Advances in Clinical Pharmacology and Therapeutics

https://acpt.cultechpub.com/acpt

E-ISSN: 2982-1460

Vol. 3, No. 2, September 2026

DOI: 10.64229/v69kej42

Review

Reproducible Multi-Source Functional Genomics Synthesis Prioritizes a Core N-Glycan Biosynthesis Module (ST6GAL1-B4GALT1-FUT8-MGAT3) in IgG Glycosylation

Arbind Kumar Choudhary[1,*], Anamitra Hait[2], Francis Britto George[3], Sudhanshu Sahu[4]

[1]Department of Pharmacology, Government Erode Medical College and Hospital, Tamil Nadu, India

[2]Indian Railways Medical Service, ICU and Internal Medicine, Chittaranjan, India

[3]Arunai Medical College and Hospital, Tiruvannamalai, Tamil Nadu, India

[4]Department of Pharmacology DRIEMS Institute of Health Sciences and Hospital Tangi, Cuttack, Odisha, India

[*]Corresponding author: Arbind Kumar Choudhary, arbindkch@gmail.com

Abstract

Background: Genetic associations with immunoglobulin G (IgG) Fc N-glycosylation involve both glycosyltransferases and immune-regulatory loci. We synthesised published genomic evidence to distinguish reproducible modules from associations shaped by gene selection. Methods: Fifty-one genes from 29 genome-wide significant loci were curated from two IgG N-glycome association studies and one network study. Over-representation across eight versioned Enrichr libraries used library-specific backgrounds, Fisher's exact tests and within-library Benjamini-Hochberg correction. Single-gene and single-locus overlaps were flagged. Robustness was assessed by leave-one-locus-out analysis and four null models (1,000 gene sets each). Results: ST6GAL1, B4GALT1, FUT8 and MGAT3 showed significant enrichment in Reactome medial trans-Golgi N-glycan antennae elongation (4/26; FDR = 1.8×10^{-4}) and KEGG N-glycan biosynthesis (4/50; FDR = 0.0015), retained across all 29 locus exclusions. The module was recovered in every fixed-core null set, indicating rediscovery of canonical enzymes. Of 2,271 tested terms, 66 were FDR-significant; their number did not exceed annotation-matched expectation (empirical P = 0.066). Convergence across four of five evidence domains was uncommon under null models (empirical P = 0.001-0.025), with exploratory lymphoid transcriptional and chromatin-regulatory findings. Metabolite associations were not FDR-supported. Conclusions: The synthesis prioritises established IgG glycosylation enzymes and exploratory regulatory candidates using a transparent framework. It does not demonstrate immune-response modulation or causality. Independent experimental and human-cohort validation is required.

Keywords

IgG glycosylation, N-glycan biosynthesis, ST6GAL1, B4GALT1, FUT8, MGAT3, Over-representation analysis, Negative control, Therapeutic antibodies

Article History

Received: 21 August 2026 Revised: 02 November 2026

Accepted: 17 November 2026 Available Online: 24 November 2026

Copyright

© 2026 by the authors. This article is published by the Cultech Publishing Sdn. Bhd. under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0): https://creativecommons.org/licenses/by/4.0/

1. Introduction

Immunoglobulin G (IgG) is the principal effector antibody of the human humoral immune response, and its biological activity is shaped not only by antigen specificity but also by the composition of a single, conserved N-linked glycan attached at asparagine 297 (Asn297) of each Fc CH2 domain. This Fc glycan is not a passive structural appendage: subtle changes in its terminal sugars—core fucose, bisecting N-acetylglucosamine (GlcNAc), galactose and sialic acid—exert decisive control over the affinity of IgG for Fcγ receptors (FcγRs) and complement component C1q, and thereby determine whether an antibody promotes or restrains inflammation [1,2]. Afucosylation enhances FcγRIIIa binding and antibody-dependent cellular cytotoxicity (ADCC) [2,3], galactosylation modulates complement-dependent cytotoxicity (CDC) [4], and Fc sialylation has shown anti-inflammatory activity in defined experimental settings [5].

Because these glycoforms are heritable and disease-associated, genome-wide association studies (GWAS) have sought the genetic determinants of IgG glycosylation. These studies revealed that IgG glycan traits are highly polygenic and that many associated loci are shared (pleiotropic) with inflammatory and autoimmune diseases, implicating both glycosyltransferase genes and a broader regulatory network of transcription factors [6,7]. Network-based reconstructions of the IgG glycosylation pathway have further refined the enzymatic reaction sequence and nominated candidate regulatory nodes [8]. Despite this progress, the translation of long GWAS gene lists into a concise, biologically interpretable and statistically disciplined model of how IgG glycosylation is controlled—and how that control relates to immune phenotypes—remains a persistent challenge. A recent review further places Fc N-glycosylation within B-cell biology, autoimmunity and anti-inflammatory therapy, providing additional biological context for the present synthesis [9].

A common pitfall in functional-genomics interpretation is the database-by-database reporting of enrichment results, in which nominally significant terms from many independent resources are presented as parallel findings without correction for the burden of multiple testing, without distinguishing robust multi-gene modules from fragile single-gene overlaps, and without integrating orthogonal lines of evidence into a unified prioritisation. Such reporting tends to over-state immune “cross-talk” and regulatory complexity while under-emphasising the small number of findings that actually survive rigorous statistical control.

Here we address this gap by re-synthesising a fully documented set of IgG glycosylation-associated genes through an integrative over-representation framework spanning seven evidence layers: GO, KEGG and Reactome pathways, CellMarker cell-type signatures, HMDB metabolite associations, PPI hubs and transcription-factor regulators. Benjamini-Hochberg FDR correction is applied within each resource, single-gene and single-locus overlaps are explicitly flagged, robustness is tested by leave-one-locus-out re-analysis, negative-control gene sets quantify the signal expected from ascertainment alone, and evidence is integrated into a transparent qualitative prioritisation. The analysis specifies the inputs, library versions, outputs and code needed for independent reproducibility. Because the input genes were themselves curated from prior IgG-glycosylation GWAS and pathway/network studies, the analysis is intended to test cross-resource convergence and prioritise candidates rather than to claim de novo pathway discovery or causal gene identification.

The analysis prioritises a compact N-glycan biosynthesis module comprising ST6GAL1, B4GALT1, FUT8 and MGAT3 as the most reproducible signal, and shows with negative controls that this signal is expected from the inclusion of the canonical IgG Fc glycosyltransferases. A secondary layer of lymphoid transcription-factor and chromatin-regulator genes shows cross-resource convergence that is uncommon under null models but remains exploratory, and cell-type and metabolite associations provide biologically plausible but predominantly exploratory context. The resulting model is therefore hypothesis-generating: established biochemical reactions and Fc-effector consequences are drawn from prior literature, whereas causal regulatory links proposed by the multi-source synthesis require independent validation.

2. Materials and Methods

2.1 Gene-Set Curation

The query set was assembled from the three source studies cited for this purpose [6-8] using explicitly defined inclusion criteria. (i) From Lauc et al. [7], every gene listed for the nine loci reaching genome-wide significance ($P < 2.27\times10^{-9}$) in the UPLC discovery meta-analysis was included (15 genes); strongly suggestive loci were excluded. (ii) From Klarić et al. [6], every candidate gene named for the 27 genome-wide significant loci ($P \leq 2.4\times10^{-9}$) was included (44 genes); the source authors had prioritised these candidates using DEPICT gene prioritisation, coding-variant annotation and SMR/HEIDI colocalisation with eQTLs in immune cells and peripheral blood, reporting the nearest gene when none of these applied. For the extended MHC locus, labelled only 'HLA region' in that study's locus table, the four genes used by the same study to label the locus in its network analysis (TCF19, GPANK1, TNXB and HLA-DRA) were used; the six

suggestive loci were excluded. (iii) From Benedetti et al. [8], the four glycosyltransferases whose gene regions were analysed (ST6GAL1, B4GALT1, FUT8 and MGAT3) were included. Symbols were harmonised to current HGNC nomenclature (e.g., SUV420H1→KMT5B), duplicates were collapsed and loci shared between studies were merged. The resulting query set contains 51 unique genes from 29 loci: seven loci were genome-wide significant in both GWAS, 11 loci contributed more than one gene (33 genes in total) and 18 loci contributed a single gene. When a locus supported more than one gene, all named genes were retained and the locus-to-gene assignment was treated as ambiguous rather than causal; every gene carries a locus identifier that is used for single-locus flags and the leave-one-locus-out analysis, so that co-localised genes cannot masquerade as independent evidence. Supplementary Table S1 provides the complete gene-level ledger (source study, locus coordinates, lead variant and P value, mapping strategy and library coverage for every gene). Because known glycogenes are present a priori in these source studies, enrichment of canonical N-glycan pathways is susceptible to ascertainment and circularity; this was quantified with negative-control analyses (Section 2.6), and the analysis is framed as cross-resource evidence synthesis and prioritisation rather than unbiased genome-wide pathway discovery.

2.2 Integrative Over-Representation Analysis

The query set was tested against eight gene-set libraries distributed by Enrichr [10], representing seven evidence layers: (i) GO_Biological_Process_2023 and GO_Molecular_Function_2023; (ii) KEGG_2021_Human; (iii) Reactome_Pathways_2024; (iv) CellMarker_2024; (v) HMDB_Metabolites [11]; (vi) PPI_Hub_Proteins; and (vii) Transcription_Factor_PPIs. Library files were downloaded from the Enrichr gene-set library endpoint on 30 October 2026; their term counts, gene universes and SHA-256 checksums are listed in Supplementary Table S2. For each library, the background universe was defined as all unique genes annotated in that library (from 3,723 genes for HMDB to 16,399 for PPI hubs), and query genes absent from a library were excluded from that library's contingency tables (represented query genes ranged from 10 for HMDB to 40 for GO BP). Three previous symbols used by some libraries (SUV420H1, AZI1 and HOLE for KMT5B, CEP131 and TMEM121) were mapped explicitly. For every term containing at least one query gene, a one-sided Fisher's exact test (hypergeometric upper tail) was computed together with the sample odds ratio (OR; Haldane-Anscombe correction only when a cell was zero); 2,271 term-level tests were performed in total. Analyses were run in Python 3.13 with NumPy 2.5, SciPy 1.18 and pandas 3.0; the analysis code (Supplementary Data S1), the exact query list, per-library universes and the complete tested-term output (Supplementary Data S2) are provided. As a sensitivity analysis, all statistics were recomputed under the Enrichr web-server convention (fixed background of 20,000 genes and the full query size); under that convention, and without symbol mapping, our implementation reproduced Enrichr web-server output exactly (identical numbers of tested terms, p-values, odds ratios and adjusted p-values in all eight libraries).

2.3 Multiple-Testing Correction and Significance Tiering

Within each library, nominal p-values were adjusted using the Benjamini-Hochberg false-discovery-rate (FDR) procedure [12] over all terms with at least one overlapping query gene (the Enrichr reporting convention; GO BP and GO MF were corrected separately). The classification rule was applied consistently throughout the text, tables and figures: FDR <0.05 = statistically significant; FDR 0.05-0.25 = suggestive/exploratory; and FDR >0.25 = not FDR-supported. A term supported by only one overlapping query gene was additionally labelled “single-gene” and treated as fragile, even when its nominal p-value was small; likewise, a term whose overlapping genes all derived from one GWAS locus was labelled “single-locus” and treated as fragile, because co-localised genes do not provide independent evidence. Infection-related or cross-species developmental annotations driven by shared immune genes were retained only as contextual findings and were not used to support the core mechanistic conclusion.

