Dunlop, Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.* **40**, 1426–1435 (2008).
58. M. G. Dunlop, S. E. Dobbins, S. M. Farrington, A. M. Jones, C. Palles, N. Whiffin, A. Tenesa, S. Spain, P. Broderick, L.-Y. Ooi, E. Domingo, C. Smillie, M. Henrion, M. Frampton, L. Martin, G. Grimes, M. Gorman, C. Semple, Y. P. Ma, E. Barclay, J. Prendergast, J.-B. Cazier, B. Olver, S. Penegar, S. Lubbe, I. Chander, L. G. Carvajal-Carmona, S. Ballereau, A. Lloyd, J. Vijayakrishnan, L. Zgaga, I. Rudan, E. Theodoratou; Colorectal Tumour Gene Identification (CORGI) Consortium, J. M. Starr, I. Deary, I. Kirac, D. Kovacević, L. A. Aaltonen, L. Renkonen-Sinisalo, J.-P. Mecklin, K. Matsuda, Y. Nakamura, Y. Okada, S. Gallinger, D. J. Duggan, D. Conti, P. Newcomb, J. Hopper, M. A. Jenkins, F. Schumacher, G. Casey, D. Easton, M. Shah, P. Pharoah, A. Lindblom, T. Liu; Swedish Low-Risk Colorectal Cancer Study Group, C. G. Smith, H. West, J. P. Cheadle; COIN Collaborative Group, R. Midgley, D. J. Kerr, H. Campbell, I. P. Tomlinson, R. S. Houlston, Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat. Genet.* **44**, 770–776 (2012).
59. T. Akbar, S. McGurnaghan, C. N. A. Palmer, S. J. Livingstone, J. Petrie, J. Chalmers, R. S. Lindsay, J. A. McKnight, D. W. M. Pearson, A. W. Patrick, J. Walker, H. C. Looker, H. M. Colhoun, Cohort profile: Scottish diabetes research network Type 1 bioresource study (SDRNT1BIO). *Int. J. Epidemiol.* **46**, 796–796i (2016).
60. M. Schoenmaker, A. J. M. de Craen, P. H. E. M. de Meijer, M. Beekman, G. J. Blauw, P. E. Slagboom, R. G. J. Westendorp, Evidence of genetic enrichment for exceptional survival using a family approach: The Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).
61. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
62. I. Ugrina, L. Klaric, Russell Alyce, F. Vuckovic, *glycanr: Tools for Analysing N-Glycan Data* (2016); https://cran.r-project.org/web/packages/glycanr/index.html.
63. Y. Li, C. J. Willer, J. Ding, P. Scheet, G. R. Abecasis, MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
64. J. Marchini, B. Howie, S. Myers, G. McVean, P. Donnelly, A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
65. B. Howie, J. Marchini, M. Stephens, Genotype imputation with thousands of genomes. *G3* **1**, 457–470 (2011).
66. R. Plomp, L. R. Ruhaak, H.-W. Uh, K. R. Reiding, M. Selman, J. J. Houwing-Duistermaat, P. E. Slagboom, M. Beekman, M. Wuhrer, Subclass-specific IgG glycosylation is associated with markers of inflammation and metabolic health. *Sci. Rep.* **7**, 12325 (2017).
67. G. Vidarsson, G. Dekkers, T. Rispens, IgG subclasses and allotypes: From structure to effector functions. *Front. Immunol.* **5**, 520 (2014).
68. A. Bondt, Y. Rombouts, M. H. J. Selman, P. J. Hensbergen, K. R. Reiding, J. M. W. Hazes, R. J. E. M. Dolhain, M. Wuhrer, Immunoglobulin G (IgG) Fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol. Cell. Proteomics* **13**, 3029–3039 (2014).
