Mapping of the gene network that regulates glycan clock of ageing

(article starts on next page)

Mapping of the gene network that regulates glycan clock of ageing

Azra Frkatović-Hodžić[1,*], Anika Mijakovac[2,*], Karlo Miškec[2,*], Arina Nostaeva[3], Sodbo Z. Sharapov[4], Arianna Landini[5], Toomas Haller[6], Erik van den Akker[7,8], Sapna Sharma[9,11], Rafael R. C. Cuadrat[10,11], Massimo Mangino[12,13], Yong Li[14], Toma Keser[15], Najda Rudman[15], Tamara Štambuk[1], Maja Pučić-Baković[1], Irena Trbojević-Akmačić[1], Ivan Gudelj[1,16], Jerko Štambuk[1], Tea Pribić[1], Barbara Radovani[1,16], Petra Tominac[1], Krista Fischer[6,17], Marian Beekman[7], Manfred Wuhrer[18], Christian Gieger[10,11], Matthias B. Schulze[11,19,20], Clemens Wittenbecher[19,21,22], Ozren Polasek[23,24], Caroline Hayward[25], James F. Wilson[5,25], Tim D. Spector[12], Anna Köttgen[14,26], Frano Vučković[1], Yurii S. Aulchenko[4,27], Aleksandar Vojta[2], Jasminka Krištić[1], Lucija Klarić[25,*], Vlatka Zoldoš[2,*], Gordan Lauc[1,15,*]

[1]Genos Glycoscience Research Laboratory, Zagreb, Croatia
[2]Department of Biology, Division of Molecular Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia
[3]Laboratory of Theoretical and Applied Functional Genomics, Novosibirsk State University, Novosibirsk, Russia
[4]MSU Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow, Russia
[5]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK
[6]Institute of Genomics, University of Tartu, Tartu, Estonia
[7]Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
[8]Department of Pattern Recognition and Bioinformatics, Delft University of Technology, Delft, The Netherlands
[9]Research Unit Molecular Endocrinology and Metabolism, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health (GmbH), Neuherberg, Germany
[10]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München –Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Munich, Germany
[11]German Center for Diabetes Research (DZD), Neuherberg, Germany
[12]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
[13]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK
[14]Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany
[15]Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia
[16]Department of Biotechnology, University of Rijeka, Rijeka, Croatia
[17]Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia
[18]Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands
[19]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany
[20]Institute of Nutritional Science, University of Potsdam, Nuthetal, Germany
[21]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
[22]SciLifeLab, Division of Food and Nutrition Science, Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden
[23]University of Split School of Medicine, Split, Croatia
[24]Algebra University College, Zagreb, Croatia
[25]MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
[26]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
[27]Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
*Equal contribution

Correspondence to: Azra Frkatović-Hodžić, Gordan Lauc; **email:** afrkatovic@genos.hr, glauc@genos.hr

ABSTRACT

Glycans are an essential structural component of immunoglobulin G (IgG) that modulate its structure and function. However, regulatory mechanisms behind this complex posttranslational modification are not well known. Previous genome-wide association studies (GWAS) identified 29 genomic regions involved in regulation of IgG glycosylation, but only a few were functionally validated. One of the key functional features of IgG glycosylation is the addition of galactose (galactosylation), a trait which was shown to be associated with ageing. We performed GWAS of IgG galactosylation (N=13,705) and identified 16 significantly associated loci, indicating that IgG galactosylation is regulated by a complex network of genes that extends beyond the galactosyltransferase enzyme that adds galactose to IgG glycans. Gene prioritization identified 37 candidate genes. Using a recently developed CRISPR/dCas9 system we manipulated gene expression of candidate genes in the *in vitro* IgG expression system. Upregulation of three genes, *EEF1A1*, *MANBA* and *TNFRSF13B*, changed the IgG glycome composition, which confirmed that these three genes are involved in IgG galactosylation in this *in vitro* expression system.

INTRODUCTION

Glycosylation is a posttranslational modification characterized by the attachment of oligosaccharide chains (glycans) to proteins or lipids [1]. Glycans linked to immunoglobulin G (IgG) are essential regulators of effector functions of both native and therapeutic monoclonal antibodies. Alternative glycosylation affects the conformation of the Fc region changing the binding affinity of IgG to downstream immune effectors. Additionally, when deglycosylated, IgG antibodies completely lose their activity in the immune system [2, 3]. Glycosylated IgG can have pro- and anti-inflammatory activity depending on the presence of specific glycan traits. This is exemplified by the removal of core fucose which modulates the affinity of IgG to FcγRIIIa receptor boosting the inflammatory potential by several folds [4, 5]. The presence of galactose (one (G1) or two (G2) galactose moieties) on IgG glycans also has an important role in the downstream immune responses mediated by IgG antibodies. Most of the new functional studies show that galactosylated IgG has increased complement and FcγR activity [6–10]. van Osch et al. [8] even proposed a mechanism where IgG galactosylation promotes the hexamerization of IgG antibodies thus enhancing the activation of the classical complement pathway. However, dozens of population studies performed on thousands of individuals demonstrated that agalactosylated IgG (zero (G0) galactose moieties) associates with inflammation, both in disease but also in advanced age (for a review see [11, 12] and references herein). The discovery that IgG galactosylation decreases with aging enabled the development of the glycan clock of ageing [13]. It was proposed that agalactosylated IgG which increases with age acts as an effector of pro-inflammatory pathological changes and therefore can be exploited not only as a biomarker, but also functional effector of ageing [14, 15].

At the population level, galactosylation is the most variable trait in the human IgG glycome [16, 17], with agalactosylated glycans comprising 6%-50% of the total IgG glycome in healthy subjects [18]. Galactosylation is also the most variable glycan trait between different strains of mice [19]. Elevated levels of agalactosylated IgG glycans were first observed in rheumatoid arthritis [20], and followed by similar discoveries in various autoimmune conditions, such as systemic lupus erythematosus (SLE), inflammatory bowel disease, active spondyloarthropathy, autoimmune vasculitis and adult periodontal disease [21–25]. Changes in levels of galactosylated IgG also occur in infectious diseases such as COVID-19 [26]. In addition, decreased IgG galactosylation levels were observed in cancer patients (for a review see [11] and references herein), potentially reflecting the defensive inflammatory response to cancer [27] and acute-phase response processes involved in cancer progression [28]. However, the question remains whether the agalactosylated IgG is just a biomarker, or whether it can functionally contribute to disease activity.

Several studies have demonstrated that the regulation of IgG glycosylation is largely influenced by genetics. Up to 75% of the variance in some IgG glycan traits can be explained by the genetic component [29]. The genetic determinants of populational variation in IgG glycosylation were explored by genome-wide association studies (GWAS) which identified at least 29 candidate genomic loci [30–34]. However, the involvement of the specific genes and their potential interactions to generate differential IgG galactosylation are still poorly understood. Here, we conducted a GWAS of IgG galactosylation phenotypes in a study that almost doubles the sample size (N=13,705) compared to previous GWAS of IgG N-glycome [33] and focused on the genes with *in silico* evidence for involvement in the IgG galactosylation process. To assess their functional role in IgG galactosylation we applied our recently developed HEK293FS IgG transient expression system based on CRISPR-dCas9 molecular tools to functionally test whether the change in the expression of associated genes affects the IgG galactosylation levels [35].

RESULTS

To enable meta-analysis of GWAS of galactosylation phenotypes measured by different analytical platforms and thereby increase the total number of samples, we first developed a protocol for data harmonisation for glycan data generated using different analytical platforms (UPLC and LC-MS). The protocol was established based on the estimated correlation of the glycan traits derived in one cohort that had glycans measured using both UPLC and LC-MS platforms. The aim was to combine IgG subclass-specific glycan information obtained from LC-MS (i.e. galactosylation measured separately on IgG1-4 subclasses) in an appropriate manner to obtain information corresponding to the total IgG galactosylation measured by UPLC (analysis of glycans from both Fab and Fc fragments from all IgG subclasses). As described in Supplementary Methods, we compared different normalization and subclass information summary approaches. In a subset of individuals that were analysed using both methods, the Pearson correlation coefficients ranged from 0.95 and 0.97 for G0, 0.73 to 0.80 for G1, and 0.90 and 0.91 for G2 trait (Appendix Table 1). Different pre-processing approaches performed similarly, so we selected median quotient normalization due to previous recommendation [36], followed by the calculation of the derived traits without IgG subclass weighting in LC-MS data.

Discovery and replication GWAS

We performed a discovery GWAS of IgG galactosylation in seven cohorts of European descent (N=13,705). The

association between HRC-imputed SNPs and three phenotypes (G0, G1 and G2) was studied under the assumption of an additive linear model. The inverse-variance meta-analysis of GWAS summary statistics resulted in 16 genome-wide significant ($p \leq 2.5 \times 10^{-8}$) genomic loci associated with at least one galactosylation trait (Table 1). Quantile-quantile plots of associations for three traits are shown in Supplementary Figure 1.

For twelve loci, the same trait-SNP association reached the significance threshold p-value ≤ 0.0031 (P ≤ 0.05/16 loci) in replication meta-analysis (N=7,775). The association between the top SNP in the HLA region (chr6:31107733-31164511) was not replicated due to the unavailability of the SNP and SNPs in LD in all four replication cohorts. However, the association of the HLA region and IgG glycan patterns was observed in the previous GWAS of IgG glycome [30, 31, 33, 34]. For top SNPs in three regions (chr6:74168723-74285118, chr4:103390496-103567348 and chr21:36564553-36665202) the association was not significant after correction for multiple testing in the replication meta-analysis, but the direction of the effect was consistent with the discovery analysis (Table 1).

Due to lack of compatibility in glycan trait definition in the current study and previous GWAS of IgG glycome [30–34], we checked for the overlap of the associated genomic regions across any previously examined IgG glycan traits. Thirteen loci overlap with previously identified and replicated genomic regions (Appendix Table 2). The remaining three associations in chr4:103390496-103567348, chr17:56404349-56418136 and chr6:74168723-74285118 are considered as novel.

Among 16 loci, three loci were G2-specific (rs3822960, rs1808192 and rs2526377), two were G1-specific (rs10896045 and rs34562254) and two were G0-specific (rs13250010 and rs199516) (Figure 1A), which is intriguing since conversion of G0 to G1, and then G2 is a consequence of the activity of the single enzyme (product of *B4GALT1* gene). Before functional evaluation of GWAS hits, careful inspection of the genomic regions was undertaken to prioritize and choose the genes with strong evidence for a potential role in a biological pathway. We applied a set of *in silico* methods to prioritise genes for functional analysis. First, we mapped associated variants to genes using mapping strategies based on position, overlap with eQTL associations and chromatin interaction as implemented in FUMA [37], and identified 107 candidate genes in total (Appendix Table 3). The number of genes prioritised by two or more approaches is shown in Figure 1B.

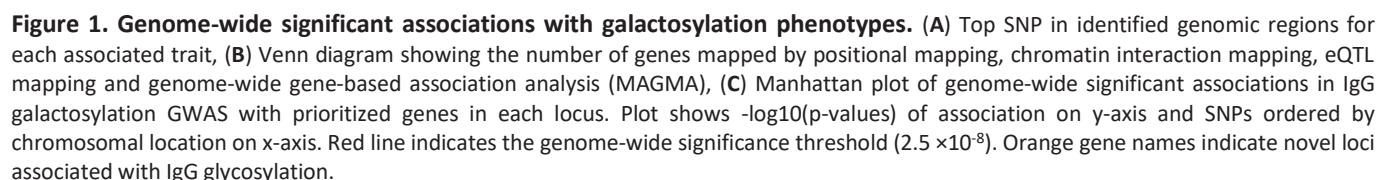**Table 1. List of genome-wide significant loci in IgG galactosylation GWAS.**

Locus	Top SNP	Trait	EA	OA	EAF	Beta	SE	P	Beta repl	SE repl	P repl	Closest gene	Prioritized genes
1:24699711-25493756	rs188468174	G0	T	C	0.015	0.693	0.048	3.20E-47	0.865	0.076	2.96E-30	*RUNX3*	*RUNX3*
2:26109539-26149988	rs10177977	G0	T	C	0.334	-0.062	0.01	3.80E-10	-0.049	0.015	1.03E-03	*KIF3C*	*KIF3C*
4:103390496-103567348	rs3774964	G2	A	G	0.37	-0.065	0.01	1.56E-11	-0.038	0.015	9.20E-03	*NFKB1*	*NFKB1; MANBA*
6:31107733-31164511	rs1265109	G0	T	G	0.295	-0.063	0.011	1.57E-09	0.013	0.019	4.97E-01	*HLA region*	*HLA region*
6:74168723-74285118	rs3822960	G2	T	C	0.306	0.061	0.01	5.84E-10	0.028	0.015	6.24E-02	*EEF1A1*	*EEF1A1; MTO1*
6:143088071-143206826	rs7758383	G2	A	G	0.486	0.092	0.009	6.09E-23	0.145	0.014	9.44E-25	*HIVEP2*	*HIVEP2*
7:150856165-150906453	rs113745074	G0	T	C	0.094	0.12	0.015	6.19E-16	0.134	0.022	8.47E-10	*SMARCD3*	*ABCF2; CHPF2; SMARCD3*
8:103542538-103550211	rs13250010	G0	T	G	0.356	0.063	0.01	7.71E-11	0.062	0.015	3.47E-05	*KB-1980E6.3*	*UBR5; RRM2B; ODF1; KB-1980E6.3*
9:32933492-33385427	rs13297246	G2	A	G	0.184	0.195	0.013	4.21E-55	0.202	0.019	2.36E-26	*B4GALT1*	*B4GALT1*
11:65555524-65555524	rs10896045	G1	A	G	0.301	0.079	0.013	2.61E-09	0.079	0.017	2.42E-06	*OVOL1*	*OVOL1; AP5B1*
17:16813994-16875636	rs34562254	G1	A	G	0.106	0.166	0.019	1.48E-18	0.092	0.024	9.95E-05	*TNFRSF13B*	*TNFRSF13B*
17:43856639-44863413	rs199516	G0	T	C	0.221	0.068	0.012	1.20E-08	0.058	0.018	1.25E-03	*WNT3*	*ARHGAP27; CRHR1; SPPL2C; MAPT; KANLS1; ARL17B; LRRC37A; NSF; WNT3*
17:45766846-45870129	rs1808192	G2	A	G	0.351	0.061	0.01	5.94E-10	0.051	0.015	6.78E-04	*TBKBP1*	*TBKBP1; TBX21*
17:56404349-56418136	rs2526377	G2	A	G	0.463	0.062	0.009	5.59E-11	0.057	0.015	8.05E-05	*BZRAP1*	*BZRAP1; SUPT4H1; RAD5C1*
17:79158040-79268562	rs2659005	G2	T	C	0.436	0.079	0.009	4.36E-17	0.097	0.016	6.99E-10	*SLC38A10*	*AZI1; ENTHD2; SLC38A10; C17orf89*
21:36564553-36665202	rs4817708	G0	T	C	0.229	0.073	0.011	2.39E-11	0.043	0.017	1.15E-02	*RUNX1*	*RUNX1*

Locus- chromosome:start-end position in GRCh37 (hg19) build; Top SNP- rsID identifier for the SNP with the strongest association with the glycan trait; Trait- trait with the lowest p-value for the top SNP in the genomic locus; EA- effect allele for which the effect is reported; OA- non-effect allele; EAF- effect allele frequency; Beta- effect estimate for the effect allele of the top SNP; SE- standard error of the effect estimate; P– p-value for the top SNP-trait association; Beta repl- effect estimate for the effect allele of the top SNP in replication analysis; SE repl- standard error of the effect estimate in the replication analysis; P repl- p-value for the top SNP in the replication analysis; Closest gene- gene found closest to the top SNP in the locus; Prioritized genes- symbol for genes prioritized in the locus.

The gene analysis in MAGMA is based on a multiple linear principal components regression model in which the statistic is computed by combining the p-values of the individual SNPs in a gene while taking into account their LD structure. In gene analysis, SNP data is aggregated to the whole gene level to test the joint association of all SNPs in the gene with the phenotype, thereby making it possible to detect effects consisting of multiple weaker SNP-phenotype associations that would otherwise be missed. The gene-based association test for G0, G1 and G2 identified 38 genes significant at FDR < 0.05 (Appendix Table 4).

Given that missense variants have a direct impact on protein function and therefore a more straightforward effect on the phenotype, we also assessed the functional consequences of associated variants using Variant Effect Predictor [38]. For two genes, *SLC38A10* and *MAPT*, variants with potentially pathogenic amino acid (aa) change were detected by both SIFT and Polyphen2 algorithms. For additional four genes, *MANBA*, *SPPL2C*, *CRHR1* and *TNFRSF13B*, the variants were predicted to result in a benign aa change. The details on position and aa changes are listed in Supplementary Table 1.

Finally, to assess the potential pleiotropic effects of the variants on IgG galactosylation and gene expression we performed colocalization analysis with the whole blood eQTLgen dataset [39]. By comparing the regional association patterns with the approximate Bayesian Factor approach in coloc [40] we identified five genes (*KIF3C*, *NFKB1*, *MIR-142*, *EEF1A1* and *MTO1*) whose expression patterns colocalize with IgG galactosylation (PP4 > 75%) (Supplementary Table 2 and Supplementary Figure 2). Previous prioritization efforts in GWAS of IgG glycosylation, as well as implications for gene's potential function in glycosylation process were taken into account

Figure 1. Genome-wide significant associations with galactosylation phenotypes. (**A**) Top SNP in identified genomic regions for each associated trait, (**B**) Venn diagram showing the number of genes mapped by positional mapping, chromatin interaction mapping, eQTL mapping and genome-wide gene-based association analysis (MAGMA), (**C**) Manhattan plot of genome-wide significant associations in IgG galactosylation GWAS with prioritized genes in each locus. Plot shows -log10(p-values) of association on y-axis and SNPs ordered by chromosomal location on x-axis. Red line indicates the genome-wide significance threshold (2.5 ×10^{-8}). Orange gene names indicate novel loci associated with IgG glycosylation.

in the final gene prioritization, which resulted in 37 credible gene candidates in the identified regions (Figure 1C). Supplementary Table 1 contains details on prioritization evidence for each gene.

IgG galactosylation and diseases

Due to existing evidence of altered IgG glycosylation patterns in multiple conditions [11], we assessed the pleiotropy of the IgG galactosylation-associated loci with a range of autoimmune, inflammatory and infectious diseases. The threshold of 75% and higher was applied for the posterior probability for colocalization of traits (PP4). A shared association pattern between SLE and monogalactosylation (PP4 = 82%) was detected in a locus on chromosome 11 (chr11:65555524), while agalactosylation-associated locus chr17:43856639-44863413 colocalized with breast cancer (PP4 = 95%), COVID-19 hospitalization (PP4 = 93%), SLE (PP4 = 89%) and schizophrenia (PP4 = 84%) (Supplementary Table 3). Because of the extensive LD pattern, there are more than twelve candidate genes in the chromosome 17 locus. It is important to note that high PP4 indicates at least one shared causal variant and low PP4 does not mean that two phenotypes do not share one, especially if the probability of both phenotypes being associated with the region but having different causal variants (PP3) is high.

Polygenic score

A polygenic score (PGS) aggregates the effects of many genetic variants into a single number which predicts genetic predisposition for the phenotype. In the standard approach, a PGS is a linear combination of linear regression effect estimates and allele counts at single-nucleotide polymorphisms (SNPs). For each of the G0, G1 and G2 traits, we created two PGS models based on genome-wide association meta-analysis (GWAMA) results. Specifically, we derived the first model based on the SBayesR [41] method, and the second model using a clumping and thresholding (C+T) method. These models were tested within the CEDAR dataset, which has aggregated genotype data and galactosylation information on 162 participants of European ancestry and which was not used in the GWAMA. The estimated SNP-based heritability (h2SNP, SBayesR package [41]) was 36.5% for the G0, 27.2% for G1, and 36.8% for the G2 trait. Our best PGS models explained 7.4% (p-value: 4.79×10^{-4}), 2.2% (p-value: 0.059) and 7.4% (p-value: 4.77×10^{-4}) of G0, G1 and G2 variance in CEDAR samples. PRS-Phenotype scatter plots and plots showing the mean of trait per group can be found in Supplementary Figures 5–7. The complete results of the implementation of SBayesR and C+T models can be found in Appendix Table 5.

Functional validation of GWAS hits

The genes for functional follow-up were selected based on novelty in GWA studies for IgG galactosylation and by reviewing the evidence for prioritization, expression in B cells (Human Protein Atlas), as well as known roles of the encoded proteins. Newly discovered GWAS hits, *EEF1A1* and *MANBA-NFKB1*, where *NFKB1* and *MANBA* are found in the same locus, were selected and functional follow-up was performed to decipher which one is indeed involved in galactosylation. We also included *HIVEP2* because it was previously assigned in the gene network as a transcription factor potentially involved in the regulation of *B4GALT1* gene expression [33], and *TNFRSF13B* for its highly specific expression in B cells and function in humoral immunity. *SLC38A10* was included as one of the genes prioritized in chr17:79158040-79268562 locus to test its potential function as a transporter protein in the IgG glycosylation pathway. We also included *KIF3C* as it was the prioritized gene in chr2:26109539-26149988 locus.