2.4 Evidence-Weighted Gene Prioritisation

Evidence weighting was used as a descriptive prioritisation framework rather than a validated prediction score, and was applied by fixed rules implemented in the deposited code. Per-layer support was coded on an ordinal 0-3 scale: 0 = no overlapping term, or gene not represented in the library; 1 = only single-gene or single-locus suggestive terms, or terms with FDR >0.25; 2 = an FDR-significant single-locus term, or a suggestive (FDR 0.05-0.25) term spanning at least two loci; and 3 = an FDR-significant term spanning at least two loci. Source-study evidence was scored 3 for the four a priori IgG Fc glycosyltransferases or for genes genome-wide significant in both GWAS and named as the sole candidate in at least one, 2 for the sole named candidate of a single-study locus (or a gene explicitly favoured over co-localised genes by the source study), and 1 for one of several genes named at a multi-gene locus. Because GO, KEGG and Reactome are correlated representations of pathway biology, they were displayed separately in the evidence map but were treated as one pathway evidence domain when assigning the final qualitative priority. Priority was assigned as High (source score 3 and FDR-significant multi-locus terms that survived every leave-one-locus-out analysis in at least two of the three

pathway resources), Medium-high (source score ≥2 and score 3 in at least two of the five non-redundant domains: pathway, cell type, metabolite, PPI hub and TF-PPI), Medium (score 3 in at least one domain) or Contextual (otherwise). Priority therefore reflects the breadth of non-redundant evidence and the presence of strong signals rather than an unqualified sum across correlated databases.

2.5 Robustness and Sensitivity Analyses

Four robustness analyses were performed. First, terms whose enrichment depended on a single query gene or a single locus were excluded from the robust interpretive set. Second, a leave-one-locus-out (jackknife) analysis removed all genes of each of the 29 loci in turn, re-ran the complete pipeline including Benjamini-Hochberg correction, and recorded how often each FDR-significant term remained significant. Third, a leave-glycosyltransferases-out analysis removed ST6GAL1, B4GALT1, FUT8 and MGAT3 and re-ran the pipeline on the remaining 47 genes to determine what signal is present beyond the canonical enzymes. Fourth, all statistics were recomputed under the Enrichr background convention (Section 2.2). Because the full tested-term universe is available, adjusted p-values were recalculated in every analysis rather than carried over (Supplementary Table S3 to Supplementary Table S5).

2.6 Negative-Control Analyses

To quantify the signal expected from ascertainment alone, the identical pipeline (same libraries, backgrounds, multiplicity families and thresholds) was applied to 1,000 null gene sets under each of four models (random seed 20261030): N1, size-matched sets drawn uniformly from all 27,418 genes annotated in any library; N2, annotation-matched sets in which each query gene was replaced by a random gene among the 100 genes closest to it in annotation degree (the number of library terms containing the gene), controlling for the tendency of GWAS-nominated genes to be well annotated; N3, glycogene-composition-matched sets containing the same number of glycosyltransferases as the query (five), drawn from a pool of 173 genes annotated to glycosyltransferase activities in GO MF 2023, plus annotation-matched non-glycosyltransferase genes; and N4, a fixed canonical-core model in which ST6GAL1, B4GALT1, FUT8 and MGAT3 were always retained and the other 47 genes were annotation-matched at random. An additional annotation-matched null of 47 genes served as the comparator for the leave-glycosyltransferases-out analysis. For each null set we recorded the number of FDR-significant terms, the number of FDR-significant multi-gene terms, the number of non-redundant evidence domains (pathway, cell type, metabolite, PPI hub and TF-PPI) containing at least one FDR-significant multi-gene term, and whether the core KEGG and Reactome N-glycan terms reached FDR <0.05. Empirical P values were computed as (1 + number of null sets ≥ observed)/(1 + 1,000).

2.7 Mechanistic Annotation and Literature Verification

Prioritised genes and enriched terms were annotated with established enzymatic activities, reactions on the IgG Fc N-glycan, documented antibody-effector consequences and disease associations. These mechanistic annotations were taken from the cited literature and were kept analytically separate from the enrichment results generated in the present study. Thus, ADCC, CDC, FcγR binding, IVIG-related anti-inflammatory activity and the biosynthetic steps catalysed by FUT8, MGAT3, B4GALT1 and ST6GAL1 are interpreted as literature-established context rather than outcomes directly measured by the current computational synthesis. Each cited statement was re-checked against the original abstract in PubMed, and metabolite identities were cross-referenced to PubChem compound identifiers (CIDs) and HMDB accessions [11,13] (Supplementary Table S6).

2.8 Visualisation

All quantitative figures (Figure 3 to Figure 6 and Figure 8) were regenerated directly from the pipeline output by the deposited plotting script, and schematic figures (Figures 1, Figure 2 and Figure 7) were redrawn to distinguish enrichment-derived findings from literature-based interpretation. No statistical results were introduced during figure preparation. Figures are presented in the order discussed in the Results: workflow (Figure 1); literature-informed core N-glycan module (Figure 2); FDR-aware enrichment landscape (Figure 3); CellMarker cell-type enrichment (Figure 4); integrated evidence heatmap (Figure 5); exploratory evidence-integration network (Figure 6); hypothesis-generating genotype-to-glycan-to-function framework (Figure 7); and negative-control analysis (Figure 8). Statistical tiering and single-gene/single-locus flags are applied consistently to all figure interpretations.

3. Results

3.1 Curated Gene Set and Analysis Overview

The documented query set comprised 51 genes from 29 loci: 15 genes from the nine Lauc et al. loci, 44 genes from the 27 Klarić et al. loci and the four glycosyltransferases analysed by Benedetti et al. (Supplementary Table S1). Forty-nine genes were represented in at least one library (NXPE1 and NXPE4 in none), with 10 (HMDB) to 40 (GO BP) query genes represented per library. Across the eight libraries, 2,271 terms were tested and 66 were FDR-significant; all 66 were multi-gene and all but one (a CellMarker B-cell term driven by two genes from the SMARCB1-DERL3-CHCHD10-VPREB3 locus) were multi-locus. Figure 1 summarises the workflow, including locus-to-gene uncertainty, resource-specific enrichment, FDR tiering, robustness checks, negative controls and evidence integration. Throughout the Results, FDR <0.05 is termed significant, FDR 0.05-0.25 suggestive/exploratory, and FDR >0.25 not FDR-supported. Two signals dominated: the multi-gene core N-glycan biosynthesis module and a lymphoid transcription-factor/chromatin-regulator layer, whose interpretation is constrained by the negative-control analysis (Section 3.11).

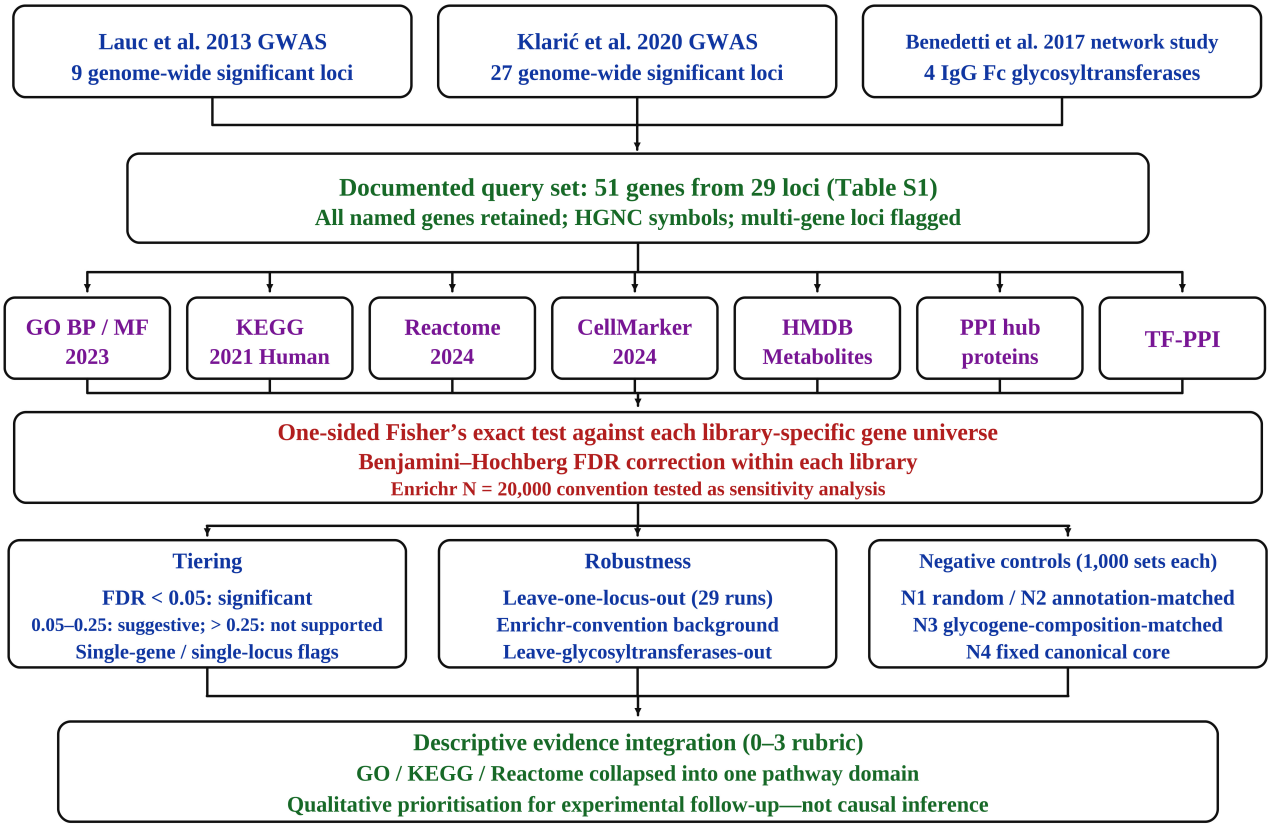


Figure 1. Reproducible multi-source functional-genomics workflow for IgG glycosylation-associated genes. Genes nominated by two GWAS and one network study (51 genes, 29 loci) were tested against eight versioned libraries with library-specific universes and Benjamini-Hochberg correction within each library, followed by single-gene/single-locus flags, leave-one-locus-out and Enrichr-convention sensitivity analyses, four negative-control null models and descriptive evidence integration. The workflow represents evidence synthesis and prioritisation rather than de novo causal discovery.

3.2 A Core Four-Enzyme N-Glycan Biosynthesis Module is Consistently and Significantly Enriched

The most robust result was the recurrent enrichment of ST6GAL1, B4GALT1, FUT8 and MGAT3 across independent pathway resources. Reactome N-glycan antennae elongation in the medial trans-Golgi (4/26 genes; OR = 75.4; FDR = 1.8×10^{-4}), KEGG N-glycan biosynthesis (4/50; OR = 31.6; FDR = 0.0015) and GO protein N-linked glycosylation (6/64, additionally including FUT6 and DERL3; FDR = 9.8×10^{-6}) were all FDR-significant and remained significant in all 29 leave-one-locus-out analyses. Related glycosylation terms—oligosaccharide biosynthesis, Golgi transport and modification, and fucosyltransferase activity (FUT6, FUT8)—were also significant, whereas KEGG various types of N-glycan biosynthesis and other types of O-glycan biosynthesis were suggestive (Table 1). These results reproduce, in a fully documented analysis, the canonical enzymes of the Fc N-glycan pathway; as shown in Section 3.11, their significance follows directly from the inclusion of these four genes in the source nominations.

