Engineering 26 (2023) 17–31

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng

Research
Glycomedicine—Review

Twelve Years of Genome-Wide Association Studies of Human Protein N-Glycosylation

Anna Timoshchuk [a], Sodbo Sharapov [a], Yurii S. Aulchenko [a,b,*]

[a] MSU Institute for Artificial Intelligence, Lomonosov Moscow State University, Moscow 119991, Russia
[b] Institute of Cytology and Genetics SB RAS, Novosibirsk 630090, Russia

ARTICLE INFO

Article history:
Received 12 October 2022
Revised 22 March 2023
Accepted 22 March 2023
Available online 19 May 2023

Keywords:
Glycome
Glycans
N-glycosylation
Genomics
Genetics
GWAS

ABSTRACT

Most human-secreted and membrane-bound proteins have covalently attached oligosaccharide chains or glycans. Glycosylation influences the physical and chemical properties of proteins, as well as their biological functions. Unsurprisingly, alterations in protein glycosylation have been implicated in a growing number of human diseases, and glycans are increasingly being considered as potential therapeutic targets, an essential part of therapeutics, and biomarkers. Although glycosylation pathways are biochemically well-studied, little is known about the networks of genes that guide the cell- and tissue-specific regulation of these biochemical reactions in humans *in vivo*. The lack of a detailed understanding of the mechanisms regulating glycome variation and linking the glycome to human health and disease is slowing progress in clinical applications of human glycobiology. Two of the tools that can provide much sought-after knowledge of human *in vivo* glycobiology are human genetics and genomics, which offer a powerful data-driven agnostic approach for dissecting the biology of complex traits. This review summarizes the current state of human populational glycogenomics. In Section 1, we provide a brief overview of the *N*-glycan's structural organization, and in Section 2, we give a description of the major blood plasma glycoproteins. Next, in Section 3, we summarize, systemize, and generalize the results from current N-glycosylation genome-wide association studies (GWASs) that provide novel knowledge of the genetic regulation of the populational variation of glycosylation. Until now, such studies have been limited to an analysis of the human blood plasma N-glycome and the N-glycosylation of immunoglobulin G and transferrin. While these three glycomes make up a rather limited set compared with the enormous multitude of glycomes of different tissues and glycoproteins, the study of these three does allow for powerful analysis and generalization. Finally, in Section 4, we turn to genes in the established loci, paying particular attention to genes with strong support in Section 5. At the end of the review, in Sections 6 and 7, we describe special cases of interest in light of new discoveries, focusing on possible mechanisms of action and biological targets of genetic variation that have been implicated in human protein N-glycosylation.

© 2023 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

N-glycans are irregular polymers of five monosaccharides—namely, mannose, galactose, fucose, sialic acid, and *N*-acetylglucosamine (GlcNAc)—that are attached to polypeptide structures, forming glycoproteins through a glycosidic linkage to the asparagine (Asn) side chain [1]. Glycosylation influences the physical and chemical properties of proteins, as well as their biological functions. N-glycoproteins exhibit great variety in their glycan part; an N-glycoprotein may have single or multiple N-glycosylation sites, and not all are necessarily glycosylated. *N*-glycans exhibit intra- and inter-protein variation: Different *N*-glycans may attach to different molecules of the same protein, and there is a wide spectrum of *N*-glycans attached to different proteins. About half of human blood plasma proteins are equipped with oligosaccharide chains ("glycans") [2]. Protein glycosylation is controlled by genetic, epigenetic, and environmental factors [3–5]. Although the glycosylation pathway is biochemically well-studied [6], little is known about the networks of genes that regulate the cell- and tissue-specific regulation of these biochemical reactions in humans *in vivo*. There is more than just academic interest in

* Corresponding author.
E-mail address: yurii@bionet.nsc.ru (Y.S. Aulchenko).

https://doi.org/10.1016/j.eng.2023.03.013
2095-8099/© 2023 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the regulation of protein N-glycosylation, as the latter's role is increasingly recognized in human health and disease [7,8]. Glycans are considered to be potential therapeutic targets [9–11], an essential part of therapeutics [12–14], and biomarkers [15–17], making glycobiology a promising field for future clinical applications [18].

Similar to other "-omics," an organism's "glycome" consists of all the glycan structures synthesized by that organism [19]. Cells of different types synthesize a subset of the glycome, depending on their differentiation state and physiological environment. Here, we use the term "glycome" to denote a quantified repertoire of glycans that a cell or tissue produces under specified conditions of time, location, and environment. We also use the term "single-protein glycome" (e.g., the "immunoglobulin-G N-glycome") to describe a quantified spectrum of glycans attached to a specific protein.

The combination of human glycomics and genomics offers a powerful "data-driven hypotheses" [20] approach for dissecting the complex human glycobiology *in vivo* in an agnostic manner [21]. Thus far, analysis of the N-glycosylation of thousands of samples is limited to a few types of available tissues and proteins. More specifically, until recently, large cohort studies of human N-glycosylation were limited to studies of the N-glycome of immunoglobulin G (IgG) and the glycome of blood plasma (the plasma N-glycome), where the latter refers to the spectrum of glycans attached to the blood plasma glycoproteins. Tens of thousands of samples have been characterized for these glycomes [22]. Recently, large-scale studies of the glycomes of other glycoproteins, such as transferrin (TF) [23], IgA [24,25], and apolipoprotein CIII [26], have become possible, but the number of samples being investigated remains limited.

Thus far, studies on the genetic regulation of the populational variation of N-glycosylation—which require a large cohort of samples measured for both their glycomes and the genomes—have been limited to the analysis of plasma glycome, IgG, and TF N-glycosylation. These three glycomes make up a rather limited set, compared with the enormous multitude of glycomes of different tissues and glycoproteins; however, the study of these three glycomes allows for powerful analysis and generalization.

The purpose of this review is to complement and extend the summary and discussion of populational glycogenomics in the reviews in Refs. [21,27]. More specifically, we provide a detailed summary of the results that have been obtained since the previous reviews were published. We summarize this new information and consider special cases of interest in light of new discoveries, focusing on possible mechanisms of action and biological targets of genetic variation that have been implicated in human protein N-glycosylation. For a detailed discussion of already-established heritability and associated loci, interested readers are referred to previous reviews and original works.

2. The N-glycosylation of IgG, TF, and the total plasma proteome

IgG is the key player of systemic humoral immunity. It is the most prevalent isotype found in the circulation and tissues. Due to its wide spectrum of activity, IgG is involved in multiple immune reactions, including complement activation, phagocytosis, and antibody-dependent cellular cytotoxicity (ADCC) [28,29]. IgG is the most abundant glycoprotein in the human bloodstream, possibly contributing close to half of all total plasma *N*-glycans (Table 1 [2,30,31]). IgG consists of two biologically different regions: the antigen-binding fragment (Fab) and the crystallizable fragment (Fc). Both of these regions can be glycosylated. The Fc fragment of IgG harbors a conservative N-glycosylation site at the Asn297 position in the CH_2 domain. Since IgG is a homodimer, each of the two heavy chains can be glycosylated at Asn297. IgG-linked *N*-glycans are either high-mannose or biantennary glycans, usually with core fucosylation, and may carry one or two galactose residuals with a possible extension of one or two sialic acids. It has been shown that Fc-linked N-glycosylation changes during pathological and physiological events and strongly affects the antibody inflammatory properties [32–36]. In contrast, little is known about Fab-linked N-glycosylation. *N*-glycans are known to occur on 15%–25% of IgG Fab fragments; they can influence antibody affinity and avidity for antigens, as well as IgG half-life [37].

TFs are blood plasma glycoproteins that are a key part of iron metabolism and are responsible for ferric-ion delivery. Iron plays a central role in many essential biochemical processes of human physiology; human cells' requirement for iron as an oxidant has given rise to a complex system that tightly regulates iron levels, tissue distribution, and bioavailability [38]. Human TF has two N-glycosylation sites, at the N432 and N630 residues, where a biantennary disialylated digalactosylated glycan structure without fucose is the most abundant glycan attached [39,40]. Together with fibrinogen, TF is one of the most abundant non-immunoglobulin human blood glycoproteins (Table 1).

There is little doubt that there are mechanisms that uniquely regulate the glycosylation of specific proteins. For example, a coding mutation that disturbs a canonical N-glycosylation amino acid sequence motif would result in the absence of *N*-glycan attachment at that site. Amino acid sequence-changing mutations outside the canonical motif that lead to changes in the N-glycosylation spectra of mutated proteins are known [41,42]. Such mutations may exhibit their effect through changes in tertiary structure that result in modified accessibility of the glycosylation site.

In addition to the considerations described above, it should be noted that the glycosylation mechanism is common to the secreted and membrane proteins produced in a specific cell. The peculiarities of this mechanism—that is, the presence or absence (or, more generally, the activity) of specific enzymes—are shared between the cells in a tissue. Thus, although unique protein- and individual cell-specific aspects can be expected, the set of rules governing the N-glycosylation of different glycoproteins in a tissue or a cell type is expected to be the same.

Table 1 shows the average concentration and likely source of major human blood plasma glycoproteins. It is clear that the majority of glycoproteins originate from just two tissues: lymphoid and liver. Therefore, it is logical to hypothesize that the glycosylation of the plasma proteome may be driven mainly by the specifics of the regulation of glycosylation in the lymphoid tissues (mainly plasma cells) and liver (hepatocytes). Under this hypothesis, it becomes clear why studies of the total plasma, IgG, and TF glycomes complement and enhance each other, as we show below.

3. Genome-wide association studies (GWASs) of protein N-glycosylation

Thanks to advances in high-throughput N-glycome profiling methods that have enabled the characterization of glycomes of thousands of samples [43,44], along with the availability of affordable and precise DNA technologies for measuring common (minor allele frequency > 1%) genetic variation, GWASs have been applied to understand the genetic determinants of the variation of the human glycome. A GWAS is a method to identify the regions of the genome (genetic loci) where variations contribute to variations in studied traits—that is, human quantitative traits and complex diseases. In brief, this method permits the reliable discovery of the effects of genetic variation (e.g., single nucleotide polymorphisms (SNPs)) on the abundance of specific glycans by comparing the genomes and N-glycomes of thousands of individuals. Due to

Table 1
Average concentration and likely origin of major human blood plasma glycoproteins. Likely source tissue and cell type are inferred as the tissues known to secrete proteins in blood and with high expression of the protein in question. (More detailed information is provided in Table S1 in Appendix A.)

Glycoprotein	Gene(s)	Tissue(s) likely to be the (primary) source of glycoprotein in blood	Likely (primary) cell type source of protein in blood	Average plasma concentration (mg·mL^{-1}) (Ref. [2])	Range of plasma concentration (mg·mL^{-1}) (Ref. [2])
IgG	*IGHG1*	Bone marrow	Plasma cells	11.8 (IgG1: 5.03, IgG2: 3.42, IgG3: 0.58, IgG4: 0.38)	7–18 (IgG1: 5.03, IgG2: 3.42, IgG3: 0.58, IgG4: 0.38)
	IGHG2	Bone marrow	Plasma cells		
	IGHG3	Bone marrow	Plasma cells		
	IGHG4	Bone marrow	Plasma cells		
Fibrinogen	*FGB*	Liver	Hepatocytes	3	2.0–4.5
	FGA	Liver	Hepatocytes		
	FGG	Liver	Hepatocytes		
IgA	*IGHA1*	Lymphoid tissues	Plasma cells	2.62 (IgA1: 2.358, IgA2: 0.262)	2.62
	IGHA2	Lymphoid tissues	Plasma cells		
TF	*TF*	Liver	Hepatocytes	2.5	2–3
IgM	*IGHM*	Spleen	Plasmoblasts (marginal zone B-cells)	1.47	0.5–2.0
Haptoglobin	*HP*	Liver	Hepatocytes	1.32	0.8–2.5
α-2-macrogobulin	*A2M*	Liver	Hepatocytes	1.2	1–2
α-1-antitrypsin	*SERPINA1*	Liver	Hepatocytes	1.1	1.1
Hemopexin	*HPX*	Liver	Hepatocytes	0.8	0.4–1.5
α-1-acid glycoprotein	*ORM1*	Liver	Hepatocytes	0.77	0.36–1.46
	ORM2	Liver	Hepatocytes		
Apolipoprotein B-100	*APOB*	Liver	Hepatocytes	0.5	0.5
α-2-heat stable-glycoprotein	*AHSG*	Liver	Hepatocytes	0.45	0.3–0.6
Ceruloplasmin	*CP*	Liver	Hepatocytes	0.355	0.15–0.96
Vitronectin	*VTN*	Liver	Hepatocytes	0.3	0.2–0.4
α-1B-glycoprotein	*A1BG*	Liver	Hepatocytes	0.22	0.22
β-2-glycoprotein 1	*APOH*	Liver	Hepatocytes	0.2	0.2
Antithrombin-III	*SERPINC1*	Liver	Hepatocytes	0.15	0.15
Histidine-rich glycoprotein	*HRG*	Liver	Hepatocytes	0.125	0.10–0.15
Apolipoprotein D	*APOD*	Brain	Glial cells	0.1	0.1
Apolipoprotein F	*APOF*	Liver	Hepatocytes	0.0845	0.073–0.096
Kininogen-1	*KNG1*	Liver	Hepatocytes	0.0725	0.055–0.090
Zinc-α-2-glycoprotein	*AZGP1*	Liver Breast	Hepatocytes Glandular cells	0.05	0.05
IgD	*IGHD*	Respiratory mucosa	IgD-secreting B-cells	0.035	< 0.003–0.400
IgE	*IGHE*	Thymus	Plasma cells	0.0003	0.0003

Source: compiled from Refs. [2,30,31].

linkage disequilibrium (LD), a GWAS typically leads to the implication of sets of correlated SNPs that show a statistically significant association with the trait of interest. The classical GWAS design assumes a "discovery" stage, in which the associated loci are identified at the genome-wide significant level, and a "replication" stage, in which association at the discovered loci is confirmed in independent samples [45]. Replication improves the reliability of GWAS findings and helps to reduce the chance that a specific observed genotype-phenotype association is a chance finding or an artifact of the design or analysis [46]. Replication is, therefore, an essential step to be taken before the complex and expensive targeted experimental follow-up studies are initiated for a specific locus. A more detailed account of the application of GWAS methodology to glycomics can be found in our recent review [21].