To functionally validate the potential role of the selected GWAS hits in IgG galactosylation, we used the newly developed transient expression system HEK293FS for IgG production with stably integrated dSaCas9-VPR or dSpCas9-KRAB expression cassettes for target gene upregulation or downregulation [35]. Therefore, the candidate genes were manipulated in dCas9-VPR or dCas9-KRAB containing monoclonal cell lines transiently transfected with a plasmid carrying genes for IgG heavy and light chains and specific gRNA. Subsequently, secreted IgG was analysed for glycan phenotype.

Utilizing dCas9-VPR cell line and three gRNAs for each locus, we successfully upregulated all selected loci except *SLC38A10* and *HIVEP2* (Supplementary Figure 3A and Supplementary Table 4). Successful downregulation in the dCas9-KRAB cell line was accomplished using three gRNAs for all selected loci except *NFKB1* and *HIVEP2* (Supplementary Figure 3B). A significant change of IgG glycome composition was observed after dCas9-VPR targeting of *HIVEP2, MANBA, TNFRSF13B* and *EEF1A1* (Figure 2 and Appendix Table 6). Interestingly, manipulation of the *HIVEP2* promotor did not upregulate *HIVEP2* transcript level sufficiently to reach statistical significance at transcript level, but it resulted in a significant increase of digalactosylated structures and a concomitant decrease of agalactosylated structures on IgG (Figure 2A). Upregulation of the *MANBA* locus resulted in a significant decrease of monogalactosylated IgG glycans, but no change in levels of agalactosylated and digalactosylated glycans was observed (Figure 2B). Targeting *TNFRSF13B* in dCas9-VPR cell line resulted

in almost 400-fold increase of its expression and a moderate increase in agalactosylated IgG glycans (Figure 2C). Upregulation of the *EEF1A1* locus resulted in a significant decrease of monogalactosylated glycans and in a concomitant increase of agalactosylated IgG glycans (Figure 2D).

DISCUSSION

Using GWAS approach we identified 16 genomic regions associated with IgG galactosylation, which is the molecular basis of the glycan clock of ageing. Thirteen out of 16 loci were associated with traits related to IgG glycosylation in our previous GWAS studies, while three were novel [30–34]. Contrary to previous studies, in which we were looking for genes that were associated with any aspect of IgG glycosylation, in this study we focused on galactosylation, a functionally well-defined element of IgG glycosylation that is the basis of increased inflammation in people with accelerated glycan ageing. This was done by converting percentages of individual glycans in the glycome into sums of glycan structures containing either one, two or no galactose residues. This enabled us to combine GWA studies for which the phenotype analysis was performed using different analytical methods. We also estimated the SNP-based heritability to be 36.5%, 27.2% and 36.8%, for G0, G1 and G2, respectively, an estimate that very closely replicates recent heritability estimate for biological age
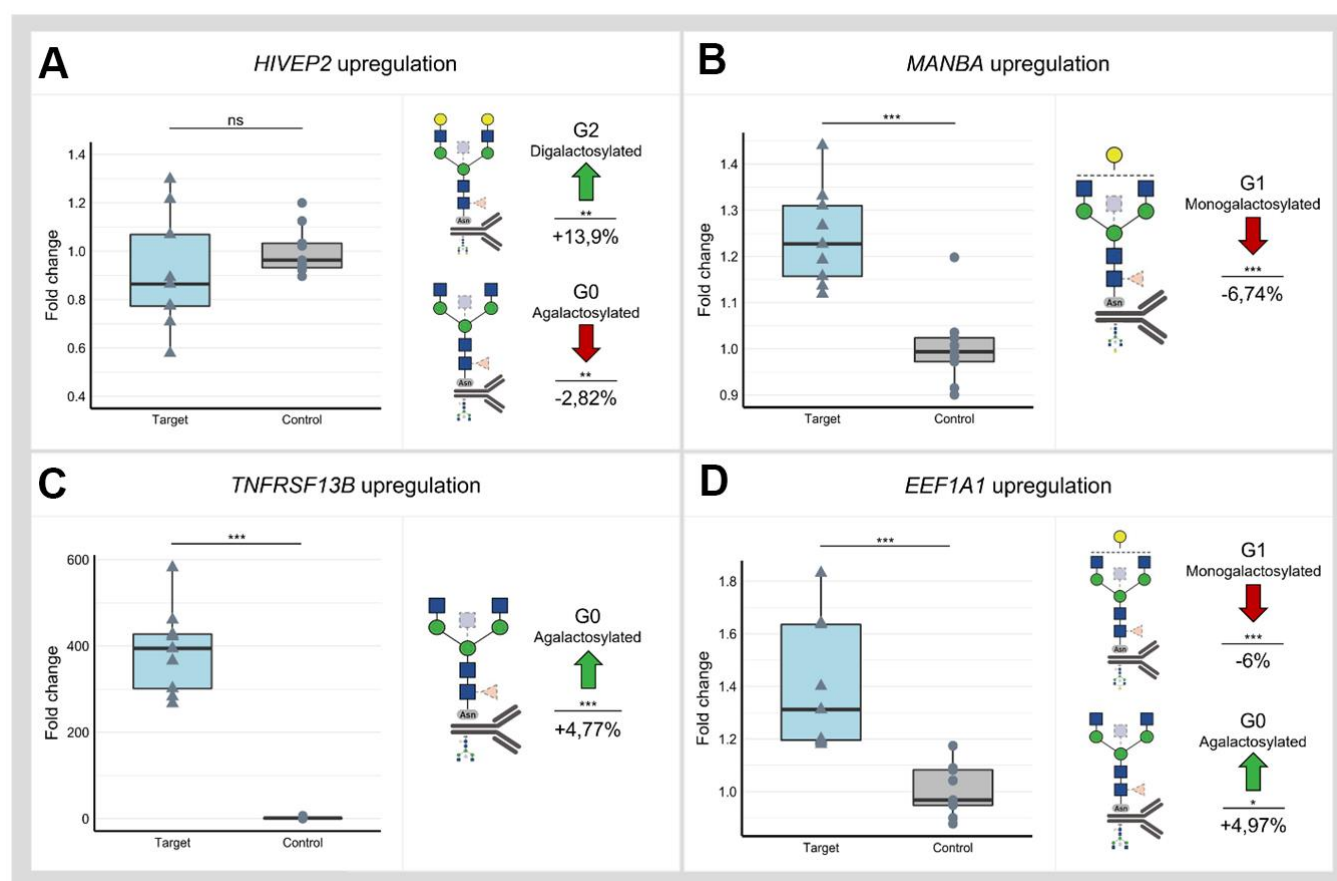

Figure 2. Changes in IgG glycan composition upon targeting of selected GWAS loci associated with IgG galactosylation (*HIVEP2, MANBA, TNFRSF13B* and *EEF1A1*) in dCas9-VPR monoclonal cell lines. Samples containing non-targeting gRNAs served as controls. Changes in transcript levels are given as fold change values and changes in IgG phenotype are given as a relative change compared to control samples. (**A**) Manipulation of the *HIVEP2* gene did not result in a statistically significant change in *HIVEP2* transcript level, however, did induce a significant increase of digalactosylated structures (G2) with a concomitant decrease of agalactosylated IgG glycan structures (G0). (**B**) Targeting of *MANBA* by dCas9-VPR elevated transcription level of this gene which resulted in decrease of monogalactosylated IgG glycan structures (G1) (**C**) Targeting *TNFRSF13B* by dCas9-VPR resulted in ~ 400-fold increase of transcript levels and significant change of IgG agalactosylated IgG glycans (G0). (**D**) Successful upregulation of the *EEF1A1* locus was followed by an increase of agalactosylated IgG glycans (G0) with a concomitant decrease of monogalactosylated IgG glycans (G1). Nominal p-value: *<0.05; **<0.01; ***<0.001; ns, not significant.

estimate based on these glycan structures [42]. Upon prioritization efforts which resulted in 37 candidate genes, we evaluated functional relevance of these genes using HEK293FS transient expression system for IgG production with stably integrated dCas9-KRAB or dCas9-VPR fusion. Results of this study expanded the list of genes for which the functional relevance for *in vitro* expression and glycosylation of IgG was confirmed with *MANBA, TNFRSF13B* and *EEF1A1* (Figure 2). Thus, out of 13 replicated GWAS hits, for six of them we have functionally confirmed that they act at IgG glycosylation in a way that is conserved between plasma cells and our artificial *in vitro* expression system in HEK293FS cells (Figure 3). It is likely that some additional GWAS hits may be plasma cell specific and will be confirmed once an efficient system for gene manipulation and IgG production in B cells is developed. However, it is important to note that IgG glycosylation can be affected at multiple other levels (e.g. B-cell selection, maturation, expansion,

etc.), thus the absence of functional confirmation of other loci in this assay does not mean that they are not implicated in IgG glycosylation in some other, more complex, manner.

The total heritability of G0, G1, G2 is not known, however, the heritability of individual IgG glycan peaks ranges between 27% and 76% [29]. More than half of the studied glycan traits have additive genetic component estimate of 50% which would imply that half of the variance in their values could be explained by common variants, an observation similar to other complex traits [29]. The constructed PGS models explained 7.4% of the variance in G0 and G2 traits in an out-of-sample testing, thus capturing about 20% of the estimated SNP heritability. However, for G1 the PGS model explained a minor proportion of SNP heritability. The explanation why for G1, that had comparable SNP heritability, the proportion of variance explained was much less may lie in a flatter distribution and hence
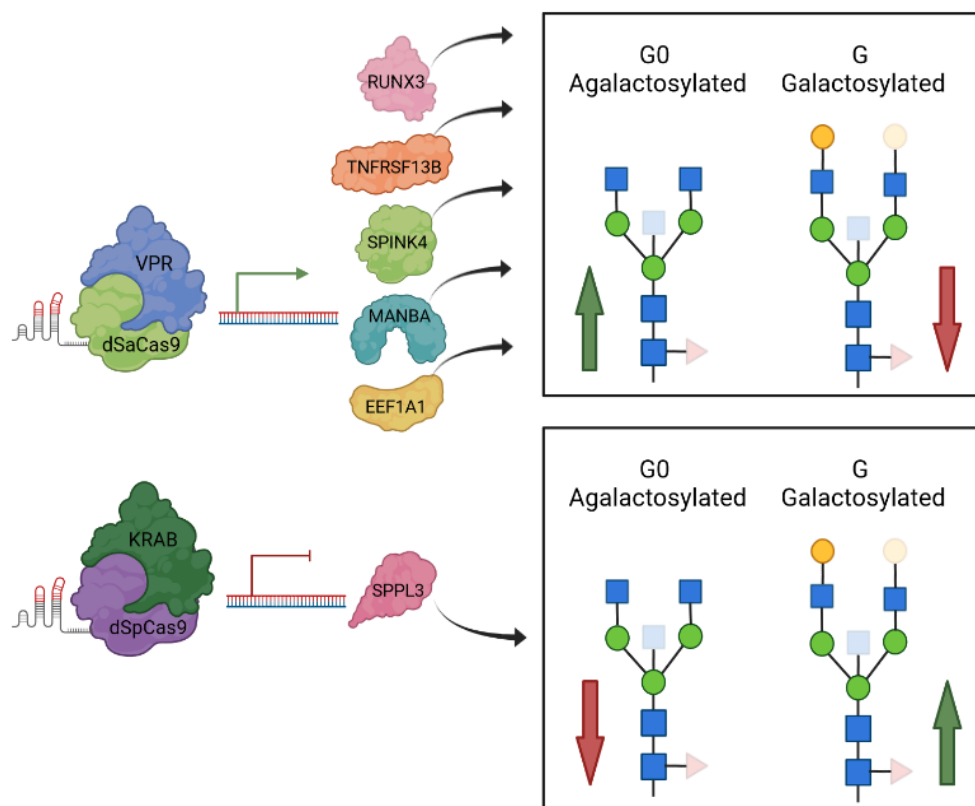

Figure 3. A graphical review of up-to-date functional validation of GWAS hits associated with IgG glycosylation using HEK293FS transient expression system with stably integrated dCas9-VPR or dCas9-KRAB fusions. When specific gRNA targeted dCas9-VPR to two estrogen associated genes, *RUNX3* and *SPINK4* (Mijakovac et al. 2021), as well as to the *TNFRSF13B, MANBA* and *EEF1A1* loci, transcription was upregulated that resulted in decreased levels of galactosylated (red arrow within box), increased levels of agalactosylated IgG glycan structures (green arrow within box) or both. Downregulated transcription of *SPPL3* resulted in an increase in galactosylated structures with a concomitant decrease in agalactosylated IgG glycan structures. Protein structures do not depict true protein structures in humans, generic protein shapes are chosen for easier visualization. Figure was created with BioRender.com. Accessed on 16 November 2021.

individually smaller genetic effects onto G1; and observation which may be described as "higher polygenicity". Indeed, while for G0 and G2 lead associations explained about 1.5% of variance, for G1 the lead variant near *TNFRSF13B* explained only about 0.5% of its variation. With smaller individual effects, one needs larger discovery GWAS to construct powerful PGS.

Given that the majority of identified candidate SNPs are located in non-coding regions, we applied several *in silico* approaches to prioritize potential causal genes in associated genomic regions, including exploring functional consequences of associated variants, pleiotropy with gene expression, gene-based association analysis, as well as evidence from previous GWA studies and prior biological knowledge of gene function. Among 37 candidate genes from 16 associated loci, for the two genes, *RUNX3* and *B4GALT1*, the effect on IgG glycosylation has recently been demonstrated using direct gene expression modulation in HEK293FS IgG transient expression system [35, 43].

Due to existing evidence of altered IgG glycosylation patterns in multiple conditions [11], we performed colocalization analysis with a range of diseases and observed a high probability for pleiotropic effects in the *AP5B1/OVOL1* locus on chromosome 11 for SLE and monogalactosylation. Interestingly, for the same locus, we observed considerable probability for colocalization (PP4=0.56) with asthma and allergy. Although this is below our pre-defined posterior probability for colocalization (PP4) threshold of PP4 > 0.75, the finding is in line with previous results linking this locus with variation of specific digalactosylated structures (FA2FG2S1), expression of the *OVOL1* gene and the risk of asthma [34]. *OVOL1* gene encodes a putative zinc finger containing transcription factor with a role in the development and differentiation of epithelial and germ cells [44]. The locus on chromosome 17 was pleiotropic for agalactosylation and SLE, breast cancer, COVID-19 hospitalization and schizophrenia. The colocalized region on chromosome 17 (chr17:43856639-44863413) contains at least twelve genes and is known for numerous copy number variants. It also includes a megabase-long inversion polymorphisms (H1 and inverted H2 forms) which contain partial duplication of *KANSL1* gene with inverted H2 isoform having a higher frequency than H1 isoform in the European population [45]. The complexity of the locus makes it challenging to prioritise a specific gene without further investigation.

The functional follow-up was based on the previously developed HEK293FS transient expression system for IgG production which was shown to be an excellent tool for studying the role of the genes identified in glycosylation GWAS [35]. Despite the fact that we could not demonstrate upregulation of *HIVEP2* mRNA using dCas9-VPR, a significant increase in digalactosylated structures with a concomitant decrease in agalactosylated structures was measured from IgG secreted from the same cells. *HIVEP2* (*HIVEP zinc protein 2*) encodes a transcription factor that interacts with numerous viral and cellular promoters, and has an established role in immunity [46, 47]. Even though we prioritised *HIVEP2,* the locus interacts with variants in six genes via chromatin interactions (Supplementary Figure 4), including *FUCA2* (alpha-L-fucosidase) which is located 0.6 megabases away. It is plausible that dCas9-VPR targeting of the *HIVEP2* promoter disrupted these chromatin interactions which in turn altered the expression of other genes mapping to this region. In the case of the locus on chromosome 4, the functional study was utilized as one of the steps to decipher which of the two genes might influence galactosylation process as both *NFKB1* and *MANBA* were prioritised with *in silico* approaches. The *MANBA/NFKB1* region was identified in previous studies as a disease susceptibility locus for primary biliary cholangitis (PBC) [48] and a recent study found a decreased level of IgG galactosylation in PBC patients [49]. The *NFKB1* gene encodes a subunit of the nuclear factor of kappa light polypeptide gene enhancer in B-cells (NF-κB) transcription factor family, known for its critical role in cell survival and inflammation [50]. The upregulation of *NFKB1* using dCas9-VPR resulted in a significant change in transcript level, but no change in IgG glycoprofile was observed. On the other hand, the upregulation of the *MANBA* gene expression resulted in a decrease of IgG monogalactosylated species. The *MANBA* gene encodes beta mannosidase which is an exoglycosidase enzyme cleaving beta-mannose residues from the non-reducing end of N-linked glycans [51]. As such, it has a known role in glycosylation process but currently we cannot speculate about the exact mechanism by which it affects IgG galactosylation. Interestingly, mutations in *MANBA* are known to affect kidney function [52], blood pressure [53] and cardiovascular disease [54], while the monogalactosylated IgG glycan (G1) (that *MANBA* affects in our *in vitro* expression system) was previously identified as the best predictor of future cardio vascular events in women [55, 56]. While at the moment we do not know putative mechanisms that could explain the link between *MANBA*, G0 and cardiovascular or renal diseases, these intriguing associations warrant future studies.

The *EEF1A1* (*Eukaryotic Translation Elongation Factor 1 Alpha 1*) gene encodes for an isomer of an alpha subunit of the eukaryotic translation elongation factor-1 complex that has a role in binding of aminoacyl tRNA to the A site on ribosomes during protein synthesis [57]. Together with EEF1A2, it is the second

most abundant protein (1%-3%) in the cell with other roles outside of protein synthesis, including protein degradation, apoptosis modulation, oncogenesis and viral pathogenesis [58, 59]. Previous study has shown that EEF1A1 also serves as a signal transducer during inflammation, where it can enhance interleukin 6 expression through STAT3 and PKCδ [60]. Interleukin 6 is implicated in different autoimmune and inflammatory diseases [61] and its high levels lead to decreased IgG galactosylation [62]. By targeting dCas9-VPR to *EEF1A1*, we increased its expression level, which resulted in a decrease of galactosylated IgG glycans. Therefore, our results are concordant with the previous findings, suggesting that *EEF1A1* upregulation could lead to pro-inflammatory modulation of IgG by decreasing its galactosylation.

A locus harbouring *TNF receptor superfamily member 13B* (*TNFRSF13B*) gene was previously identified in the multivariate IgG glycosylation GWAS [34], while here it was associated with IgG monogalactosylation. *TNFRSF13B* encodes a transmembrane activator calcium modulator and cyclophilin ligand interactor (TACI) protein, a lymphocyte-specific member of the tumour necrosis factor (TNF) receptor superfamily, which has a role in signalling pathway leading to B cell differentiation and antibody production [63, 64]. Common genetic variation in *TNFRSF13B* locus was previously associated with different levels of immunoglobulins in serum [65], while rare deleterious variants were implicated in common variable immunodeficiencies and IgA deficiencies [66, 67]. Therefore, we speculate that *TNFRSF13B* might impact glycosylation profile of the sample by controlling secretion of certain IgG glycoforms. Despite extensive increase in *TNFRSF13B* expression upon activation with dCas9-VPR, we observed only a moderate increase in IgG agalactosylation and no changes in mono- or digalactosylation. Our experimental setting might be limited by the B-cell specific function of *TNFRSF13B* and thus, we might not be able to observe changes comparable to changes in the gene expression level.

The role of four other prioritized genes, *NFKB1*, *KIF3C*, *HIVEP2* and *SLC38A10,* could not be validated in our HEK293FS transient expression system. Even though this cell system represents a good model for studies of IgG glycosylation regulation, it has some limitations. HEK293FS cells do not inherently secrete immunoglobulins and their epigenetic context is not equivalent to plasma and B cells. In addition, alternative galactosylation might depend on gene expression in other cell types and their interactions, hence we might not observe the natural effects after manipulation of gene expression in this model system, due to the lack of the right biological context.

In conclusion, in the present study we mapped 16 genetic loci regulating glycan structures that are not only the basis of the glycan clock of ageing, but also functional effectors of inflammation at multiple levels. Furthermore, we were able to confirm the functional role of three genes (*MANBA, TNFRSF13B* and *EEF1A1*) in the IgG galactosylation pathway by use of the HEK293FS transient expression system with stably integrated dCas9 fusion designed for this purpose. Figure 3 summarizes all GWAS hits with previously unknown link to IgG glycosylation, whose functional relevance has been shown using this cell system. Further research is needed to fully elucidate functional mechanism behind their role in ageing and to reveal the complete network of gene interactions regulating the complex process of IgG glycosylation. However, this study is an important step in this direction since it clearly indicates that regulation of IgG galactosylation extends far beyond the expression of the galactosyltransferase gene that adds galactose.