69. International Genetics of Ankylosing Spondylitis Consortium (IGAS), A. Cortes, J. Hadler, J. P. Pointon, P. C. Robinson, T. Karaderi, P. Leo, K. Cremin, K. Pryce, J. Harris, S. Lee, K. B. Joo, S.-C. Shim, M. Weisman, M. Ward, X. Zhou, H.-J. Garchon, G. Chiocchia, J. Nossent, B. A. Lie, Ø. Førre, J. Tuomilehto, K. Laiho, L. Jiang, Y. Liu, X. Wu, L. A. Bradbury, D. Elewaut, R. Burgos-Vargas, S. Stebbings, L. Appleton, C. Farrah, J. Lau, T. J. Kenna, N. Haroon, M. A. Ferreira, J. Yang, J. Mulero, J. L. Fernandez-Sueiro, M. A. Gonzalez-Gay, C. Lopez-Larrea, P. Deloukas, P. Donnelly; Australo-Anglo-American Spondyloarthritis Consortium (TASC); Groupe Française d'Etude Génétique des Spondylarthrites (GFEGS); Nord-Trøndelag Health Study (HUNT); Spondyloarthritis Research Consortium of Canada (SPARCC); Wellcome Trust Case Control Consortium 2 (WTCCC2), P. Bowness, K. Gafney,
71. J. Z. Liu, S. van Sommeren, H. Huang, S. C. Ng, R. Alberts, A. Takahashi, S. Ripke, J. C. Lee, L. Jostins, T. Shah, S. Abedian, J. H. Cheon, J. Cho, N. E. Daryani, L. Franke, Y. Fuyuno, A. Hart, R. C. Juyal, G. Juyal, W. H. Kim, A. P. Morris, H. Poustchi, W. G. Newman, V. Midha, T. R. Orchard, H. Vahedi, A. Sood, J. J. Y. Sung, R. Malekzadeh, H.-J. Westra, K. Yamazaki, S.-K. Yang; International Multiple Sclerosis Genetics Consortium; International IBD Genetics Consortium, J. C. Barrett, A. Franke, B. Z. Alizadeh, M. Parkes, T. B K, M. J. Daly, M. Kubo, C. A. Anderson, R. K. Weersma, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
72. C. J. Willer, E. M. Schmidt, S. Sengupta, G. M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, M. L. Buchkovich, S. Mora, J. S. Beckmann, J. L. Bragg-Gresham, H.-Y. Chang, A. Demirkan, H. M. Den Hertog, R. Do, L. A. Donnelly, G. B. Ehret, T. Esko, M. F. Feitosa, T. Ferreira, K. Fischer, P. Fontanillas, R. M. Fraser, D. F. Freitag, D. Gurdasani, K. Heikkilä, E. Hyppönen, A. Isaacs, A. U. Jackson, Å. Johansson, T. Johnson, M. Kaakinen, J. Kettunen, M. E. Kleber, X. Li, J. Luan, L.-P. Lyytikäinen, P. K. E. Magnusson, M. Mangino, E. Mihailov, M. E. Montasser, M. Müller-Nurasyid, I. M. Nolte, J. R. O'Connell, C. D. Palmer, M. Perola, A.-K. Petersen, S. Sanna, R. Saxena, S. K. Service, S. Shah, D. Shungin, C. Sidore, C. Song, R. J. Strawbridge, I. Surakka, T. Tanaka, T. M. Teslovich, G. Thorleifsson, E. G. Van den Herik, B. F. Voight, K. A. Volcik, L. L. Waite, A. Wong, Y. Wu, W. Zhang, D. Absher, G. Asiki, I. Barroso, L. F. Been, J. L. Bolton, L. L. Bonnycastle, P. Brambilla, M. S. Burnett, G. Cesana, M. Dimitriou, A. S. F. Doney, A. Döring, P. Elliott, S. E. Epstein, G. I. Eyjolfsson, B. Gigante, M. O. Goodarzi, H. Grallert, M. L. Gravito, C. J. Groves, G. Hallmans, A.-L. Hartikainen, C. Hayward, D. Hernandez, A. A. Hicks, H. Holm, Y.-J. Hung, T. Illig, M. R. Jones, P. Kaleebu, J. J. P. Kastelein, K.-T. Khaw, E. Kim, N. Klopp, P. Komulainen, M. Kumari, C. Langenberg, T. Lehtimäki, S.