The identification of loci associated with the glycome provides starting points for a better understanding of the glycobiology and its connection with human health and disease. However, next steps are not trivial, as neither functional polymorphism nor the affected gene can be immediately determined through a GWAS. A locus that is identified in a GWAS could encompass dozens of genes; moreover, an unknown functional element in the locus may regulate a gene that is located far beyond the associated region [47]. We touch upon the challenges of linking an associated region to a biological function in the sections that follow. Below, we summarize the gene-mapping results from studies of the N-glycosylation of IgG, TF, and the total plasma proteome (plasma glycome).

Five GWASs of IgG N-glycosylation measured in up to 8000 people have been published so far [48–52]. Of the 33 genome-wide significant loci discovered in these GWASs, 29 were replicated in independent samples. The first GWAS of the IgG glycome measured with liquid chromatography identified associated variants in or near four glycosyltransferase genes and in five additional loci that did not contain genes with a known role in protein glycosylation [48]. Shen et al. [49] developed and applied multivariate GWAS discovery and replication methods to extend the list by five additional loci, only one of which contained a glycosyltransferase gene. A number of loci were replicated, and one additional locus that did not contain a glycosyltransferase gene was discovered by Wahl et al. [50] in a study that used liquid chromatography coupled with electrospray mass spectrometry (LC-ESI-MS) to measure IgG N-glycosylation.

While the abovementioned studies were limited to a discovery-stage sample size of about 2000 individuals, a recent work by Klarić et al. [51] investigated associations between the IgG glycome and genome in a discovery sample of more than 8000 individuals. The results were replicated in 2400 samples with the IgG glycome measured via ultra-high-performance liquid chromatography (UHPLC) and validated in about 1800 samples with the IgG glycome measured using LC-ESI-MS. Overall, this work brought the number of loci that were significantly associated genome-wide with the IgG glycome to 27, 22 of which were replicated in independent samples.

The summary statistics generated by Klarić et al. were reanalyzed by Shadrina et al. [52] using multivariate summary-based methods for GWAS discovery and replication [53]. Thanks to the multivariate methodology and extended replication sample (N = 3147), this analysis resulted in six new loci, five of which were replicated, and the replication of two previously discovered loci.

Five of 29 replicated loci associated with IgG N-glycosylation contain glycosyltransferases; these are responsible for core (*FUT8*) and antennary (*FUT6*) fucosylation, the addition of bisecting GlcNAc (*MGAT3*), and terminal modifications such as galactosylation (*B4GALT1*) and sialylation (*ST6GAL1*). Although the other 24 loci contain genes that fall into multiple classes, a group of genes involved in the biology of antibody-producing cells, such as *RUNX3*, *RUNX1*, *IKZF3*, *IKZF1*, and *TNFRSF13B*, should be noted. Other notable associations include one with the human leukocyte antigen (*HLA*) region, as well as an association with the region that contains genes encoding immunoglobulin heavy chain (*IGH*)—although this region also contains a secondary candidate (*TMEM121*, implicated through colocalization of its expression quantitative trait locus (eQTL)). The candidate genes prioritized in the implicated loci are discussed in more detail in the next section.

Compared with studies of IgG N-glycosylation, genetic studies of the N-glycosylation of the total human plasma proteome have somewhat lagged behind in terms of the number of studies, sample sizes, and—consequently—the number of loci identified. Thus far, three GWASs of the human plasma N-glycome, measured in up to 3500 people using liquid-chromatography-based techniques, have identified 16 loci [54–56]. An overwhelming majority (15) of these loci were replicated in independent samples, either in the original works or in a study by Sharapov et al. [57], which used the largest replication sample size to date, of 4800 people. The study had very high (> 95%) expected power to replicate all but one locus (chromosome 22, near *KREMEN1*). For the latter locus, the minor allele frequency of the associated insertion/deletion polymorphism rs140053014 was low (about 2%) in the studied populations, which complicated its imputations, hence reducing the number of available samples and power to about 10% and making the replication results inconclusive.

Eleven of the 15 replicated loci overlap with loci discovered in studies of the genetic control of IgG N-glycosylation. These loci contain genes encoding glycosyltransferases: *FUT8*, *FUT6*, *MGAT3*, *ST6GAL1*, and *B4GALT1*; and other genes: *HLA*, *IGH*, *IKZF1*, and *RUNX3*. This may not be particularly surprising, given that immunoglobulins make up a major fraction of the plasma glycoproteome (Table 1). In this context, the four plasma-glycome specific loci, which were not discovered in the GWAS of IgG N-glycosylation despite their larger sample size, may be of particular interest. Three of these contain genes encoding a glycosyltransferase: *MGAT5*, whose product adds GlcNAc, facilitating further branching of biantennary glycans; *ST3GAL4*, whose product facilitates terminal sialylation; and *B3GAT1*, encoding a glucuronyltransferase. In addition, a region containing the gene *HNF1A* is specifically associated with the plasma but not the IgG N-glycome. The gene was demonstrated to be a master regulator of fucosylation in liver [54]. The GWAS of the plasma N-glycome is reviewed in more detail in Ref. [21].

A recent GWAS of TF N-glycosylation [23], despite a relatively small sample size (discovery sample size of N = 948; replication N = 952), has brought important conceptual advances. Among the eight loci implicated at the genome-wide significance level and replicated, one (containing the *TF* itself) has not been reported in previous studies of N-glycosylation. Five loci associated with TF N-glycosylation contain genes encoding glycosyltransferases, which are already implicated in the control of total plasma (*B3GAT1*, *MGAT5*, and *ST3GAL4*) or IgG and total plasma (*FUT8* and *FUT6*) N-glycosylation. Additional shared loci include *HNF1A*, which is shared between TF and total plasma N-glycosylation; and *NXPE2/4*, which is shared between TF and IgG N-glycosylation. After a meta-analysis of the discovery and replication samples, two more loci achieved genome-wide significance. One of these significant—but not (yet) replicated—loci contained a gene encoding a member of the forkhead family of transcription factors, *FOXI1*, and another region was located near the gene *MSR1*.

Landini et al. [23] also specifically studied the colocalization of TF and IgG N-glycosylation association patterns in the *FUT8* and the *FUT6* regions, which are associated with both TF and IgG N-glycosylation. They concluded that different causative variants are most likely responsible for the association of these loci with TF and IgG N-glycosylation. Six loci reported in Refs. [23,51] that are associated with changes in human protein glycosylation contain more than one independently associated variant (marked with an asterisk in Fig. 1).

To summarize, nine published GWASs established 34 replicated loci associated with variation in the N-glycosylation of the TF, IgG, or total plasma proteome.

Methodologically, from the results discussed above, we can conclude that—given a sufficiently large replication sample—the glycomic loci discovered in the GWASs have a high replication rate. For example, for the plasma N-glycome, 15 of the 16 loci reported at genome-wide significance have been replicated and one replication was inconclusive [57], while 29 of the 33 IgG N-glycosylation loci have been replicated so far [52]. These results highlight the general robustness of the GWAS findings, as well as the high standards adopted by the community studying the genetic regulation of protein N-glycosylation.

At least in single-protein IgG N-glycosylation studies, we can also conclude that the multivariate methods of genetic analysis demonstrate superior power. Studies of IgG N-glycosylation have empirically demonstrated that the application of multivariate analysis is equivalent to increasing the sample size by 20%–50% [49,52]. This significant increase in power is likely due to large correlations being observed between different glycans—correlations that are not only large but also biologically meaningful [58], which may provide an ideal case for multivariate analysis [59].

An account of the established loci is summarized in Fig. 1. The domination of loci established in studies of IgG N-glycosylation is obvious in the figure, as is the absence of plasma-glycome-specific loci. While the former observation is likely to be a statistical artefact, the latter probably reflects the true underlying biology.

Indeed, the sample sizes of IgG N-glycosylation studies have reached 8000, while the maximal sizes of plasma glycome and TF N-glycosylation studies are 3500 and 1900, respectively. Empirically, we know that, for a complex trait, the number of loci discovered is roughly proportional to the sample size [60,61]. Therefore, one would expect that the number of loci established for IgG N-glycosylation (29 loci) is roughly twice that established for the plasma glycome (15 loci) and four times that established for TF N-glycosylation (eight loci)—which is indeed the case. Moreover, we may speculate that the yield of loci per fixed sample size is roughly the same for all three studied glycomes, which may reflect similarities between their genetic effect size distributions and the relative role played by genetic and environmental variation.

The observation that, thus far, any (replicated) locus known for total plasma N-glycosylation can be found to act on IgG and/or TF N-glycosylation is likely to reflect the underlying biology. This observation corroborates the hypothesis that the total plasma glycome could perhaps be conceived as a superposition of the glycomes of proteins secreted to the blood by two primary contributors—namely, lymphoid tissues (mainly plasma cells) and liver (hepatocytes)—and mainly driven by the specifics of the regulation of N-glycosylation in these tissues. However, the

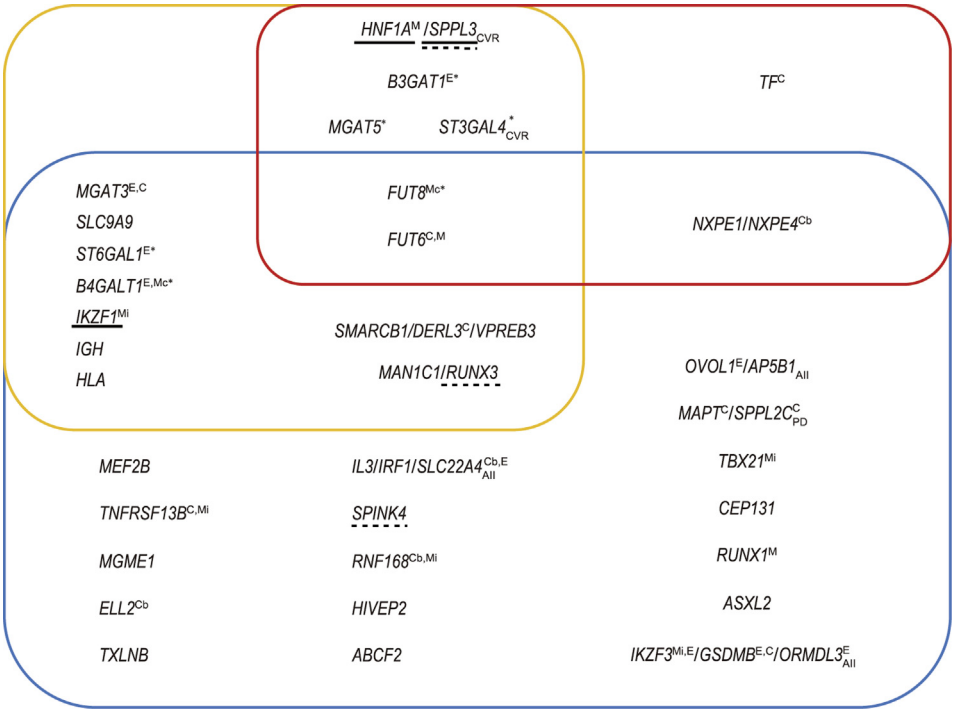

Fig. 1. Summary of 34 loci robustly associated with N-glycosylation. Loci are named after the proposed candidate gene(s). Upper left (yellow rectangle): loci associated with the plasma glycome; upper right (red rectangle): loci associated with TF N-glycosylation; bottom (blue rectangle): loci associated with IgG N-glycosylation. Gene names in blue are glycosyltransferase genes, in green are transcription regulator genes, and in black are other relevant genes; yellow indicates that the mechanism has not been hypothesized, and the gene indicated is the nearest gene. Underlined genes have experimental support for causation (solid line: classical assays; dashed line: transient system based on HEK-293F cells with stably integrated clustered regularly interspaced short palindromic repeat (CRISPR)/dead CRISPR-associated endonuclease 9 (dCas9)). A lower index denotes colocalization of the locus with: AII: an autoimmune or inflammatory disease; CVR: a cardiovascular disease or its risk factors; PD: Parkinson's disease. A superscript indicates evidence linking variation in the locus specifically to the function of this gene: C: coding variant in the gene; Cb: coding variant predicted to be benign; E: colocalization with eQTL for this gene; M: the relevant monogenic phenotype is known; Mc: monogenic phenotype of congenital disorder of glycosylation; Mi: monogenic phenotype of immunodeficiency. Loci marked with * contain more than one independently associated genetic variant.

glycosylation machinery in other tissues may also have an effect on the glycosylation of plasma proteins [62].