MATERIALS AND METHODS

Information about sample numbers, sex and age for each cohort is provided in Appendix Table 7. Further information on genotyping, genotype imputation and quality control can be found in Appendix Table 8.

Glycan measurements

An overview of cohorts and technologies used for IgG glycome measurement (LC-MS or UPLC), as well as the published studies where IgG glycome analysis for the corresponding cohort was described in more detail, are shown in Appendix Table 7.

Briefly, for UPLC-based analysis, IgG was isolated from blood plasma samples using protein G plates, followed by neutralization and denaturation of the protein. N-glycans were enzymatically released from IgG and fluorescently labelled with 2-aminobenzamide dye, followed by separation and quantification by hydrophilic interaction UPLC which resulted in 24 peaks (GP1-GP24), most of which represent a single glycan structure. In cases where multiple glycan structures are found under one peak, the most abundant structure is considered. The list of all glycans and their proportion under each peak can be found in Pučić et al. [18].

For LC-MS-based glycan analysis, IgG was isolated from blood plasma samples by affinity chromatography binding to protein G plates and treated with trypsin to allow cleavage of IgG at specific amino acid sites. The cleavage resulted in different glycopeptides across IgG subclasses, thereby enabling subclass-specific glycan measurements. IgG subclass separation and detection

were performed using a nano-LC system coupled with quadrupole-TOF-MS. IgG2 and IgG3 subclasses have the same tryptic glycopeptide moieties, thus the separation of the subclass-specific glycopeptides was not possible. LC-MS quantification resulted in 50 values which correspond to 20 glycans measured on IgG1, 20 glycans on IgG2/3 and 10 glycans on IgG4. The list of glycan structures measured by UPLC and LC-MS along with their description is listed in Appendix Tables 9, 10.

Glycan data pre-processing

Prior to genetic analysis, the best approach for harmonization of glycan measurements by UPLC and LC-MS was chosen (Supplementary Methods) to allow for meta-analysis of GWAS summary statistics from all cohorts. To reduce the impact of experimental variation on the downstream analysis, glycan measurements were normalized using median quotient normalization and batch corrected using empirical Bayes method [68] implemented in ComBat function of the "sva" package [69] in statistical software R [70]. Median quotient normalization was applied due to negligible differences between different normalization approaches in the comparison and further supported by previous recommendation [36]. Next, three derived traits were calculated as the percentage of structures in the total IgG glycome containing zero, one or two galactose units as G0, G1 and G2, respectively (Appendix Table 11). In this way, the subclass-specific glycan values from LC-MS measurements were summarized to allow for comparison with total glycan data obtained with UPLC. The glycan data was pre-processed centrally in Genos for all cohorts except Leiden Longevity Study (LLS), for which the glycan data was pre-processed by the phenotype provider from the Leiden University Medical Center.

Genome-wide association study

Discovery genome-wide association study was performed in seven cohorts of European descent (CROATIA-Vis, CROATIA-Korcula, CROATIA-Split, TwinsUK, EPIC, VIKING, ORCADES) with a total sample size of 13,705. Quantile normalization was applied to the derived traits in each cohort. For CROATIA, VIKING and ORCADES cohorts, first linear mixed model was fit to adjust for age, sex and cohort-specific covariates while also accounting for the genomic kinship. This step was performed using polygenic function in GenABEL R package [71], followed by testing the association between SNPs and phenotype using the RegScan software v0.5 [72]. For TwinsUK cohort, the derived traits were adjusted for age and sex and the association analysis was carried out

using GEMMA [73] including kinship matrix to correct for family structure. SNPTEST software was used for association testing of age- and sex-adjusted glycan values with genetic variants in EPIC cohort. GWAS was performed on Haplotype Reference Consortium (HRC) [74] imputed genotypes, assuming an additive linear model of association. Quality control of study-specific summary statistics was performed prior to meta-analysis using EasyQC R package as described in Winkler et al. [75].

After quality control, the GWAS summary statistics for seven cohorts were pooled using the inverse-variance weighted method for meta-analysis as implemented in METAL software [76]. METAL software also allowed for the estimation of genomic control (GC) to correct test statistics to account for population stratification in the studies included in the meta-analysis (lambda GC median =1.00; range 0.98-1.01). Given that we performed GWAS for three correlated traits, we used 5 x $10^{-8}/2= 2.5 \times 10^{-8}$ as a genome-wide significance threshold, where 2 denotes the number of PCs that explain 99% of the variation in the three galactosylation traits.

Replication analysis

Following the discovery GWAS, we performed replication meta-analysis to validate results from the discovery phase. Four independent cohorts of European descent were used: KORA F4, LLS, EGCUT and GCKD, with a total sample size of 7,775. Glycan measurements for KORA F4 and LLS cohort were quantified with LC-MS, while EGCUT and GCKD glycan measurements were obtained with UPLC. The strongest glycan trait-SNP pair from each locus in the discovery analysis was meta-analysed using the fixed-effect inverse variance method. The significance threshold was set to p-value < 0.05/16 = 0.0031, where 16 is the number of associated genomic regions. We also checked the consistency of effect direction between the discovery and replication analyses in cases where the top SNP did not formally replicate.

Genomic loci definition

Genomic loci associated with galactosylation phenotypes were defined using FUMA v. 1.3.6b [37]. SNP2GENE function in FUMA first identifies independent (LD $r^2 < 0.6$) significant SNPs with MAF > 0.01 to determine borders of genomic locus. LD estimates were inferred from reference panel derived from 10,000 subjects of European descent in UK Biobank [77]. The maximum distance of 250kb was used for merging LD blocks into a single genomic locus. SNPs in LD ($r^2 > 0.6$) with independent SNPs

within 250 kb distance were selected as associated SNPs and considered in further analysis.

Gene prioritization

Since we aimed to demonstrate *in vitro* the functional activity of the proteins encoded by genes potentially causal for IgG galactosylation, we first employed multiple *in silico* strategies to prioritise candidate genes. Briefly, we used the following criteria: gene mapping (prioritising based on genomic position, overlap with eQTLs and chromatin interactions), functional consequence-based prioritisation (genes whose missense variants were associated with galactosylation), co-localizetion of SNP associations with gene expression in whole blood and with galactosylation, and gene-based analysis.

Gene mapping

Genes in the significantly associated genomic regions were mapped using three approaches implemented in FUMA: positional mapping, eQTL mapping and chromatin interaction mapping. Genes were positionally mapped based on ANNOVAR [78] annotations and the maximum distance of <10 kb between associated variants and genes. The eQTL mapping was based on overlap of galactosylation-associated variants and eQTLs in B and T cells, including Database of Immune Cell Expression (DICE) [79], Fairfax et al. [80] and CEDAR [81] datasets. Only significant eQTL signals at FDR < 0.05 were considered. Chromatin interaction mapping was performed using the Hi-C data derived from B cell line (GM12878) [82] and the suggested FDR value $< 1 \times 10^{-6}$ for significant chromatin interaction was used.

Functional consequences

To assess potential functional consequences of galactosylation associated variants we used SIFT and Polyphen-2 algorithms, as implemented in the Variant Effect Predictor (VEP) v97 by Ensembl [38].

Colocalization with gene expression in whole blood

The colocalization of galactosylation GWAS signals and gene expression was estimated using Approximate Bayes Factor (ABF) method [83], as implemented in the coloc package [40] in R. The method outputs five posterior probabilities, one for each of the five hypotheses: H0) no association with either of the two traits, H1) association with trait 1, but not with trait 2, H2) association with trait 2, but not with trait 1, H3) association with both trait 1 and trait 2, but two independent causal variants and H4) association with

both trait 1 and trait 2 and one shared causal variant. Whole blood eQTL data was obtained from the publicly available eQTLgen dataset which was derived from 31,684 individuals across 37 cohorts [39]. Tissue- or cell-specific eQTL data (e.g., B cell or T cell) were not used due to low number of SNPs in the associated regions which overlap in galactosylation GWAS summary statistics and tissue-specific eQTL data to allow for reliable colocalization test. The posterior probabilities for colocalization with cis-eQTLs were computed for each gene found in galactosylation-associated genomic loci using SNP p-values and MAF. The default values for prior probabilities were used ($p1 = 1 \times 10^{-4}$, $p2 = 1 \times 10^{-4}$ and $p12 = 1 \times 10^{-5}$). The threshold of 75% for PP4 (probability of the same shared variant for two traits) indicated positive colocalization and strong support for prioritization of the gene in the given genomic locus.

Gene-based association test

Genome-wide gene-based association test for all three traits was performed by MAGMA v1.08 [84] to evaluate the joint effects of variants in 18,655 protein-coding genes while accounting for LD between those variants. In such test, SNP data is aggregated to the whole gene level to test the joint association of SNPs in the gene with the phenotype to allow detection of effects comprising of multiple weaker SNP-phenotype associations that would potentially be missed. MAGMA uses p-values derived from variant-based analyses as input and implements the SNP-wise mean model which derives a mean χ^2 statistic for SNPs in the gene and a p-value by using a known approximation of sampling distribution. Only genes significant at FDR < 0.05 were considered.

Colocalization with diseases and traits

The ABF colocalization method was used to explore the pleiotropy between IgG galactosylation and complex diseases for which there is previous evidence of aberrant IgG glycosylation [11] and traits with shared associated variants in the GWAS Catalog, as obtained from GWAS Catalog query (accessed in June 2021). The full list of traits and links for summary statistics download is available in Supplementary Table 5. The default values for prior probabilities were used ($p1 = 1 \times 10^{-4}$, $p2 = 1 \times 10^{-4}$ and $p12 = 1 \times 10^{-5}$). The posterior probabilities by ABF were computed using beta and standard error values if available in the dataset, otherwise, SNP p-values and MAF were used. The threshold of 75% for PP4 (probability of the shared underlying causal variant for two traits) was used for colocalization and evidence of high confidence for pleiotropy between IgG galactosylation and disease or trait.

Polygenic score

SbayesR method reweights the effect of each variant according to the marginal estimate of its effect size, statistical strength of association, the degree of correlation between the variant and other variants nearby, and tuning parameters. This method requires a compatible LD matrix file computed using individual-level data from a reference population. For these analyses, we used publicly available shrunk sparse GCTB LD matrix including 2.8 million pruned common variants from the full UK Biobank (UKB) European ancestry ($n \approx 450,000$) data set and computed from a random set of 50,000 individuals of European ancestry from the UKB data set [41, 77]. SbayesR was run for each chromosome separately, and with the default parameters except for the number of iterations (N=3000) and p-value (0.9) (Appendix Table 12).

Clumping and thresholding model was built using a p-value and linkage disequilibrium-driven clumping procedure in PLINK version 1.90b (--clump) [85]. In brief, the algorithm forms clumps around SNPs with association p-values less than a provided threshold (p-value=5×10^{-8}). Each clump contains all SNPs within 250 kilobases of the index SNP that are also in linkage disequilibrium with the index SNP as determined by a provided pairwise correlation (r^2=0.2) threshold in the linkage disequilibrium reference. The algorithm iteratively cycles through all index SNPs, beginning with the smallest p-value, only allowing each SNP to appear in one clump. The final output should contain the most significantly disease-associated SNP for each linkage disequilibrium-based clump across the genome. The prediction accuracy was defined as the proportion of the variance of a phenotype that is explained by PGS values (R^2). To calculate PGS based on the PGS model we used PLINK2 software, where PGS values were calculated as a weighted sum of allele counts. Out-of-sample prediction accuracy was evaluated using samples from the CEDAR cohort that was not used for discovery or replication.

Functional validation of GWAS hits associated with galactosylation trait

Plasmid constructs for targeted upregulation/ downregulation of prioritized GWAS hits

Seven genes strongly associated with IgG glycosylation were selected for the follow-up functional validation. A newly developed transient expression system with stably integrated CRISPR/dCas9 fusions dSaCas9-VPR (for gene expression upregulation) and dSpCas9-KRAB (for gene expression downregulation) [35] was used for *in vitro* validation. Three guide RNAs (gRNAs) were designed using the CRISPick online tool

(Broad Institute) for adequate dCas9 orthologue (dSaCas9 or dSpCas9) and each selected locus (*KIF3C, NFKB1, MANBA, SLC38A10, TNFRSF13B, EEF1A1* and *HIVEP2*). Specific gRNAs were assembled with genes coding for IgG light and heavy chains, CBh promoter and bGH terminator using Golden Gate enzymes in three steps as described in Mijakovac et al. [35]. Non-targeting gRNAs were assembled in the same manner. Plasmids pORF-hp21, pORF-hp27 (Invivogen, San Diego, CA, USA) and p3SVLT were used for enhanced IgG production [86]. Specific gRNA sequences targeting each gene are shown in Appendix Table 13.

Transfection of monoclonal dCas9-VPR/dCas9-KRAB HEK293FS cell lines

The main approach was to use suspension cell lines with integrated expression control machinery based on dCas9 fusions with IgG-producing plasmid bearing gRNA expression cassettes, as described in Mijakovac et al. [35]. Briefly, suspension-adapted monoclonal cell lines with stably integrated dSaCas9-VPR (dCas9-VPR) or dSpCas9-KRAB (dCas9-KRAB) were grown until they reached the appropriate density for cell transfection with gRNA-IgG bearing plasmids as well as plasmids used for enhanced protein production in mass ratio gRNA and IgG chain bearing plasmid, p3SVLT, pORF-hp21 (Invivogen, San Diego, CA, USA) and pORF-hp27 (Invivogen, San Diego, CA, USA) 0.69/0.01/0.05/0.25. Transfections were performed with 293fectin transfection reagent diluted in Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific, Waltham, MA, USA). Five days following transfection, cells were centrifugated and used for gene expression analysis, while supernatant was used for IgG isolation and glycan analysis.

Reverse transcription and quantitative real-time PCR (qPCR)

Total RNA was isolated from cell pellets with Rneasy Mini Kit (Qiagen, Hilden, Germany) and 50 ng of RNA was converted to cDNA using the PrimeScript Rtase (TaKaRa, Kusatsu, Japan) and random hexamer primers (Invitrogen, Waltham, MA, USA). All samples were treated with Turbo Dnase (Invitrogen, Waltham, MA, USA) to remove any remaining DNA. Gene transcripts were detected with SYBR Green Gene Expression Assay using the 7500 Fast Real-Time PCR System. Primer sequences are listed in Appendix Table 14. The mean value of nine replicates was normalized to *HPRT1* expression which was used as an endogenous control. Fold changes in transcript expression compared to non-targeting control were analysed using the ΔΔCt method [87].

Glycan measurements

IgG glycan measurements were done as described previously [35] with a few modifications described below. Briefly, this method consisted of five main steps: 1) IgG isolation from HEK293FS transient expression system cell culture supernatants using Protein G Agarose fast flow beads (Merck, Darmstadt, Germany); 2) Enzymatic release of N-glycans from isolated IgG using PNGase F (Promega, Madison, WI, USA); 3) fluorescent labelling of released IgG glycans with procainamide hydrochloride (Thermo Fisher Scientific, Waltham, MA, USA); 4) clean-up of labelled IgG glycans using hydrophilic interaction liquid chromatography solid-phase extraction (HILIC-SPE) on a 0.2 μm Supor filter plate (Pall Corporation, Port Washington, NY, USA) and 5) separation of fluorescently labelled glycans by hydrophilic interaction chromatography on a Waters Acquity UPLC instrument (Waters, Milford, MA, USA). The IgG isolation step was modified from that previously described in the way that all washing steps after incubation of samples with Protein G beads as well as elution of IgG were done on an Orochem filter plate (Orochem Technologies Inc., Naperville, IL, USA). The obtained UPLC chromatograms were all separated in the same manner into 24 peaks and the amount of glycans in each peak was expressed as a percentage of total integrated area.

Statistical analysis

Statistical analysis of RT-QPCR and IgG glycan data was performed using R [70] and GraphPad (GraphPad Software, San Diego, CA, USA). RT-QPCR was done in two technical replicates and biological replicates were pooled from independent experiments. Group differences were determined using a two-tailed t-test on ΔΔCt values. IgG glycans were measured as a percentage of area under chromatogram peaks (Supplementary Table 6). Derived glycan traits (G0, G1 and G2) were calculated as a sum of glycan peaks containing zero, one or two galactose units. Initial round of experiments involved both up- and down-regulation of the genes and was used as pilot study to determine whether the expression of genes changed and if there were significant changes in glycan profile by applying Student's T-test. Five genes (*HIVEP2, MANBA, EEF1A1, TNFRSF13B* and *NFKB1*) were selected and the experiments for gene expression upregulation were repeated. To combine results of testing differences in control and treated samples by Student's T-test across multiple experimental runs, meta-analysis approach was applied using metafor R package [88].

Data and materials availability

The full summary statistics for discovery meta-analysis for G0, G1 and G2 traits is deposited at Zenodo (https://www.doi.org/10.5281/zenodo.7227999). The summary statistics for gene expression levels in different tissues and cell types are available in the eQTL catalogue project, DICE project and eQTLGen consortium. The summary statistics for other complex traits are available in publicly available resources, as listed in Supplementary Table 5. Additional data related to this paper may be requested from the authors.

AUTHOR CONTRIBUTIONS

AFH- data analysis and interpretation, visualisation, writing- original draft preparation, writing- review and editing. KM- functional study experiments, data analysis and interpretation, visualisation, writing- original draft preparation, writing- review and editing. AM- functional study experiments, data analysis and interpretation, visualisation, writing- original draft preparation, writing- review and editing. AN- PGS calculation, visualisation, writing- original draft preparation, writing- review and editing. SZS- PGS calculation, data analysis. AL- data analysis, TH- performing GWAS in EGCUT cohort; EA- performing GWAS in LLS cohort. SS- performing GWAS in KORA F4 cohort. RRCC- performing GWAS in EPIC cohort. MM- performing GWAS in EPIC cohort. YL- performing GWAS in GCKD cohort. TK- IgG glycan quantification. NR- IgG glycan quantification. TŠ- IgG glycan quantification, writing- review and editing. MPB- IgG glycan quantification, writing- review and editing. ITA- IgG glycan quantification, writing- review and editing. IG-IgG glycan quantification. JŠ- IgG glycan quantification. TP- IgG glycan quantification. BR- IgG glycan quantification. PT- IgG glycan quantification. KF- genomic and demographic data provider for EGCUT cohort. MB- genomic and demographic data provider for LLS cohort. MW- genomic and demographic data provider for LLS cohort, writing- review and editing. CG- genomic and demographic data provider for KORA F4 cohort. MBS- genomic and demographic data provider for EPIC cohort, writing- review and editing. CW- genomic and demographic data provider for EPIC cohort, writing-review and editing. OP- genomic and demographic data provider for CROATIA cohorts. CH- genomic and demographic data provider for ORCADES cohort. JFW- genomic and demographic data provider for VIKING cohort, writing- review and editing. TDS- genomic and demographic data provider for TwinsUK cohort. AK- genomic and demographic data provider for GCKD cohort, writing- review and editing. FV- IgG glycan data quality control, supervision, YSA- PGS

calculation supervision, writing- original draft preparation, writing- review and editing. AV- data analysis, visualization, supervision of functional study. JK- IgG glycan quantification in functional study, IgG glycan quantification in cohorts, supervision, writing-original draft preparation, writing- review and editing. LK- conceptualisation, supervision, data interpretation, writing—original draft preparation, writing- review and editing, VZ- supervision for the functional study, data interpretation, writing- original draft preparation, writing- review and editing., GL- conceptualisation, supervision, data interpretation, glycan data provider, writing-original draft preparation, writing- review and editing.

CONFLICTS OF INTEREST

YSA is a cofounder and a co-owner of PolyOmica and PolyKnomics, private organizations providing research services in the field of quantitative, computational and statistical genomics. GL is the founder and CEO of Genos Ltd, a private research organization that specializes in high throughput glycomics analysis and has several patents in this field. AFH, TŠ, MPB, ITA, IG, JŠ, TP, BR, PT, FV and JK are employees of Genos Ltd. The remaining authors declare no Conflicts of interest.

ETHICAL STATEMENT AND CONSENT

The participating cohorts were approved by local research ethics committees for academic and commercial use of the study. The informed consent was obtained from the participants of the study.