Table 1. Glycosylation-module enrichment across GO, KEGG and Reactome (primary analysis; library-specific gene universes).

Pathway (database)	Overlap	Genes	OR	Nominal p	FDR	Tier / flag	LOO retained
N-glycan antennae elongation in the medial trans-Golgi (Reactome)	4/26	B4GALT1; FUT8; MGAT3; ST6GAL1	75.4	6.7×10^{-7}	1.8×10^{-4}	Significant	29/29
Protein N-linked glycosylation (GO:0006487)	6/64	B4GALT1; DERL3; FUT6; FUT8; MGAT3; ST6GAL1	44.4	1.8×10^{-8}	9.8×10^{-6}	Significant	29/29
N-glycan biosynthesis (KEGG)	4/50	B4GALT1; FUT8; MGAT3; ST6GAL1	31.6	1.8×10^{-5}	0.0015	Significant	29/29
Oligosaccharide biosynthetic process (GO:0009312)	3/22	B4GALT1; FUT6; FUT8	62.5	2.8×10^{-5}	0.0041	Significant	29/29
Protein N-linked glycosylation via asparagine (GO:0018279)	3/23	DERL3; FUT8; ST6GAL1	59.3	3.2×10^{-5}	0.0041	Significant	29/29
Protein glycosylation (GO:0006486)	5/147	B4GALT1; COG7; FUT6; FUT8; MGAT3	14.6	4.6×10^{-5}	0.0041	Significant	29/29
Transport to the Golgi and subsequent modification (Reactome)	5/186	B4GALT1; COG7; FUT8; MGAT3; ST6GAL1	11.7	1.4×10^{-4}	0.012	Significant	26/29
N-glycan antennae elongation (Reactome)	2/15	B4GALT1; ST6GAL1	59.6	7.5×10^{-4}	0.028	Significant	26/29
Fucosyltransferase activity (GO:0008417)	2/11	FUT6; FUT8	86.6	3.9×10^{-4}	0.040	Significant	27/29
Asparagine N-linked glycosylation (Reactome)	5/306	B4GALT1; COG7; FUT8; MGAT3; ST6GAL1	7.0	0.0014	0.045	Significant	23/29
Various types of N-glycan biosynthesis (KEGG)	2/39	B4GALT1; FUT8	18.1	0.0069	0.138	Suggestive	—
Other types of O-glycan biosynthesis (KEGG)	2/47	B4GALT1; ST6GAL1	14.8	0.0099	0.138	Suggestive	—
Galactosyltransferase activity (GO:0008378)	1/10	B4GALT1	42.0	0.027	0.102	Suggestive; single-gene	—

Note: Overlap = query genes in term/term size. FDR <0.05 is significant; FDR 0.05-0.25 is suggestive/exploratory; FDR >0.25 is not FDR-supported. Single-gene and single-locus terms are flagged as fragile. LOO retained = number of the 29 leave-one-locus-out analyses in which an FDR-significant term remained significant.

Table 2 separates the present-study statistical evidence from established biochemical and antibody-functional knowledge. The four prioritised enzymes act on the conserved Fc N-glycan: FUT8 adds core α1,6-fucose, MGAT3 adds bisecting GlcNAc, B4GALT1 catalyses galactosylation and ST6GAL1 catalyses terminal α2,6-sialylation. These reactions and their associations with Fc effector behaviour are literature-established context rather than causal effects demonstrated by the current analysis. Figure 2 illustrates this distinction using solid arrows for biosynthetic steps and dashed links for literature-based functional interpretation.

Table 2. Annotation of the enrichment-prioritised ST6GAL1-B4GALT1-FUT8-MGAT3 module, separating evidence generated by the present analysis from literature-established enzymatic and antibody-functional consequences. Literature statements were re-checked against the original abstracts; bisecting GlcNAc is described as acting indirectly, through inhibition of core fucosylation, and not as altering Fcγ-receptor binding directly [14].

Gene	Enzyme / activity	Reaction on IgG Fc N-glycan	Literature-established functional consequence (re-checked against original abstracts)	Present-analysis evidence / interpretation
FUT8	α-1,6-fucosyltransferase 8	Adds core α1,6-fucose to the innermost GlcNAc (donor GDP-fucose)	Absence of core fucose increases FcγRIIIa affinity up to about 50-fold and enhances ADCC, through carbohydrate-carbohydrate contacts with the receptor glycan [2,3].	Member of all three jackknife-robust core terms (Reactome, KEGG, GO) and of fucosyltransferase activity (with FUT6). ADCC was not measured.
MGAT3	β-1,4-mannosyl-glycoprotein 4-β-N-acetylglucosaminyltransferase (GnT-III)	Adds bisecting GlcNAc to the β-mannose of the trimannosyl core (donor UDP-GlcNAc)	Bisecting GlcNAc suppresses subsequent terminal modification of N-glycans [15], and a bisected core cannot be core-fucosylated by FUT8. In homogeneous trastuzumab glycoforms, bisecting GlcNAc did not alter binding to any Fcγ receptor; the enhanced ADCC of antibodies from MGAT3-overexpressing cells is attributed to reduced core fucosylation, i.e. an indirect effect [14]. Earlier *in vitro* GnT-III remodelling reported increased ADCC at high bisecting-GlcNAc content [4].	Member of all three jackknife-robust core terms. FcγR binding and ADCC were not measured.
B4GALT1	β-1,4-galactosyltransferase 1	Transfers galactose to antennary GlcNAc (donor UDP-galactose)	Degalactosylation of a therapeutic IgG1 reduced CDC by about half [4]; low IgG galactosylation is present years before a diagnosis of rheumatoid arthritis [16].	Member of all three core terms; suggestive UDP association in HMDB. CDC was not measured.
ST6GAL1	β-galactoside α-2,6-sialyltransferase 1	Adds terminal α2,6-linked sialic acid to galactose (donor CMP-Neu5Ac)	Fc sialylation confers anti-inflammatory activity on IgG [5], with subclass-specific Fc residues and DC-SIGN/SIGN-R1 dependence [17]; circulating ST6GAL1 can sialylate IgG independently of B cells and modulates B-cell immunoglobulin production [18,19].	Member of all three core terms. Anti-inflammatory activity was not measured.

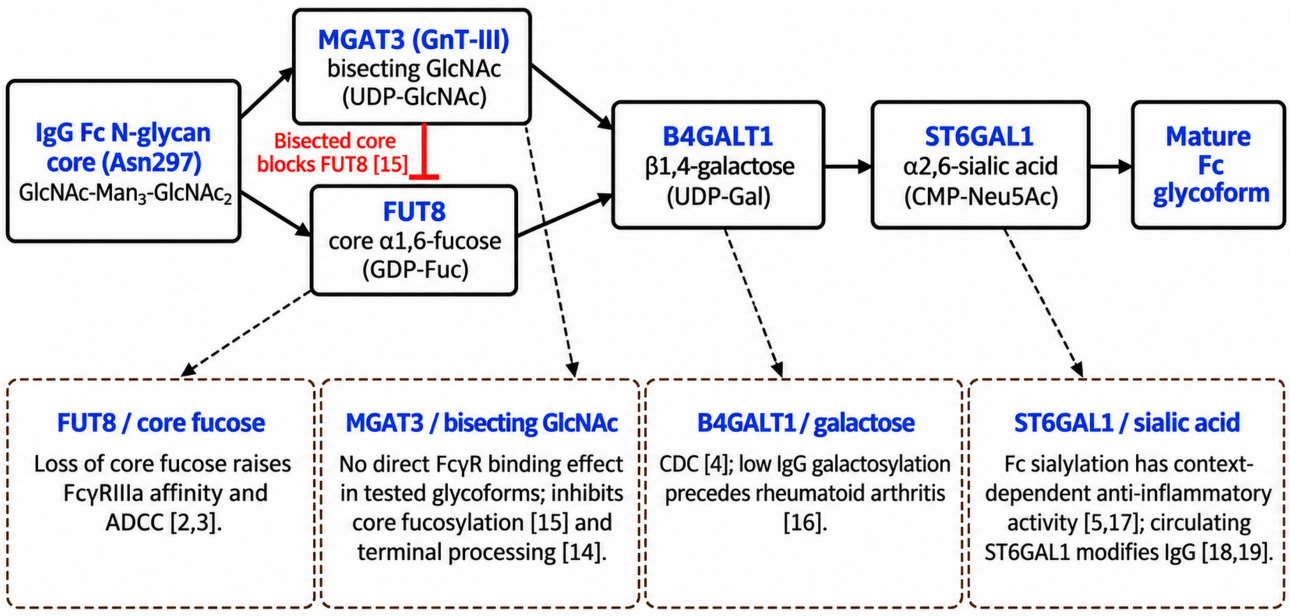
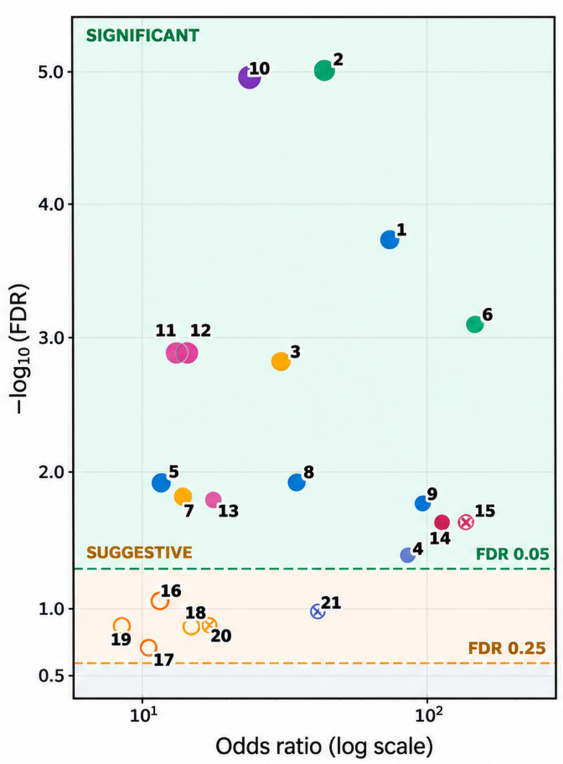


Figure 2. Literature-informed core N-glycan biosynthesis module prioritised by the present enrichment analysis. Solid arrows indicate established biosynthetic steps (order simplified); the red bar indicates that a bisected core is not a substrate for FUT8 [14]. Dashed links indicate literature-based functional consequences that were not measured in the present computational study. Nucleotide-sugar donors are shown in parentheses (PubChem CIDs in Supplementary Table S6).