4. Candidate genes in the established loci

Although replication establishes a locus as a reliable starting point for further investigation, neither a GWAS nor replication answers the question of what potential mechanism of action by which variation in the locus leads to variation of the trait. The presence of LD limits the precision of mapping of a functional polymorphism with classical statistical genetic methods. A GWAS makes it possible to establish a list of polymorphisms that contains the causative ones with high probability; however, for a locus, such a list may include tens or even hundreds of candidate polymorphisms. Only extremely large studies can map associations to just a few SNPs per locus [63], if the effect of the locus is large enough, and the LD in the region is not excessive. However, even when it is possible to limit the list to a single causative variant, this still only rarely allows a direct answer to the question of what the biological mechanism of action of this polymorphism is. It might be possible to unequivocally determine the gene target that is affected, but only when the variant is reliably predicted to change the sequence of the protein encoded by the gene. At the same time, the majority of the genetic variation involved in complex traits is regulatory [64]. Although it is often the case that a gene nearest to the variant with the strongest association is indeed the target of that variation [65], it is generally not straightforward to link a regulatory variant to the gene or set of genes being regulated. For example, a functional element perturbed by the associated SNP may affect a target located millions of base pairs apart and may exhibit its action only in a specific tissue [47]. Even more eccentrically, a functional element could target a gene located on a different chromosome through trans-regulatory mechanisms [66,67].

The identification of causative polymorphisms and genes that explain association is critical in translating the statistical findings of a GWAS into actionable biological knowledge. Targeted functional experimentation is the gold standard to establish the mechanism of action; however, given the multitude of hypotheses that could be suggested for a locus, *in silico* functional investigation plays a large role in prioritizing possible mechanisms. *In silico* methods utilize a large body of accumulated knowledge on the potential consequences of changes in DNA [68,69] and the functional role of different genomic elements [70,71], as well as data accumulated in omics studies, including those investigating the genetic control of omics traits [72–76]. The latter data are especially useful when hypothesizing the potential molecular genetic mechanism underlying a specific association. Summary-statistics-based methods of colocalization analysis [74,77,78] can be used to address the question of whether the association of two traits with the same genetic region may be due to pleiotropic effects of the same genetic variant(s) or is instead a consequence of the association of two traits with two distinct—albeit close—polymorphisms. Summary-statistics-based *in silico* investigation is facilitated by the availability of (online) databases and tools integrating relevant information [79–83].

In this section, we classify 34 loci that are replicably associated with protein N-glycosylation, according to the function of the proposed candidate gene. For six loci, a candidate gene could not be suggested; in these cases, the region is named after the gene(s)

nearest to the most-associated polymorphism (gene names are in yellow in Fig. 1). For five loci, multiple candidate genes with different functions are known, and these are classified in several categories. It should be noted, that in principle, for complex traits, an association to a locus does not have to be due to single causative gene, as multiple candidates in a locus are not necessarily mutually exclusive.

Eight loci contain glycosyltransferase (gene names are in blue in Fig. 1) as the primary candidate gene for the respective loci. It was found that, in eight glycosyltransferases, genetic variation in regions including *MGAT3*, *ST6GAL1*, and *B4GALT1* is associated with variation in IgG and total plasma N-glycosylation; genetic variation near *B3GAT1*, *MGAT5*, and *ST3GAL4* relates to TF and total plasma N-glycosylation; and genetic variation near *FUT8* and *FUT6* is associated with all three studied glycomes. Most of the candidate glycosyltransferases are transcribed in both liver and plasma cells (Fig. 2(a) [84]), although there are several notable exceptions.

MGAT3 is preferentially expressed in plasma cells and *MGAT5* is expressed in hepatocytes (Fig. 2(a)). This is consistent with the fact that IgG-attached glycans are biantennary, with optionally present bisecting GlcNAc (the reaction facilitated by an enzyme encoded by *MGAT3*), while the glycans attached to the glycoproteins produced by liver are often multiantennary (*MGAT5* plays a role in glycan branching). Higher expression of α-2,3 sialyltransferase-encoding *ST3GAL4* in liver, with minimal expression in plasma cells, is consistent with the fact that the locus is implicated in plasma and TF, but not in IgG N-glycosylation GWAS, and the almost exclusive α-2,6 linked sialylation of IgG is a known fact [85]. A stronger expression of galactosyltransferase gene *B4GALT1* in liver compared with plasma cells may seem inconsistent with observations that galactose is present in IgG as well as liver glycoprotein *N*-glycans, and that the locus is implicated only in IgG N-glycosylation. However, it should be noted that the expression of *B4GALT1* in plasma cells, although lower than in hepatocytes, is still rather high—that is, it is higher than the expression levels of *MGAT3* in plasma cells. Secondly, it should be considered that the tissue specificity of gene regulation may lead to the presence of a glycomic quantitative trait locus (QTL) in one tissue but its absence in another tissue. The latter may also explain why the region near *ST6GAL1* seems to be associated only with variation of the IgG N-glycome, despite being expressed to an equally high degree in hepatocytes and plasma cells, and encoding a sialyltransferase facilitating α-2,6 bonds in glycans attached to liver-secreted proteins and immunoglobulins.

In addition to glycosyltransferases, another locus (including the *RUNX3* candidate gene) contains the gene *MAN1C1*, which encodes a glycosidase. This gene is expressed with low tissue specificity and encodes a protein involved in the early Golgi stages of the transition from high-mannose to hybrid glycan structures. Given the functional network results of Ref. [51] and the spectrum of associated glycoforms, it is more likely that the causative gene in the locus is *RUNX3*.

A large group of prioritized candidate genes code for (tissue-specific) transcription regulators (gene names are in green in Fig. 1). For plasma and TF N-glycosylation, *HNF1A* (located close to another candidate, *SPPL3*) is an example of such a gene. For IgG N-glycosylation, eight loci contain candidate genes encoding the transcription factor involved in B-cell differentiation and function. These include *RUNX1* [86], *RUNX3*, *IKZF1* [87], and *IKZF3* [88], which play a central role in B-cell maturation and differentiation; the gene *MEF2B*, which is mutated in about 15% of follicular lymphoma and which diffuses large B-cell lymphoma and plays a role in germinal center development [89]; the gene *ASXL2*, with a product whose target genes strongly overlap with those of RUNX1 and is required for normal hematopoiesis [90]; and the transcription elongation factor ELL2, which directs immunoglobulin secretion in plasma cells [91]. The gene *TNFRSF13B*, located in a recently established locus [52], encodes a lymphocyte-specific member of the tumor necrosis factor receptor superfamily known as the transmembrane activator and calcium modulator and cyclophilin ligand interactor (TACI) protein. Signaling through this protein activates several transcription factors, including nuclear factor of activated T cell (NFAT), activator protein (AP)-1, and nuclear factor-κB (NF-κB) [92]. TACI plays a crucial role in humoral immunity, acting in synergy with other signaling networks to promote B-cell differentiation into plasma cells and immunoglobulin production [93,94].

Many of the suggested candidates that encode transcription factors are differentially expressed in plasma cells and hepatocytes (Fig. 2(b)). The group with stronger expression in plasma cells includes *IKZF1*, *IKZF3*, *RUNX3*, *TNFRSF13B*, and (perhaps) *IRF1*. The gene *HNF1A* has minimal expression in plasma cells but is strongly expressed in hepatocytes. This pattern of expression is consistent with—but is perhaps also partly driven by—the GWAS findings: Indeed, the first group was discovered in a GWAS of IgG N-glycosylation, while the locus containing *HNF1A* was implicated only in plasma and TF N-glycosylation.

Three loci contain genes with products implicated in B-cell immunity. These include the IgG heavy chain locus (14q32.33) (unsurprisingly, the *IGHG1* is expressed very strongly in plasma cells but not in hepatocytes; Fig. 2(c)); the gene *VPREB3* (near *DERL3* and *SMARCB1*) coding for a protein, which likely participates in the transport and assembly of the pre-B cell receptor [95]; and the recently suggested [52] gene *RNF168*, which encodes an E3 ubiquitin ligase protein involved in DNA double-strand break repair and immunoglobulin class switch recombination [96].

Five loci contain genes involved in (not B-cell) immune function: the *HLA* region, which contains multiple immune-related genes; a locus that contains a gene coding for T-box transcription factor (TBX21), whose mouse homolog regulates the expression of interferon γ (IFNγ) and controls T helper 1 (Th1) lineage commitment [97]; the gene *HIVEP2* [98,99]; and *IRF1*, which codes for interferon regulatory factor 1. The region including *IRF1* also encompasses the gene *IL3* that codes for interleukin 3, a cytokine with an important function in the differentiation and proliferation of hematopoietic and lymphoid cell lineages; and *SLC22A4*, which encodes OCTN1, the primary transporter of ergothioneine, an amino acid with anti-oxidant and anti-inflammatory properties [100]. In the same locus with the B-lineage specific transcription factor IKZF3 is *GSDMB*, which encodes a member of a protein family with a role in pyroptosis [101,102], and the gene *ORMDL3*, whose product has been shown to modulate T-lymphocyte activation via controlling the influx of calcium ions in the endoplasmic reticulum upon antigen binding [103].

Four loci contain genes involved in the functioning of the endoplasmic reticulum, Golgi apparatus, or endosomes. These include a gene that potentially affects protein glycosylation via regulation of endosomal pH (*SLC9A9*); the gene *DERL3* (near *VPREB3* and *SMARCB1*), which is involved in the degradation of misfolded glycoproteins in the endoplasmic reticulum and is preferentially expressed in plasma cells (Fig. 2(c)); the gene *SPPL3* (near *HNF1A*), which encodes an intramembrane protease that cleaves glycosidases and glycosyltransferases [104]; and a recently [52] established locus containing *AP5B1*, a gene whose expression is enriched in bone marrow and that encodes a subunit of an adaptor protein complex, AP-5 [105], which is probably involved in facilitating the late endosome-to-Golgi retrieval [106].