FUNDING

REFERENCES

1. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012; 13:448–62. https://doi.org/10.1038/nrm3383 PMID:22722607

2. Wada R, Matsui M, Kawasaki N. Influence of N-glycosylation on effector functions and thermal stability of glycoengineered IgG1 monoclonal antibody with homogeneous glycoforms. MAbs. 2019; 11:350–72. https://doi.org/10.1080/19420862.2018.1551044 PMID:30466347

3. Zheng K, Bantog C, Bayer R. The impact of glycosylation on monoclonal antibody conformation and stability. MAbs. 2011; 3:568–76. https://doi.org/10.4161/mabs.3.6.17922 PMID:22123061

4. Masuda K, Kubota T, Kaneko E, Iida S, Wakitani M, Kobayashi-Natsume Y, Kubota A, Shitara K, Nakamura K. Enhanced binding affinity for FcgammaRIIIa of fucose-negative antibody is sufficient to induce maximal antibody-dependent cellular cytotoxicity. Mol Immunol. 2007; 44:3122–31. https://doi.org/10.1016/j.molimm.2007.02.005 PMID:17379311

5. Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, Weikert SHA, Presta LG. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcgamma RIII and antibody-dependent cellular toxicity. J Biol Chem. 2002; 277:26733–40. https://doi.org/10.1074/jbc.M202069200

PMID:11986321

6. Dekkers G, Treffers L, Plomp R, Bentlage AEH, de Boer M, Koeleman CAM, Lissenberg-Thunnissen SN, Visser R, Brouwer M, Mok JY, Matlung H, van den Berg TK, van Esch WJE, et al. Decoding the Human Immunoglobulin G-Glycan Repertoire Reveals a Spectrum of Fc-Receptor- and Complement-Mediated-Effector Activities. Front Immunol. 2017; 8:877.
https://doi.org/10.3389/fimmu.2017.00877
PMID:28824618

7. Dekkers G, Rispens T, Vidarsson G. Novel Concepts of Altered Immunoglobulin G Galactosylation in Autoimmune Diseases. Front Immunol. 2018; 9:553.
https://doi.org/10.3389/fimmu.2018.00553
PMID:29616041

8. van Osch TLJ, Nouta J, Derksen NIL, van Mierlo G, van der Schoot CE, Wuhrer M, Rispens T, Vidarsson G. Fc Galactosylation Promotes Hexamerization of Human IgG1, Leading to Enhanced Classical Complement Activation. J Immunol. 2021; 207:1545–54.
https://doi.org/10.4049/jimmunol.2100399
PMID:34408013

9. Peschke B, Keller CW, Weber P, Quast I, Lünemann JD. Fc-Galactosylation of Human Immunoglobulin Gamma Isotypes Improves C1q Binding and Enhances Complement-Dependent Cytotoxicity. Front Immunol. 2017; 8:646.
https://doi.org/10.3389/fimmu.2017.00646
PMID:28634480

10. Wei B, Gao X, Cadang L, Izadi S, Liu P, Zhang HM, Hecht E, Shim J, Magill G, Pabon JR, Dai L, Phung W, Lin E, et al. Fc galactosylation follows consecutive reaction kinetics and enhances immunoglobulin G hexamerization for complement activation. MAbs. 2021; 13:1893427.
https://doi.org/10.1080/19420862.2021.1893427
PMID:33682619

11. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. Cell Immunol. 2018; 333:65–79.
https://doi.org/10.1016/j.cellimm.2018.07.009
PMID:30107893

12. Shkunnikova S, Mijakovac A, Sironic L, Hanic M, Lauc G, Kavur MM. IgG glycans in health and disease: Prediction, intervention, prognosis, and therapy. Biotechnol Adv. 2023; 67:108169.
https://doi.org/10.1016/j.biotechadv.2023.108169
PMID:37207876

13. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, Pučić-Baković M, Novokmet M, Mangino M, Thaqi K, Rudan P, Novokmet N, Sarac J, et al. Glycans are a novel biomarker of chronological and biological ages. J Gerontol A Biol Sci Med Sci. 2014; 69:779–89.
https://doi.org/10.1093/gerona/glt190
PMID:24325898

14. Dall'Olio F, Vanhooren V, Chen CC, Slagboom PE, Wuhrer M, Franceschi C. N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. Ageing Res Rev. 2013; 12:685–98.
https://doi.org/10.1016/j.arr.2012.02.002
PMID:22353383

15. Franceschi C, Capri M, Monti D, Giunta S, Olivieri F, Sevini F, Panourgia MP, Invidia L, Celani L, Scurti M, Cevenini E, Castellani GC, Salvioli S. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. Mech Ageing Dev. 2007; 128:92–105.
https://doi.org/10.1016/j.mad.2006.11.016
PMID:17116321

16. Huhn C, Selman MHJ, Ruhaak LR, Deelder AM, Wuhrer M. IgG glycosylation analysis. Proteomics. 2009; 9:882–913.
https://doi.org/10.1002/pmic.200800715
PMID:19212958

17. Gornik O, Pavić T, Lauc G. Alternative glycosylation modulates function of IgG and other proteins - implications on evolution and disease. Biochim Biophys Acta. 2012; 1820:1318–26.
https://doi.org/10.1016/j.bbagen.2011.12.004
PMID:22183029

18. Pucić M, Knezević A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redzić I, Campbell H, Wright A, Hastie ND, et al. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Proteomics. 2011; 10:M111.010090.
https://doi.org/10.1074/mcp.M111.010090
PMID:21653738

19. Krištić J, Zaytseva OO, Ram R, Nguyen Q, Novokmet M, Vučković F, Vilaj M, Trbojević-Akmačić I, Pezer M, Davern KM, Morahan G, Lauc G. Profiling and genetic control of the murine immunoglobulin G glycome. Nat Chem Biol. 2018; 14:516–24.
https://doi.org/10.1038/s41589-018-0034-3
PMID:29632412

20. Parekh RB, Dwek RA, Sutton BJ, Fernandes DL, Leung A, Stanworth D, Rademacher TW, Mizuochi T, Taniguchi T, Matsuta K. Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. Nature. 1985; 316:452–7.
https://doi.org/10.1038/316452a0 PMID:3927174

---

21. Tomana M, Schrohenloher RE, Koopman WJ, Alarcón GS, Paul WA. Abnormal glycosylation of serum IgG from patients with chronic inflammatory diseases. Arthritis Rheum. 1988; 31:333–8.
https://doi.org/10.1002/art.1780310304
PMID:3358797

22. Novak J, Tomana M, Shah GR, Brown R, Mestecky J. Heterogeneity of IgG glycosylation in adult periodontal disease. J Dent Res. 2005; 84:897–901.
https://doi.org/10.1177/154405910508401005
PMID:16183787

23. Leirisalo-Repo M, Hernandez-Munoz HE, Rook GAW. Agalactosyl IgG is elevated in patients with active spondyloarthropathy. Rheumatol Int. 1999; 18:171–6.
https://doi.org/10.1007/s002960050080
PMID:10399791

24. Holland M, Yagi H, Takahashi N, Kato K, Savage COS, Goodall DM, Jefferis R. Differential glycosylation of polyclonal IgG, IgG-Fc and IgG-Fab isolated from the sera of patients with ANCA-associated systemic vasculitis. Biochim Biophys Acta. 2006; 1760:669–77.
https://doi.org/10.1016/j.bbagen.2005.11.021
PMID:16413679

25. Parekh RB, Roitt IM, Isenberg DA, Dwek RA, Ansell BM, Rademacher TW. Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease activity. Lancet. 1988; 1:966–9.
https://doi.org/10.1016/s0140-6736(88)91781-3
PMID:2896829

26. Siekman SL, Pongracz T, Wang W, Nouta J, Kremsner PG, da Silva-Neto PV, Esen M, Kreidenweiss A, Held J, Trapé ÁA, Fendel R, de Miranda Santos IKF, Wuhrer M, and ImmunoCovid Consortium. The IgG glycome of SARS-CoV-2 infected individuals reflects disease course and severity. Front Immunol. 2022; 13:993354.
https://doi.org/10.3389/fimmu.2022.993354
PMID:36389824

27. Bones J, Byrne JC, O'Donoghue N, McManus C, Scaife C, Boissin H, Nastase A, Rudd PM. Glycomic and glycoproteomic analysis of serum from patients with stomach cancer reveals potential markers arising from host defense response mechanisms. J Proteome Res. 2011; 10:1246–65.
https://doi.org/10.1021/pr101036b
PMID:21142185

28. Arnold JN, Saldova R, Hamid UMA, Rudd PM. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. Proteomics. 2008; 8:3284–93.
https://doi.org/10.1002/pmic.200800163
PMID:18646009

29. Menni C, Keser T, Mangino M, Bell JT, Erte I, Akmačić I, Vučković F, Pučić Baković M, Gornik O, McCarthy MI, Zoldoš V, Spector TD, Lauc G, Valdes AM. Glycosylation of immunoglobulin g: role of genetic and epigenetic influences. PLoS One. 2013; 8:e82558.
https://doi.org/10.1371/journal.pone.0082558
PMID:24324808

30. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, Novokmet M, Polašek O, Gornik O, Krištić J, Keser T, Vitart V, Scheijen B, et al. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet. 2013; 9:e1003225.
https://doi.org/10.1371/journal.pgen.1003225
PMID:23382691

31. Shen X, Klarić L, Sharapov S, Mangino M, Ning Z, Wu D, Trbojević-Akmačić I, Pučić-Baković M, Rudan I, Polašek O, Hayward C, Spector TD, Wilson JF, et al. Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. Nat Commun. 2017; 8:447.
https://doi.org/10.1038/s41467-017-00453-3
PMID:28878392

32. Wahl A, van den Akker E, Klaric L, Štambuk J, Benedetti E, Plomp R, Razdorov G, Trbojević-Akmačić I, Deelen J, van Heemst D, Slagboom PE, Vučković F, Grallert H, et al. Genome-Wide Association Study on Immunoglobulin G Glycosylation Patterns. Front Immunol. 2018; 9:277.
https://doi.org/10.3389/fimmu.2018.00277
PMID:29535710

33. Klarić L, Tsepilov YA, Stanton CM, Mangino M, Sikka TT, Esko T, Pakhomov E, Salo P, Deelen J, McGurnaghan SJ, Keser T, Vučković F, Ugrina I, et al. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. Sci Adv. 2020; 6:eaax0301.
https://doi.org/10.1126/sciadv.aax0301
PMID:32128391

34. Shadrina AS, Zlobin AS, Zaytseva OO, Klarić L, Sharapov SZ, D Pakhomov E, Perola M, Esko T, Hayward C, Wilson JF, Lauc G, Aulchenko YS, Tsepilov YA. Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. Hum Mol Genet. 2021; 30:1259–70.
https://doi.org/10.1093/hmg/ddab072
PMID:33710309

35. Mijakovac A, Miškec K, Krištić J, Vičić Bočkor V, Tadić V, Bošković M, Lauc G, Zoldoš V, Vojta A. A Transient Expression System with Stably Integrated CRISPR-dCas9 Fusions for Regulation of Genes Involved in Immunoglobulin G Glycosylation. CRISPR J. 2022;

5:237–53.
https://doi.org/10.1089/crispr.2021.0089
PMID:35021898

36. Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, Liu L, Trbojević-Akmačić I, Razdorov G, Štambuk J, Klarić L, Ugrina I, Selman MHJ, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. Nat Commun. 2017; 8:1483.
https://doi.org/10.1038/s41467-017-01525-0
PMID:29133956

37. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017; 8:1826.
https://doi.org/10.1038/s41467-017-01261-5
PMID:29184056

38. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. Genome Biol. 2016; 17:122.
https://doi.org/10.1186/s13059-016-0974-4
PMID:27268795

39. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, Brugge H, Oelen R, de Vries DH, et al, and BIOS Consortium, and i2QTL Consortium. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021; 53:1300–10.
https://doi.org/10.1038/s41588-021-00913-z
PMID:34475573

40. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014; 10:e1004383.
https://doi.org/10.1371/journal.pgen.1004383
PMID:24830394

41. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, Wang H, Zheng Z, Magi R, Esko T, Metspalu A, Wray NR, Goddard ME, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. Nat Commun. 2019; 10:5086.
https://doi.org/10.1038/s41467-019-12653-0
PMID:31704910

42. Mijakovac A, Frkatović A, Hanić M, Ivok J, Martinić Kavur M, Pučić-Baković M, Spector T, Zoldoš V, Mangino M, Lauc G. Heritability of the glycan clock of biological age. Front Cell Dev Biol. 2022; 10:982609.
https://doi.org/10.3389/fcell.2022.982609
PMID:36619858

43. Mijakovac A, Jurić J, Kohrt WM, Krištić J, Kifer D, Gavin KM, Miškec K, Frkatović A, Vučković F, Pezer M, Vojta A, Nigrović PA, Zoldoš V, Lauc G. Effects of Estradiol on Immunoglobulin G Glycosylation: Mapping of the Downstream Signaling Mechanism. Front Immunol. 2021; 12:680227.
https://doi.org/10.3389/fimmu.2021.680227
PMID:34113353

44. Nair M, Teng A, Bilanchone V, Agrawal A, Li B, Dai X. Ovol1 regulates the growth arrest of embryonic epidermal progenitor cells and represses c-myc transcription. J Cell Biol. 2006; 173:253–64.
https://doi.org/10.1083/jcb.200508196
PMID:16636146

45. Boettger LM, Handsaker RE, Zody MC, McCarroll SA. Structural haplotypes and recent evolution of the human 17q21.31 region. Nat Genet. 2012; 44:881–5.
https://doi.org/10.1038/ng.2334 PMID:22751096

46. Yamashita J, Iwamura C, Mitsumori K, Hosokawa H, Sasaki T, Takahashi M, Tanaka H, Kaneko K, Hanazawa A, Watanabe Y, Shinoda K, Tumes D, Motohashi S, Nakayama T. Murine Schnurri-2 controls natural killer cell function and lymphoma development. Leuk Lymphoma. 2012; 53:479–86.
https://doi.org/10.3109/10428194.2011.625099
PMID:21936769

47. Schumann K, Raju SS, Lauber M, Kolb S, Shifrut E, Cortez JT, Skartsis N, Nguyen VQ, Woo JM, Roth TL, Yu R, Nguyen MLT, Simeonov DR, et al. Functional CRISPR dissection of gene networks controlling human regulatory T cell identity. Nat Immunol. 2020; 21:1456–66.
https://doi.org/10.1038/s41590-020-0784-4
PMID:32989329

48. Hitomi Y, Nakatani K, Kojima K, Nishida N, Kawai Y, Kawashima M, Aiba Y, Nagasaki M, Nakamura M, Tokunaga K. NFKB1 and MANBA Confer Disease Susceptibility to Primary Biliary Cholangitis via Independent Putative Primary Functional Variants. Cell Mol Gastroenterol Hepatol. 2019; 7:515–32.
https://doi.org/10.1016/j.jcmgh.2018.11.006
PMID:30528300

49. Zeng X, Li S, Tang S, Li X, Zhang G, Li M, Zeng X, Hu C. Changes of Serum IgG Glycosylation Patterns in Primary Biliary Cholangitis Patients. Front Immunol. 2021; 12:669137.
https://doi.org/10.3389/fimmu.2021.669137
PMID:34248947

50. Cartwright T, Perkins ND, L Wilson C. NFKB1: a suppressor of inflammation, ageing and cancer. FEBS J. 2016; 283:1812–22.
https://doi.org/10.1111/febs.13627 PMID:26663363

51. Zhu M, Lovell KL, Patterson JS, Saunders TL, Hughes ED, Friderici KH. Beta-mannosidosis mice: a model for the

human lysosomal storage disease. Hum Mol Genet. 2006; 15:493–500.
https://doi.org/10.1093/hmg/ddi465
PMID:16377659

52. Gu X, Yang H, Sheng X, Ko YA, Qiu C, Park J, Huang S, Kember R, Judy RL, Park J, Damrauer SM, Nadkarni G, Loos RJF, et al. Kidney disease genetic risk variants alter lysosomal beta-mannosidase (*MANBA*) expression and disease severity. Sci Transl Med. 2021; 13:eaaz1458.
https://doi.org/10.1126/scitranslmed.aaz1458
PMID:33441424

53. Guo Y, Rist PM, Daghlas I, Giulianini F, Kurth T, Chasman DI, and International Headache Genetics Consortium, and 23andMe Research Team. A genome-wide cross-phenotype meta-analysis of the association of blood pressure with migraine. Nat Commun. 2020; 11:3368.
https://doi.org/10.1038/s41467-020-17002-0
PMID:32632093

54. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, Ma'ayan A. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford). 2016; 2016:baw100.
https://doi.org/10.1093/database/baw100
PMID:27374120

55. Wittenbecher C, Štambuk T, Kuxhaus O, Rudman N, Vučković F, Štambuk J, Schiborn C, Rahelić D, Dietrich S, Gornik O, Perola M, Boeing H, Schulze MB, Lauc G. Plasma *N*-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study. Diabetes Care. 2020; 43:661–8.
https://doi.org/10.2337/dc19-1507 PMID:31915204

56. Birukov A, Plavša B, Eichelmann F, Kuxhaus O, Hoshi RA, Rudman N, Štambuk T, Trbojević-Akmačić I, Schiborn C, Morze J, Mihelčić M, Cindrić A, Liu Y, et al. Immunoglobulin G N-Glycosylation Signatures in Incident Type 2 Diabetes and Cardiovascular Disease. Diabetes Care. 2022; 45:2729–36.
https://doi.org/10.2337/dc22-0833
PMID:36174116

57. Browne GJ, Proud CG. Regulation of peptide-chain elongation in mammalian cells. Eur J Biochem. 2002; 269:5360–8.
https://doi.org/10.1046/j.1432-1033.2002.03290.x
PMID:12423334

58. Li D, Wei T, Abbott CM, Harrich D. The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. Microbiol Mol Biol Rev. 2013; 77:253–66.
https://doi.org/10.1128/MMBR.00059-12

PMID:23699257

59. Abbas W, Kumar A, Herbein G. The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. Front Oncol. 2015; 5:75.
https://doi.org/10.3389/fonc.2015.00075
PMID:25905039

60. Schulz I, Engel C, Niestroj AJ, Kehlen A, Rahfeld JU, Kleinschmidt M, Lehmann K, Roßner S, Demuth HU. A non-canonical function of eukaryotic elongation factor 1A1: regulation of interleukin-6 expression. Biochim Biophys Acta. 2014; 1843:965–75.
https://doi.org/10.1016/j.bbamcr.2014.01.022
PMID:24487064

61. Hirano T. IL-6 in inflammation, autoimmunity and cancer. Int Immunol. 2021; 33:127–48.
https://doi.org/10.1093/intimm/dxaa078
PMID:33337480

62. Troelsen LN, Jacobsen S, Abrahams JL, Royle L, Rudd PM, Narvestad E, Heegaard NH, Garred P. IgG glycosylation changes and MBL2 polymorphisms: associations with markers of systemic inflammation and joint destruction in rheumatoid arthritis. J Rheumatol. 2012; 39:463–9.
https://doi.org/10.3899/jrheum.110584
PMID:22247351

63. Ozcan E, Garibyan L, Lee JJY, Bram RJ, Lam KP, Geha RS. Transmembrane activator, calcium modulator, and cyclophilin ligand interactor drives plasma cell differentiation in LPS-activated B cells. J Allergy Clin Immunol. 2009; 123:1277–86.e5.
https://doi.org/10.1016/j.jaci.2009.03.019
PMID:19427685

64. Castigli E, Wilson SA, Elkhal A, Ozcan E, Garibyan L, Geha RS. Transmembrane activator and calcium modulator and cyclophilin ligand interactor enhances CD40-driven plasma cell differentiation. J Allergy Clin Immunol. 2007; 120:885–91.
https://doi.org/10.1016/j.jaci.2007.06.012
PMID:17689597

65. Jonsson S, Sveinbjornsson G, de Lapuente Portilla AL, Swaminathan B, Plomp R, Dekkers G, Ajore R, Ali M, Bentlage AEH, Elmér E, Eyjolfsson GI, Gudjonsson SA, Gullberg U, et al. Identification of sequence variants influencing immunoglobulin levels. Nat Genet. 2017; 49:1182–91.
https://doi.org/10.1038/ng.3897
PMID:28628107

66. Martinez-Gallo M, Radigan L, Almejún MB, Martínez-Pomar N, Matamoros N, Cunningham-Rundles C. TACI mutations and impaired B-cell function in subjects with CVID and healthy heterozygotes. J Allergy Clin Immunol. 2013; 131:468–76.

https://doi.org/10.1016/j.jaci.2012.10.029
PMID:23237420

67. Poodt AEJ, Driessen GJA, de Klein A, van Dongen JJM, van der Burg M, de Vries E. TACI mutations and disease susceptibility in patients with common variable immunodeficiency. Clin Exp Immunol. 2009; 156:35–9.
https://doi.org/10.1111/j.1365-2249.2008.03863.x
PMID:19210517

68. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8:118–27.
https://doi.org/10.1093/biostatistics/kxj037
PMID:16632515

69. Leek JT. svaseq: removing batch effects and other unwanted noise from sequencing data. Nucleic Acids Res. 2014; 42:e161.
https://doi.org/10.1093/nar/gku864 PMID:25294822

70. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing. 2017.

71. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007; 23:1294–6.
https://doi.org/10.1093/bioinformatics/btm108
PMID:17384015

72. Haller T, Kals M, Esko T, Mägi R, Fischer K. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. Brief Bioinform. 2015; 16:39–44.
https://doi.org/10.1093/bib/bbt066
PMID:24008273

73. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44:821–4.
https://doi.org/10.1038/ng.2310 PMID:22706312

74. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, et al, and Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016; 48:1279–83.
https://doi.org/10.1038/ng.3643
PMID:27548312

75. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, Ferreira T, Fall T, Graff M, Justice AE, Luan J, Gustafsson S, Randall JC, et al, and Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014; 9:1192–212.
https://doi.org/10.1038/nprot.2014.071
PMID:24762786

76. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1.
https://doi.org/10.1093/bioinformatics/btq340
PMID:20616382

77. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018; 562:203–9.
https://doi.org/10.1038/s41586-018-0579-z
PMID:30305743

78. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164.
https://doi.org/10.1093/nar/gkq603
PMID:20601685

79. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, Ha B, Altay G, Greenbaum JA, McVicker G, Seumois G, Rao A, Kronenberg M, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell. 2018; 175:1701–15.e16.
https://doi.org/10.1016/j.cell.2018.10.022
PMID:30449622

80. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014; 343:1246949.
https://doi.org/10.1126/science.1246949
PMID:24604202

81. Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B, Crins F, Docampo E, Elansary M, Gori AS, Lecut C, Mariman R, Mni M, et al, and International IBD Genetics Consortium. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun. 2018; 9:2427.
https://doi.org/10.1038/s41467-018-04365-8
PMID:29930244

82. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–80.
https://doi.org/10.1016/j.cell.2014.11.021
PMID:25497547

83. Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009; 33:79–86.
https://doi.org/10.1002/gepi.20359
PMID:18642345

84. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015; 11:e1004219. https://doi.org/10.1371/journal.pcbi.1004219 PMID:25885710

85. Chang CC, Chow CC, Tellier LCA, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7. https://doi.org/10.1186/s13742-015-0047-8 PMID:25722852

86. Vink T, Oudshoorn-Dickmann M, Roza M, Reitsma JJ, de Jong RN. A simple, robust and highly efficient transient expression system for producing antibodies. Methods. 2014; 65:5–10. https://doi.org/10.1016/j.ymeth.2013.07.018 PMID:23872061

87. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25:402–8. https://doi.org/10.1006/meth.2001.1262 PMID:11846609

88. Viechtbauer W. Conducting meta-analyses in {R} with the {metafor} package. J Stat Softw. 2010; 36:1–48. https://doi.org/10.18637/jss.v036.i03

SUPPLEMENTARY MATERIALS

Supplementary Methods

Studied cohorts

Recruitment of participants, sample collection, genotyping and phenotyping in the cohorts used in the galactosylation GWAS was performed by staff members at the University of Zagreb, Croatia, University of Split Medical School, Croatia, UK, United Kingdom, King's College London, UK, German Institute of Human Nutrition, Germany, University of Tartu, Estonia, Leiden University Medical Centre (LUMC), Netherlands, Helmholtz Zentrum München – German Research Center for Environmental Health, Germany, and the nine study centers of the German Chronic Kidney Disease (GCKD) study.

IgG N-glycan quantification was performed by Genos Glycoscience Research Laboratory, Zagreb, Croatia and LUMC, Leiden, Netherlands.

TwinsUK

TwinsUK is a national registry of 12,000 volunteer twins in the UK. The cohort consists of 83% female subjects with a nearly equal number of monozygotic (51%) and dizygotic (49%) twin pairs. With an aim to study the genetics of healthy ageing and complex diseases, a sample of 7000 twins was assessed for a range of clinical, biochemical, behavioural and socio-economic characteristics. Moreover, several omics' datasets for the TwinsUK dataset are available including genome-wide SNP data. The study participants provided informed consent and ethical approval was obtained for academic and commercial use of the study [1].

The European Prospective Investigation into Cancer and Nutrition

The European Prospective Investigation into Cancer and Nutrition (EPIC) -Potsdam is a prospective cohort study that includes 27548 participants recruited from the Potsdam population in Germany from 1994 to 1998 [2]. Participants' age at recruitment ranges between 35 and 65, and the number of male and female subjects is 10904 and 16644, respectively. Initial data collection consisted of anthropometric measurements and blood sample collection used for omics' data derivation [3]. Ethical approval was obtained from the ethics committee in Germany and all participants provided informed consent [2].

CROATIA Vis, CROATIA-Split and CROATIA-Korcula

CROATIA-Vis, CROATIA-Korcula and CROATIA-Split cohorts were collected as part of the "10001 Dalmatians" study, a study of Croatian island isolates with participants from six Adriatic islands (Korčula, Vis, Lastovo, Susak, Rab, Mljet) and the city of Split. The aim of the study is to investigate genetic and environmental determinants in health and disease by using the advantage of genetically isolated populations. In the recruitment process, a total of 1008 participants aged 19-93 were recruited for the CROATIA-Vis cohort in villages of Vis and Komiza during 2003 and 2004, 1012 subjects aged 18-85 were recruited in 2009-2010 in the city of Split for CROATIA-Split cohort and data on CROATIA-Korcula subjects (aged 18-98) was collected from the island of Korcula, specifically from the town of Korcula and three villages including Lumbarda, Zrnovo and Racisce. Participants were assessed for a number of anthropometric and physiological measurements, and they donated overnight fasting blood samples which were later used for DNA analysis, biochemical measurements and molecular marker assessment [4]. The study participants provided signed informed consent. Ethical approval was obtained for each cohort from ethics committees in Croatia and Scotland.

The Orkney Complex Disease Study

The Orkney Complex Disease Study (ORCADES) is a family-based cohort collected with the goal of identifying genetic risk factors in complex diseases in the population of isolated Orkney Island in northern Scotland. The recruitment started in 2005 and lasted for six years during which 2080 participants were recruited. The subjects were included if they had at least two of their grandparents who were Orcadian. The initial data collection included cardiovascular measurements and fasting blood sample collection, followed by subsequent visits for cognitive function assessment, eye measurements and DEXA scans. The study was approved by ethics committees in Scotland and all participants gave informed consent [5].

Leiden Longevity Study

Leiden Longevity Study (LLS) is a family-based cohort from the Dutch population which was intended for studies of human longevity. Nonagenarian siblings (individuals having a sibling older than 89 years for men and 91 years for women) and their offspring and offspring's spouses (serve as controls) were included in the study if they were European. Initial recruitment

started in 2002 and ended in 2006 during which blood samples were collected for assessment of plasma parameters and genetic material extraction. A total of 3359 subjects were included: 944 long-lived proband siblings, 1671 offspring and 744 controls (offspring's spouses). This study was approved by the ethics committee of LUMC, Netherlands. Signed informed consent was given by all participants [6]. In the current study, a subset of 1190 participants including only offspring and their spouses is used.

The Cooperative Health Research in the Augsburg Region F4

The Cooperative Health Research in The Augsburg Region (KORA) F4 is a population-based study that was conducted between 2006 and 2008 as a follow-up of the KORA S4 study [7]. Participants were randomly selected from the population registry in the Augsburg region and two other neighbouring counties in Germany. The data collection included standard medical and physical examinations. A total of 3080 participants (51% females) aged 32-86 years were recruited [8]. Ethical approval for the study was obtained from the Ethics committee of Bavarian Chamber of Physicians, Germany. All participants provided signed informed consent before entering the study.

The Viking Health Study - Shetland

The Viking Health Study - Shetland (VIKING) is an epidemiologic study initiated to explore genetic risk factors for complex diseases. The cohort consists of individuals from an isolated population of Shetland in northern Scotland and the main criteria for recruitment was to have at least two grandparents from Shetland. Between 2013 and 2015, 2105 participants were recruited. A large number of distant relatives makes the VIKING cohort fit for the identification of rare genetic variants influencing the disease risk [9]. During initial data collection, data on health-related phenotypes and environmental parameters were collected and participants donated a fasting blood sample.

The Estonian Genome Center of the University of Tartu

The Estonian Genome Center of the University of Tartu (EGCUT) Biobank is a volunteer-based cohort of 52 thousand adult subjects aged ≥ 18 from the Estonian population. The recruitment was conducted throughout Estonia via general practitioners' offices and medical personnel during the 2002-2012 period. Besides completing a questionnaire on topics such as lifestyle, diet and clinical diagnostics, participants also donated blood samples. The cohort was utilized in the exploration

of more than 200 traits including anthropometric traits, common and rare diseases, blood biochemistry, as well as lifestyle and personality traits. The data on study participants is being continuously updated via follow-up health checks using electronic health registries and re-examinations [10].

German Chronic Kidney Disease study

German Chronic Kidney Disease (GCKD) study is an ongoing prospective study of kidney disease patients who are under nephrologist care in Germany [11]. The current sample size of 5217 makes it one of the largest chronic kidney disease cohort in the world. The subject enrolment was undertaken between 2010 and 2012 via nephrologist practice and outpatient care units of nine study centers throughout different regions in Germany. The mean age of study subjects is 60 years, with 40% of the participants being female. The data collection includes collecting information on sociodemographic factors, medical and family history, as well as obtaining blood samples in a standardized manner, which we immediately processed and stored frozen in a central biobank until measurement of the glycans.

Genotyping, genotype QC and genotype imputation

Genotyping was performed using commercially available SNP genotyping arrays, followed by genotype calling in Illumina and Genome Browser software. Quality control (QC) of genotype data was performed to exclude SNPs and samples with low quality including removal of 1) SNPs with low call rate, 2) SNPs violating the assumptions of Hardy-Weinberg Equilibrium (HWE) and 3) SNPs with low minor allele frequency (MAF) < 1%. Details on SNP exclusion criteria for each cohort and imputation to Haplotype Reference Consortium (HRC) [12] panel are shown in Appendix Table 8. The genotypes were mapped to Genome Reference Consortium GRCh37 (hg19) build.

IgG N-glycome analysis

IgG N-glycan quantification by ultra-performance liquid chromatography

Ultra-performance liquid chromatography (UPLC) is used for quantification of glycan structures attached to Fc (constant region) and Fab (variable region) portions of IgG without the possibility to distinguish them. Detailed protocol for UPLC quantification of IgG N-glycans is published elsewhere [13].

Briefly, IgG was isolated from blood plasma samples using Protein G plates (BIA Separations, Ajdovščina, Slovenia). After filtration, plates were extensively washed to remove unwanted proteins and IgG was

released from protein G monoliths using 0.1 M formic acid. Eluates were collected in a 96-well plate and neutralized with neutralization buffer (1 M ammonium bicarbonate) to pH 7.0 to maintain the stability of IgG. IgG samples were dried and denatured using SDS detergent and incubated at 65° C for 10 minutes. N-glycans from IgG were released using recombinant N-glycosidase F followed by fluorescent labelling with 2-aminobenzamide (2-AB) dye. Hydrophilic interaction liquid chromatography (HILIC) based solid-phase extraction (SPE) was used to remove excess protein, reagents and fluorescent label. Fluorescently labelled N-glycans were separated by hydrophilic interaction UPLC on Waters Aquity UPLC H-class instrument (Waters, Milford, MA) with Waters bridged ethylene hybrid (BEH) glycan chromatography column. A linear gradient of 75 to 62% ACN in a 20-min analytical run was used to separate different glycan structures. The retention times for individual glycans were converted to glucose units based on hydrolysed and 2-AB labelled glucose oligomers which were used as external standards for calibration of the system. Data processing was done in two ways depending on the cohort, 1) automatic integration as described in Agakova et al. [14] or 2) using Empower 3 software with an automated processing method with traditional integration algorithm, followed by manual correction of each chromatogram to maintain the same integration intervals in all samples. The resulting chromatograms were separated into 24 peaks where the amount of glycans was expressed as % of the total integrated area in the corresponding peak (GP1-GP24). Total separation of each glycan structure is not possible using the described method, thus resulting in multiple glycan structures being detected under ten peaks. Glycan structures in each peak are listed in Appendix Table 9.

Glycan quantification by liquid chromatography coupled with mass spectrometry

The full name of the method is reverse-phase nano-liquid-chromatography-sheath-flow-electrospray-mass spectrometry (LC-ESI-MS) but in this study, we refer to it as LC-MS. The detailed protocol for analysis of IgG N-glycans using LC-MS is described in Selman et al. [15]. Briefly, IgG was isolated by affinity chroma-tography binding to protein G 96-well plates (BIA Separation, Ajdovščina, Slovenia) and treated with trypsin overnight at 37° C which allowed cleavage of IgG at specific amino acid sites. The cleavage by trypsin resulted in different glycopeptides due to the difference of amino acid sequence in different IgG subclasses, thereby enabling subclass-specific glycan measurements. IgG subclass separation was performed using the Ultimate 3000 HPLC system (Dionex Corporation, Sunnyvale, CA). The SPE trap column was conditioned

with mobile phase A and samples were loaded and separated on Ascentis Express C18 nano-LC column (Supelco, Bellefonte, USA) conditioned with mobile phase A and 95% ACN. For detection of separated subclass-specific glycopeptides, the HPLC system was coupled to a Dionex Ultimate UV detector and interfaced to a quadrupole-TOF-MS mass spectrometer (Bruker Daltonics, Bremen, Germany) with a standard ESI source (Bruker Daltonics, Bremen, Germany) and a sheath-flow ESI sprayer (Agilent Technologies, Santa Clara, USA). The mass spectra were recorded in a range between 300 and 2000 m/z with two averages at the frequency of 1Hz. The analysis time for one sample was 16 minutes. The calibration of LCMS datasets was done internally using a list of known glycopeptides and datasets were exported to the open mzXML format by Bruker DataAnalysis 4.0 software, followed by alignment to a master dataset of a typical sample. In-house software "Xtractor2D" was used to extract pre-defined features such as peak maximum or peak area in specific retention time and mass windows. Relative intensities of subclass-specific glycopeptides were obtained by integrating and summing three isotopic peaks. The obtained intensities were then normalized to the total IgG subclass-specific glycopeptide intensities. IgG2 and IgG3 subclasses have the same tryptic glycopeptide moieties, thus not enabling the separation of the subclass-specific glycopeptides. Here, obtained measurements are simply referred to as IgG2/3. LC-MS quantification results in 50 values which refer to 20 glycans measured on IgG1, 20 glycans on IgG2/3 and 10 glycans on IgG4. All glycans measured on IgG4 are fucosylated structures since the nonfucosylated glycans are hard to distinguish from the glycans found on IgG1 [16]. The list of glycans measured by LC-MS and their description is listed in Appendix Table 10.

IgG glycan data harmonization

Previously, there were no GWA meta-analyses of IgG N-glycan patterns using GWAS of both UPLC- and LC-MS-derived IgG N-glycan traits, therefore making it necessary to first assess the correlation of the data and methods which should be applied in pre-processing step to make UPLC and LC-MS glycan traits comparable. For this purpose, we used the CROATIA-Vis cohort (n=661) as both UPLC and LC-MS IgG N-glycan measurements are available in the same samples.

We aimed to combine IgG subclass information obtained from LC-MS in an appropriate manner to obtain information corresponding to total IgG glycome values measured by UPLC. Pre-processing of IgG glycome data consists of normalization of the data and batch correction to remove the effects of experimental variation. Normalization procedure is necessary to

remove unwanted variation between the samples and allows quantitative comparison of the samples [17]. We tested the following three normalization types: total area normalization, largest peak normalization and median quotient normalization, all of which can be applied using "glycanr" [18] package in R software [19]. We tested different normalizations both across the total glycome and per IgG subclass in LC-MS data.

Due to varying laboratory conditions during IgG N-glycan measurement, it was necessary to perform batch correction to remove non-biological, experimental variation. The batch correction was performed using ComBat function in R package "sva" [20]. We first log-transform the data as ComBat function implements empirical Bayes method for batch correction [21] which assumes a normal distribution of the data, followed by batch correction with ComBat() where the batch is denoted by the plate on which the samples were analysed, and lastly, exponential transformation of the values to the original scale.

We calculated derived traits from the initial traits for purpose of data harmonization and to enable a more straightforward interpretation of the GWAS results so that the discovered genomic loci can be directly linked to the addition of one or two galactose residues to the IgG N-glycan chain: agalactosylation (G0), monogalactosylation (G1) and digalactosylation (G2). Formulas for calculation are listed in Appendix Table 11.

Since LC-MS data quantifies glycans attached to different IgG subclasses, in order to calculate different galactosylation traits we also incorporated the approximation of the IgG glycan subclass response factor (RF) to represent the IgG subclass concentration relative to other subclasses. RF is defined as the ratio between the concentration of the analyte and instrument response to the analyte. Using unpublished in-house experimental data, the subclass-specific response factors were approximated as follows: IgG1 with RF=1, IgG2/IgG3 with RF=2 and IgG4 with RF=1. IgG subclasses are present in different quantities in human serum; hence we also incorporated relative concentrations of each IgG subclass in the calculation of derived traits. The following relative measurements were indicated in the literature: 66% for IgG1, 30% for IgG2/IgG3 and 4% for IgG4 [22]. The subclass-specific glycan measurements were weighted by the corresponding concentration and response factor before trait calculation.

Pre-processing of glycan data

Glycan data was pre-processed centrally in Genos for all cohorts except the LLS cohort for which the glycan data was pre-processed by a colleague from Leiden

University Medical Centre. It is important to note that glycan data for the CROATIA-Korcula cohort was obtained in three instances (2010, 2013 and 2017) and each dataset was pre-processed separately and treated as an individual cohort in downstream analysis. Additionally, the TwinsUK cohort was analysed in four separate batches. Due to differences in sample collection, batches 1 and 2 were treated as one dataset and batches 3 and 4 were treated as the second dataset.

Data points in the 99.9^{th} percentile were removed and considered as technical outliers. Next, based on the results of the previous data harmonization assessment, median quotient normalization was applied on both UPLC and LC-MS glycan data across 24 and 50 glycan measurements, respectively and the batch correction was applied. Prior to the genetic association test, galactosylation traits in all cohorts were transformed using rank-based inverse normal transformation (mean=0, standard deviation=1).