3.3 Pathway Enrichment Landscape

Figure 3 provides an FDR-aware enrichment landscape across representative pathway, cell-type, metabolite and regulatory terms. The core N-glycan terms and GO protein N-linked glycosylation occupy the FDR-significant region together with a group of regulatory terms (the HDAC2 PPI hub, NOTCH1 and SMARCA4 TF-PPIs, RUNX1/RUNX3 Reactome terms and T-cell differentiation terms). Metabolite, Th1/Th2 and inflammatory-bowel-disease terms fall in the suggestive band, and single-gene or single-locus terms are flagged to prevent visual over-interpretation.



TERM KEY 21 representative terms

#	Source · term (overlap/size)		OR	FDR
1	Reactome	N-glycan antennae elongation, medial trans-Golgi (4/26)	75.4	1.8e-04
2	GO BP	Protein N-linked glycosylation (6/64)	44.4	9.8e-06
3	KEGG	N-glycan biosynthesis (4/50)	31.6	0.0015
4	GO MF	Fucosyltransferase activity (2/11)	86.6	0.040
5	Reactome	Transport to Golgi & subsequent modification (5/186)	11.7	0.012
6	GO BP	Regulation of CD4+ alpha-beta T-cell differentiation (3/11)	148.5	8.1e-04
7	KEGG	Th17 cell differentiation (4/107)	14	0.015
8	Reactome	RUNX1 interacts with co-factors (3/37)	35.3	0.012
9	Reactome	RUNX3 regulates p14-ARF (2/10)	96.9	0.017
10	PPI hub	HDAC2 (7/214)	24.8	1.1e-05
11	TF-PPI	NOTCH1 (6/204)	14.6	0.0013
12	TF-PPI	SMARCA4 (6/219)	13.5	0.0013
13	TF-PPI	GATA1 (3/69)	17.9	0.016
14	CellMarker	Effector T cell, blood (human) (2/8)	113.5	0.023
15	CellMarker	B cell, colorectum (human) (2/7)	136.2	0.023
16	KEGG	Th1 and Th2 cell differentiation (3/92)	11.7	0.086
17	KEGG	Inflammatory bowel disease (2/65)	10.6	0.192
18	HMDB	Guanosine diphosphate (GDP) (2/63)	15	0.133
19	HMDB	Uridine 5'-diphosphate (UDP) (2/107)	8.6	0.133
20	HMDB	UDP-GlcNAc (1/25)	17.1	0.133
21	GO MF	Galactosyltransferase activity (1/10)	42	0.102

Source: revised ACPT-60 manuscript, Tables 1–6 and original Figure 3.
FDR: Benjamini-Hochberg within each library; x-axis on logarithmic scale.

MARKER GUIDE ● FDR < 0.05 ○ FDR 0.05–0.25 ✖ Single-gene / single-locus
Bubble area ∝ number of overlapping genes; resource identity is coded by colour.

Figure 3. FDR-aware enrichment landscape across representative pathway, cell-type, metabolite and regulatory terms (primary analysis). Bubble area reflects the number of overlapping query genes; the odds-ratio axis is logarithmic; dashed horizontal lines correspond to FDR 0.05 and 0.25; numbered labels correspond to the term key; and X-shaped markers denote single-gene or single-locus terms.

3.4 Immune and Inflammatory Pathways: Exploratory T-Cell and Disease-Pleiotropy Context

Immune-pathway results were driven by a small number of lymphoid transcription-factor, MHC class II and cytokine-receptor genes (Table 3). GO regulation of CD4+ αβ T-cell differentiation (HLA-DRA, RUNX1, RUNX3; FDR = 8.1×10^{-4}) was significant and retained in all leave-one-locus-out analyses, whereas KEGG Th17 cell differentiation (HLA-DRA, IL6ST, RUNX1, TBX21; FDR = 0.015) and several two-gene RUNX1/RUNX3 T-cell terms lost significance when a single contributing locus was removed. Th1/Th2 differentiation, inflammatory bowel disease and antigen processing and presentation were suggestive only. These terms reflect the known pleiotropy of IgG-glycosylation loci with inflammatory and autoimmune disease [6,7] and the lymphoid identity of several nominated transcription factors; they do not demonstrate that the present analysis detected immune-pathway regulation of IgG glycan assembly, and immune-response modulation was not tested.

Table 3. Immune/inflammatory enrichment and contextual associations (primary analysis). The tier column distinguishes significant, suggestive and fragile results; LOO = number of 29 leave-one-locus-out analyses in which an FDR-significant term remained significant. Interpretation is literature-linked context and not a demonstration of immune modulation.

Pathway / term	Source	Overlap	Genes	OR	Nominal p	FDR	Tier	LOO	Interpretation / literature context
Regulation of CD4+ αβ T-cell differentiation (GO:0043370)	GO BP	3/11	HLA-DRA; RUNX1; RUNX3	148.5	3.0×10^{-6}	8.1×10^{-4}	Significant	29/29	RUNX-family T-cell lineage regulation; does not show control of IgG glycan assembly [20]
Regulation of CD8+ αβ T-cell differentiation (GO:0043376)	GO BP	2/7	RUNX1; RUNX3	154.2	1.5×10^{-4}	0.0071	Significant	27/29	Two-gene RUNX1/RUNX3 signal; lost when either locus is removed
Positive regulation of T-cell activation (GO:0050870)	GO BP	4/107	HLA-DRA; IL6ST; SMARCB1; SMARCD3	15.7	2.0×10^{-4}	0.0071	Significant	26/29	Immune context; four loci
Regulation of adaptive immune response (GO:0002819)	GO BP	2/18	IL6ST; IRF1	48.2	0.0011	0.024	Significant	27/29	Cytokine (IL-6 family) and interferon-regulatory context
Th17 cell differentiation (KEGG)	KEGG	4/107	HLA-DRA; IL6ST; RUNX1; TBX21	14.0	3.5×10^{-4}	0.015	Significant	25/29	Shared with inflammatory-disease loci [6]; not robust to removal of single loci
Th1 and Th2 cell differentiation (KEGG)	KEGG	3/92	HLA-DRA; RUNX3; TBX21	11.7	0.0031	0.086	Suggestive	—	Exploratory T-helper context
Inflammatory bowel disease (KEGG)	KEGG	2/65	HLA-DRA; TBX21	10.6	0.018	0.192	Suggestive	—	Literature-linked pleiotropy with inflammatory disease [6,7]
Antigen processing and presentation (KEGG)	KEGG	2/78	HLA-DRA; RFXANK	8.7	0.026	0.200	Suggestive	—	MHC class II context (HLA-DRA, RFXANK)

3.5 Cell-Type Analysis Points to Immune Contexts, with Caveats

Six CellMarker terms were FDR-significant, each with two to eight overlapping genes (Table 4; Figure 4). The most consistent were human effector T cells (RUNX3, TBX21; FDR = 0.023) and B-cell signatures from human and mouse kidney (FDR = 0.030), the human signature including ST6GAL1 and FUT8 within a large 667-gene marker set. The human colorectal B-cell term (FDR = 0.023) was driven by DERL3 and VPREB3 from a single locus and is therefore treated as fragile, and developmental or cross-species terms (e.g., embryonic mouse neurons via RUNX1/RUNX3) are retained only as annotation background. No cell-type term was retained in all leave-one-locus-out analyses, and cell-type enrichment does not identify the cells in which IgG is glycosylated.

Table 4. CellMarker 2024 enrichment with species, overlap structure and evidence status. Six terms are FDR-significant; one (colorectal B cell) is single-locus and fragile, and no cell-type term is robust to removal of every individual locus. LOO as in Table 3.

Cell type (species)	Overlap	Genes	OR	Nominal p	FDR	Evidence status	LOO	Interpretation
Effector T cell, blood (Human)	2/8	RUNX3; TBX21	113.5	2.6×10^{-4}	0.023	Significant	27/29	T-cell transcription factors (T-bet/RUNX3); not a site of IgG synthesis
B cell, colorectum (Human)	2/7	DERL3; VPREB3	136.2	1.9×10^{-4}	0.023	Significant; single-locus	28/29	Both genes from the SMARCB1-DERL3-CHCHD10-VPREB3 locus; fragile
B cell, kidney (Mouse)	3/60	DERL3; TBX21; VPREB3	18.3	8.2×10^{-4}	0.030	Significant	27/29	Cross-species B-cell context
Myeloid cell, skin (Human)	2/14	HLA-DRA; TBX21	56.7	8.2×10^{-4}	0.030	Significant	26/29	Antigen-presentation/myeloid context
B cell, kidney (Human)	8/667	DERL3; FUT8; GGA2; HLA-DRA; SMARCB1; ST6GAL1; TCF19; VPREB3	4.7	8.3×10^{-4}	0.030	Significant	23/29	Large marker set (667 genes); B-cell context including ST6GAL1 and FUT8
Neuron, embryo (Mouse)	2/18	RUNX1; RUNX3	42.5	0.0014	0.041	Significant	25/29	Developmental RUNX annotation; retained as annotation background
M1 macrophage, lung (Human)	2/23	HLA-DRA; IRF1	32.4	0.0023	0.058	Suggestive	—	Interferon/antigen-presentation context
B cell, bone marrow (Human)	2/38	HLA-DRA; VPREB3	18.9	0.0061	0.113	Suggestive	—	B-lineage context (VPREB3, HLA-DRA)
Antibody-secreting cell, peripheral blood (Human)	1/5	IL6ST	82.9	0.015	0.113	Suggestive; single-gene	—	Biologically relevant but single-gene (IL6ST)

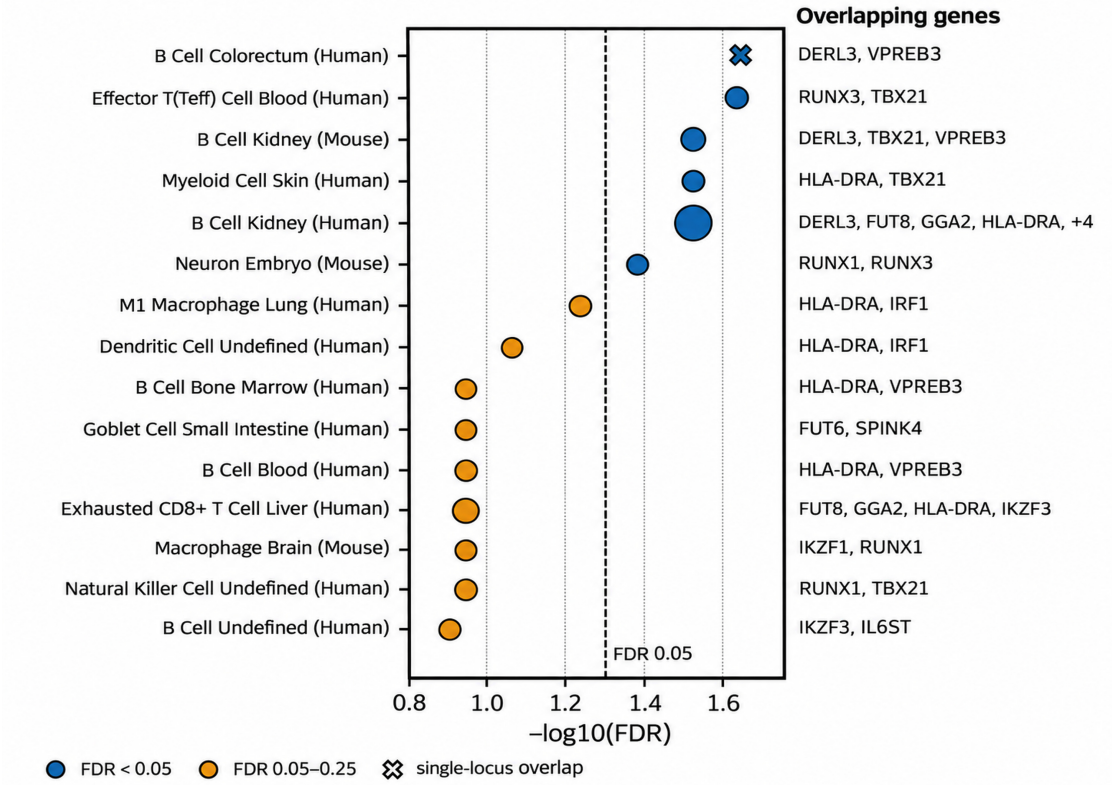


Figure 4. Cell-type marker enrichment (CellMarker 2024; multi-gene terms with FDR ≤0.15). Marker size indicates the number of overlapping query genes, colour the FDR tier, and the X marker a single-locus overlap; overlapping genes are listed on the right. The figure is descriptive and does not imply that enriched cell types are sites of IgG synthesis.