In the same functional group, Frkatovic er al. [27] suggested the candidate gene *SPPL2C* (near *MAPT*). This gene encodes an intramembrane-cleaving aspartic protease that impairs vesicular transport and possibly causes retention of cargo proteins in the endoplasmic reticulum [107], thereby potentially affecting cellular processes such as protein glycosylation through mislocalization of

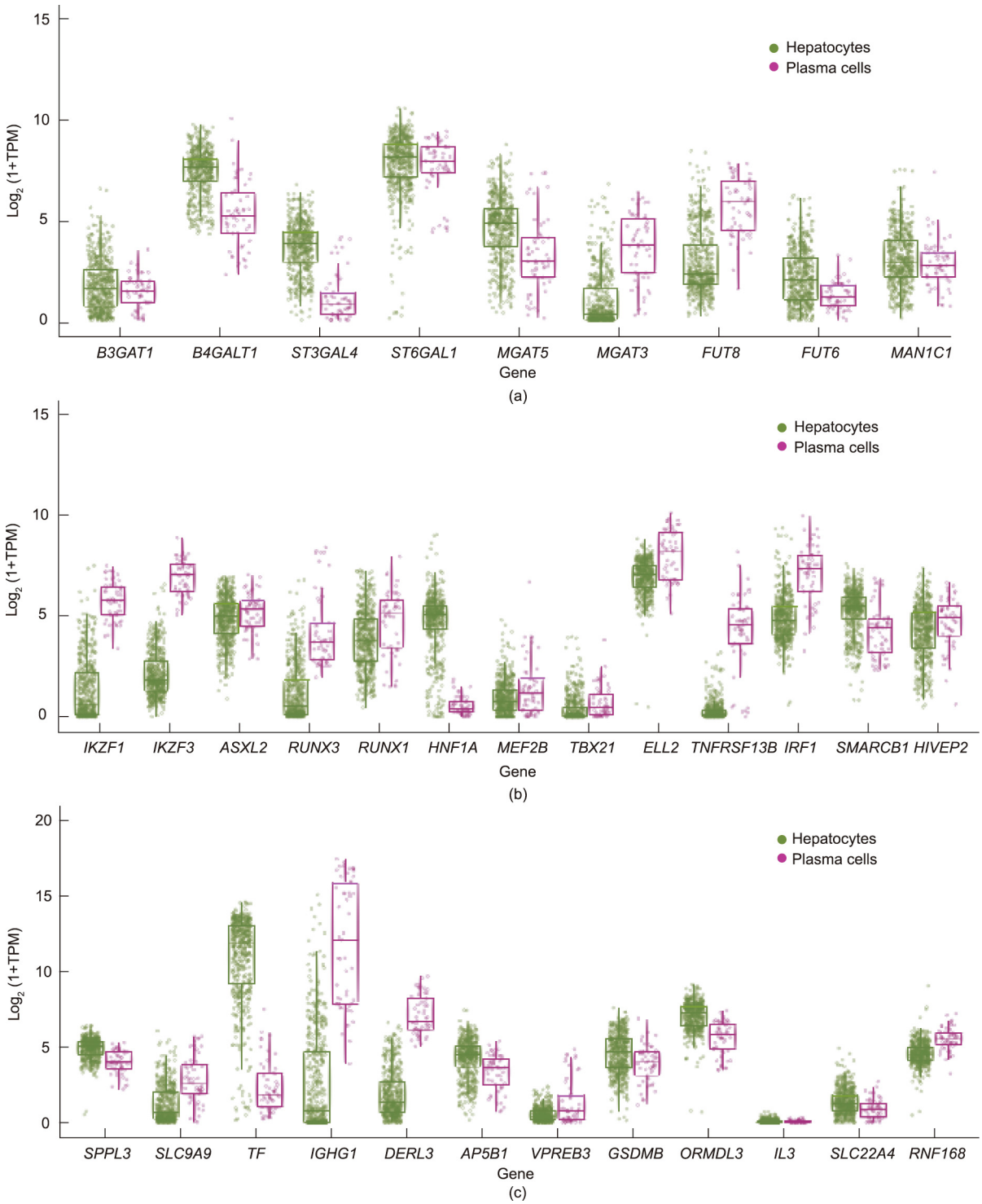

Fig. 2. Expression of candidate genes in hepatocytes (green) and bone marrow plasma cells (pink). Candidate genes are grouped by function: (a) glycosyltransferases and a glycosidase; (b) transcription regulators; (c) other. TPM: transcripts per million. Reproduced using data reported in Ref. [84], with methods and code identical to those used to generate Fig. 5 in Ref. [23].

glycan-modifying enzymes. Another member of the same gene family, *SPPL3*, which is expressed with low tissue specificity, is known to regulate cellular N-glycosylation [104,108]. However, the expression of *SPPL2C* is enriched in testes, with very low expression in other tissues. Another possible candidate for this group is the serine protease inhibitor encoded by the gene *SPINK4*. However, its expression is enriched in intestine and limited in lymphoid tissues; its possible mechanism of action remains unclear, and experimental validation is not fully conclusive [109]. Given the somewhat contradictory information regarding the possible role of genes *SPPL2C* and *SPINK4*, they are indicated in yellow in Fig. 1.

Two genes of notice do not belong to any class outlined above. One is *SMARCB1* (near to *VPREB3* and *DERL3*), which is also known as *BAF47*. This gene encodes a core subunit of the adenosine triphosphate (ATP)-dependent chromatin remodeling complex,

which plays a key role in the regulation of gene transcription [110]. Finally, for TF N-glycosylation, the gene *TF* itself is the candidate gene in the locus established by Ref. [23]. In line with known biology, the gene *TF* is strongly expressed in hepatocytes but not in plasma cells (Fig. 2(c)).

To conclude, even from the somewhat superficial account of candidate genes, it is clear that populational variation of protein N-glycosylation is regulated by genetic variation affecting: ① glycosyltransferases involved in the fucosylation, branching, and terminal modifications of *N*-glycans; ② tissue-specific regulators involved in the differentiation, maturation, and regulation of cells (hepatocytes and plasma cells) producing relevant glycoproteins; ③ perhaps more speculatively, the cellular machinery involved in the production and transportation of the secreted glycoproteins; and ④ the production and structure of the studied glycoproteins (TF and IgG) themselves.

5. Candidate genes with strong support

Confidence that the genes discussed in the previous section are indeed the ones that determine the association between N-glycosylation and genetic variation in the established loci varies significantly. For example, for loci containing glycosyltransferases (except, perhaps, *B3GAT1*), the evidence is quite complete and compelling. It is known that these glycosyltransferases are involved in protein N-glycosylation, and that the changes in N-glycosylation associated with the locus are consistent with the known enzymatic activity of the product encoded by the candidate. For a few genes, their effects on N-glycosylation have been verified in cell line and/or model animal experiments. For some genes, mutations leading to a monogenic phenotype that includes (or could be linked to) changes in N-glycosylation are known. Some genes, however, are proposed based on known involvement in the same class of processes, such as immune system functioning; for such candidates, the level of confidence is low. Moreover, even when a strong candidate is present in an established locus, unless the mechanism through which an associated variant links to that gene's function can be proposed and validated, our understanding of the mechanism of action of the locus is incomplete.

In this section, we selectively summarize evidence for candidate genes with strong support from monogenic phenotypes and/or experimental studies. We discuss potential mechanisms of action and highlight possible inconsistencies. Key facts and references are presented in Table 2 [23,51,52,54,56,108,111].

As mentioned earlier, for all but one (*B3GAT1*) glycosyltransferase-containing locus, the changes in N-glycosylation associated with the locus are consistent with the known enzymatic activity of the product encoded by the candidate. For half of glycosyltransferases (*MGAT3*, *ST6GAL1*, *B4GALT1*, and *B3GAT1*), the likely mechanism of genetic variation transformation into function is transcription regulation, as suggested by a colocalization between a glycomic GWAS and respective eQTL [51,52,54,107] (Fig. 1 and Table 2).

For one glycosyltransferase encoded by *FUT6*, a coding variant may explain the observed association with the plasma and TF N-glycome [23,56]. Indeed, in people of European descent, the C allele (with a frequency of 4%) of the polymorphism with the strongest association, rs7255720 [56], is in perfect (r^2 = 1) LD with rs17855739-T, which codes for inactivating E247K substitution in the *FUT6* catalytic domain, leading to the phenotype of (plasma) fucosyltransferase 6 (FUT6) deficiency when homozygous [112]. Within this phenotype, plasma glycoproteins produced in the liver of affected individuals lack α3-fucosylation [113]. Otherwise, the deleterious alleles of *FUT6* may reach high frequency in some populations without apparent clinical consequences.

Recently, Puan et al. [114] demonstrated that FUT6 deficiency compromises basophil function through the selective abrogation of their sialyl-Lewis^x expression; FUT6-deficient basophils rolling on E-selectin are severely impaired. The researchers also demonstrated that *FUT6* null allele carriers exhibit elevated blood basophil counts and reduced itch sensitivity against insect bites, concluding that FUT6 deficiency dampens the basophil-mediated allergic response in the periphery. This conclusion was also supported by an observation of lower IgE titers and reduced eosinophil counts.

The loci containing the FUT genes *FUT6* and *FUT8* have been implicated in both IgG and TF N-glycosylation, suggesting the important role of these genes in the N-glycosylation of proteins produced in different tissues [23,51]. However, a biologically important conclusion from the work of Landini et al. [23] is that the details of the mechanisms of action of a "common" locus may not coincide for two glycoproteins, as reflected by different causative variants associated with the N-glycosylation of each of them. This finding suggests that, although some enzymes are involved in N-glycosylation in several tissues, their genetic regulation in these tissues may differ.

A possible mechanism has been suggested for the locus containing *FUT8*. It is known that the expression of *FUT8* is regulated by the Ikaros family zinc finger 1 (IKZF1) [51] and hepatocyte nuclear factor 1-α (HNF1A) [54] transcription factors in plasma cells and hepatocytes, respectively. Many variants in the *FUT8* region are independently associated, and it can be speculated that some groups may affect the binding of HNF1A and consequently change the fucosylation of liver-secreted glycoproteins, while others may affect the binding of IKZF1 and consequently change the fucosylation of IgG. Thus, perhaps the most parsimonious biological explanation in this case is that different causative variants may affect the binding of different transcription factors in different tissues and therefore regulate the glycosylation of the tissue-specific proteins—that is, the TF produced in liver and the IgG produced by plasma cells in a tissue-specific manner. Although this explanation aligns with general knowledge about transcription regulation (i.e., Ref. [74]), verification of this hypothesis would require further *in silico* studies and experimental functional validation.

For the *FUT6* locus, the mechanism should be at least partially different than the described above mechanism for the *FUT8* locus. In plasma glycome and TF N-glycosylation GWASs [23,56,57], the variants with strongest association were in strong LD with inactivating E247K polymorphism that leads to dysfunctional FUT6 (described in more detail above). In IgG N-glycosylation GWASs, however, the SNPs with the strongest association (rs35222081, rs874232, and rs12986368) are distinctly different [23,51,52] and are in weak (r^2 < 0.1) LD with coding polymorphism. The fact that deactivating FUT6 mutation does not show an association with IgG N-glycosylation may suggest that a different gene in the region is responsible for association in the case of IgG N-glycosylation.

The evidence for the gene *B3GAT1*, which is associated with variation in plasma glycome and TF N-glycosylation [23,56], is somewhat contradictory. On the one hand, the evidence supporting this gene is strong: The candidate encodes a glycosyltransferase, and its role has been reinforced by a colocalization analysis that suggests that the same functional variant may explain the association with plasma glycan levels as well as the expression of *B3GAT1* in liver (Fig. S1 in Appendix A). However, *B3GAT1* is a member of the glucuronyltransferase gene family, encoding glucuronyl transferase, which acts in the biosynthesis of the carbohydrate epitope human natural killer-1 (HNK-1), adding a glucuronic acid (GlcA) to the terminal *N*-acetyllactosamine (LacNAc) disaccharide to form the HNK-1 epitope precursor [115,116]. The HNK-1 epitope is expressed on a subset of human lymphocytes, including natural killer cells. This epitope has not been reported to exist on

Table 2
Summary of evidence for candidate gene mapping to the regions implicated by GWASs. Only genes with a role in N-glycosylation strongly supported by a monogenic phenotype and/or experiments are listed. The probability of being loss-of-function (LoF) intolerant (pLI) and observed (o)/expected (e) are the Genome Aggregation Database (gnomAD) [111] constraint metric, with pLI values close to 1 and o/e values close to 0 denoting stronger constraint. Blank cells indicate an absence of evidence at the time of manuscript preparation.