Supplementary References

1. Moayyeri A, Hammond CJ, Hart DJ, Spector TD. The UK Adult Twin Registry (TwinsUK Resource). Twin Res Hum Genet. 2013; 16:144–9.
https://doi.org/10.1017/thg.2012.89 PMID:23088889

2. Boeing H, Korfmann A, Bergmann MM. Recruitment procedures of EPIC-Germany. European Investigation into Cancer and Nutrition. Ann Nutr Metab. 1999; 43:205–15.
https://doi.org/10.1159/000012787 PMID:10592369

3. Bergmann MM, Bussas U, Boeing H. Follow-up procedures in EPIC-Germany--data quality aspects. European Prospective Investigation into Cancer and Nutrition. Ann Nutr Metab. 1999; 43:225–34.
https://doi.org/10.1159/000012789 PMID:10592371

4. Campbell H, Carothers AD, Rudan I, Hayward C, Biloglav Z, Barac L, Pericic M, Janicijevic B, Smolej-Narancic N, Polasek O, Kolcic I, Weber JL, Hastie ND, et al. Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. Hum Mol Genet. 2007; 16:233–41.
https://doi.org/10.1093/hmg/ddl473 PMID:17220173

5. McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008; 83:359–72.
https://doi.org/10.1016/j.ajhg.2008.08.007 PMID:18760389

6. Schoenmaker M, de Craen AJ, de Meijer PHE, Beekman M, Blauw GJ, Slagboom PE, Westendorp RGJ. Evidence

of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet. 2006; 14:79–84.
https://doi.org/10.1038/sj.ejhg.5201508
PMID:16251894

7. Holle R, Happich M, Löwel H, Wichmann HE, and MONICA/KORA Study Group. KORA--a research platform for population based health research. Gesundheitswesen. 2005 (Suppl 1); 67:S19–25.
https://doi.org/10.1055/s-2005-858235
PMID:16032513

8. Rathmann W, Kowall B, Tamayo T, Giani G, Holle R, Thorand B, Heier M, Huth C, Meisinger C. Hemoglobin A1c and glucose criteria identify different subjects as having type 2 diabetes in middle-aged and older populations: the KORA S4/F4 Study. Ann Med. 2012; 44:170–7.
https://doi.org/10.3109/07853890.2010.531759
PMID:21091229

9. Halachev M, Meynert A, Taylor MS, Vitart V, Kerr SM, Klaric L, Consortium SGP, Aitman TJ, Haley CS, Prendergast JG, Pugh C, Hume DA, Harris SE, Liewald DC, et al. Increased ultra-rare variant load in an isolated Scottish population impacts exonic and regulatory regions. PLoS Genet. 2019; 15:e1008480.
https://doi.org/10.1371/journal.pgen.1008480
PMID:31765389

10. Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, Perola M, Ng PC, Mägi R, Milani L, Fischer K, Metspalu A. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. Int J Epidemiol. 2015; 44:1137–47.
https://doi.org/10.1093/ije/dyt268 PMID:24518929

11. Titze S, Schmid M, Köttgen A, Busch M, Floege J, Wanner C, Kronenberg F, Eckardt KU, and GCKD study investigators. Disease burden and risk profile in referred patients with moderate chronic kidney disease: composition of the German Chronic Kidney Disease (GCKD) cohort. Nephrol Dial Transplant. 2015; 30:441–51.
https://doi.org/10.1093/ndt/gfu294 PMID:25271006

12. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GR, Durbin R, L Price A. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016; 48:1443–8.
https://doi.org/10.1038/ng.3679 PMID:27694958

13. Pucić M, Knezević A, Vidic J, Adamczyk B, Novokmet M, Polasek O, Gornik O, Supraha-Goreta S, Wormald MR, Redzić I, Campbell H, Wright A, Hastie ND, et al. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. Mol Cell Proteomics. 2011; 10:M111.010090.
https://doi.org/10.1074/mcp.M111.010090
PMID:21653738

14. Agakova A, Vučković F, Klarić L, Lauc G, Agakov F. Automated Integration of a UPLC Glycomic Profile. Methods Mol Biol. 2017; 1503:217–33.
https://doi.org/10.1007/978-1-4939-6493-2_17
PMID:27743370

15. Selman MHJ, Derks RJE, Bondt A, Palmblad M, Schoenmaker B, Koeleman CAM, van de Geijn FE, Dolhain RJE, Deelder AM, Wuhrer M. Fc specific IgG glycosylation profiling by robust nano-reverse phase HPLC-MS using a sheath-flow ESI sprayer interface. J Proteomics. 2012; 75:1318–29.
https://doi.org/10.1016/j.jprot.2011.11.003
PMID:22120122

16. Huffman JE, Pučić-Baković M, Klarić L, Hennig R, Selman MHJ, Vučković F, Novokmet M, Krištić J, Borowiak M, Muth T, Polašek O, Razdorov G, Gornik O, et al. Comparative performance of four methods for high-throughput glycosylation analysis of immunoglobulin G in genetic and epidemiological research. Mol Cell Proteomics. 2014; 13:1598–610.
https://doi.org/10.1074/mcp.M113.037465
PMID:24719452

17. Karaman I. Preprocessing and Pretreatment of Metabolomics Data for Statistical Analysis. Adv Exp Med Biol. 2017; 965:145–61.
https://doi.org/10.1007/978-3-319-47656-8_6
PMID:28132179

18. Ugrina I, Klaric L, Vuckovic F, Russell A. glycanr: Tools for Analysing N-Glycan Data [Internet]. 2018.

19. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing. 2017.

20. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012; 28:882–3.
https://doi.org/10.1093/bioinformatics/bts034
PMID:22257669

21. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007; 8:118–27.
https://doi.org/10.1093/biostatistics/kxj037
PMID:16632515

22. Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. Front Immunol. 2014; 5:520.
https://doi.org/10.3389/fimmu.2014.00520
PMID:25368619
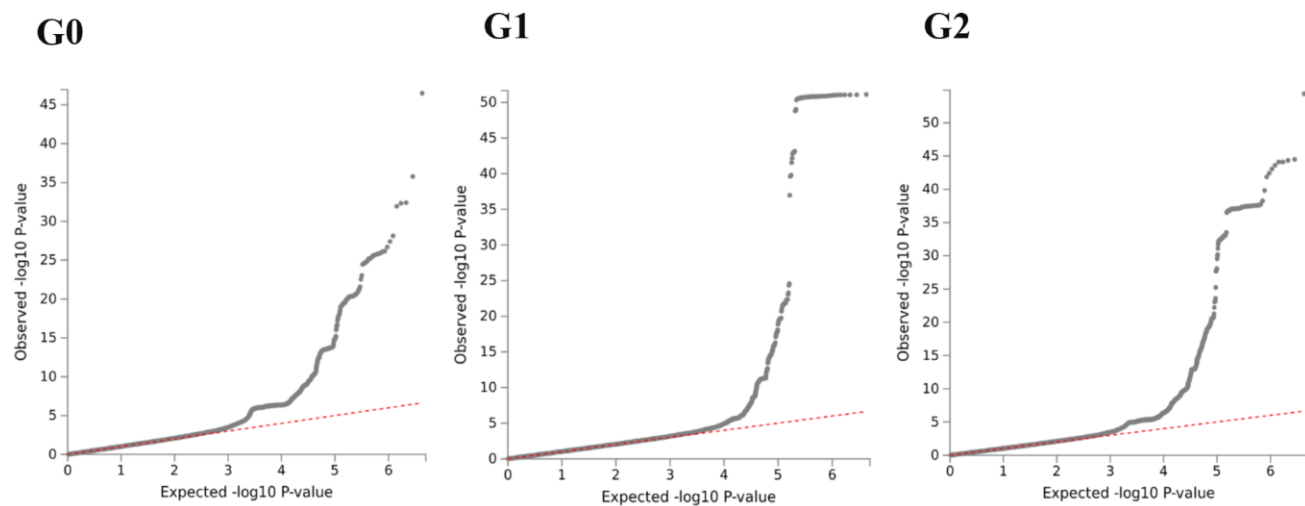
Supplementary Figure 1. QQ plots of association analysis for G0, G1 and G2 traits. Genomic inflation (lambda) for meta-analysis was 1.02, 1.01 and 1.03 for G0, G1 and G2, respectively.
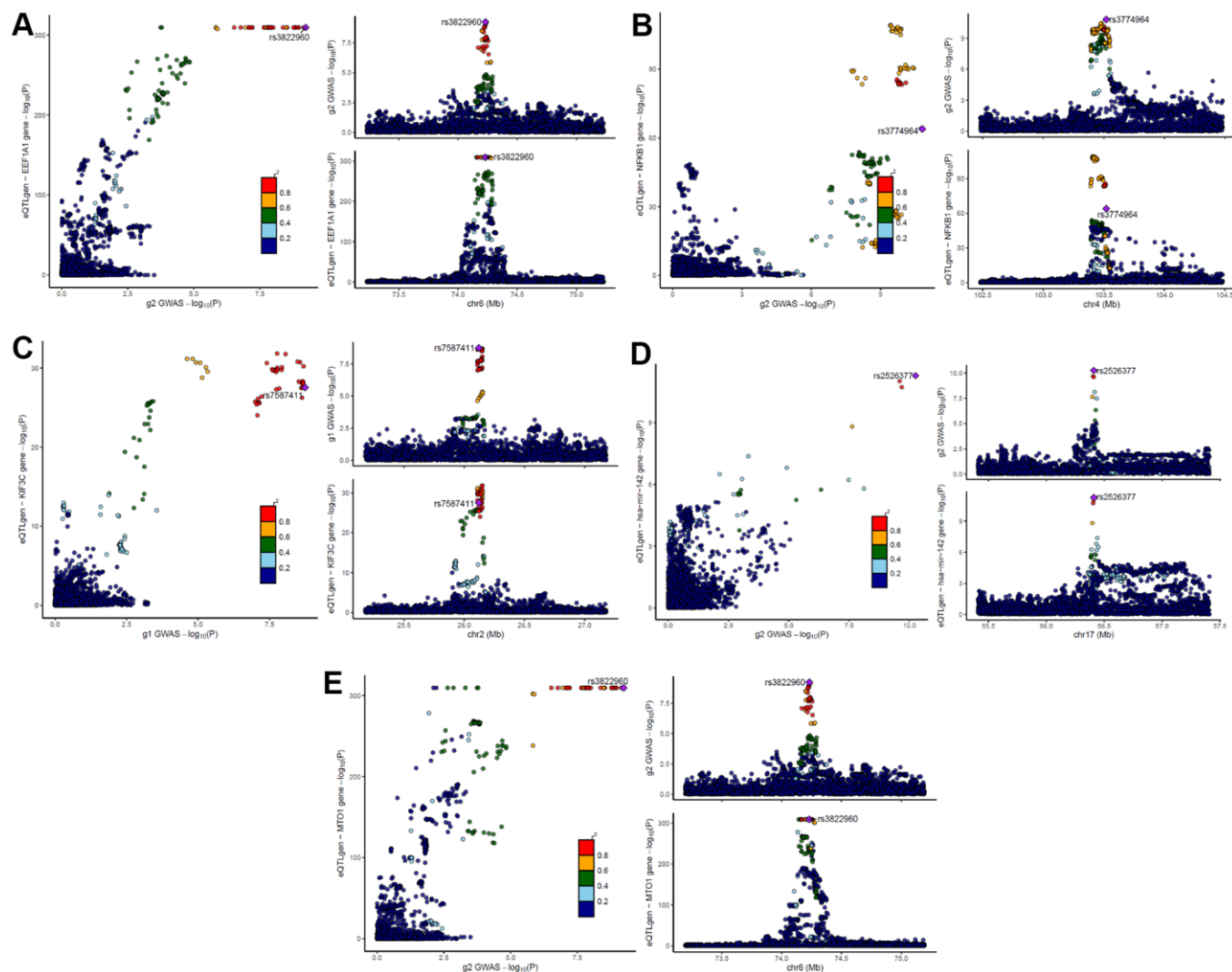
Supplementary Figure 2. IgG galactosylation and eQTL association plots at loci with colocalization (PP4 > 75%). Association plot for IgG galactosylation GWAS and eQTL are shown in the top right and bottom right corners, respectively; x-axis shows the chromosomal position, y-axis shows -log10(p-value) of the association. Correlation of -log10(p-values) for IgG galactosylation GWAS and eQTL are shown on left side. (**A**) Colocalization in *EEF1A1* locus (prioritized genes *EEF1A1* and *MTO1*), (**B**) Colocalization in *NFKB1* locus (prioritized genes *NFKB1* and *MANBA*), (**C**) Colocalization in *KIF3C* locus (prioritized gene *KIF3C*), (**D**) Colocalization in *hsa-mir-142* locus (prioritized genes *BZRAP1*, *SUPT4H1* and *RAD5C1*), (**E**) Colocalization in *MTO1* locus (prioritized genes *EEF1A1* and *MTO1*).
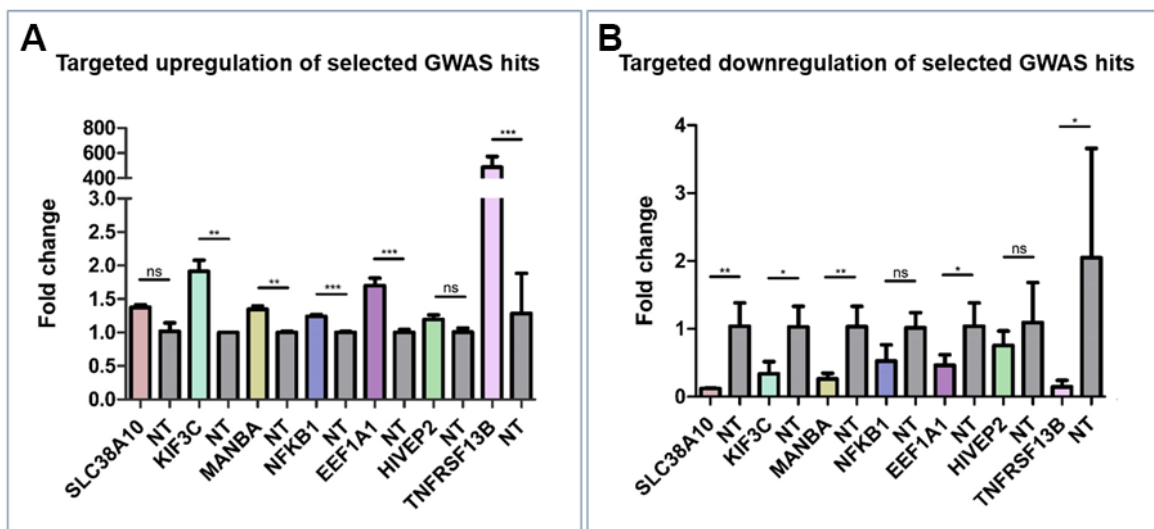
Supplementary Figure 3. Upregulation and downregulation of selected GWAS hits. Selected GWAS hits *associated* with IgG galactosylation (*SLC38A10, KIF3C, MANBA, NFKB1, EEF1A1, HIVEP2* and *TNFRSF13B*) were targeted in dCas9-VPR and dCas9-KRAB monoclonal cell lines. Samples containing non-targeting gRNAs served as controls. Changes of transcript levels are given as fold change values. (**A**) Targeting of selected GWAS hits with specific gRNA molecules in dCas9-VPR cell line resulted in significant increase of transcript levels of all loci except for *SLC38A10* and *HIVEP2*. (**B**) Targeting of the same genes in dCas9-KRAB call line resulted in significant downregulation of all genes except for *NFKB1* and *HIVEP2.* Statistical significance: *<0.05; **<0.01; ***<0.001; ns, not significant.
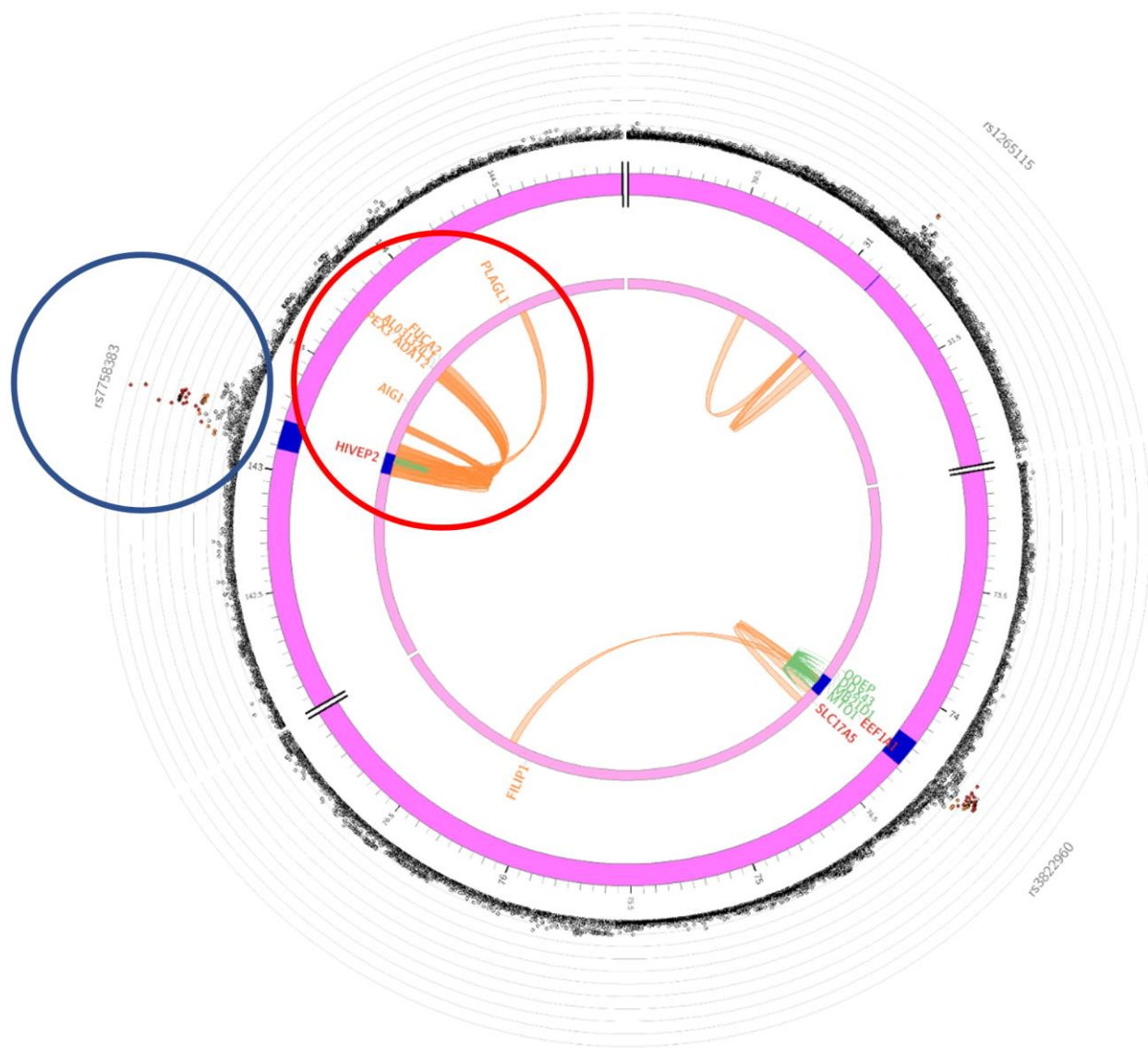
Supplementary Figure 4. Chromatin interaction map based on Hi-C data for GM12878 cell line (GSE87112) for chr6:143088071-143206826 locus indicating *FUCA2* and 4 other genes mapped via chromatin interaction with genetic variants in *HIVEP2* locus. Blue circle denotes the defined genomic locus; red circle denotes genes mapped via chromatin interaction mapping by FUMA.
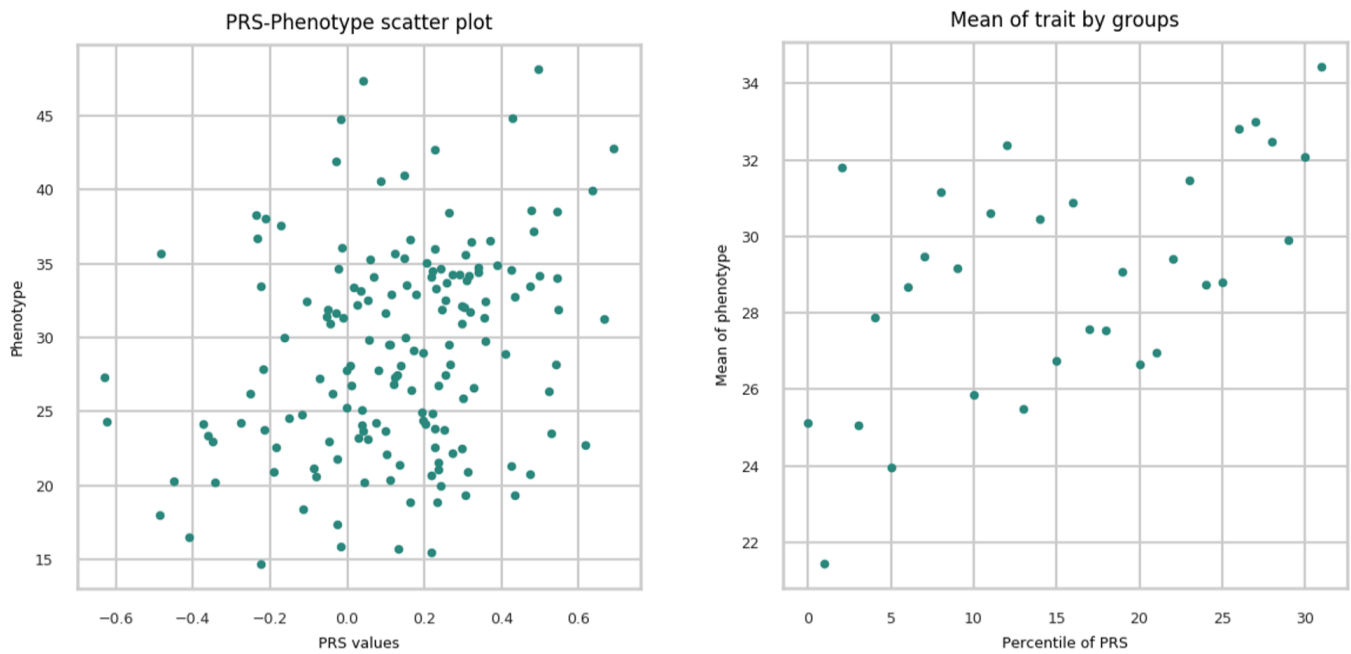
Supplementary Figure 5. PRS-Phenotype scatter plot (G0) and mean of trait by groups. The graph on the left-hand side shows the dependence of phenotype values (G0) on PRS values. The graph on the right-hand side is data stratified into 30 groups according to the PRS percentile. The Y-axis shows the average value of the trait for the group, the X-axis shows the group number.
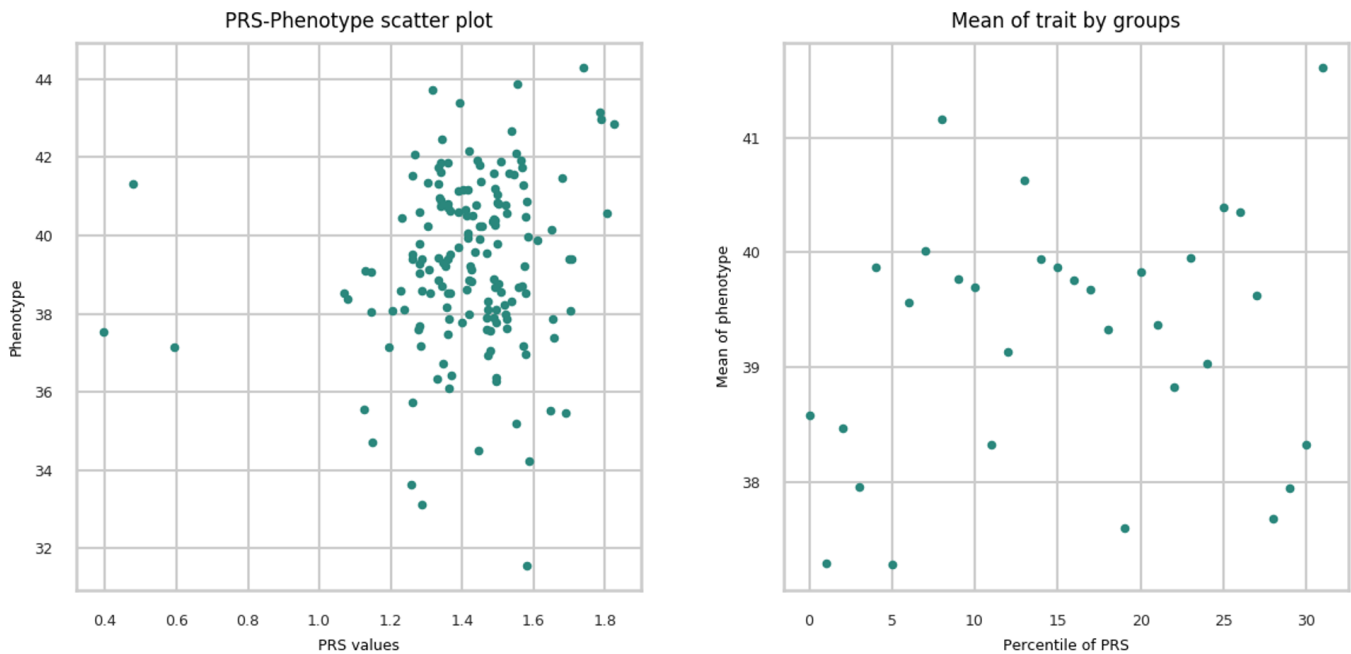
Supplementary Figure 6. PRS-Phenotype scatter plot (G1) and mean of trait by groups. The graph on the left-hand side shows the dependence of phenotype values (G1) on PRS values. The graph on the right-hand side shows data stratified into 30 groups according to the PRS percentile. The Y-axis shows the average value of the trait for the group, the X-axis shows the group number.
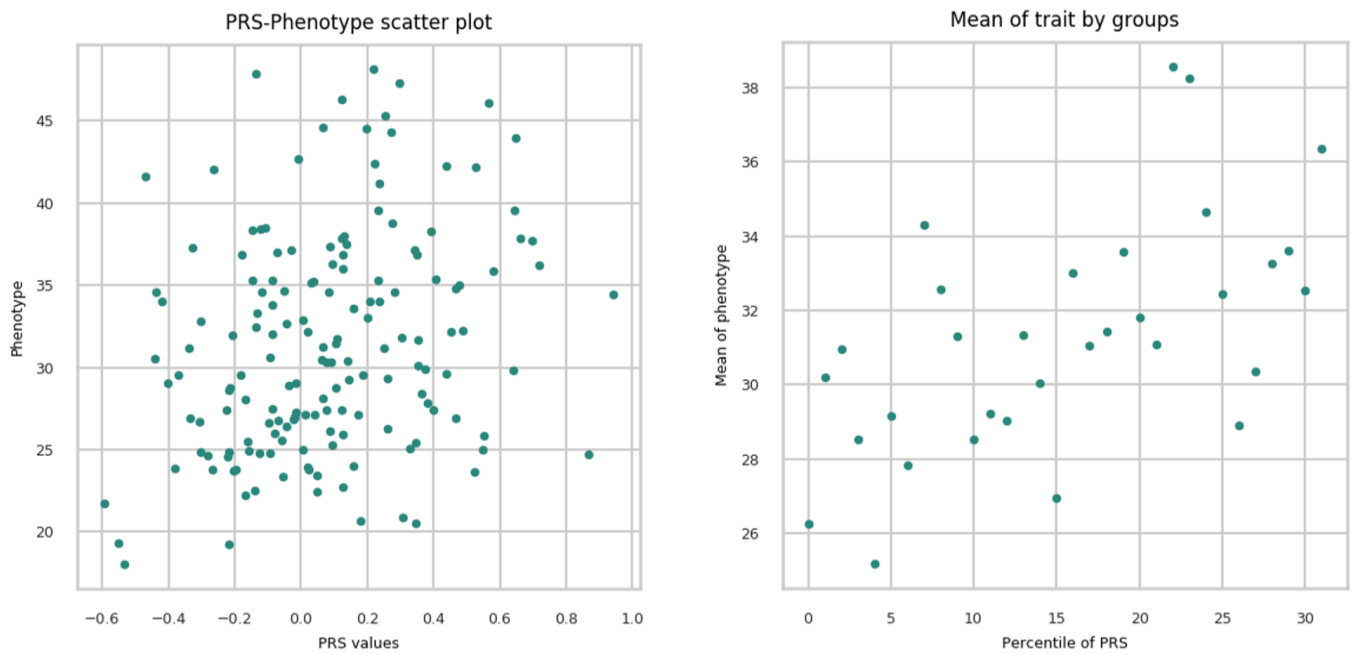
Supplementary Figure 7. PRS-Phenotype scatter plot (G2) and mean of trait by groups. The graph on the left-hand side shows the dependence of phenotype values (G2) on PRS values. The graph on the right-hand sideshows data stratified into 30 groups according to the PRS percentile. The Y-axis shows the average value of the trait for the group, the X-axis shows the group number.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1–4, 6.