-Y. Lin, J. Lindström, R. J. F. Loos, F. Mach, W. L. McArdle, C. Meisinger, B. D. Mitchell, G. Müller, R. Nagaraja, N. Narisu, T. V. M. Nieminen, R. N. Nsubuga, I. Olafsson, K. K. Ong, A. Palotie, T. Papamarkou, C. Pomilla, A. Pouta, D. J. Rader, M. P. Reilly, P. M. Ridker, F. Rivadeneira, I. Rudan, A. Ruokonen, N. Samani, H. Scharnagl, J. Seeley, K. Silander, A. Stančáková, K. Stirrups, A. J. Swift, L. Tiret, A. G. Uitterlinden, L. J. van Pelt, S. Vedantam, N. Wainwright, C. Wijmenga, S. H. Wild, G. Willemsen, T. Wilsgaard, J. F. Wilson, E. H. Young, J. H. Zhao, L. S. Adair, D. Arveiler, T. L. Assimes, S. Bandinelli, F. Bennett, M. Bochud, B. O. Boehm, D. I. Boomsma, I. B. Borecki, S. R. Bornstein, P. Bovet, M. Burnier, H. Campbell, A. Chakravarti, J. C. Chambers, Y.-D. I. Chen, F. S. Collins, R. S. Cooper, J. Danesh, G. Dedoussis, U. de Faire, A. B. Feranil, J. Ferrières, L. Ferrucci, N. B. Freimer, C. Gieger, L. C. Groop, V. Gudnason, U. Gyllensten, A. Hamsten, T. B. Harris, A. Hingorani, J. N. Hirschhorn, A. Hofman, G. K. Hovingh, C. A. Hsiung, S. E. Humphries, S. C. Hunt, K. Hveem, C. Iribarren, M.-R. Järvelin, A. Jula, M. Kähönen, J. Kaprio, A. Kesäniemi, M. Kivimaki, J. S. Kooner, P. J. Koudstaal, R. M. Krauss, D. Kuh, J. Kuusisto, K. O. Kyvik, M. Laakso, T. A. Lakka, L. Lind, C. M. Lindgren, N. G. Martin, W. März, M. I. McCarthy, C. A. McKenzie, P. Meneton, A. Metspalu, L. Moilanen, A. D. Morris, P. B. Munroe, I. Njølstad, N. L. Pedersen, C. Power, P. P. Pramstaller, J. F. Price, B. M. Psaty, T. Quertermous, R. Rauramaa, D. Saleheen, V. Salomaa, D. K. Sanghera, J. Saramies, P. E. H. Schwarz, W. H.-H. Sheu, A. R. Shuldiner, A. Siegbahn, T. D. Spector, K. Stefansson, D. P. Strachan, B. O. Tayo, E. Tremoli, J. Tuomilehto, M. Uusitupa, C. M. van Duijn, P. Vollenweider, L. Wallentin, N. J. Wareham, J. B. Whitfield, B. H. R. Wolffenbuttel, J. M. Ordovas, E. Boerwinkle, C. N. A. Palmer, U. Thorsteinsdottir, D. I. Chasman, J. I. Rotter, P. W. Franks, S. Ripatti, L. A. Cupples, M. S. Sandhu, S. S. Rich, M. Boehnke, P. Deloukas, S. Kathiresan, K. L. Mohlke, E. Ingelsson, G. R. Abecasis; Global Lipids Genetics Consortium, Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
73. A. R. Wood, T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, N. Amin, M. L. Buchkovich, D. C. Croteau-Chonka, F. R. Day, Y. Duan, T. Fall, R. Fehrmann, T. Ferreira, A. U. Jackson, J. Karjalainen, K. S. Lo, A. E. Locke, R. Mägi, E. Mihailov, E. Porcu, J. C. Randall, A. Scherag, A. A. E. Vinkhuyzen, H.-J. Westra, T. W. Winkler, T. Workalemahu, J. H. Zhao, D. Absher, E. Albrecht, D. Anderson, J. Baron, M. Beekman, A. Demirkan, G. B. Ehret, B. Feenstra, M. F. Feitosa, K. Fischer, R. M. Fraser, A. Goel, J. Gong, A. E. Justice, S. Kanoni, M. E. Kleber, K. Kristiansson, U. Lim, V. Lotay, J. C. Lui, M. Mangino, I. Mateo Leach, C. Medina-Gomez, M. A. Nalls, D. R. Nyholt, C. D. Palmer, D. Pasko, S. Pechlivanis, I. Prokopenko, J. S. Ried, S. Ripke, D. Shungin, A. Stancáková, R. J. Strawbridge, Y. J. Sung, T. Tanaka, A. Teumer, S. Trompet, S. W. van der Laan, J. van Setten, J. V. Van Vliet-Ostaptchouk, Z. Wang, L. Yengo, W. Zhang,