Candidate gene	pLI	o/e LoF (90% CI)	Link between an associated variant and the gene	OMIM	Monogenic phenotype	Could the phenotype be directly or indirectly related to N-glycosylation?	Direct experimental support
Glycosyltransferases gene							
MGAT5	1.00	0.07 (0.03–0.18)	—	—	—	—	—
ST3GAL4	0.04	0.31 (0.17–0.61)	Yes, eQTL colocalization [56]	—	—	—	—
FUT6	—	—	Yes, coding variant (for plasma and TF N-glycosylation) [23,56]	613852	Fucosyltransferase 6 (FUT6) deficiency; autosomal recessive; the Lewis negative phenotype is found on red cells of 95% of these FUT-deficient individuals	Yes, directly	—
FUT8	0.01	0.3 (0.18–0.51)	—	618005	Congenital disorder of glycosylation with defective fucosylation 1 (CDGF1), FUT8-congenital disorder of glycosylation (FUT8-CDG); autosomal recessive	Yes, directly	—
MGAT3	0.15	0.28 (0.14–0.63)	Yes, eQTL colocalization found in Ref. [51]	—	—	—	—
ST6GAL1	0.23	0.25 (0.12– 0.57)	Yes, eQTL colocalization found in Ref. [51]	—	—	—	—
B4GALT1	0.04	0.33 (0.17–0.70)	Yes, eQTL colocalization found in Ref. [51]	607091	Congenital disorder of glycosylation, type IId (CDG2D), β-1,4-galactosyltransferase 1-CDG (B4GALT1-CDG); autosomal recessive	Yes, directly	—
Non-glycosyltransferases gene							
HNF1A	0.96	0.14 (0.07–0.33)	—	600496	Maturity-onset diabetes of the young, type 3 (MODY-3); autosomal recessive	Yes, directly	[54]
IKZF1	1.00	0 (0–0.16)	—	616873	Common variable immunodeficiency, type 13 (CVID13); autosomal dominant	Yes, through hypogammaglobulinemia and decreased number of B cells	[51]
IKZF3	0.98	0.09 (0.04–0.29)	Yes, eQTL colocalization found in Ref. [51]	619437	Immunodeficiency 84 (IMD84); autosomal dominant (provisional)	Yes, through decreased B cells	—
TNFRSF13B	0	1.73 (1.14–1.96)	Yes, eQTL colocalization found in Ref. [51]	240500	Common variable immunodeficiency, type 2 (CVID2)	Yes, through hypogammaglobulinemia with low serum IgG, IgM, and IgA	—
TNFRSF13B	0	1.73 (1.14–1.96)	Yes, eQTL colocalization found in Ref. [51]	609529	IgA deficiency, type 2 (IGAD2)	Perhaps, through IgA deficiency	—
RUNX1	0.65	0.19 (0.09–0.44)	—	601399	Platelet disorder, familial, with associated myeloid malignancy (FPDMM); autosomal dominant	Perhaps, through thrombocytopenia and platelet disorder	—
RUNX3	0.84	0.13 (0.05–0.42)	—	—	—	—	[108]
RNF168	0.84	0.84 (0.60–1.20)	Maybe (coding variant), see Ref. [52]	611943	Radiosensitivity, immunodeficiency, dysmorphic features, learning difficulties (RIDDLE) syndrome; autosomal recessive	Yes, through hypogammaglobulinemia	—

CI: confidence interval; OMIM: Online Mendelian Inheritance in Man.

plasma proteins. Although Huffman et al. [55] reported the presence of GlcA on some glycans contained in the DG13 plasma glycan peak associated with the *B3GAT1* region, these results have not yet been replicated; it is unclear whether this observation may fully explain the observed association.

The locus at 12q24 contains two strong candidate genes, *SPPL3* and *HNF1A*. The gene *SPPL3* encodes an intramembrane aspartyl protease that cleaves glycosidases and glycosyltransferases, and that—when suppressed—leads to hyperglycosylation [104]. No monogenic phenotypes are known for this gene. For *HNF1A*, Lauc et al. [54] demonstrated that it and its downstream target *HNF4A* regulate the expression of key FUTs and fucose biosynthesis genes, showing that *HNF1A* is both necessary and sufficient to drive the expression of these genes in hepatic cells. The researchers suggested that *HNF1A* reciprocally regulates core versus antennary fucosylation, with increased expression of *HNF1A* repressing core fucosylation (*FUT8*) while activating the FUTs involved in antennary fucosylation (e.g., *FUT6*). In line with this mechanism of action, rare mutations in *HNF1A* that lead to maturity-onset diabetes of the young, type 3 (MODY-3) are characterized by an increased level of antennary fucosylation on tri- and tetra-antennary plasma *N*-glycans [16,117], as well as a general increase in α1-3,4 fucosylation levels in blood plasma [118]. A mechanism through which genetic variation in 12q24 links to the function of the candidate gene(s) has not yet been suggested.

Two loci containing genes *IKZF1* and *IKZF3* that belong to the Ikaros zinc finger transcription factors family are associated with the IgG-glycome [51], while the *IKZF3* locus is also associated with the total plasma glycome [56]. Both genes have been reported to harbor mutations that cause monogenic forms of immunodeficiencies, which are characterized by abnormalities in B-cell functioning. Mutations in *IKZF1* lead to the common variable

immunodeficiency-13, an autosomal dominant primary immunodeficiency disorder associated with hypogammaglobulinemia and decreased numbers of B cells [119]. Mutations in *IKZF3* cause immunodeficiency-84 (IMD84), an autosomal dominant primary immunologic disorder associated with low levels of B cells and impaired early B-cell development [120].

By exploiting the multidimensional nature of glycomic data, Klarić et al. [51] have built a network connecting associated loci. In this network, the closely located nodes are loci with a similar spectrum of associated glycans, suggesting a functional similarity of the causal genes in these loci. The presence of glycosyltransferases in the network allows for functional inferences via an interesting approach. Given two connected loci containing genes encoding a glycosyltransferase and a transcription factor, it could be suggested that the transcription factor may regulate the expression of the glycosyltransferase. In particular, it follows from the network deduced in Ref. [51] that loci containing *IKZF1* and *IKZF3* may regulate *FUT8*, an enzyme responsible for the core fucosylation of IgG-attached *N*-glycans. Moreover, the researchers showed that, in the *IKZF3* locus, the same genetic variant(s) may have an effect on both IgG fucosylation and the expression of *IKZF3* in peripheral blood. It was demonstrated that the knockdown of *IKZF1* in the lymphoblastoid cell line MATAT-6 leads to decreased expression of *IKZF3*, increased expression of *FUT8*, and increased core fucosylation of IgG [51], confirming the suggested hypotheses.

In two loci, genes belonging to the Runt-related family—namely, *RUNX1* and *RUNX3*—have been highlighted as candidates that may be responsible for the association with IgG N-glycosylation [51]. These genes, which are coding transcription factors, play an important role in hematopoietic cell differentiation [121,122]. Mijakovac et al. [109] proposed that IgG galactosylation and the transcription factor genes *RUNX1* and *RUNX3* are regulated by estradiol. Subsequent analysis using clustered regularly interspaced short palindromic repeat (CRISPR)/dead CRISPR-associated endonuclease 9 (dCas9) molecular tools in the HEK-293FS-derived cell line, which is optimized for secreting a high quantity of native IgG antibodies, provided evidence that the upregulation of *RUNX3* leads to a decrease in the level of IgG galactosylation. Although the role of *RUNX3* in the regulation of the IgG glycome is sufficiently supported, the exact mechanism underlying the genetic association of *RUNX3* remains unclear.

A locus harboring the gene *TNFRSF13B* was discovered by means of a multivariate genetic association analysis of the IgG glycome by Shadrina et al. [52]. *TNFRSF13B* encodes the TACI protein, a lymphocyte-specific member of the tumor necrosis factor receptor superfamily. Rare *TNFRSF13B* mutations can increase the risk of asthma symptoms [123]; moreover, several common variants in this gene have been reported to be associated with circulating IgG levels [124,125], while coding variants in *TNFRSF13B* are associated with common variable immunodeficiency and selective IgA deficiency [126].

Another locus discovered and replicated in Ref. [52] contains the gene *RFN168*, which encodes an E3 ubiquitin ligase protein involved in DNA double-strand break repair and immunoglobulin class switch recombination [96]. The association of common genetic variation in the *RNF168* locus with IgM levels was previously reported [125], and rare mutations in the gene *RNF168* lead to autosomal recessive radiosensitivity, immunodeficiency, dysmorphic features, learning difficulties (RIDDLE) syndrome, the features of which include decreased immunoglobulin production [127].

Another strong candidate for discovered and replicated in Ref. [23] locus is the gene *TF*. As mentioned above, TF itself is an N-glycoprotein with two N-glycosylation sites. Genetic variation in the *TF* locus is associated with levels of carbohydrate-deficient TF [41] and with abundances of TF-linked *N*-glycans [23]. A cis- protein quantitative trait locus (pQTL) exists for TF, explaining about 40% of the population variation of its serum levels [128]. The pQTL tagged by rs8177240 [128] is in low ($r^2 < 0.1$) LD with the sentinel TF N-glycosylation SNP, rs6785596 [23]. On the other hand, the sentinel N-glycosylation SNP is in a reasonably strong ($r^2 < 0.57$) LD with one of the SNPs, rs1799899, associated with the levels of carbohydrate-deficient TF [41]. The latter SNP codes for G277S amino acid change in the TF. It can be speculated that this amino acid substitution may change the protein tertiary structure and the accessibility of the N-glycosylation site. Whether or not the association of TF N-glycosylation with the *TF* region may be fully explained by the LD between the sentinel SNP reported by Landini et al. [23] and the SNPs reported by Kutalik et al. [41] remains open.

To conclude, for about half of the 34 loci associated with variation in N-glycosylation, the biological mechanism of action may be hypothesized based on the presence of a strong candidate gene. For about half of these loci, a mechanism linking genetic variation to the target gene can also be hypothesized. The large number of loci for which the mechanism of action is difficult to guess can be a reach source of knowledge about human glycobiology.

6. Shared genetic determinants of N-glycosylation and complex disease: Pleiotropy at specific loci

Some loci that are associated with protein glycosylation also show an association with complex human diseases [51,52,56]. The association of a locus with two traits may be due to the pleiotropic effect of a genetic variant on both traits or the independent effects of different genetic variants that are in LD with each other by two traits. It should be noted that, with perfect LD between two variants, the latter hypothesis becomes statistically indistinguishable from the former. Given this fundamental restriction, methods have been developed to distinguish between these two hypotheses [77,78]. Such methods compare the patterns of regional association for the two traits associated with the same region. In the case of sufficient dissimilarity, the hypothesis of pleiotropy is rejected. A similarity suggests the possibility of pleiotropy.

In a case in which a locus has a pleiotropic effect on a glycan level and a disease, the option of the glycan may be investigated as a biomarker of the disease [129]. This logic formed the basis for the discovery of an increased level of antennary fucosylation of plasma proteins as a functional biomarker of one of the most prevalent MODY-3, which is due to mutations in the gene *HNF1A*. In the first plasma glycome GWAS, a locus containing the gene *HNF1A* was associated with changes in the composition of the human plasma N-glycome; this gene was confirmed to shift the balance between core and antennary fucosylation in a cell line experiment [54]. MODY-3 is often misdiagnosed as type 1 diabetes (T1D) or type 2 diabetes (T2D) [130]. The antennary fucosylation biomarkers have been successfully used to separate people with MODY-3 from people with T1D and T2D [118,131] (Fig. 3).

Klarić et al. [51] demonstrated the pleiotropic effects of the IgG N-glycome-associated locus 17q12-21 (containing genes *ORMDL3*, *GSDMB*, and *IKZF3*) on a range of autoimmune and inflammatory diseases. It is notable that 17q12-21 is one of the best-known asthma-associated loci [132]. The asthma risk alleles have been associated with decreased levels of afucosylated, agalactosylated, and monogalactosylated glycans and increased levels of core fucosylated glycans [51]. This locus was also pleiotropically associated with the risk of autoimmune and inflammatory diseases such as rheumatoid arthritis, Crohn's disease, ulcerative colitis, and primary biliary cirrhosis, as demonstrated by colocalization analysis.

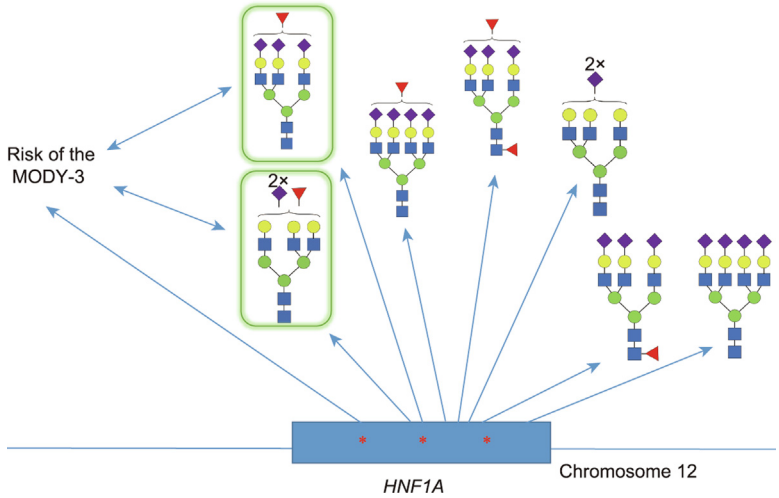

Fig. 3. Significant associations between glycan traits and the locus containing gene HNF1A, obtained by means of GWASs [56]. The framed glycans are the best-performing single glycan traits for classifying MODY-3. The figure was compiled based on results and data reported in Refs. [118,131].