Supplementary Table 1. Gene prioritization table with overview of evidence used to priotize a gene in a given locus.

Supplementary Table 2. Results of the colocalization analysis for IgG galactosylation signals and eQTL signals for gene in the discovered genomic loci.

Supplementary Table 3. Results of the colocalization analysis for IgG galactosylation signals and disease signals in the discovered genomic loci.

Supplementary Table 4. Raw data results of RT-qPCR for up (VPR) and down (KRAB) regulation of genes KIF3C, NFKB1, MANBA, SLC38A10, TNFRSF13B, and HIVEP2.

Supplementary Table 5. Overview of traits and diseases in colocalization analysis.

Trait	First author	Accession number GWAS catalog	N total	Cases	Control	Case prop	Download link
Adult-onset asthma	Ferreira MAR et al	GCST007799	327253	26582	300671	0.081	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/FerreiraMAR_30929738_GCST007799
Primary billiary cirrhosis	Cordell HJ	GCST003129	13239	2764	10475	0.209	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/CordellHJ_26394269_GCST003129/harmonised/26394269-GCST003129-EFO_1001486-Build37.f.tsv.gz
Asthma	Han Y	GCST010042	303859	64538	239321	0.212	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/HanY_32296059_GCST010042
Systemic Lupus Erythematosus	Julia A	GCST005831	16966	4943	8483	0.291	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/JuliaA_29848360_GCST005831
Type I diabetes	Forgetta V	GCST010681	24840	9266	15574	0.373	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ForgettaV_32005708_GCST010681
IgG level	Scepanovic P	GCST006357	1000	NA	NA	NA	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/ScepanovicP_30053915_GCST006357
Schizophrenia	Pardinas AF	GCST006803	105318	40675	64643	0.386	https://walters.psycm.cf.ac.uk/clozuk_pgc2.meta.sumstats.txt.gz
Rheumatoid arthritis	Eyre S	GCST005569	47580	13838	33742	0.290	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/EyreS_23143596_GCST005569
Type II diabetes	Mahajan A	GCST007518	298957	48286	250671	0.162	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/MahajanA_29632382_GCST007518_T2D_European.BMIadjusted.txt
Allergic Disease	Ferreira	GCST005038	360838	180129	180709	0.499	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/FerreiraMA_29083406_GCST005038
Alzheimer's disease	Kunkle	NA	63926	21982	41944	0.344	https://www.niagads.org/datasets/ng00075
Lymphocyte percentage of white blood cells	Astle WJ	GCST004632	171748	NA	NA	NA	ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/AstleWJ_27863252_GCST004632
Total Cholesterol	Willer CJ	GCST002221	94595	NA	NA	NA	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST002001-GCST003000/GCST002221/harmonised/24097068-GCST002221-EFO_0004574-build37.f.tsv.gz
Crohn's disease	Liu JZ	GCST003044	20883	5956	14927	0.285	http://www.ibdgenetics.org/downloads.html
Ulcerative colitis	Liu JZ	GCST003045	27432	6968	20464	0.254	http://www.ibdgenetics.org/downloads.html
Inflammatory bowel disease	Liu JZ	GCST003043	34652	12882	21770	0.372	http://www.ibdgenetics.org/downloads.html
HDL cholesterol	Willer CJ	GCST002223	94595	NA	NA	NA	http://ftp.ebi.ac.uk/pub/databases/gwas/summar

Trait	First author	Accession No	N total	Cases	Controls	Case prop	Download link
LDL	Willer CJ	GCST002222	94595	NA	NA	NA	y_statistics/GCST002001-GCST003000/GCST002223/harmonised/24097068-GCST002223-EFO_0004612-build37.f.tsv.gz http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST002001-GCST003000/GCST002222/harmonised/24097068-GCST002222-EFO_0004611-build37.f.tsv.gz
Juvenile idiopathic arthritis	Hinks A	GCST005528	15872	2816	13056	0.177	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST005001-GCST006000/GCST005528/harmonised/23603761-GCST005528-EFO_1001999-Build37.f.tsv.gz
Osteoarthritis	Tachmazidou I	GCST007092	417596	39427	378169	0.094	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST007001-GCST008000/GCST007092/harmonised/30664745-GCST007092-EFO_0002506-build37.f.tsv.gz
Chronic kidney disease	Wuttke M	GCST008065	625219	64164	561055	0.103	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST008001-GCST009000/GCST008065/CKD_overall_EA_JW_20180223_nstud23.dbgap.txt.gz
Hypertension	Zhu Z	GCST007610	458554	144793	313761	0.316	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST007001-GCST008000/GCST007610/ZhuZ_30940143_ukbb.bolt_460K_selfRepWhite.doctor_highbloodpressure.assoc.gz
Thyroid cancer	Zhou W	GCST008371	407757	358	407399	0.001	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST008001-GCST009000/GCST008371/PheCode_193_SAIGE_MACge20.txt.vcf.gz
Lung cancer	Rashkin SR	GCST90011812	412835	2485	410350	0.006	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011001-GCST90012000/GCST90011812/harmonised/32887889-GCST90011812-EFO_0001071-Build37.f.tsv.gz
Ovarian cancer	Rashkin SR	GCST90011821	411609	1259	410350	0.003	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011001-GCST90012000/GCST90011821/harmonised/32887889-GCST90011821-EFO_0001075-Build37.f.tsv.gz
Colorectal cancer	Zhou W	GCST008372	387318	4562	382756	0.012	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST008001-GCST009000/GCST008372/PheCode_153_SAIGE_MACge20.txt.vcf.gz
Breast cancer	Rashkin SR	GCST90011804	428231	17881	410350	0.042	http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90011001-GCST90012000/GCST90011804/harmonised/32887889-GCST90011804-EFO_0000305-Build37.f.tsv.gz
COVID-19 (Hospitalized vs. Population)	COVID-19 Host Genetics Initiative	NA	1887658	9986	1877672	0.005	https://storage.googleapis.com/covid19-hg-public/20201215/results/20210107/COVID19_HGI_B2_ALL_eur_leave_23andme_20210107.b37.txt.gz

Trait- tested trait or disease; First author- first author of the publication describing the GWAS of the trait or disease; Accession No- accession number in GWAS Catalog; N total- total number of subjects in the study; Cases- number of case subjects in the study; Controls- number of control subjects in the study; Case prop- proportion of the cases in the total sample; Download link- link for download of the GWAS summary statistics for the given trait or disease.

Supplementary Table 6. Raw experimental data of 24 glycan traits after gene upregulation (VPR) shown as a percentage of area under chromatogram peaks replicated in two experiments (exp1 and exp2).

Appendix Tables

Please browse Full Text version to see the data of Supplementary Table 2.

Appendix Table 1. Pearson's correlation coefficients for UPLC and LC-MS derived glycan traits after applying different normalization types and weighting of the subclass-specific values in LC-MS measurements.

UPLC	Total area	Largest peak	Median quotient	Total area	Total area	Total area	Total area	Total area	Total area	Total area
LCMS	Total area	Largest peak	Median quotient	Total area with trait subclass average	Total area with weighted subclass average	Total area per subclass with weighted average	Total area per subclass	Total area per subclass with trait average	Total area with subclass concentration applied to raw values	Total area with response factor applied to raw values
G0	0.96	0.96	0.96	0.95	0.97	0.97	0.95	0.95	0.96	0.95
G1	0.78	0.77	0.78	0.75	0.80	0.79	0.75	0.75	0.79	0.73
G2	0.91	0.91	0.91	0.91	0.91	0.91	0.90	0.90	0.91	0.90

Combination of normalization types as applied in UPLC- and LC- MS-derived data is shown on the top.

Appendix Table 2. List of IgG galactosylation associated loci and comparison of GWAS summary statistics for significantly associated genomic regions in current GWAS with previous GWAS of IgG N-glycome.

Appendix Table 3. List of genes mapped by positional, eQTL and chromatin interaction mapping in FUMA.

Mapping strategy	Mapped genes
Positional mapping	*GRHL3,STPG1,NIPAL3,RCAN3,SRRM1,RUNX3,ASXL2,KIF3C,NFKB1,MANBA,MB21D1,MTO1,EEF1A1, HIVEP2,GBX1,ABCF2,CHPF2,SMARCD3,KB1980E6.3,APTX,DNAJA1,SMU1,B4GALT1,SPINK4,BAG1,C HMP5,NFX1,AQP7,AP5B1,OVOL1,TNFRSF13B,CRHR1,SPPL2C,MAPT,STH,KANSL1,ARL17B,LRRC37A, NSF,WNT3,KPNB1,TBKBP1,TBX21,BZRAP1,SUPT4H1,AZI1,ENTHD2,C17orf89,SLC38A10,RUNX1*
eQTL mapping	*MANBA,EEF1A1,NSF,MTO1,OOEP,DNAJA1,RAD51C,TMEM176A,TBKBP1,SUPT4H1,GPS1,CISD2,APTX ,LRRC37A,WNT3,B4GALT1,LRRC37A2,C17orf89,KB-1980E6.3,SMU1,OVOL1,KANSL1,ARHGAP27,ARL17A,ARL17B*
Chromatin interaction mapping	*CNR2,PNRC2,SRSF10,MYOM3,IL22RA1,IFNLR1,GRHL3,STPG1,NIPAL3,RCAN3,SRRM1,SYF2,DAPP1,D NAJB14,H2AFZ,SLC39A8,NFKB1,UBE2D3,CISD2,SLC9B1,SLC9B2,BDH2,CENPE,EEF1A1,SLC17A5,FIL IP1,HIVEP2,AIG1,ADAT2,AL031320.1,PEX3,FUCA2,PLAGL1,RRM2B,UBR5,ODF1,ACO1,NDUFB6,GVQ W1,APTX,DNAJA1,SMU1,B4GALT1,SPINK4,NFX1,AQP3,NOL6,RNASEH2C,ZNF287,MPRIP,PLD6,RP114 5M22.4,FLCN,STH,KANSL1,LRRC37A,RPRML,KPNB1,TBKBP1,TBX21,VEZF1,MKS1,LPO,MPO,RNF43, HSF5,MTMR4,RCAN1*

Appendix Table 4. Results of genome-wide gene-based association test by MAGMA (FDR<0.05).

Trait	Gene	Chr	Start	End	N SNPs	Z	P	P adjusted
g0	RUNX3	1	25226002	25291612	113	4.207	1.29E-05	9.68E-03
g2	RUNX3	1	25226002	25291612	113	4.109	1.99E-05	1.33E-02
g0	NFKB1	4	103422486	103538459	276	6.163	3.57E-10	1.33E-06
g0	MANBA	4	103552660	103682151	235	4.202	1.33E-05	9.77E-03
g2	NFKB1	4	103422486	103538459	276	6.506	3.86E-11	2.70E-07
g2	MANBA	4	103552660	103682151	235	4.749	1.02E-06	1.24E-03
g0	EEF1A1	6	74225473	74233520	19	4.370	6.21E-06	5.61E-03
g0	HIVEP2	6	143072604	143266338	486	4.459	4.11E-06	4.05E-03
g1	HIVEP2	6	143072604	143266338	485	7.134	4.86E-13	9.07E-09
g2	MTO1	6	74171301	74218959	118	4.824	7.05E-07	9.39E-04
g2	EEF1A1	6	74225473	74233520	19	5.809	3.14E-09	8.37E-06
g2	HIVEP2	6	143072604	143266338	486	6.109	5.00E-10	1.65E-06
g0	ABCF2	7	150904923	150924316	70	5.770	3.97E-09	1.01E-05
g0	CHPF2	7	150929575	150935908	16	4.928	4.16E-07	6.30E-04
g0	SMARCD3	7	150935850	150974982	113	6.517	3.58E-11	2.70E-07
g2	ABCF2	7	150904923	150924316	70	4.621	1.91E-06	2.14E-03
g2	CHPF2	7	150929575	150935908	16	3.890	5.02E-05	2.84E-02
g2	SMARCD3	7	150935850	150974982	113	5.292	6.06E-08	1.13E-04
g0	KB-1980E6.3	8	103540968	103550896	60	6.235	2.25E-10	9.55E-07
g2	KB-1980E6.3	8	103540968	103550896	60	5.039	2.34E-07	3.86E-04
g0	DNAJA1	9	33025209	33039905	55	3.791	7.49E-05	3.99E-02
g0	SMU1	9	33041762	33076665	103	4.738	1.08E-06	1.28E-03
g0	B4GALT1	9	33104080	33167354	238	6.800	5.22E-12	4.87E-08
g0	SPINK4	9	33218363	33248565	103	5.628	9.11E-09	2.22E-05
g0	BAG1	9	33247818	33264761	22	4.264	1.01E-05	7.94E-03
g0	CHMP5	9	33264940	33281977	21	5.408	3.20E-08	6.39E-05
g0	NFX1	9	33290509	33371155	115	6.090	5.64E-10	1.76E-06
g1	B4GALT1	9	33104080	33167354	239	7.576	1.79E-14	5.00E-10
g2	DNAJA1	9	33025209	33039905	55	4.353	6.72E-06	5.85E-03
g2	SMU1	9	33041762	33076665	103	5.461	2.37E-08	4.91E-05
g2	B4GALT1	9	33104080	33167354	239	6.109	5.00E-10	1.65E-06
g2	SPINK4	9	33218363	33248565	103	6.352	1.06E-10	4.95E-07
g2	BAG1	9	33247818	33264761	22	4.350	6.79E-06	5.85E-03
g2	CHMP5	9	33264940	33281977	21	5.902	1.79E-09	5.01E-06
g2	NFX1	9	33290509	33371155	115	7.088	6.80E-13	9.52E-09
g1	OVOL1	11	65554493	65564690	25	4.144	1.71E-05	1.20E-02
g0	ARHGAP27	17	43471275	43511787	103	3.784	7.73E-05	4.08E-02
g0	PLEKHM1	17	43513266	43568115	97	3.762	8.44E-05	4.42E-02
g0	CRHR1	17	43699267	43913194	957	4.787	8.47E-07	1.09E-03
g0	SPPL2C	17	43922256	43924438	18	4.712	1.23E-06	1.43E-03
g0	MAPT	17	43971748	44105700	682	4.878	5.37E-07	7.70E-04
g0	STH	17	44076616	44077060	1	4.884	5.20E-07	7.66E-04
g0	KANSL1	17	44107282	44302733	743	4.979	3.20E-07	5.11E-04
g0	ARL17B	17	44352150	44439130	14	4.583	2.30E-06	2.43E-03
g0	NSF	17	44668035	44834830	67	4.967	3.40E-07	5.29E-04
g0	WNT3	17	44839872	44910520	112	5.383	3.67E-08	7.09E-05
g0	TBX21	17	45810610	45823485	38	4.367	6.31E-06	5.61E-03
g0	AZI1	17	79163393	79196799	130	5.152	1.29E-07	2.33E-04

g0	ENTHD2	17	79202077	79212891	30	6.226	2.39E-10	9.55E-07
g0	C17orf89	17	79213039	79215081	6	5.998	9.98E-10	2.94E-06
g0	SLC38A10	17	79218800	79269347	185	5.490	2.01E-08	4.33E-05
g1	TNFRSF13B	17	16832849	16875432	180	6.406	7.49E-11	4.20E-07
g2	CRHR1	17	43699267	43913194	957	4.294	8.78E-06	7.23E-03
g2	SPPL2C	17	43922256	43924438	18	4.207	1.30E-05	9.68E-03
g2	MAPT	17	43971748	44105700	682	4.394	5.57E-06	5.20E-03
g2	STH	17	44076616	44077060	1	4.424	4.85E-06	4.61E-03
g2	KANSL1	17	44107282	44302733	743	4.515	3.16E-06	3.28E-03
g2	ARL17B	17	44352150	44439130	14	4.182	1.44E-05	1.05E-02
g2	NSF	17	44668035	44834830	67	4.608	2.03E-06	2.23E-03
g2	WNT3	17	44839872	44910520	112	5.056	2.14E-07	3.62E-04
g2	ITGB3	17	45331212	45421658	287	3.940	4.08E-05	2.38E-02
g2	ITGB3	17	45331263	45421658	287	3.940	4.08E-05	2.38E-02
g2	TBX21	17	45810610	45823485	38	4.771	9.15E-07	1.14E-03
g2	BZRAP1	17	56378592	56406152	82	4.075	2.30E-05	1.47E-02
g2	SUPT4H1	17	56422539	56430454	22	4.088	2.18E-05	1.42E-02
g2	AZI1	17	79163393	79196799	130	6.462	5.18E-11	3.22E-07
g2	ENTHD2	17	79202077	79212891	30	7.584	1.68E-14	5.00E-10
g2	C17orf89	17	79213039	79215081	6	6.366	9.68E-11	4.93E-07
g2	SLC38A10	17	79218800	79269347	185	7.012	1.18E-12	1.32E-08
g0	RUNX1	21	36160098	37376965	3508	4.784	8.60E-07	1.09E-03
g2	RUNX1	21	36160098	37376965	3506	4.440	4.51E-06	4.35E-03

Trait- galactosylation trait which was tested; Gene- symbol of the gene; Chr- chromosome on which gene is found; Start- starting position for the gene sequence in GRCh37(hg19) build; End- end position for the gene sequence; N SNPs- number of SNPs in the gene; Z- Z-score of the association; P- unadjusted P value of the association; P adjusted- adjusted p value using Benjamini and Hochberg method.