J. Lindström, S. Lobbens, M. Lorentzon, Y. Lu, V. Lyssenko, P. K. E. Magnusson, A. Mahajan, M. Maillard, W. L. McArdle, C. A. McKenzie, S. McLachlan, P. J. McLaren, C. Menni, S. Merger, L. Milani, A. Moayyeri, K. L. Monda, M. A. Morken, G. Müller, M. Müller-Nurasyid, A. W. Musk, N. Narisu, M. Nauck, I. M. Nolte, M. M. Nöthen, L. Oozageer, S. Pilz, N. W. Rayner, F. Renstrom, N. R. Robertson, L. M. Rose, R. Roussel, S. Sanna, H. Scharnagl, S. Scholtens, F. R. Schumacher, H. Schunkert, R. A. Scott, J. Sehmi, T. Seufferlein, J. Shi, K. Silventoinen, J. H. Smit, A. V. Smith, J. Smolonska, A. V. Stanton, K. Stirrups, D. J. Stott, H. M. Stringham, J. Sundström, M. A. Swertz, A.-C. Syvänen, B. O. Tayo, G. Thorleifsson, J. P. Tyrer, S. van Dijk, N. M. van Schoor, N. van der Velde, D. van Heemst, F. V. A. van Oort, S. H. Vermeulen, N. Verweij, J. M. Vonk, L. L. Waite, M. Waldenberger, R. Wennauer, L. R. Wilkens, C. Willenborg, T. Wilsgaard, M. K. Wojczynski, A. Wong, A. F. Wright, Q. Zhang, D. Arveiler, S. J. L. Bakker, J. Beilby, R. N. Bergman, S. Bergmann, R. Biffar, J. Blangero, D. I. Boomsma, S. R. Bornstein, P. Bovet, P. Brambilla, M. J. Brown, H. Campbell, M. J. Caulfield, A. Chakravarti, R. Collins, F. S. Collins, D. C. Crawford, L. A. Cupples, J. Danesh, U. de Faire, H. M. den Ruijter, R. Erbel, J. Erdmann, J. G. Eriksson, M. Farrall, E. Ferrannini, J. Ferrières, I. Ford, N. G. Forouhi, T. Forrester, R. T. Gansevoort, P. V. Gejman, C. Gieger, A. Golay, O. Gottesman, V. Gudnason, U. Gyllensten, D. W. Haas, A. S. Hall, T. B. Harris, A. T. Hattersley, A. C. Heath, C. Hengstenberg, A. A. Hicks, L. A. Hindorff, A. D. Hingorani, A. Hofman, G. K. Hovingh, S. E. Humphries, S. C. Hunt, E. Hypponen, K. B. Jacobs, M.-R. Jarvelin, P. Jousilahti, A. M. Jula, J. Kaprio, J. J. P. Kastelein, M. Kayser, F. Kee, S. M. Keinanen-Kiukaanniemi, L. A. Kiemeney, J. S. Kooner, C. Kooperberg, S. Koskinen, P. Kovacs, A. T. Kraja, M. Kumari, J. Kuusisto, T. A. Lakka, C. Langenberg, L. Le Marchand, T. Lehtimäki, S. Lupoli, P. A. F. Madden, S. Männistö, P. Manunta, A. Marette, T. C. Matise, B. McKnight, T. Meitinger, F. L. Moll, G. W. Montgomery, A. D. Morris, A. P. Morris, J. C. Murray, M. Nelis, C. Ohlsson, A. J. Oldehinkel, K. K. Ong, W. H. Ouwehand, G. Pasterkamp, A. Peters, P. P. Pramstaller, J. F. Price, L. Qi, O. T. Raitakari, T. Rankinen, D. C. Rao, T. K. Rice, M. Ritchie, I. Rudan, V. Salomaa, N. J. Samani, J. Saramies, M. A. Sarzynski, P. E. H. Schwarz, S. Sebert, P. Sever, A. R. Shuldiner, J. Sinisalo, V. Steinthorsdottir, R. P. Stolk, J.-C. Tardif, A. Tönjes, A. Tremblay, E. Tremoli, J. Virtamo, M.-C. Vohl; Electronic Medical Records and Genomics (eMERGE) Consortium; MIGen Consortium; PAGE Consortium; LifeLines Cohort Study, P. Amouyel, F. W. Asselbergs, T. L. Assimes, M. Bochud, B. O. Boehm, E. Boerwinkle, E. P. Bottinger, C. Bouchard, S. Cauchi, J. C. Chambers, S. J. Chanock, R. S. Cooper, P. I. W. de Bakker, G. Dedoussis, L. Ferrucci, P. W. Franks, P. Froguel, L. C. Groop, C. A. Haiman, A. Hamsten, M. G. Hayes, J. Hui, D. J. Hunter, K. Hveem, J. W. Jukema, R. C. Kaplan, M. Kivimaki, D. Kuh, M. Laakso, Y. Liu, N. G. Martin, W. März, M. Melbye, S. Moebus, P. B. Munroe, I. Njølstad, B. A. Oostra, C. N. A. Palmer, N. L. Pedersen, M. Perola, L. Pérusse, U. Peters, J. E. Powell, C. Power, T. Quertermous, R. Rauramaa, E. Reinmaa, P. M. Ridker, F. Rivadeneira, J. I. Rotter, T. E. Saaristo, D. Saleheen, D. Schlessinger, P. E. Slagboom, H. Snieder, T. D. Spector, K. Strauch, M. Stumvoll, J. Tuomilehto, M. Uusitupa, P. van der Harst, H. Völzke, M. Walker, N. J. Wareham, H. Watkins, H.-E. Wichmann, J. F. Wilson, P. Zanen, P. Deloukas, I. M. Heid, C. M. Lindgren, K. L. Mohlke, E. K. Speliotes, U. Thorsteinsdottir, I. Barroso, C. S. Fox, K. E. North, D. P. Strachan, J. S. Beckmann, S. I. Berndt, M. Boehnke, I. B. Borecki, M. I. McCarthy, A. Metspalu, K. Stefansson, A. G. Uitterlinden, C. M. van Duijn, L. Franke, C. J. Willer, A. L. Price, G. Lettre, R. J. F. Loos, M. N. Weedon, E. Ingelsson, J. R. O'Connell, G. R. Abecasis, D. I. Chasman, M. E. Goddard, P. M. Visscher, J. N. Hirschhorn, T. M. Frayling, Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