The autoimmune disease risk allele has been associated with increased levels of afucosylated, agalactosylated, and monogalactosylated glycans and decreased levels of fucosylated glycans [51]. In line with these observations, it is known that, although the same SNPs are associated with both asthma and autoimmune disease in this locus, the risk alleles associated with autoimmune disease are the opposite of those associated with asthma risk [132].

Furthermore, Shadrina et al. [52] demonstrated the pleiotropic effect of another IgG-glycome-associated locus containing genes *OVAL1* and *AP5B1* on asthma. This locus was also found to have pleiotropic effects on dermatological and allergic diseases [133]. While it is tempting to speculate about the biological mechanisms that link or are shared between IgG N-glycosylation and asthma, a generalization may be somewhat premature. Several observations, including the association of the two loci with different spectra of glycans and subtypes of asthma, suggest that the mechanism determining the pleiotropic effects of the locus that includes *ORMDL3*, *GSDMB*, and *IKZF3* and the locus that includes *OVOL1* and *AP5B1* on asthma and the IgG N-glycome may be different (see Ref. [52] for a more detailed discussion).

It can be speculated that the association of the same allele with the risk of autoimmune disease and the increased abundance of afucosylated agalactosylated glycans at the *ORMDL3-GSDMB-IKZF3* locus is biologically plausible, as IgG carrying such glycans is known to be pro-inflammatory [33] and to have a strong ability to induce ADCC [33]. Consistent with this speculation, in the *IRF1-SLC22A4* locus, the Crohn's disease risk allele has been demonstrated to be associated with increased levels of afucosylated agalactosylated glycans and decreased levels of fucosylation [51]. A colocalization analysis by Klarić et al. [51] also suggested possible pleiotropy between variants affecting glycans and the risk of Crohn's disease at the locus containing the gene *IRF7* and the risk of Parkinson's disease at the locus containing the gene *MAPT*.

To conclude, a number of loci associated with IgG N-glycosylation may have pleiotropic effects on the risk of autoimmune and inflammatory disease—particularly inflammatory bowel disease and asthma. In addition, colocalization with Parkinson's disease was observed at the *MAPT* locus. For the *HNF1A* locus associated with TF N-glycosylation, pleiotropic effects on C-reactive protein (CRP), low-density lipoprotein (LDL), and total cholesterol levels have been shown [23].

7. Shared genetic determinants of N-glycosylation and complex disease: Results from Mendelian randomization (MR) analyses

Although changes in protein glycosylation have been observed in many complex human diseases [29], little is known about a causal relationship between glycosylation and complex diseases. Randomized controlled trials are the gold standard of experimental design in epidemiology for testing the causal relationship between risk factor and disease. However, the application of such a design is challenging in the case of protein glycosylation. Here, genetics offers a MR approach [80,134] that mimics a randomized controlled trial using genetic variants associated with a risk factor as instrumental variables. If genetic variants shift the level of risk factor and this shift has an effect on disease outcome, then it provides evidence that the risk factor has a causal effect on the disease outcome. In this case, the risk factor may be suggested as a therapeutic target for manipulation, which may change the risk of the disease. Reverse causality testing is also relevant because, in some cases, the manifestation of disease has a causal effect on molecular phenotypes, such as cardiovascular risk and high-density lipoprotein (HDL) [135]. In this case, the molecular phenotype may be suggested as a potential biomarker of the disease. It should be mentioned that, although the idea behind MR is relatively simple, there are several assumptions, the violation of which may lead to false-positive results; thus, the results of MR should be treated with caution [134].

Thus far, the MR approach has been used to study the causal relationship between IgG glycome traits and 12 common diseases associated with the IgG N-glycome [136,137], as well as between the total plasma N-glycome, the IgG N-glycome, and prostate cancer [138]. Systemic lupus erythematosus (SLE) showed a causal effect on the abundance of *N*-glycans with bisecting GlcNAc in the total IgG N-glycome [137], suggesting this glycan trait as a biomarker of SLE. The results of the analysis are rather robust, as a large number (36) of SLE-associated genetic variants, large sample sizes, and a range of methods were used to come to a conclusion. The abundance of bisected *N*-glycans attached to IgG is associated with the pro-inflammatory state [139] and is elevated in patients with SLE [140], although the mechanism is not yet clear. It is possible that an increased level of bisected glycans in SLE is a consequence of lower fucosylation of IgG glycans, since bisection and core fucosylation of *N*-glycans are competing processes [141],

although no causal relationship between IgG fucosylation and SLE has been observed. IgG-linked high-mannose glycan peak 5 (GP5; Man5) and bisected GP13 in samples from people of East Asian descent showed a causative association with T2D outcome [136], which is in line with the known correlation between these glycan traits and T2D [142].

The same approach was used to test the causal effect of the total plasma N-glycome and IgG N-glycome traits on prostate cancer [138], where decreased high-mannose glycan (Man9) level, increased ratio of disialylated to trisialylated glycans, and increased level of four-branched heavily galactosylated and sialylated glycans (attached to α-1-acid protein—an acute-phase protein) were shown to cause a lower risk of prostate cancer. However, the latter is contradictory to the observed elevated levels of this glycan with a higher level of low-grade inflammation, and future work is needed to better understand the relationship [143]. It should be noted that the obtained results were driven by one or two instrumental variables, suggesting the possible presence of weak instrumental bias [144]. More powerful GWASs are needed to confirm or refute the postulated relation.

To conclude, there is genetic support for the hypothesis that the N-glycosylation of IgG may serve as a biomarker for autoimmune disease, such as SLE, while the plasma N-glycome may provide biomarkers for endocrine disease and cancers. This hypothesis is consistent with and confirms the results of (prospective) epidemiologic studies [29,145].

8. Conclusions

Protein glycosylation complements other layers of omics phenotypes in studies aimed at discovering the mechanisms of human health and disease. The glycosylation of human proteins is implicated in many human diseases and pathological states, making the glycome a rich source of biomarkers and drug targets. Compared with the proteome and metabolome, examples of glycome-based biomarkers and drug targets are relatively rare. This is partly due to insufficient knowledge about the mechanisms of human glycome regulation *in vivo.* While the biochemical bases of glycan biosynthesis are well studied, little is known about how this network is regulated, especially in a cell-, tissue-, and environment-specific manner.

A combination of high-throughput glycomics, genetics, and genomics offers a "data-driven hypotheses" approach to discover the regulation of protein glycosylation in an agnostic manner [20,21]. At present, genetic studies of human blood plasma, IgG, and TF N-glycosylation have established the link to genetic variation at 34 loci. From the analysis of candidate genes in these loci, it follows that this genetic variation affects glycosyltransferases involved in the fucosylation, branching, and terminal modifications of *N*-glycans and tissue-specific regulators involved in the differentiation, maturation, and regulation of the cells (hepatocytes and plasma cells) producing the relevant glycoproteins. The variations affecting the cellular machinery involved in the production and transportation of the secreted glycoproteins and the production and structure of the studied glycoproteins (TF and IgG) themselves are also becoming apparent. For about half of the 34 loci associated with variation in N-glycosylation, the biological mechanism of action may be hypothesized based on the presence of a strong candidate gene; moreover, for about half of these loci, a mechanism linking genetic variation to the target gene can also be hypothesized. The large number of loci for which a mechanism of action is difficult to hypothesize may provide a reach source of novel knowledge about human glycobiology. Hypotheses about the mechanism of action must be targeted in functional wet-lab experiments.

A great deal of effort has been made to establish and understand the mechanisms linking variation in N-glycosylation to human health and disease. It has been demonstrated that a number of loci associated with IgG N-glycosylation may have pleiotropic effects on the risk of autoimmune and inflammatory disease, while plasma N-glycome loci may have pleiotropic effects on endogenous cardiovascular risk factors. The candidate genes in these loci provide a starting point for research on the common etiology of N-glycosylation and complex diseases. Genetics have provided support for the hypothesis that the N-glycosylation of IgG may serve as a biomarker for autoimmune disease, while the plasma N-glycome may provide biomarkers for endocrine disease and cancer. These findings confirm the results from epidemiologic studies that have established parts of the N-glycome as risk factors and/or biomarkers of cardiovascular, metabolic, and autoimmune diseases and cancers [29,145]. However, it should be noted that glycogenomic studies have not only supported but also proposed novel biomarkers of human diseases [16,48,137]. Some of these biomarkers have demonstrated high discriminative power and are *en route* to clinical practice [118].

9. Perspectives

Glycogenetics and glycogenomics are evolving rapidly. We expect at least three directions for further studies of the genetic control of human protein glycosylation. First, compared with studies of other complex traits, the sample sizes used in genetic analyses of human N-glycomics are still relatively small, which limits the application of quantitative genetics methods. We expect more powerful GWASs with increased sample size in the future. For complex human traits, it is expected that doubling the sample size will roughly double the number of discovered genetic loci [60,61]. An increased sample size will lead to more precise estimates of genetic effect, thereby providing more reliable input for quantitative genetics analyses such as MR, colocalization, genetic correlation, and others, allowing for a better understanding of the role of the glycome in disease etiology.

Second, we expect the rise of studies focused on single-protein glycosylation. The availability of material for these studies will depend on advances in high-throughput glycomics technologies [22]. Scaling up conventional glycomics to large-scale glycoprotein analysis using, for example, liquid chromatography-tandem mass spectrometry (LC-MS/MS) instrumentation is important for the advancement toward functional glycomics that permit highly detailed site-, protein-, and cell-specific elucidation of the *N*-glycans attached to proteins. Therefore, it is logical to assume that a deeper understanding of the mechanism of action of specific GWAS loci will at least partly depend on developments in proteomics and glycoproteomics. Such studies will not only lead to better interpretation of observed genetic associations and deeper dissection of tissue-specific regulation networks but will also facilitate translation to clinical applications.

Third, progress in high-throughput glycomics and the genetics of glycosylation will be complemented and enhanced by advances in multi-omics approaches. Of specific interest are techniques for measuring the transcriptomes of single cells [146]. Such technology makes it possible to decompose transcription in a complex mixture of cells, thus providing potential for aligning glycosylation with specific—potentially rare—cell types. The availability of samples with measured multi-omics layers will have a synergistic effect on our understanding of human glycobiology and will improve the integration of the glycome into the system biology of cells, tissues, and organisms. It will also facilitate the discovery of novel biomarkers, drug targets, and therapeutics with better lines of evidence.

Thus, the continuous study of the genetic control of protein N-glycosylation should provide further insights into our knowledge of protein glycosylation biology, which will help in the development of new approaches for fighting glycome-dependent forms of autoimmune, cancer, and other diseases.

Acknowledgments

We are grateful to Anna Soplenkova and Denis Maslov for help with tables and figures.

This work is an output of a research project implemented as part of the Research Program at the Moscow State University (MSU) Institute for Artificial Intelligence.

Compliance with ethics guidelines

Anna Timoshchuk, Sodbo Sharapov, and Yurii S. Aulchenko declare that they have no conflict of interest or financial conflicts to disclose.

Yurii S. Aulchenko is co-founder and co-owner of PolyKnomics BV, a research-intensive SME providing services and developing products in the field in quantitative and statistical genetics and computational genomics.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eng.2023.03.013.