Appendix Table 5. Results of implementation of SBayesR and C+T models for G0, G1 and G2.

PRS model	Parameters	G0			G1			G2		
		R2	R spearman	PRS p-value	R2	R spearman	PRS p-value	R2	R spearman	PRS p-value
SbayesR	PRS+sex+age	0.3345	0.5861	4.68E-04	0.0159	0.1424	0.4978	0.3407	0.5776	1.76E-04
	PRS+age	0.3108	0.5796	2.70E-04	0.0056	0.1240	0.4347	0.3232	0.5745	1.29E-04
	PRS+sex	0.1076	0.2977	8.55E-04	0.0146	0.1063	0.5009	0.1008	0.2934	6.55E-04
	PRS	0.0736	0.2528	4.79E-04	0.0038	0.0574	0.4367	0.0737	0.2801	4.77E-04
C+T	PRS+sex+age	0.3372	0.6198	3.31E-04	0.0363	0.2043	0.0524	0.3423	0.5843	1.45E-04
	PRS+age	0.3020	0.6086	7.86E-04	0.0245	0.1817	0.0559	0.3191	0.5784	2.16E-04
	PRS+sex	0.0935	0.2892	3.29E-03	0.0346	0.1756	0.0543	0.0946	0.2835	1.18E-03
	PRS	0.0450	0.1999	6.74E-03	0.0222	0.1200	0.0586	0.0600	0.2217	1.67E-03
Models	age+sex	0.2807	0.5379		0.0130	0.1276		0.2791	0.5199	
	age	0.2506	0.5346		0.0018	0.1037		0.2577	0.5202	
	sex	0.0428	0.1702		0.0118	0.1026		0.0324	0.1512	

PRS Model - method that was used to build model; Parameters - sets of parameters for a linear regression model with the trait as the outcome; R2 - part of trait variance that a model explains; R Spearman - the Spearman correlation coefficient between observed and predicted values of the trait; PRS p-value - the two-tailed p-value for the t-stats of PRS param.

Appendix Table 6. Meta-analysis results of testing differences in G0, G1 and G2 traits in treated (upregulation) and control cell line.

Experiment	Glycan trait	Estimate	SE	p-value	CI lower	CI upper
	GP_G0	2.484280	0.966877	0.010188	0.589236	4.379323
VPR_EEF1A1	GP_G1	-2.246458	0.487330	0.000004	-3.201608	-1.291309
	GP_G2	-0.257913	0.352768	0.464710	-0.949325	0.433499
	GP_G0	-1.408927	0.518379	0.006569	-2.424931	-0.392924
VPR_HIVEP2	GP_G1	0.117828	0.443605	0.790535	-0.751621	0.987278
	GP_G2	1.079269	0.379913	0.004500	0.334654	1.823884
	GP_G0	2.106936	1.137043	0.063883	-0.121627	4.335500
VPR_MANBA	GP_G1	-2.521960	0.755039	0.000837	-4.001809	-1.042112
	GP_G2	-0.176833	0.265544	0.505459	-0.972904	0.343625
	GP_G0	0.823330	1.339071	0.538653	-1.801201	3.447862
VPR_NFKB1	GP_G1	-0.886318	1.267960	0.484545	-3.371473	1.598837
	GP_G2	0.022321	0.755465	0.976429	-1.458363	1.503006
	GP_G0	2.384531	0.622566	0.000128	1.164323	3.604739
VPR_TNFRSF13B	GP_G1	-1.515340	0.902826	0.093261	-3.284846	0.254166
	GP_G2	-0.706994	0.360817	0.050063	-1.414181	0.000194

Trait values were transformed to obtain normal distribution using Box-Cox method and analyzed using Student's t test. Estimate- effect size estimates; SE- standard error of estimate; p-value- p value of the meta-analysis; CI-confidence interval.

Appendix Table 7. Overview of platforms used for quantification of IgG N-glycans and sample demographics.

Cohort	Platform for glycan analysis	Reference	N total available (glycans)	N female	N male	median age (min-max)	N in GWAS
TwinsUK	UPLC	Menni et al.	4624	4282	342	54	4477
CROATIA-Korcula	UPLC and LC-MS	Pučić et al.	2478	1148	1330	57	2436
CROATIA-Split	LC-MS	Not published	973	383	590	52	920
CROATIA-Vis	LC-MS	Pučić et al.	683	394	289	57	675
VIKING	UPLC	Landini et al.	1086	645	441	51	1071
ORCADES	UPLC	Krištić et al.	1786	1082	704	54	1720
LLS	LC-MS	Ruhaak et al.	1841	974	867	59	1190
KORA F4	LC-MS	Wahl et al.	1823	935	888	62	1167
EGCUT	UPLC	Trbojević-Akmačić et al.	1108	516	592	69	483
GCKD	UPLC	Not published	4933	1965	2968	63	4933
EPIC	UPLC	Not published	3600	61.20%	38.80%	50	2406

Cohort- name of the study; Platform- platform used for quantification of IgG N-glycans; Reference- published study in which the IgG N-glycome analysis for the cohort was described; N total (glycans)- number of subjects in the cohort with measured glycans; N female- number of female subjects; N male- number of male subjects; median age- median age of participants in the cohort; N in GWAS- number of pariticipants in GWAS.

Appendix Table 8. Overview of genotyping arrays and imputation.

Cohort	Genotyping platform(s)	ID call rate	SNP call rate	HWE p	MAF	Imputation tool
TwinsUK	Illumina HumanHap300; Illumina HumanHap610Q	>95%	> 97% (MAF ≥ 5%); > 99% (1% ≤ MAF < 5%)	>1e-6	≥1%	MACH (Michigan Imputation Server v1.0.2)
	Human660W-Quad_v1_A	>97% >99%	>95%	>1e-3		Eagle2/minimac 3
EPIC	HumanCoreExome-12v1-0_B					Eagle2/minimac 3
	Illumina InfiniumOmniExpressExome-8v1-3_A DNA Analysis BeadChip				zCall threshold=7	Eagle2/minimac 3
LLS	Illumina660 W; Illumina OmniExpress	>95%	>95%	>1e-4	≥1%	IMPUTE2
CROATIA-Korcula 1,2,3	Illumina HumanHap s370CNV DUO/QUAD Phase 1(1); Illumina HumanOmniExpress Exome (2 and 3)	>97%	>98%	>1e-6	≥1%	SHAPEIT2/Sanger
CROATIA-Split	Illumina HumanHap 370CNV QUAD Phase I; Illumina HumanOmniExpress Exome	≥97%	>98%	>1e-6	≥1%	SHAPEIT2/Sanger
CROATIA-Vis	Illumina HumanHap300v1 BeadChip	>97%	>98%	>1e-6	≥1%	SHAPEIT2/Sanger
VIKING	Illumina HumanOmniExpress Exome	>97%	>98%	>1e-6	≥1% omni markers; ≥0.01% exome markers	SHAPEIT2/Sanger
EGCUT	Illumina GSAv1.0, GSAv2.0, GSAv2.0_EST	>95%	>95%	>1e-4	≥1%	Beagle v.28Sep18.793
KORA F4	Affymetrix Axiom	>97%	>98%	>5e-6	≥1%	SHAPEIT/IMPUTE
ORCADES	HumanHap300v2 Phase 1	>97%	>98%	>1e-6	≥1%	SHAPEIT2/Sanger
GCKD	Illumina Omni2.5Exome BeadChip	>97%	>96%	>1e-5	>1%	Eagle/minimac3

Cohort- name of the participating study; Genotyping platform- SNP array used for genotyping in the given cohort; ID call rate-ID call rate; SNP call rate- SNP call rate; HWE p- p value of HWE test used for filtering; MAF- minor allele frequency threshold used for filtering; Imputation tool- software used for imputation of genetic data.

Appendix Table 9. Glycan names and description of the most abundant glycan structure in each peak as measured by ultra-performance liquid chromatography; glycan name- name of the glycan; description- structure of the glycan (F- core fucosylated, A1- monoantennary, A2- diantennary, B- bisecting GlcNAc, G1- monogalactosylated, G2- digalactosylated, S1- monosialylated, S2- disialylated, M5- pentamannose, [3] or [6]- position of the attached galactose).

	Glycan name	Description
	GP1	FA1 glycan
	GP2	A2 glycan
	GP3	Structure not determined
	GP4	FA2 glycan
	GP5	M5 glycan
	GP6	FA2B glycan
	GP7	A2G1 glycan
	GP8	FA2[6]G1 glycan
	GP9	FA2[3]G1 glycan
	GP10	FA2[6]BG1 glycan
	GP11	FA2[3]BG1 glycan
UPLC measured glycan traits	GP12	A2G2 glycan
	GP13	A2BG2 glycan
	GP14	FA2G2 glycan
	GP15	FA2BG2 glycan
	GP16	FA2G1S1 glycan
	GP17	A2G2S1 glycan
	GP18	FA2G2S1 glycan
	GP19	FA2BG2S1 glycan
	GP20	Structure not determined
	GP21	A2G2S2 glycan
	GP22	A2BG2S2 glycan
	GP23	FA2G2S2 glycan
	GP24	FA2BG2S2 glycan

Appendix Table 10. Glycan names and description of glycan structure in each peak as measured by liquid chromatography coupled with mass spectrometry; glycan name- name of the glycan in LC-MS data; description- structure of the glycan (F- core fucosylated, N- bisecting GlcNAc, G1- monogalactosylated, G2-digalactosylated, S1- monosialylated).

Glycan class	Glycan name	Description
	LC_IGP1	IgG1_G0
	LC_IGP2	IgG1_G0F
	LC_IGP3	IgG1_G0FN
	LC_IGP4	IgG1_G0N
	LC_IGP5	IgG1_G1
	LC_IGP6	IgG1_G1F
	LC_IGP7	IgG1_G1FN
	LC_IGP8	IgG1_G1FNS1
	LC_IGP9	IgG1_G1FS1
Total IgG1 glycans	LC_IGP10	IgG1_G1N
	LC_IGP11	IgG1_G1NS1
	LC_IGP12	IgG1_G1S1
	LC_IGP13	IgG1_G2
	LC_IGP14	IgG1_G2F
	LC_IGP15	IgG1_G2FN
	LC_IGP16	IgG1_G2FNS1
	LC_IGP17	IgG1_G2FS1
	LC_IGP18	IgG1_G2N
	LC_IGP19	IgG1_G2NS1
	LC_IGP20	IgG1_G2S1
	LC_IGP87	IgG2_G0
	LC_IGP88	IgG2_G0F
	LC_IGP89	IgG2_G0FN
	LC_IGP90	IgG2_G0N
	LC_IGP91	IgG2_G1
	LC_IGP92	IgG2_G1F
	LC_IGP93	IgG2_G1FN
	LC_IGP94	IgG2_G1FNS1
	LC_IGP95	IgG2_G1FS1
Total IgG2 glycans	LC_IGP96	IgG2_G1N
	LC_IGP97	IgG2_G1NS1
	LC_IGP98	IgG2_G1S1
	LC_IGP99	IgG2_G2
	LC_IGP100	IgG2_G2F
	LC_IGP101	IgG2_G2FN
	LC_IGP102	IgG2_G2FNS1
	LC_IGP103	IgG2_G2FS1
	LC_IGP104	IgG2_G2N
	LC_IGP105	IgG2_G2NS1
	LC_IGP106	IgG2_G2S

	LC_IGP173	IgG4_G0F
	LC_IGP174	IgG4_G0FN
	LC_IGP175	IgG4_G1F
	LC_IGP176	IgG4_G1FN
Total IgG4 glycans	LC_IGP177	IgG4_G1FNS1
	LC_IGP178	IgG4_G1FS1
	LC_IGP179	IgG4_G2F
	LC_IGP180	IgG4_G2FN
	LC_IGP181	IgG4_G2FNS1
	LC_IGP182	IgG4_G2FS1

Appendix Table 11. Formulas used for calculation of G0, G1 and G2 traits from LC-MS and UPLC measured glycan data.

IgG N-glycan trait	Description	LCMS formula	UPLC formula
G0	Percentage of agalactosylated structures in total IgG N-glycome	(LC_IGP1+LC_IGP2+LC_IGP3+LC_IGP4+LC_IGP87+LC_IGP88+LC_IGP89+LC_IGP90+LC_IGP173+LC_IGP174)/SUM(ALL)*100	SUM(GP1+GP2+GP4+GP5+GP6)/SUM(GP1:GP24)*100
G1	Percentage of monogalactosylated structures in total IgG N-glycome	(SUM(LC_IGP5:LC_IGP12)+SUM(LC_IGP91:LC_IGP98)+SUM(LC_IGP175:LC_IGP178))/SUM(ALL)*100	(GP7+GP8+GP9+GP10+GP11+GP16)/SUM(GP1:GP24)*100
G2	Percentage of digalactosylated structures in total IgG N-glycome	(SUM(LC_IGP13:LC_IGP20)+SUM(LC_IGP99:LC_IGP106)+SUM(LC_IGP179:LC_IGP182))/SUM(ALL)*100	(GP12+GP13+GP14+GP15+GP17+GP18+GP19+GP21+GP22+GP23+GP24)/SUM(GP1:GP24)*100

Appendix Table 12. Parameters for SBayesR model for PRS calculation.

--pi	0.95,0.02,0.02,0.01
--gamma	0.0,0.01,0.1,1
--hsq	0.5
--chain-length	3000
--burn-in	1000
--unscale-genotype	yes
--exclude-mhc	yes
--robust	yes
--p-value	0.9

Appendix Table 13. gRNA sequences for targeted upregulation/downregulation.

sgRNA molecule	Target site sequence (5'→3')	Experiment
KIF3C_A-g1	GTGTCCGAGGATGGTGGCTCG	
KIF3C_A-g2	GTTCCCGGCATGCTCGGGCTCG	
KIF3C_A-g3	GTTGTAGTCCTGCCTTTGTTAA	
NFKB1_A-g1	GTATGCTCTCTCGACGTCAGT	
NFKB1_A-g2	GTCGAGAGAGCATACAGACAG	
NFKB1_A-g3	GGGCGGGAGGTAGGGTCGCTG	
MANBA_A-g1	GACCCGCGACGCCCGGCTTTCA	
MANBA_A-g2	GAGCGCTCAGCTGACCTAGGG	
MANBA_A-g3	GTCAGTACATACAACTGGAGGC	
SLC38A10_A-g1	GGTTTGGTGGGACGCGAGCCA	
SLC38A10_A-g2	GTCCAAACCGAGGCCTCGCGGG	CRISPRa (dSaCas9)
SLC38A10_A-g3	GCCAGCTCGTTGAACTGGGGCC	
TNFRSF13B_A-g1	GTGTGTACCCCATGGGACGCC	
TNFRSF13B_A-g2	GTTAGTGCTTGGGCTGCACTTC	
TNFRSF13B_A-g3	GCTTGCGATGAGGTATTGACTA	
EEF1A1_A-g1	GCCTCAGTGATGACGAGAGCGG	
EEF1A1_A-g2	GTGAGGCTCCGGTGCCCGTCA	
EEF1A1_A-g3	GTCGGCAATTGAACCGGTGCC	
HIVEP2_A-g1	GTCATCTGCATAGATCCGCGGC	
HIVEP2_A-g2	GGAGCTTGTTTCACTTTATAT	
HIVEP2_A-g3	GCTGCTTTGCGTAATGTCCTGG	
KIF3C_G-g1	GCCAGCTAAGCCTCCCGACTG	
KIF3C_G-g2	GCGGCATGCTCGGGCTCGGTG	
KIF3C_G-g3	GCCTGTCATCCCGGAGCGGT	
NFKB1_G-g1	GGAAGAGGAGGTTTCGCCAC	
NFKB1_G-g2	GGCGGCCCGACCGACGGGAA	
NFKB1_G-g3	GCGCCGCCTTACTGGCCGGG	
MANBA_G-g1	GCTGTAACTGAGCTCCGCGG	
MANBA_G-g2	GTCGATCTCTCCACATCTCGG	
MANBA_G-g3	GCCTGATCCAGGTATGGCGTG	
SLC38A10_G-g1	GTCCCACCAAACCCTACCCCC	
SLC38A10_G-g2	GGTCACAGGTCCCAACGTCC	CRISPRi (dSpCas9)
SLC38A10_G-g3	GGTGTCCCTAGCTGGCTCTG	
TNFRSF13B_G-g1	GCTGGACCTTGCAACCCCCA	
TNFRSF13B_G-g2	GCCCAGGCCACTCATTACTC	
TNFRSF13B_G-g3	GCGAGGTGGCCGGAGCCGTG	
EEF1A1_G-g1	GGTGCCACCAGATTCGCACG	
EEF1A1_G-g2	GTTGCGAAAAAGAACGTTCA	
EEF1A1_G-g3	GAAGGGCCATAACCCGTAAAG	
HIVEP2_G-g1	GCTTTGCGTAATGTCCTGGG	
HIVEP2_G-g2	GGCCCGTTGTCAGGTAGCGG	
HIVEP2_G-g3	GCGACGTGCACCGGCGCGCA	
NT-sgRNA	GTAGGCGCGCCGCTCTCTAC	Non-targeting control

Appendix Table 14. qPCR primer sequences for SYBR detection of gene expression.

Primer name	Sequence (5'→3')
KIF3C_FW1	CATGGAGAGGCTCATGCGAT
KIF3C_REV1	TGGTTGTCACTCATGGTCCG
NFKB1_FW1	GAAGCACGAATGACAGAGGC
NFKB1_REV1	GCTTGGCGGATTAGCTCTTTT
MANBA_FW1	TGAGCTGCGTTTCCAGTCAG
MANBA_REV1	ACATGGCATTCACCCTTCTGC
SLC38A10_FW1	GGTCCCAACACAGGCTACTG
SLC38A10_REV1	GACACCAAACAGCATCCACC
TNFRSF13B_FW1	GCAAGGAGCAAGGCAAGTTC
TNFRSF13B_REV1	TCCTTGGTACCTTCCCGAGT
EEF1A1_FW1	TTGGACACGTAGATTCGGGC
EEF1A1_REV1	CGTTCACGCTCAGCTTTCAG
HIVEP2_FW1	GGGACACAGCTCTAGGACAAA
HIVEP2_REV1	TGCGGTGGATATTGCTTCTCT