74. N. Pankratz, G. W. Beecham, A. L. DeStefano, T. M. Dawson, K. F. Doheny, S. A. Factor, T. H. Hamza, A. Y. Hung, B. T. Hyman, A. J. Ivinson, D. Krainc, J. C. Latourelle, L. N. Clark, K. Marder, E. R. Martin, R. Mayeux, O. A. Ross, C. R. Scherzer, D. K. Simon, C. Tanner, J. M. Vance, Z. K. Wszolek, C. P. Zabetian, R. H. Myers, H. Payami, W. K. Scott, T. Foroud; PD GWAS Consortium, Meta-analysis of Parkinson's Disease: Identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).

75. H. J. Cordell, Y. Han, G. F. Mells, Y. Li, G. M. Hirschfield, C. S. Greene, G. Xie, B. D. Juran, D. Zhu, D. C. Qian, J. A. B. Floyd, K. I. Morley, D. Prati, A. Lleo, D. Cusi; Canadian-US PBC Consortium; Italian PBC Genetics Study Group; UK-PBC Consortium, M. E. Gershwin, C. A. Anderson, K. N. Lazaridis, P. Invernizzi, M. F. Seldin, R. N. Sandford, C. I. Amos, K. A. Siminovitch, International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* **6**, 8019 (2015).