References

[1] Chauhan JS, Rao A, Raghava GPS. *In silico* platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. PLoS One 2013;8(6):e67008.
[2] Clerc F, Reiding KR, Jansen BC, Kammeijer GSM, Bondt A, Wuhrer M. Human plasma protein N-glycosylation. Glycoconj J 2016;33(3):309–43.
[3] Knezević A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, et al. Variability, heritability and environmental determinants of human plasma N-glycome. J Proteome Res 2009;8(2):694–701.
[4] Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol 2012;13(7):448–62.
[5] Lauc G, Vojta A, Zoldoš V. Epigenetic regulation of glycosylation is the quantum mechanics of biology. Biochim Biophys Acta 2014;1840(1):65–70.
[6] Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, et al. KEGG as a glycome informatics resource. Glycobiology 2006;16(5):63R–70R.
[7] Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol 2019;15(6):346–66.
[8] Lauc G, Trbojević-Akmačić I. The role of glycosylation in health and disease. Berlin: Springer International Publishing; 2021.
[9] Dwek RA, Butters TD, Platt FM, Zitzmann N. Targeting glycosylation as a therapeutic approach. Nat Rev Drug Discov 2002;1(1):65–75.
[10] Rodríguez E, Schetters STT, van Kooyk Y. The tumour glyco-code as a novel immune checkpoint for immunotherapy. Nat Rev Immunol 2018;18(3):204–11.
[11] Pagan JD, Kitaoka M, Anthony RM. Engineered sialylation of pathogenic antibodies *in vivo* attenuates autoimmune disease. Cell 2018;172(3). 564–77.e13.
[12] Johannssen T, Lepenies B. Glycan-based cell targeting to modulate immune responses. Trends Biotechnol 2017;35(4):334–46.
[13] Paderi J, Prestwich GD, Panitch A, Boone T, Stuart K. Glycan therapeutics: resurrecting an almost pharma-forgotten drug class. Adv Ther 2018;1(8):1800082.
[14] Van Landuyt L, Lonigro C, Meuris L, Callewaert N. Customized protein glycosylation to improve biopharmaceutical function and targeting. Curr Opin Biotechnol 2019;60:17–28.
[15] Adamczyk B, Tharmalingam T, Rudd PM. Glycans as cancer biomarkers. Biochim Biophys Acta 2012;1820(9):1347–53.
[16] Thanabalasingham G, Huffman JE, Kattla JJ, Novokmet M, Rudan I, Gloyn AL, et al. Mutations in HNF1A result in marked alterations of plasma glycan profile. Diabetes 2013;62(4):1329–37.
[17] Shinohara Y, Furukawa JI, Miura Y. Glycome as biomarkers. In: General methods in biomarker research and their applications. Dordrecht: Springer Netherlands; 2015. p. 111–40.
[18] Wang W. Glycomedicine: the current state of the art. Engineering. In press.
[19] Henrissat B, Surolia A, Stanley P. A genomic view of glycobiology. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, editors. Essentials of glycobiology. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2017.
[20] Van Helden P. Data-driven hypotheses. EMBO Rep 2013;14(2):104.
[21] Krištić J, Sharapov SZ, Aulchenko YS. Quantitative genetics of human protein N-glycosylation. In: Lauc G, Trbojević-Akmačić I, editors. The role of glycosylation in health and disease. Springer; 2021. p. 151–71.
[22] De Haan N, Pučić-Baković M, Novokmet M, Falck D, Lageveen-Kammeijer G, Razdorov G, et al. Developments and perspectives in high-throughput protein glycomics: enabling the analysis of thousands of samples. Glycobiology 2022;32(8):651–63.
[23] Landini A, Trbojević-Akmačić I, Navarro P, Tsepilov YA, Sharapov SZ, Vučković F, et al. Genetic regulation of post-translational modification of two distinct proteins. Nat Commun 2022;13(1):1586.
[24] Momčilović A, de Haan N, Hipgrave Ederveen AL, Bondt A, Koeleman CAM, Falck D, et al. Simultaneous immunoglobulin A and G glycopeptide profiling for high-throughput applications. Anal Chem 2020;92(6):4518–26.
[25] Dotz V, Visconti A, Lomax-Browne HJ, Clerc F, Hipgrave Ederveen AL, Medjeral-Thomas NR, et al. O- and N-glycosylation of serum immunoglobulin A is associated with IgA nephropathy and glomerular function. J Am Soc Nephrol 2021;32(10):2455–65.
[26] Demus D, Naber A, Dotz V, Jansen BC, Bladergroen MR, Nouta J, et al. Large-scale analysis of apolipoprotein CIII glycosylation by ultrahigh resolution mass spectrometry. Front Chem 2021;9:678883.
[27] Frkatovic A, Zaytseva OO, Klaric L. Genetic regulation of immunoglobulin G glycosylation. Exp Suppl 2021;112:259–87.
[28] Vidarsson G, Dekkers G, Rispens T. IgG subclasses and allotypes: from structure to effector functions. Front Immunol 2014;5:520.
[29] Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. Cell Immunol 2018;333:65–79.
[30] Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347(6220):1260419.
[31] Uhlén M, Karlsson MJ, Hober A, Svensson AS, Scheffel J, Kotol D, et al. The human secretome. Sci Signal 2019;12(609):eaaz0274.
[32] Bondt A, Selman MHJ, Deelder AM, Hazes JMW, Willemsen SP, Wuhrer M, et al. Association between galactosylation of immunoglobulin G and improvement of rheumatoid arthritis during pregnancy is independent of sialylation. J Proteome Res 2013;12(10):4522–31.
[33] Parekh R, Roitt I, Isenberg D, Dwek R, Rademacher T. Age-related galactosylation of the N-linked oligosaccharides of human serum IgG. J Exp Med 1988;167(5):1731–6.
[34] Ruhaak LR, Uh HW, Beekman M, Koeleman CAM, Hokke CH, Westendorp RGJ, et al. Decreased levels of bisecting GlcNAc glycoforms of IgG are associated with human longevity. PLoS One 2010;5(9):e12566.
[35] Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcγ RIII and antibody-dependent cellular toxicity. J Biol Chem 2002;277(30):26733–40.
[36] Ferrara C, Grau S, Jäger C, Sondermann P, Brünker P, Waldhauer I, et al. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIII and antibodies lacking core fucose. Proc Natl Acad Sci USA 2011;108(31):12669–74.
[37] Van de Bovenkamp FS, Hafkenscheid L, Rispens T, Rombouts Y. The emerging importance of IgG Fab glycosylation in immunity. J Immunol 2016;196(4):1435–41.
[38] McClain DA, Sharma NK, Jain S, Harrison A, Salaye LN, Comeau ME, et al. Adipose tissue transferrin and insulin resistance. J Clin Endocrinol Metab 2018;103(11):4197–208.
[39] Spik G, Bayard B, Fournet B, Strecker G, Bouquelet S, Montreuil J, et al. Complete structure of two carbohydrate units of human serotransferrin. FEBS Lett 1975;50(3):296–9.
[40] Karlsson I, Ndreu L, Quaranta A, Thorsén G. Glycosylation patterns of selected proteins in individual serum and cerebrospinal fluid samples. J Pharm Biomed Anal 2017;145:431–9.
[41] Kutalik Z, Benyamin B, Bergmann S, Mooser V, Waeber G, Montgomery GW, et al. Genome-wide association study identifies two loci strongly affecting transferrin glycosylation. Hum Mol Genet 2011;20(18):3710–7.
[42] Rose RJ, van Berkel PHC, van den Bremer ETJ, Labrijn AF, Vink T, Schuurman J, et al. Mutation of Y407 in the CH3 domain dramatically alters glycosylation and structure of human IgG. MAbs 2013;5(2):219–28.
[43] Huffman JE, Pučić-Baković M, Klarić L, Hennig R, Selman MHJ, Vučković F, 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(6):1598–610.
[44] Reiding KR, Bondt A, Hennig R, Gardner RA, O'Flaherty R, Trbojević-Akmačić I, et al. High-throughput serum N-glycomics: method comparison and application to study rheumatoid arthritis and pregnancy-associated changes. Mol Cell Proteomics 2019;18(1):3–15.
[45] Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLOS Comput Biol 2012;8(12):e1002822.
[46] Kraft P, Zeggini E, Ioannidis JPA. Replication in genome-wide association studies. Stat Sci 2009;24(4):561–73.
[47] Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gómez-Marín C, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 2014;507(7492):371–5.

[48] Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, et al. Loci associated with N-glycosylation of human immunoglobulin G show pleiotropy with autoimmune diseases and haematological cancers. PLoS Genet 2013;9(1):e1003225.
[49] Shen X, Klarić L, Sharapov S, Mangino M, Ning Z, Wu D, et al. Multivariate discovery and replication of five novel loci associated with immunoglobulin G N-glycosylation. Nat Commun 2017;8(1):447.
[50] Wahl A, van den Akker E, Klaric L, Štambuk J, Benedetti E, Plomp R, et al. Genome-wide association study on immunoglobulin G glycosylation patterns. Front Immunol 2018;9:277.
[51] Klarić L, Tsepilov YA, Stanton CM, Mangino M, Sikka TT, Esko T, et al. Glycosylation of immunoglobulin G is regulated by a large network of genes pleiotropic with inflammatory diseases. Sci Adv 2020;6(8):eaax0301.
[52] Shadrina AS, Zlobin AS, Zaytseva OO, Klarić L, Sharapov SZ, Pakhomov D, et al. Multivariate genome-wide analysis of immunoglobulin G N-glycosylation identifies new loci pleiotropic with immune function. Hum Mol Genet 2021;30(13):1259–70.
[53] Ning Z, Tsepilov YA, Sharapov SZ, Wang Z, Grishenko AK, Feng X, et al. Nontrivial replication of loci detected by multi-trait methods. Front Genet 2021;12:627989.
[54] Lauc G, Essafi A, Huffman JE, Hayward C, Knežević A, Kattla JJ, et al. Genomics meets glycomics-the first GWAS study of human N-glycome identifies HNF1α as a master regulator of plasma protein fucosylation. PLoS Genet 2010;6(12):e1001256.
[55] Huffman JE, Knezevic A, Vitart V, Kattla J, Adamczyk B, Novokmet M, et al. Polymorphisms in B3GAT1, SLC9A9 and MGAT5 are associated with variation within the human plasma N-glycome of 3533 European adults. Hum Mol Genet 2011;20(24):5000–11.
[56] Sharapov SZ, Tsepilov YA, Klaric L, Mangino M, Thareja G, Shadrina AS, et al. Defining the genetic control of human blood plasma N-glycome using genome-wide association study. Hum Mol Genet 2019;28(12):2062–77.
[57] Sharapov SZ, Shadrina AS, Tsepilov YA, Elgaeva EE, Tiys ES, Feoktistova SG, et al. Replication of fifteen loci involved in human plasma protein N-glycosylation in 4,802 samples from four cohorts. Glycobiology 2020;31(2):82–8.
[58] Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. Nat Commun 2017;8(1):1483.
[59] Tsepilov YA, Sharapov SZ, Zaytseva OO, Krumsiek J, Prehn C, Adamski J, et al. A network-based conditional genetic association analysis of the human metabolome. Gigascience 2018;7(12):giy137.
[60] Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. Am J Hum Genet 2012;90(1):7–24.
[61] Canela-Xandri O, Rawlik K, Tenesa A. An atlas of genetic associations in UK Biobank. Nat Genet 2018;50(11):1593–9.
[62] Li X, Wang H, Zhu Y, Cao W, Song M, Wang Y, et al. Heritability enrichment of immunoglobulin G N-glycosylation in specific tissues. Front Immunol 2021;12:741705.
[63] Huang H, Fang M, Jostins L, Umićević Mirkov M, Boucher G, Anderson CA, et al.; IIBDGC. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 2017;547(7662):173–8.
[64] Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet 2018;19(8):491–504.
[65] Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al.; GTEx Consortium. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet 2018;50(7):956–67.
[66] Dekker J, Misteli T. Long-range chromatin interactions. Cold Spring Harb Perspect Biol 2015;7(10):a019356.
[67] Statello L, Guo CJ, Chen LL, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 2021;22(2):96–118.
[68] McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. Genome Biol 2016;17(1):122.
[69] Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 2018;34(3):511–3.
[70] Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, et al.; ENCODE Project Consortium. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 2020;583(7818):699–710.
[71] Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–30.
[72] Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al.; MuTHER Consortium. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46(6):543–50.
[73] Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium, LDACC—Analysis Working Group, Statistical Methods groups—Analysis Working Group, eGTEx groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead Analysts, LDACC, NIH Program Management, Biospecimen Collection, Pathology, eQTL Manuscript Working Group. Genetic effects on gene expression across human tissues. Nature 2017;550(7675):204–213.
[74] Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charloteaux B, et al.; IIBDGC. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun 2018;9(1):2427.
[75] Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature 2018;558(7708):73–9.
[76] Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet 2021;53(12):1712–21.
[77] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10(5):e1004383.
[78] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48(5):481–7.
[79] Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8(1):1826.
[80] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. eLife 2018;7:e34408.
[81] Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. Bioinformatics 2019;35(22):4851–3.
[82] Shashkova TI, Gorev DD, Pakhomov ED, Shadrina AS, Sharapov SZ, Tsepilov YA, et al. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP|homo database of 70 billion genetic associations of human traits. Vavilovskii Zhurnal Genet Selektsii 2020;24(8):876–84.
[83] Shashkova TI, Pakhomov ED, Gorev DD, Karssen LC, Joshi PK, Aulchenko YS. PheLiGe: an interactive database of billions of human genotype-phenotype associations. Nucleic Acids Res 2021;49(D1):D1347–50.
[84] Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun 2018;9(1):1366.
[85] Zhang Z, Shah B, Richardson J. Impact of Fc *N*-glycan sialylation on IgG structure. MAbs 2019;11(8):1381–90.
[86] Niebuhr B, Kriebitzsch N, Fischer M, Behrens K, Günther T, Alawi M, et al. Runx1 is essential at two stages of early murine B-cell development. Blood 2013;122(3):413–23.
[87] Sellars M, Reina-San-Martin B, Kastner P, Chan S. Ikaros controls isotype selection during immunoglobulin class switch recombination. J Exp Med 2009;206(5):1073–87.
[88] Wang JH, Avitahl N, Cariappa A, Friedrich C, Ikeda T, Renold A, et al. Aiolos regulates B cell activation and maturation to effector state. Immunity 1998;9(4):543–53.
[89] Brescia P, Schneider C, Holmes AB, Shen Q, Hussein S, Pasqualucci L, et al. MEF2B instructs germinal center development and acts as an oncogene in B cell lymphomagenesis. Cancer Cell 2018;34(3). 453–65.e9.
[90] Micol JB, Pastore A, Inoue D, Duployez N, Kim E, Lee SCW, et al. ASXL2 is essential for haematopoiesis and acts as a haploinsufficient tumour suppressor in leukemia. Nat Commun 2017;8(1):15429.
[91] Martincic K, Alkan SA, Cheatle A, Borghesi L, Milcarek C. Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat Immunol 2009;10(10):1102–9.
[92] Von Bülow GU, Bram RJ. NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. Science 1997;278(5335):138–41.
[93] 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(4):885–91.
[94] 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(6). 1277–86.e5.
[95] Rosnet O, Blanco-Betancourt C, Grivel K, Richter K, Schiff C. Binding of free immunoglobulin light chains to VpreB3 inhibits their maturation and secretion in chicken B cells. J Biol Chem 2004;279(11):10228–36.
[96] Ramachandran S, Chahwan R, Nepal RM, Frieder D, Panier S, Roa S, et al. The RNF8/RNF168 ubiquitin ligase cascade facilitates class switch recombination. Proc Natl Acad Sci USA 2010;107(2):809–14.
[97] Szabo SJ, Kim ST, Costa GL, Zhang X, Fathman CG, Glimcher LH. A novel transcription factor, T-bet, directs Th1 lineage commitment. Cell 2000;100(6):655–69.
[98] Nakayama T, Kimura MY. Memory Th1/Th2 cell generation controlled by Schnurri-2. Adv Exp Med Biol 2010;684:1–10.
[99] Staton TL, Lazarevic V, Jones DC, Lanser AJ, Takagi T, Ishii S, et al. Dampening of death pathways by Schnurri-2 is essential for T-cell development. Nature 2011;472(7341):105–9.
[100] Cheah IK, Halliwell B. Ergothioneine, recent developments. Redox Biol 2021;42:101868.
[101] Ding J, Wang K, Liu W, She Y, Sun Q, Shi J, et al. Pore-forming activity and structural autoinhibition of the gasdermin family. Nature 2016;535(7610):111–6.