3.6 Metabolite Associations are not FDR-Supported

Only 10 query genes were represented in the HMDB library, and no metabolite term reached FDR <0.05 (Table 5). The two multi-gene associations were guanosine diphosphate (FUT6, FUT8) and uridine 5′-diphosphate (B4GALT1, MGAT3), the nucleotides released from the GDP-fucose, UDP-galactose and UDP-GlcNAc donors (FDR = 0.133); CMP (ST6GAL1), UDP-GlcNAc (MGAT3), N-acetyl-D-glucosamine and D-glucose (B4GALT1) were single-gene associations. The nucleotide-sugar donors themselves (GDP-fucose, UDP-galactose and CMP-Neu5Ac) are not represented as terms in this library, so the layer cannot test donor availability directly. Metabolite identities were verified against PubChem and HMDB [11,13] (Table 5; Supplementary Table S6). UDP-GlcNAc and related hexosamine-pathway metabolites have established roles in glycan synthesis [21,22], but the analysis does not show that substrate availability regulates IgG glycosylation, and the metabolite layer is interpreted strictly as exploratory context.

Table 5. Glycosylation-relevant metabolite associations (HMDB_Metabolites) with verified HMDB accessions and PubChem compound identifiers. No metabolite term is FDR-significant; single-gene overlaps are identified, and biochemical roles are literature-derived [21,22] rather than demonstrated here. Donor nucleotide sugars and their identifiers are listed in Supplementary Table S6.

Metabolite	HMDB ID	PubChem CID	Genes	Role in glycosylation	OR	Nominal p	FDR	Evidence status
Guanosine diphosphate (GDP)	HMDB0001201	135398619	FUT6; FUT8	Nucleotide released from the GDP-fucose donor used by FUT8/FUT6	15.0	0.012	0.133	Suggestive
Uridine 5′-diphosphate (UDP)	HMDB0000295	6031	B4GALT1; MGAT3	Nucleotide released from UDP-Gal (B4GALT1) and UDP-GlcNAc (MGAT3) donors	8.6	0.032	0.133	Suggestive
Cytidine 5′-monophosphate (CMP; 'C5P')	HMDB0000095	6131	ST6GAL1	Nucleotide released from the CMP-Neu5Ac donor used by ST6GAL1	15.2	0.073	0.133	Suggestive; single-gene
N-Acetyl-D-glucosamine (GlcNAc)	HMDB0000215	439174	B4GALT1	Acceptor/salvage precursor of UDP-GlcNAc	22.8	0.050	0.133	Suggestive; single-gene
UDP-N-acetylglucosamine (UDP-GlcNAc)	HMDB0000290	445675	MGAT3	Hexosamine-pathway end product; GlcNAc donor for MGAT3	17.1	0.065	0.133	Suggestive; single-gene
D-Glucose	HMDB0000122	5793	B4GALT1	Upstream carbon source for nucleotide-sugar synthesis	16.4	0.068	0.133	Suggestive; single-gene
L-Carnitine	HMDB0000062	10917	SLC22A4	SLC22A4 (OCTN1) transporter substrate; not glycosylation-related	37.4	0.032	0.133	Suggestive; single-gene

3.7 Regulatory Networks: Transcription Factors and Chromatin Remodelers (Exploratory)

Transcriptional and chromatin-regulatory terms formed the largest group of significant results outside the core module (Table 6). The HDAC2 hub was significant in both the PPI-hub (7/214; FDR = 1.1×10^{-5}) and TF-PPI (7/387; FDR = 0.0022) libraries, linking IKZF1, IKZF3, RUNX1, RUNX3, SMARCB1, RFXANK and TCF19; NOTCH1 and SMARCA4 (BRG1) TF-PPI terms were also significant and, like HDAC2, retained in all leave-one-locus-out analyses. SIN3A, HDAC1, HDAC4, SMARCC1, GATA1, CBFB and the Reactome RUNX1 co-factor and RUNX3-p14-ARF terms were significant but lost significance when single contributing loci (most often RUNX1, IKZF1 or SMARCB1) were removed. These associations are concordant with the source study's proposal that IKZF1 regulates FUT8 and that RUNX1/RUNX3 together with SMARCB1 regulate MGAT3 [6], but they arise from protein-protein-interaction and annotation co-membership among transcription factors that were themselves nominated partly by network-based prioritisation; they do not demonstrate that any of these regulators controls ST6GAL1, B4GALT1, FUT8 or MGAT3 expression.

Table 6. Transcriptional/chromatin-regulatory enrichment (primary analysis). HDAC2 (both libraries), NOTCH1 and SMARCA4 are significant and robust to removal of any single locus; the remaining regulatory terms are significant but locus-dependent. All are exploratory and do not establish regulation of the glycosyltransferases. LOO as in Table 3.

Regulator / term	Library	Overlap	Genes	OR	Nominal p	FDR	Tier	LOO	Interpretation
HDAC2	PPI hub	7/214	IKZF1; IKZF3; RFXANK; RUNX1; RUNX3; SMARCB1; TCF19	24.8	7.1×10^{-8}	1.1×10^{-5}	Significant	29/29	Histone-deacetylase hub linking IKZF1/IKZF3, RUNX1/RUNX3, SMARCB1 [23]
HDAC4	PPI hub	4/127	IKZF1; IKZF3; RFXANK; RUNX3	21.1	7.0×10^{-5}	0.0055	Significant	28/29	Class IIa HDAC context
SIN3A	PPI hub	4/183	IKZF1; IKZF3; RUNX1; SMARCB1	14.5	2.9×10^{-4}	0.012	Significant	25/29	Co-repressor scaffold
HDAC1	PPI hub	5/346	IKZF1; RUNX1; RUNX3; SMARCB1; TCF19	9.8	3.2×10^{-4}	0.012	Significant	26/29	NuRD/SIN3 context
NOTCH1	TF-PPI	6/204	IKZF1; IKZF3; RUNX1; RUNX3; SMARCD3; TAB1	14.6	1.9×10^{-5}	0.0013	Significant	29/29	Lymphocyte fate-decision context
SMARCA4 (BRG1)	TF-PPI	6/219	IKZF1; IRF1; RUNX1; SMARCB1; SMARCD3; TBX21	13.5	2.8×10^{-5}	0.0013	Significant	29/29	SWI/SNF chromatin remodelling (SMARCB1, SMARCD3 subunits)
HDAC2	TF-PPI	7/387	IKZF1; IKZF3; RFXANK; RUNX1; RUNX3; SMARCB1; TCF19	9.4	7.5×10^{-5}	0.0022	Significant	29/29	Concordant with PPI-hub layer
SMARCC1	TF-PPI	4/114	IKZF1; RUNX1; SMARCB1; SMARCD3	15.3	3.1×10^{-4}	0.0069	Significant	26/29	SWI/SNF core subunit
GATA1	TF-PPI	3/69	IKZF1; RUNX1; SMARCB1	17.9	0.0010	0.016	Significant	26/29	Haematopoietic TF context [24]
RUNX1 interacts with co-factors (Reactome)	Reactome	3/37	RUNX1; SMARCB1; SMARCD3	35.3	1.4×10^{-4}	0.012	Significant	26/29	Consistent with proposed RUNX1/SMARCB1 regulation of MGAT3 [6]; not tested here
RUNX3 regulates p14-ARF (Reactome)	Reactome	2/10	RUNX1; RUNX3	96.9	3.2×10^{-4}	0.017	Significant	27/29	Two-gene RUNX1/RUNX3 overlap; no RUNX→glycogene regulation demonstrated [20,25]
CBFB	TF-PPI	2/34	RUNX1; RUNX3	23.2	0.0045	0.045	Significant	25/29	RUNX heterodimer partner; two-gene

3.8 Integrated, Evidence-Weighted Gene Prioritisation

Table 7 and Figure 5 summarise the rule-based, descriptive prioritisation (scores for all 51 genes in Supplementary Table S7). ST6GAL1, B4GALT1, FUT8 and MGAT3 were the only genes classified as High priority: each has the strongest source-study support and belongs to FDR-significant multi-locus terms that survived every leave-one-locus-out analysis in all three pathway resources. RUNX1, RUNX3, IKZF1, SMARCB1, TBX21 and RFXANK were classified as Medium-high because they reached score 3 in at least two non-redundant evidence domains, chiefly the PPI-hub and TF-PPI layers, which are themselves biased towards well-studied transcription factors. Genes at ambiguous multi-gene loci (e.g., IKZF3,

SMARCD3, IRF1, HLA-DRA, IL6ST and DERL3) were capped at Medium (17 genes in total), and 24 genes had no domain scoring 3 (Contextual). The ordinal scores in Figure 5 are descriptive evidence categories and must not be interpreted as probabilities or calibrated causal scores.