E. W. Karlson, A. Taniguchi, R. Yamada, M. Kubo, J. S. Liu, S.-C. Bae, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, S. Raychaudhuri, B. E. Stranger, P. L. De Jager, L. Franke, P. M. Visscher, M. A. Brown, H. Yamanaka, T. Mimori, A. Takahashi, H. Xu, T. W. Behrens, K. A. Siminovitch, S. Momohara, F. Matsuda, K. Yamamoto, R. M. Plenge, Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376–381 (2014).

77. P. Kheradpour, M. Kellis, Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* **42**, 2976–2987 (2014).

78. A. Mathelier, X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, W. W. Wasserman, JASPAR 2014: An extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**, D142–D147 (2014).

79. I. V. Kulakovskiy, I. E. Vorontsov, I. S. Yevshin, R. N. Sharipov, A. D. Fedorova, E. I. Rumynskiy, Y. A. Medvedeva, A. Magana-Mora, V. B. Bajic, D. A. Papatsenko, F. A. Kolpakov, V. J. Makeev, HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* **46**, D252–D259 (2018).

80. J.-V. Turatsinze, M. Thomas-Chollier, M. Defrance, J. van Helden, Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. *Nat. Protoc.* **3**, 1578–1588 (2008).

81. D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. García Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).

82. N.-L. Sim, P. Kumar, J. Hu, S. Henikoff, G. Schneider, P. C. Ng, SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).

83. S. Das, M. Miller, D. H. Broide, (2017); https://linkinghub.elsevier.com/retrieve/pii/S006527761730038X, pp. 1–52.

Acknowledgments: We thank S. Kerr for proofreading and for the useful comments on the manuscript, and J. E. Huffman for sharing knowledge and experience with glycosylation GWAS. ORCADES DNA extractions and genotyping were performed at the Genetics Core of the Clinical Research Facility, University of Edinburgh. We acknowledge the invaluable contributions of the research nurses in Orkney, the administrative team in Edinburgh, and the people of Orkney. COGS thanks the participants in all the studies who contributed to this piece of work and all the recruitment teams and collaborators who made these studies possible. We acknowledge the excellent technical support from M. Walker, and we are grateful to R. Wilson, D. Markie, and all those who continue to contribute to recruitment, data collection, and data curation for the Study of Colorectal Cancer in Scotland studies. The work was supported by funding for the infrastructure and staffing of the Edinburgh CRUK Cancer Research Centre. We acknowledge the expert support on sample preparation from the Genetics Core of the Clinical Research Facility, University of Edinburgh. We also acknowledge all the staff of several institutions in Croatia that supported the CROATIA_Vis and CROATIA_Korcula, fieldwork, including, but not limited to, the University of Split and Zagreb Medical Schools, Institute for Anthropological Research in Zagreb, and the Croatian Institute for Public Health in Split. Genotyping was performed in the Genetics Core of the Clinical Research Facility, University of Edinburgh. We are grateful to the Russian Federal Science and Technology Program of Development of Genetic Technologies. **Funding:** The work of L.K. was supported by the European Commission MIMOmics (contract #305280) and an RCUK Innovation Fellowship from the National Productivity Investment Fund (MR/R026408/1). The work of Y.A.T. was supported by the Russian Ministry of Education and Science under the 5-100 Excellence Programme and also was supported by the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics SB RAS (project number 0324-2019-0040). The work of Y.S.A. was funded by PolyOmica and also was supported by the

European Union Framework Programme 6 EUROSPAN project (contract number LSHG-CT-2006-018947). The CROATIA_Vis and CROATIA_Korcula studies were funded by grants from the Medical Research Council (United Kingdom), European Commission Framework 6 project EUROSPAN (contract number LSHG-CT-2006-018947), Croatian Science Foundation (grant 8875), and the Republic of Croatia Ministry of Science, Education and Sports (216-1080315-0302). EGCUT was funded by Estonian Research Council Grant IUT20-60 and PUT1660, EU H2020 grant 692145, and European Union through the European Regional Development Fund (project number 2014-2020.4.01.15-0012) GENTRANSMED. The Leiden Longevity Study was financially supported by the European Union (Seventh Framework Programme HighGlycan project grant number 278535, IDEAL grant number 259679, and MIMOmics grant number 305280), the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Netherlands Consortium for Healthy Ageing (grants 05040202 and 050-060-810), in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO) Unilever Colworth, and by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007 and 184.033.111). Glycan analysis performed at Genos Ltd. was supported by the HighGlycan FP7 grant (contract #278535), the European Structural and Investment Funds IRI grant (#KK.01.2.1.01.0003), and the Croatian National Centre of Research Excellence in Personalized Healthcare grant (number KK.01.1.1.01.0010). The work of M.G.D. was supported by the C348/A18927 grant. **Author contributions:** L.K. and Y.A.T. contributed to the design of the study, carried out the statistical analyses, produced the figures, and interpreted the results; C.M.S., C.D., J.K., P.D., J.M., and B.J. performed the in vitro functional follow-up experiments and interpreted the results; L.K., Y.A.T., E.P., and S.Z.S. performed in silico functional follow-up; L.K., Y.A.T., C.M.S., G.L., C.H., and Y.S.A. wrote the manuscript; L.K., A.K.G., Y.A.T., and S.Z.S. contributed to the GWAS quality control and meta-analysis; F.V., I.U., L.K., H.-W.U., and R.P. contributed to the phenotype preprocessing; T.K., I.G., J.Š., J.K., R.P., M.P.-B., T.P., M.V., and I.T.-A. contributed to the IgG N-glycome measurements; L.K., M.M., T.T.S., T.E., P.S., J.De., M.L.B., S.J.M., and M.T. performed the specific cohort GWAS and contributed to the interpretation of the results; A.R. contributed to the bioinformatics analyses; L.K. prepared the shinyApp; S.M.F., E.T., H.-W.U., M.B., E.P.S., M.W., H.M.C., M.G.D., M.P., K.F., O.P., H.C., I.R., J.F.W., V.V., V.Z., T.S., Y.S.A., G.L., and C.H. designed the specific studies and contributed to the interpretation of the results; E.L., J.Dm., and M.G. designed the CEDAR study and contributed to the interpretation of the results; Y.S.A., G.L., and C.H. conceived and oversaw the study and contributed to the design and interpretation of the results; and all coauthors contributed to the final manuscript. **Competing interests:** Y.S.A. is a founder and co-owner of PolyOmica, a private organization providing services, research, and development in the field of computational and statistical (gen)omics. G.L. is a founder and owner of the Genos Glycoscience Research Laboratory, and I.G., J.Š., J.K., M.P.-B., M.V., I.T.-A., F.V., and I.U. are employees of the said company. The other authors declare that they have no competing interests. **Data and materials availability:** The datasets supporting the conclusions of this article are included within the article, and its additional files and interactive shinyApp are hosted at https://shiny.igmm.ed.ac.uk/igg_glycans_gwas/. Full summary statistics from the IgG N-glycome GWAS for 77 glycan traits are available at the DataShare (https://datashare.is.ed.ac.uk/handle/10283/705 and https://doi.org/10.7488/ds/2481) and for the interactive exploration at the GWAS archive (http://gwasarchive.org). Additional data related to this paper may be requested from the authors.