[102] Zhou Z, He H, Wang K, Shi X, Wang Y, Su Y, et al. Granzyme A from cytotoxic lymphocytes cleaves GSDMB to trigger pyroptosis in target cells. Science 2020;368(6494):eaaz7548.
[103] Carreras-Sureda A, Cantero-Recasens G, Rubio-Moscardo F, Kiefer K, Peinelt C, Niemeyer BA, et al. ORMDL3 modulates store-operated calcium entry and lymphocyte activation. Hum Mol Genet 2013;22(3):519–30.
[104] Voss M, Künzel U, Higel F, Kuhn PH, Colombo A, Fukumori A, et al. Shedding of glycan-modifying enzymes by signal peptide peptidase-like 3 (SPPL3) regulates cellular N-glycosylation. EMBO J 2014;33(24):2890–905.
[105] Hirst J, Barlow LD, Francisco GC, Sahlender DA, Seaman MNJ, Dacks JB, et al. The fifth adaptor protein complex. PLoS Biol 2011;9(10):e1001170.
[106] Hirst J, Itzhak DN, Antrobus R, Borner GHH, Robinson MS. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. PLoS Biol 2018;16(1):e2004411.
[107] Papadopoulou AA, Müller SA, Mentrup T, Shmueli MD, Niemeyer J, Haug-Kröper M, et al. Signal peptide peptidase-like 2c impairs vesicular transport and cleaves SNARE proteins. EMBO Rep 2019;20(3):e46451.
[108] Mijakovac A, Miškec K, Krištić J, Vičić Bočkor V, Tadić V, Bošković M, et al. A transient expression system with stably integrated CRISPR-dCas9 fusions for regulation of genes involved in immunoglobulin G glycosylation. CRISPR J 2022;5(2):237–53.
[109] Mijakovac A, Jurić J, Kohrt WM, Krištić J, Kifer D, Gavin KM, et al. Effects of estradiol on immunoglobulin G glycosylation: mapping of the downstream signaling mechanism. Front Immunol 2021;12:680227.
[110] Kalimuthu SN, Chetty R. Gene of the month: SMARCB1. J Clin Pathol 2016;69(6):484–9.
[111] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al.; gonmAD Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581(7809):434–443.
[112] Mollicone R, Reguigne I, Fletcher A, Aziz A, Rustam M, Weston BW, et al. Molecular basis for plasma alpha(1,3)-fucosyltransferase gene deficiency (FUT6). J Biol Chem 1994;269(17):12662–71.
[113] Brinkman-van der Linden ECM, Mollicone R, Oriol R, Larson G, van den Eijnden DH, van Dijk W. A missense mutation in the *FUT6* gene results in total absence of α3-fucosylation of human α1-acid glycoprotein. J Biol Chem 1996;271(24):14492–5.
[114] Puan KJ, San Luis B, Yusof N, Kumar D, Andiappan AK, Lee W, et al.; 23andMe Research Team. FUT6 deficiency compromises basophil function by selectively abrogating their sialyl-Lewis x expression. Commun Biol 2021;4(1):832.
[115] Oka S, Terayama K, Kawashima C, Kawasaki T. A novel glucuronyltransferase in nervous system presumably associated with the biosynthesis of HNK-1 carbohydrate epitope on glycoproteins. J Biol Chem 1992;267(32):22711–4.
[116] Mitsumoto Y, Oka S, Sakuma H, Inazawa J, Kawasaki T. Cloning and chromosomal mapping of human glucuronyltransferase involved in biosynthesis of the HNK-1 carbohydrate epitope. Genomics 2000;65(2):166–73.
[117] Juszczak A, Pavić T, Vučković F, Bennett AJ, Shah N, Pape Medvidović E, et al. Plasma fucosylated glycans and C-reactive protein as biomarkers of HNF1A-MODY in young adult-onset nonautoimmune diabetes. Diabetes Care 2019;42(1):17–26.
[118] Demus D, Urbanowicz PA, Gardner RA, Wu H, Juszczak A, Štambuk T, et al. Development of an exoglycosidase plate-based assay for detecting α1-3,4 fucosylation biomarker in individuals with HNF1A-MODY. Glycobiology 2022;32(3):230–8.
[119] Kuehn HS, Boisson B, Cunningham-Rundles C, Reichenbach J, Stray-Pedersen A, Gelfand EW, et al. Loss of B cells in patients with heterozygous mutations in IKAROS. N Engl J Med 2016;374(11):1032–43.
[120] Yamashita M, Kuehn HS, Okuyama K, Okada S, Inoue Y, Mitsuiki N, et al. A variant in human AIOLOS impairs adaptive immunity by interfering with IKAROS. Nat Immunol 2021;22(7):893–903.
[121] Woolf E, Brenner O, Goldenberg D, Levanon D, Groner Y. Runx3 regulates dendritic epidermal T cell development. Dev Biol 2007;303(2):703–14.
[122] Michaud J, Wu F, Osato M, Cottles GM, Yanagida M, Asou N, et al. *In vitro* analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. Blood 2002;99(4):1364–72.
[123] Janzi M, Melén E, Kull I, Wickman M, Hammarström L. Rare mutations in TNFRSF13B increase the risk of asthma symptoms in Swedish children. Genes Immun 2012;13(1):59–65.
[124] Liao M, Ye F, Zhang B, Huang L, Xiao Q, Qin M, et al. Genome-wide association study identifies common variants at TNFRSF13B associated with IgG level in a healthy Chinese male population. Genes Immun 2012;13(6):509–13.
[125] Jonsson S, Sveinbjornsson G, de Lapuente Portilla AL, Swaminathan B, Plomp R, Dekkers G, et al. Identification of sequence variants influencing immunoglobulin levels. Nat Genet 2017;49(8):1182–91.
[126] Castigli E, Wilson S, Garibyan L, Rachid R, Bonilla F, Schneider L, et al. Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. Nat Genet 2007;39(4):430–1.
[127] Stewart GS, Panier S, Townsend K, Al-Hakim AK, Kolas NK, Miller ES, et al. The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. Cell 2009;136(3):420–34.
[128] Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, et al. Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. Am J Hum Genet 2009;84(1):60–5.
[129] Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurle MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet 2020;52(10):1122–31.
[130] Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S. Maturity-onset diabetes of the young (MODY): how many cases are we missing? Diabetologia 2010;53(12):2504–8.
[131] Demus D, Jansen BC, Gardner RA, Urbanowicz PA, Wu H, Štambuk T, et al. Interlaboratory evaluation of plasma *N*-glycan antennary fucosylation as a clinical biomarker for HNF1A-MODY using liquid chromatography methods. Glycoconj J 2021;38(3):375–86.
[132] Stein MM, Thompson EE, Schoettler N, Helling BA, Magnaye KM, Stanhope C, et al. A decade of research on the 17q12-21 asthma locus: piecing together the puzzle. J Allergy Clin Immunol 2018;142(3):749-64.e3.
[133] Christou MA, Ntritsos G, Markozannes G, Koskeridis F, Nikas SN, Karasik D, et al. A genome-wide scan for pleiotropy between bone mineral density and nonbone phenotypes. Bone Res 2020;8(1):26.
[134] Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.
[135] Hewing B, Moore KJ, Fisher EA. HDL and cardiovascular risk: time to call the plumber? Circ Res 2012;111(9):1117–20.
[136] Wang B, Liu D, Song M, Wang W, Guo B, Wang Y. Immunoglobulin G *N*-glycan, inflammation and type 2 diabetes in East Asian and European populations: a Mendelian randomization study. Mol Med 2022;28(1):114.
[137] Zaytseva OO, Sharapov SZ, Perola M, Esko T, Landini A, Hayward C, et al. Investigation of the causal relationships between human IgG N-glycosylation and 12 common diseases associated with changes in the IgG N-glycome. Hum Mol Genet 2022;31(10):1545–59.
[138] Liu D, Zhu J, Zhao T, Sharapov S, Tiys E, Wu L. Associations between genetically predicted plasma *N*-glycans and prostate cancer risk: analysis of over 140,000 European descendants. Pharm Genomics Pers Med 2021;14:1211–20.
[139] Umaña P, Jean-Mairet J, Moudry R, Amstutz H, Bailey JE. Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibody-dependent cellular cytotoxic activity. Nat Biotechnol 1999;17(2):176–80.
[140] Vučković F, Krištić J, Gudelj I, Teruel M, Keser T, Pezer M, et al. Association of systemic lupus erythematosus with decreased immunosuppressive potential of the IgG glycome. Arthritis Rheumatol 2015;67(11):2978–89.
[141] Ikeda Y, Ihara H, Tsukamoto H, Gu J, Taniguchi N. Mannosyl(β-1,4-)-glycoprotein β-1,4-*N*-acetylglucosaminyltransferase (MGAT3); β1,4-*N*-acetylglucosaminyltransferase III (GnT-III, GlcNAcT-III). In: Handbook of glycosyltransferases and related genes. Tokyo: Springer Japan; 2014. p. 209–22.
[142] Rudman N, Gornik O, Lauc G. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. FEBS Lett 2019;593(13):1598–615.
[143] Novokmet M, Lukić E, Vučković F, Đurić Ž, Keser T, Rajšl K, et al. Changes in IgG and total plasma protein glycomes in acute systemic inflammation. Sci Rep 2014;4(1):4347.
[144] Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. Int J Epidemiol 2011;40(3):755–64.
[145] Dotz V, Wuhrer M. N-glycome signatures in human plasma: associations with physiology and major diseases. FEBS Lett 2019;593(21):2966–76.
[146] Dworkin LA, Clausen H, Joshi HJ. Applying transcriptomics to study glycosylation at the cell type level. iScience 2022;25(6):104419.