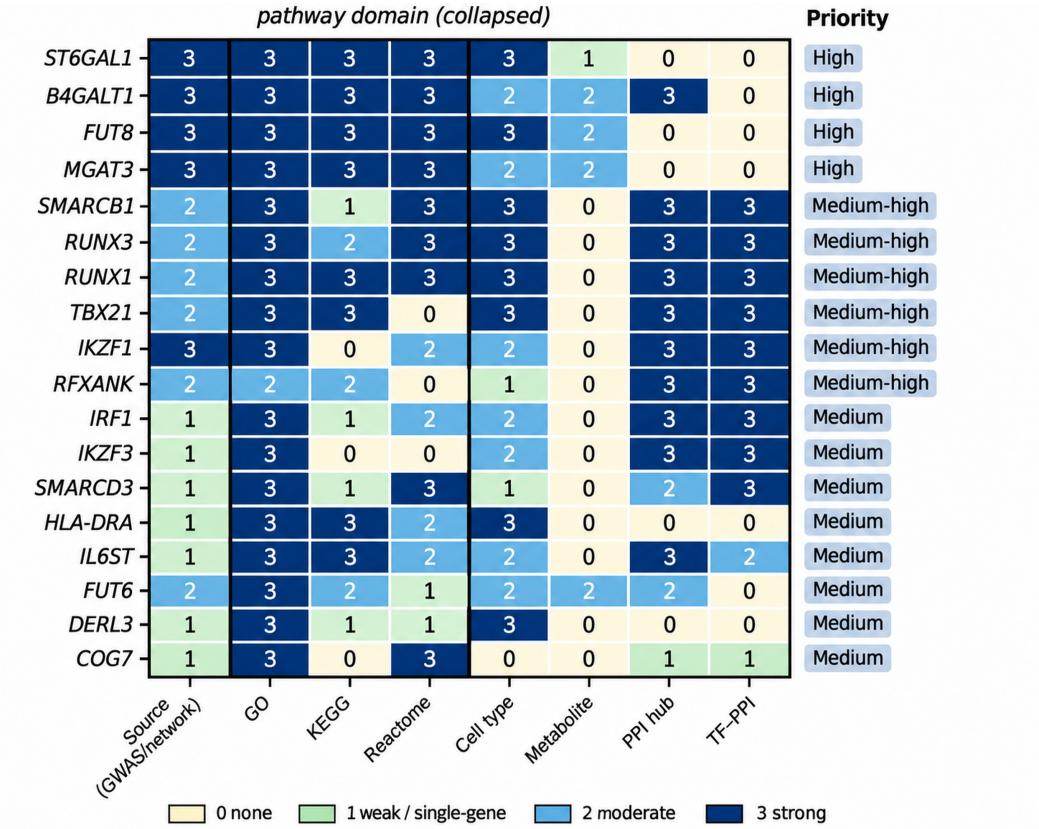


Figure 5. Integrated evidence map for prioritised IgG glycosylation-associated genes (18 of 51 genes shown; all genes in Supplementary Table S7). Scores 0-3 follow the pre-specified rubric (Section 2.4); GO, KEGG and Reactome are displayed separately for transparency but were collapsed into one pathway domain for the priority class shown on the right.

Table 7. Integrated evidence-weighted prioritisation of selected candidate genes (rule-based; Section 2.4). Values in parentheses give leave-one-locus-out retention. Priority categories are qualitative and reflect source-study support together with the breadth of non-redundant, robust evidence; they are not statistically calibrated prediction scores. Scores for all 51 genes are given in Supplementary Table S7.

Gene	Primary role	Source score	FDR-supported evidence (LOO retention)	Exploratory / contextual evidence	Literature context	Priority / interpretation
ST6GAL1	α2,6-sialyltransferase	3	Core Reactome, KEGG and GO N-glycan terms (29/29)	CellMarker B cell, kidney; HMDB CMP (single-gene)	Anti-inflammatory sialylated IgG; extracellular sialylation [5,17-19]	High; validation candidate, not a proven causal driver
B4GALT1	β1,4-galactosyltransferase	3	Core N-glycan terms (29/29)	HMDB UDP (suggestive); TUBB PPI hub	Galactosylation, complement and RA risk [4,16]	High; validation candidate
FUT8	α1,6-fucosyltransferase	3	Core N-glycan terms (29/29); fucosyltransferase activity	HMDB GDP (suggestive); CellMarker B cell, kidney	Core fucose controls FcγRIIIa binding and ADCC [2,3]; IKZF1-dependent expression proposed [6]	High; validation candidate
MGAT3	Bisecting GlcNAc transferase	3	Core N-glycan terms (29/29)	HMDB UDP and UDP-GlcNAc (exploratory)	Indirect effect through inhibition of core fucosylation [14,15]	High; validation candidate

Gene	Primary role	Source score	FDR-supported evidence (LOO retention)	Exploratory / contextual evidence	Literature context	Priority / interpretation
RUNX1	Runt-related TF	2	HDAC2, NOTCH1, SMARCA4 (29/29); GO CD4+ T-cell differentiation (29/29)	Reactome RUNX terms; KEGG Th17 (locus-dependent)	Proposed MGAT3 regulation with SMARCB1 [6]	Medium-high; exploratory regulator
RUNX3	Runt-related TF	2	HDAC2, NOTCH1 (29/29); GO CD4+ T-cell differentiation (29/29)	CellMarker effector T cell; RUNX3-p14-ARF	T-cell and dendritic-cell lineage regulation [20,25]; proposed MGAT3 regulation [6]	Medium-high; exploratory regulator
IKZF1	Ikaros lymphoid TF	3	HDAC2, NOTCH1, SMARCA4 (29/29)	SIN3A, GATA1, HDAC1/HDAC4 (locus-dependent)	Ikaros recruits NuRD/HDAC complexes [23]; knock-down lowers FUT8 *in vitro* [6]	Medium-high; exploratory regulator
SMARCB1	SWI/SNF core subunit	2	HDAC2, SMARCA4 (29/29); GO regulation of cell differentiation (29/29)	RUNX1 co-factors; SMARCC1/SMARCC2	Favoured over DERL3 at the MGAT3-correlated locus [6]	Medium-high; exploratory regulator
TBX21	T-bet TF	2	SMARCA4 (29/29)	KEGG Th17; CellMarker effector T cell; Th1/Th2 (suggestive)	Th1 lineage context	Medium-high; exploratory
RFXANK	MHC class II transactivator subunit	2	HDAC2 in PPI-hub and TF-PPI layers (29/29)	KEGG antigen processing (suggestive)	MHC class II regulation	Medium-high; exploratory
IKZF3	Aiolos lymphoid TF	1	HDAC2, NOTCH1 (29/29)	SIN3A, HDAC4 (locus-dependent)	Multi-gene ORMDL3-GSDMB-IKZF3-ZPBP2 locus	Medium (ambiguous locus)
IRF1	Interferon-regulatory TF	1	SMARCA4 (29/29)	Adaptive immune response; CellMarker macrophage/dendritic	Multi-gene IRF1-SLC22A4 locus	Medium (ambiguous locus)
HLA-DRA	MHC class II α chain	1	GO CD4+ T-cell differentiation (29/29)	KEGG Th17; antigen presentation (suggestive)	Extended MHC; locus-to-gene highly ambiguous	Medium (ambiguous locus)
FUT6	α1,3-fucosyltransferase	2	GO protein N-linked glycosylation (29/29); fucosyltransferase activity	HMDB GDP (suggestive)	Not part of the IgG Fc biantennary pathway	Medium; contextual glycogene

3.9 Mechanistic Synthesis

Integrating the evidence layers supports a hypothesis-generating framework in which the ST6GAL1-B4GALT1-FUT8-MGAT3 module represents the principal glycosylation signal, while lymphoid transcription-factor/chromatin, cell-type and metabolite associations provide secondary context. Figure 6 displays the exploratory evidence-integration network; edges connect genes that co-occur in at least one FDR-significant multi-locus term, with width proportional to the number of shared terms, and should not be interpreted as direct molecular interactions. The network separates into a glycosylation

cluster and a lymphoid-regulatory cluster that are joined only through broad annotation terms such as large B-cell marker sets. Figure 7 separates established biochemical/functional directions from annotation- and literature-supported hypotheses using solid and dashed arrows, respectively.

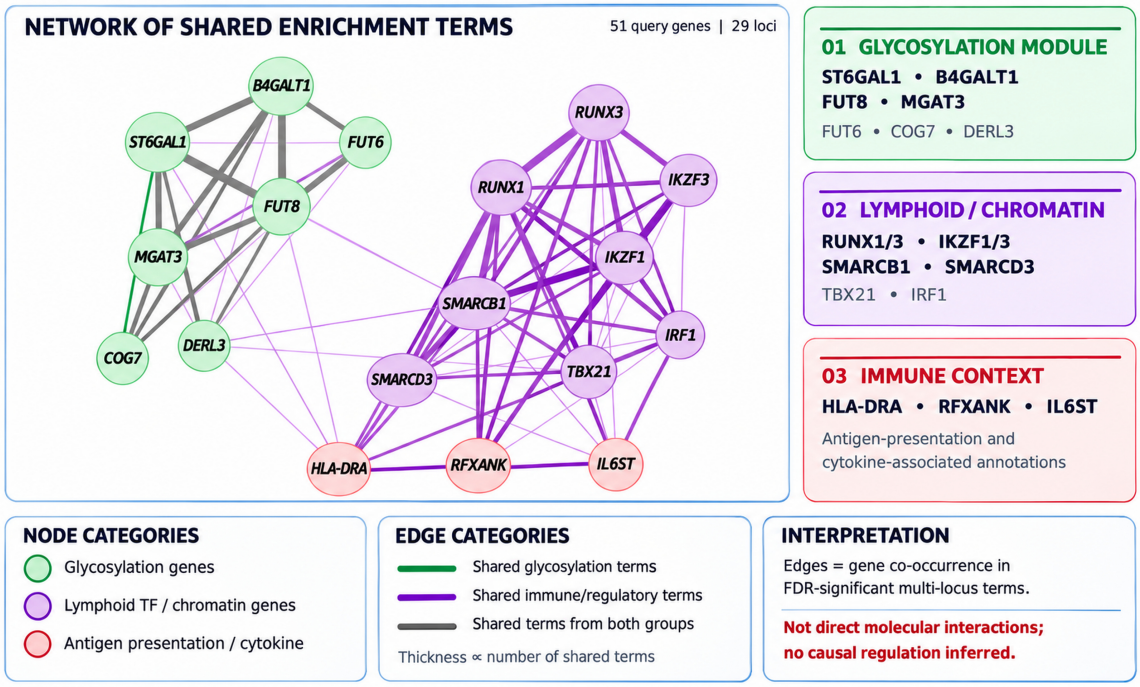


Figure 6. Exploratory evidence-integration network of prioritised IgG glycosylation-associated genes. Edges join genes that co-occur in at least one FDR-significant multi-locus term (green, glycosylation terms; purple, immune/regulatory terms; grey, both); edge width is proportional to the number of shared terms. Edges do not represent direct molecular interactions, and no regulation of ST6GAL1, B4GALT1, FUT8 or MGAT3 by the regulatory genes is inferred.