Submitted 15 February 2019
Accepted 19 November 2019
Published 19 February 2020
10.1126/sciadv.aax0301

Citation: L. Klarić, Y. A. Tsepilov, C. M. Stanton, M. Mangino, T. T. Sikka, T. Esko, E. Pakhomov, P. Salo, J. Deelen, S. J. McGurnaghan, T. Keser, F. Vučković, I. Ugrina, J. Krištić, I. Gudelj, J. Štambuk, R. Plomp, M. Pučić-Baković, T. Pavić, M. Vilaj, I. Trbojević-Akmačić, C. Drake, P. Dobrinić, J. Mlinarec, B. Jelušić, A. Richmond, M. Timofeeva, A. K. Grishchenko, J. Dmitrieva, M. L. Bermingham, S. Z. Sharapov, S. M. Farrington, E. Theodoratou, H.-W. Uh, M. Beekman, E. P. Slagboom, E. Louis, M. Georges, M. Wuhrer, H. M. Colhoun, M. G. Dunlop, M. Perola, K. Fischer, O. Polasek, H. Campbell, I. Rudan, J. F. Wilson, V. Zoldoš, V. Vitart, T. Spector, Y. S. Aulchenko, G. Lauc, C. Hayward, Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. *Sci. Adv.* **6**, eaax0301 (2020).

ScienceAdvances

Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases

Lucija Klaric, Yakov A. Tsepilov, Chloe M. Stanton, Massimo Mangino, Timo Tõnis Sikka, Tõnu Esko, Eugene Pakhomov, Perttu Salo, Joris Deelen, Stuart J. McGurnaghan, Toma Keser, Frano Vuckovic, Ivo Ugrina, Jasminka Kristic, Ivan Gudelj, Jerko Stambuk, Rosina Plomp, Maja Pucic-Bakovic, Tamara Pavic, Marija Vilaj, Irena Trbojevic-Akmacic, Camilla Drake, Paula Dobrinic, Jelena Mlinarec, Barbara Jelusic, Anne Richmond, Maria Timofeeva, Alexander K. Grishchenko, Julia Dmitrieva, Mairead L. Bermingham, Sodbo Zh. Sharapov, Susan M. Farrington, Evropi Theodoratou, Hae-Won Uh, Marian Beekman, Eline P. Slagboom, Edouard Louis, Michel Georges, Manfred Wuhrer, Helen M. Colhoun, Malcolm G. Dunlop, Markus Perola, Krista Fischer, Ozren Polasek, Harry Campbell, Igor Rudan, James F. Wilson, Vlatka Zoldos, Veronique Vitart, Tim Spector, Yurii S. Aulchenko, Gordan Lauc and Caroline Hayward

Sci Adv **6** (8), eaax0301.
DOI: 10.1126/sciadv.aax0301