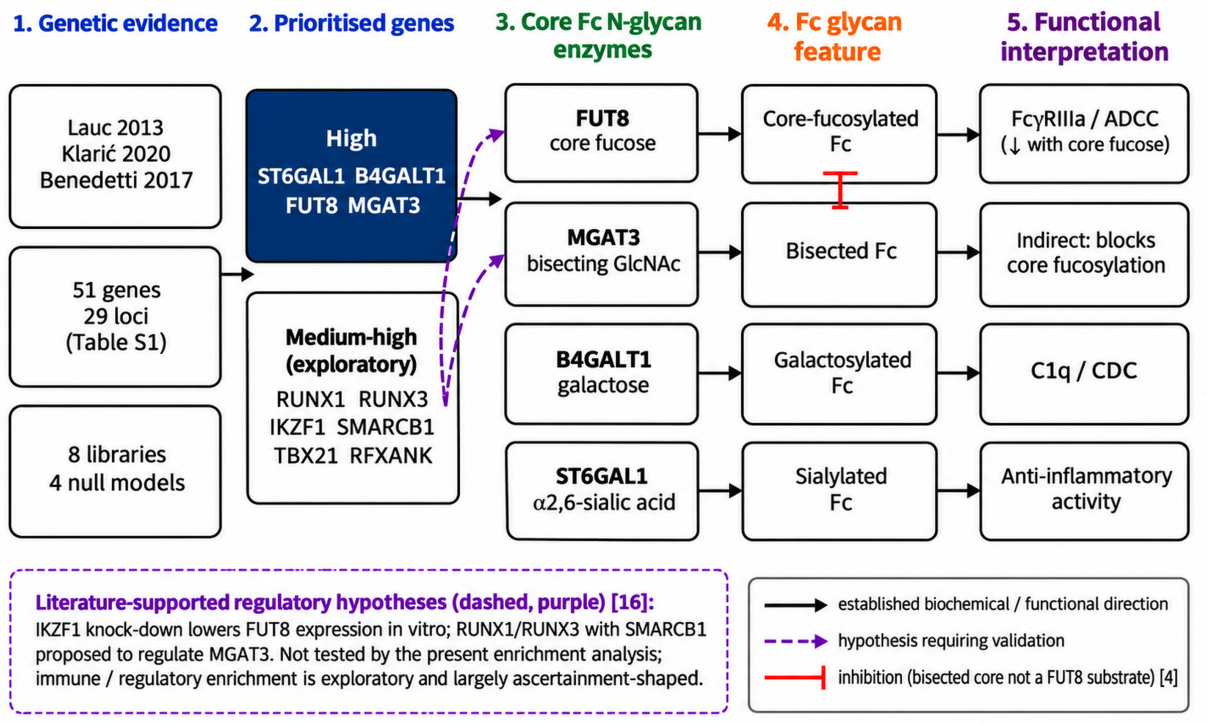


Figure 7. Hypothesis-generating genotype-to-glycan-to-function framework. Solid arrows denote established biochemical or functional direction from prior literature; the red bar denotes prevention of core fucosylation by bisection [14]; dashed purple arrows denote regulatory hypotheses proposed by the source study [6] that require functional and independent-cohort validation.

Statistical caveat. Of 2,271 tested terms, 66 met FDR <0.05, but only 13 of these remained significant after removal of any single locus, and the total number of significant terms was within the range expected for annotation-matched random gene sets (Section 3.11). Findings with FDR 0.05-0.25 are described as suggestive/exploratory, whereas those with FDR >0.25 are not FDR-supported and are retained only for descriptive context. Single-gene and single-locus terms are additionally treated as fragile. Accordingly, statements concerning immune cross-talk, metabolic control, chromatin regulation or disease pleiotropy are hypothesis-generating and require direct functional and independent-cohort validation.

3.10 Robustness and Sensitivity Analyses

Excluding single-gene and single-locus terms removed one of the 66 significant terms (the colorectal B-cell term) and all single-gene metabolite, CellMarker and GO MF associations (e.g., galactosyltransferase activity) from the robust interpretive set (Supplementary Table S3). In the leave-one-locus-out analysis, 13 significant terms were retained in all 29 runs—the KEGG and Reactome N-glycan terms, five GO glycosylation/biosynthesis terms, GO regulation of CD4+ αβ T-cell differentiation, GO regulation of cell differentiation, the HDAC2 PPI hub and the NOTCH1, SMARCA4 and HDAC2 TF-PPI terms—and a further 23 were retained in at least 27 runs (Supplementary Table S4). Under the Enrichr background convention, 59 terms were significant and the core module remained the strongest pathway signal (Reactome medial trans-Golgi FDR = 1.4×10^{-4}; KEGG N-glycan biosynthesis FDR = 6.7×10^{-4}); the main difference was fewer significant PPI-hub terms (2 versus 11), because that convention counts query genes absent from a library in the background calculation (Supplementary Table S5). Removing the four canonical glycosyltransferases abolished the KEGG and Reactome N-glycan signals but left 49 significant terms, dominated by the HDAC2, NOTCH1 and SMARCA4 regulatory hubs and T-cell differentiation terms.

3.11 Negative-Control Analysis Quantifies Ascertainment

The negative controls (Figure 8; Supplementary Table S8) addressed whether the framework adds signal beyond what the composition of the query set guarantees. First, the core N-glycan module is explained by ascertainment: KEGG N-glycan biosynthesis and the Reactome medial trans-Golgi term reached FDR <0.05 in 100% of fixed-core (N4) null sets that retained ST6GAL1, B4GALT1, FUT8 and MGAT3, but in only 0-0.2% of random or annotation-matched sets (N1, N2) and 2.4-2.6% of glycogene-composition-matched sets (N3). The module therefore reflects the specific canonical IgG Fc enzymes nominated by the source studies rather than glycogene content in general, but its recovery is not independent evidence. Second, the number of significant terms is not unusual: the observed 66 terms lay within the heavy-tailed null distributions (N1 mean 25.7 and 95th percentile 75.1, empirical P = 0.053; N2 mean 14.3 and 95th percentile 80.0, P = 0.066; N3 P = 0.154; N4 P = 0.056), and the 49 terms remaining after removal of the canonical enzymes were likewise within the annotation-matched null for 47 genes (mean 21.6; P = 0.093). Third, the breadth of convergence was unusual: the query set had FDR-significant multi-gene terms in four of five non-redundant evidence domains (pathway, cell type, PPI hub and TF-PPI), a pattern observed in 0% of N1, 1.1% of N2, 2.4% of N3 and 1.6% of N4 null sets (empirical P = 0.001-0.025), and this persisted after removal of the canonical enzymes (1.6% of null sets; P = 0.017). Thus, beyond rediscovering the canonical module, the framework detects more cross-resource convergence among the remaining, largely lymphoid-regulatory genes than expected for comparably annotated random genes; because many of these genes were nominated by network-based prioritisation in the source GWAS, this convergence is also partly shaped by ascertainment and is reported as exploratory.

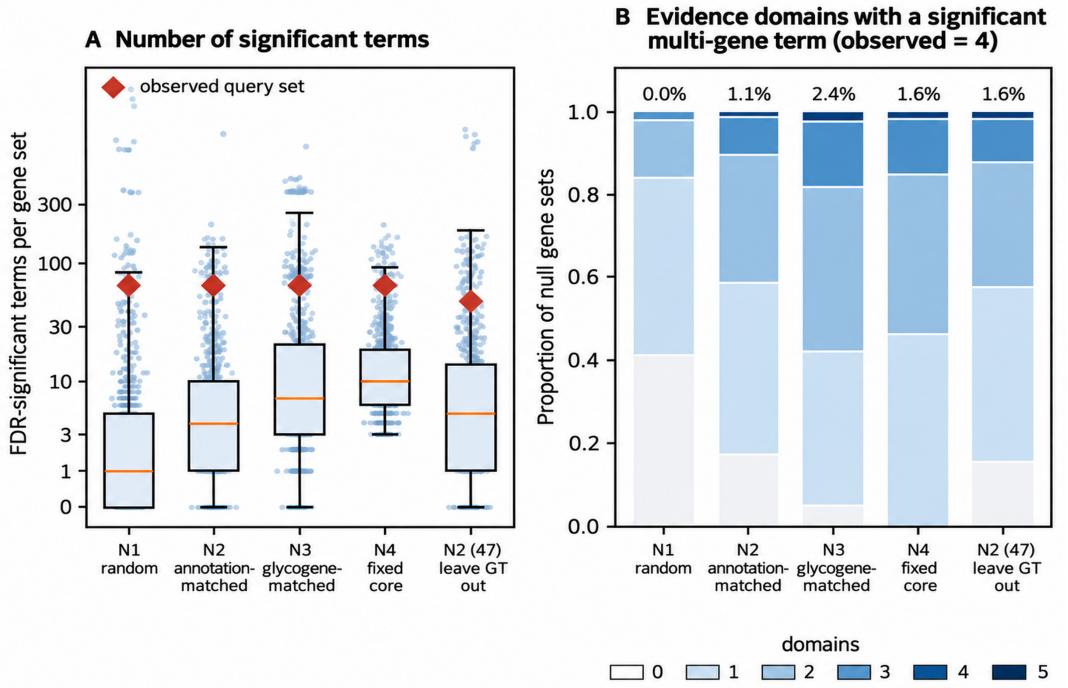


Figure 8. Negative-control analysis (1,000 null gene sets per model). (A) Number of FDR-significant terms per gene set (log scale) for size-matched random (N1), annotation-matched (N2), glycogene-composition-matched (N3) and fixed-canonical-core (N4) null models, and for an annotation-matched null for the 47 genes remaining after removal of the four canonical glycosyltransferases; red diamonds mark the observed query sets. (B) Distribution of the number of non-redundant evidence domains containing an FDR-significant multi-gene term; percentages give the proportion of null sets reaching the observed value of four. The probability that KEGG N-glycan biosynthesis reaches FDR <0.05 under each model is given below the panels.

4. Discussion

This integrative synthesis, now based on a fully documented query set and versioned libraries, prioritises a compact and biologically coherent glycosylation module after explicit multiple-testing control, robustness analysis and evidence integration. ST6GAL1, B4GALT1, FUT8 and MGAT3 jointly occur in the KEGG and Reactome N-glycan terms and in GO protein N-linked glycosylation, all of which survived removal of any single locus, making this four-gene module the strongest cross-resource result. The negative-control analysis makes the limits of this result explicit: whenever these four enzymes are present in a gene set, the module is recovered, so its significance quantifies rediscovery of the canonical Fc N-glycan enzymes nominated by the source GWAS and network study rather than independent discovery. The module should therefore be interpreted as convergence within a curated IgG-glycosylation candidate set, not as proof that the four genes are newly identified causal drivers.

The biological coherence of the module is nevertheless strong and is supported by prior biochemical literature. FUT8 adds core fucose, whose absence increases FcγRIIIa affinity and ADCC [2,3]; MGAT3 (GnT-III) adds bisecting GlcNAc, which suppresses subsequent terminal modification [15] and prevents FUT8-mediated core fucosylation of the bisected core, so that its influence on FcγRIIIa binding and ADCC is indirect rather than a direct effect of the bisecting residue on Fcγ-receptor binding [14]; B4GALT1 catalyses antennary galactosylation, which supports CDC [4]; and ST6GAL1 adds terminal α-2,6-linked sialic acid associated with anti-inflammatory IgG activity [5,17]. These established reactions provide a plausible mechanistic explanation for the recurrent pathway enrichment. However, the present computational analysis did not measure enzyme activity, Fc glycan flux, FcγR binding, ADCC, CDC or anti-inflammatory function, so these consequences must remain explicitly literature-derived rather than being attributed to the enrichment analysis itself.

ST6GAL1 is particularly relevant because prior work indicates that sialylation need not be restricted to intracellular B-cell processing: circulating ST6GAL1 can contribute to extracellular IgG sialylation and also influences B-cell immunoglobulin biology [18,19]. These observations strengthen the rationale for prioritising ST6GAL1 for functional experiments, but they do not establish that the genetic associations synthesised here alter circulating ST6GAL1 activity or determine individual anti-inflammatory IgG phenotypes.

The prioritised module has potential relevance to clinical pharmacology and antibody engineering because Fc fucosylation, galactosylation and sialylation can modulate Fc-receptor and complement interactions. Nevertheless, the translational implication of the present study is candidate prioritisation rather than demonstration of a therapeutic effect. Functional follow-up should therefore include controlled perturbation of ST6GAL1, B4GALT1, FUT8 and MGAT3, quantitative glycan profiling, and direct Fc-effector assays before any claim is made that manipulating the prioritised module will predictably tune ADCC, CDC or anti-inflammatory activity.

The metabolite layer adds a biologically plausible but statistically unsupported substrate dimension. GDP, UDP, CMP, UDP-GlcNAc and N-acetyl-D-glucosamine participate in the synthesis or turnover of the nucleotide-sugar donors used by the four enzymes [21,22], but the HMDB associations in this analysis have FDR = 0.133, several depend on one gene, only 10 query genes are represented in the library, and the donors themselves are not represented as HMDB terms. They therefore do not validate a rate-limiting metabolic mechanism for IgG glycosylation. A direct test would require experimental manipulation of hexosamine-pathway flux or nucleotide-sugar availability together with quantitative measurement of IgG Fc glycan branching and terminal modifications.

The immune, cell-type and regulatory layers require similarly cautious interpretation. The documented query set contains several lymphoid transcription factors and chromatin regulators (RUNX1, RUNX3, IKZF1, IKZF3, SMARCB1, SMARCD3, TBX21 and IRF1) that converge on HDAC2/SIN3A, SWI/SNF and NOTCH1 interaction hubs and on T-cell differentiation terms. HDAC2, NOTCH1, SMARCA4 and CD4+ T-cell differentiation were robust to removal of any single locus, and the breadth of cross-resource convergence exceeded that of annotation-matched random genes. This pattern is consistent with the hypothesis of Klarić et al. that IKZF1 regulates FUT8 and that RUNX1/RUNX3 with SMARCB1 regulate MGAT3 [6]. However, PPI-hub and TF-PPI libraries are dominated by well-studied transcriptional regulators, several of these genes were prioritised in the source study partly by gene-set-based methods, and the enrichment reflects shared interaction partners rather than regulation of glycosyltransferase genes. No direct evidence in the present analysis shows that any of these factors transcriptionally regulates ST6GAL1, B4GALT1, FUT8 or MGAT3, and the cell-type terms do not identify the cells in which IgG is glycosylated. Immune-response modulation was not tested and should not be inferred from these results.

Our findings are concordant with earlier genetic and network studies showing that IgG glycosylation is polygenic, pleiotropic and organised around known glycosylation enzymes and regulatory loci [6-8]. More recent work reinforces both the biological importance and the complexity of IgG glycosylation: population-scale synthesis has highlighted the extensive inter-individual and disease-associated variation in IgG glycans [26], while a contemporary review of glycosylation genetics places these associations within the broader genetic architecture of mammalian glycan synthesis [27]. Recent pharmacology and glycoengineering literature also emphasises the therapeutic potential of manipulating antibody glycoforms [28,29]. *In vivo* delivery of B4GALT1- and ST6GAL1-encoding mRNA has shown therapeutic effects in experimental rheumatoid arthritis models [30], whereas recent disease studies demonstrate altered IgG glycosylation in systemic lupus erythematosus [31], asymmetric Fc glycan pairing in human IgG1 [32], continuing

advances in antibody glycoengineering [33], and infection-associated IgG glycan remodelling [34]. These newer studies strengthen the rationale for validation but do not convert the present enrichment findings into causal evidence.

Several limitations should be acknowledged. First, the analysis starts from a GWAS/network-derived gene set and is sensitive to locus-to-gene mapping choices, gene-set boundaries and prior inclusion of known glycosyltransferases; we retained every gene named by the source studies, flagged single-locus overlaps and used leave-one-locus-out analysis, but alternative mapping rules (e.g., nearest gene only or eQTL-colocalised genes only) could alter downstream enrichment. Second, enrichment statistics depend on the library-specific gene universe and database version; the eight library versions, universes and checksums are reported, and results under the Enrichr default background are provided for comparison. Third, GO, KEGG and Reactome contain correlated pathway annotations; they were collapsed into one domain for prioritisation, and the total count of significant terms, dominated by redundant GO terms, was not unusual relative to null models. Fourth, the negative controls show that the core module is recovered by construction once the canonical enzymes are included, and that the regulatory signal, although broader than expected for random genes, may still reflect ascertainment through network-based gene prioritisation in the source GWAS. Fifth, several cell-type, metabolite and regulatory results are single-gene, single-locus or not FDR-supported. Finally, all proposed causal and regulatory relationships are hypotheses generated from annotation convergence rather than direct experimental observations. Independent human-cohort replication, allele-resolved expression/QTL analyses, gene knockout or overexpression, quantitative Fc-glycan profiling, nucleotide-sugar manipulation and direct Fc-effector assays are required to establish causality and directionality.

5. Conclusion

A reproducible multi-source synthesis of 51 genes from 29 IgG N-glycosylation loci prioritises ST6GAL1, B4GALT1, FUT8 and MGAT3 as a compact, recurrent and locus-robust N-glycan biosynthesis module. Negative controls show that this module is recovered whenever the canonical Fc glycosyltransferases are included, so it represents a rigorously documented rediscovery rather than a new finding. A secondary layer of lymphoid transcription-factor and chromatin-regulator genes shows cross-resource convergence beyond annotation-matched expectation but remains exploratory, and cell-type and metabolite findings are suggestive or fragile. The established biosynthetic roles of the four enzymes offer a coherent literature-based framework for interpreting the enrichment, but the present in-silico study does not demonstrate immune-response modulation, that the module acts as a causal molecular rheostat, that any transcription factor regulates these glycosyltransferases, or that nucleotide-sugar availability controls IgG glycosylation *in vivo*. The principal contribution is a transparent, fully reproducible prioritisation framework, with code, inputs and complete outputs, for targeted functional experiments, independent-cohort validation and subsequent evaluation of Fc glycoengineering strategies.

Acknowledgements

The authors acknowledge the investigators who generated the published GWAS, glycoproteomics and functional-genomics datasets synthesised in this study, and the curators of the GO, KEGG, Reactome, CellMarker and HMDB resources. No additional individual met authorship criteria.

Funding

This work did not receive a specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Ethics Statement

Not applicable. This manuscript is a computational synthesis of previously published genome-wide association and functional-genomics resources. No new human participants, human tissues or animals were recruited or studied.

Informed Consent

Not applicable. No new human participants were enrolled. All source data were obtained from published studies that had already received their own institutional ethical approvals.

Data Availability Statement

All data and code needed to reproduce every statistic, table and figure are provided as Supplementary Materials: the query-gene ledger (Supplementary Table S1 and Data S2), library versions, universes and SHA-256 checksums (Supplementary Table S2), the complete tested-term outputs of the primary, Enrichr-convention and leave-glycosyltransferases-out analyses (Data S2), leave-one-locus-out and negative-control results (Supplementary Table S4 and Supplementary S8; Data S2), and the analysis and plotting code with instructions (Data S1). The Enrichr library files

are publicly available from the Enrichr gene-set library endpoint (https://maayanlab.cloud/Enrichr/) and are identified by name and checksum. No independent public repository accession has been assigned; the code and complete outputs are designated for submission as Supplementary Data S1 and Supplementary Data S2.

Author Contributions

A.K.C.: Conceptualisation, methodology, formal analysis, investigation, writing – original draft, supervision, and project administration. A.H.: Writing – review and editing. F.B.G.: Writing – review and editing. S.S.: Methodology; writing – review and editing , manuscript revision and corrections; reference formatting. All authors have read and approved the final version of the article and agreed to the author order.

Conflict of Interest

The authors declare no competing interests.

Generative AI Statement

Generative AI tools were used solely for language editing, grammatical correction, and improvement of textual clarity. No generative AI tools were used for research design, data analysis, code generation, figure preparation, or scientific interpretation. The authors reviewed and verified the scientific content and take full responsibility for the accuracy and integrity of the work

References

[1] Arnold JN, Wormald MR, Sim RB, Rudd PM, Dwek RA. The impact of glycosylation on the biological function and structure of human immunoglobulins. Annual Review of Immunology, 2007, 25, 21-50. DOI: 10.1146/annurev.immunol.25.022106.141702

[2] 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 FcγRIII and antibody-dependent cellular toxicity. The Journal of Biological Chemistry, 2002, 277(30), 26733-26740. DOI: 10.1074/jbc.M202069200

[3] 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. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(31), 12669-12674. DOI: 10.1073/pnas.1108455108

[4] Hodoniczky J, Zheng YZ, James DC. Control of recombinant monoclonal antibody effector functions by Fc N-glycan remodeling *in vitro*. Biotechnology Progress, 2005, 21(6), 1644-1652. DOI: 10.1021/bp050228w

[5] Kaneko Y, Nimmerjahn F, Ravetch JV. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. Science, 2006, 313(5787), 670-673. DOI: 10.1126/science.1129594

[6] 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. Science Advances, 2020, 6(8), eaax0301. DOI: 10.1126/sciadv.aax0301

[7] 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 Genetics, 2013, 9(1), e1003225. DOI: 10.1371/journal.pgen.1003225

[8] 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. Nature Communications, 2017, 8(1), 1483. DOI: 10.1038/s41467-017-01525-0

[9] Trzos S, Link-Lenczowski P, Pocheć E. The role of N-glycosylation in B-cell biology and IgG activity. The aspects of autoimmunity and anti-inflammatory therapy. Frontiers in Immunology, 2023, 14, 1188838. DOI: 10.3389/fimmu.2023.1188838

[10] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Research, 2016, 44(W1), W90-W97. DOI: 10.1093/nar/gkw377

[11] Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: The Human Metabolome Database for 2022. Nucleic Acids Research, 2022, 50(D1), D622-D631. DOI: 10.1093/nar/gkab1062

[12] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 1995, 57(1), 289-300. DOI: 10.1111/j.2517-6161.1995.tb02031.x

[13] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. Nucleic Acids Research, 2023, 51(D1), D1373-D1380. DOI: 10.1093/nar/gkac956

[14] Bosman GP, Stoof ID, Bastiaansen HP, Quarles van Ufford L, Dobruchowska JM, Langenbach JWH, et al. Bisecting N-acetylglucosamine of the N-glycan of immunoglobulin G does not affect binding to Fc gamma receptors. ACS Chemical Biology, 2025, 20(3), 680-689. DOI: 10.1021/acschembio.4c00807

[15] Nakano M, Mishra SK, Tokoro Y, Sato K, Nakajima K, Yamaguchi Y, et al. Bisecting GlcNAc is a general suppressor of terminal modification of N-glycan. Molecular & Cellular Proteomics, 2019, 18(10), 2044-2057. DOI: 10.1074/mcp.RA119.001534

[16] Gudelj I, Salo PP, Trbojević-Akmačić I, Albers M, Primorac D, Perola M, et al. Low galactosylation of IgG associates with higher risk for future diagnosis of rheumatoid arthritis during 10 years of follow-up. Biochimica et Biophysica Acta - Molecular Basis of Disease, 2018, 1864(6 Pt A), 2034-2039. DOI: 10.1016/j.bbadis.2018.03.018

[17] Sneed SL, Strandberg EA, Laureano AFS, Vattepu R, Sun Y, Tran TT, et al. Molecular determinants of sialylated IgG anti-inflammatory activity. Proceedings of the National Academy of Sciences of the United States of America, 2025, 122(20), e2411600122. DOI: 10.1073/pnas.2411600122

[18] Irons EE, Punch PR, Lau JTY. Blood-borne ST6GAL1 regulates immunoglobulin production in B cells. Frontiers in Immunology, 2020, 11, 617. DOI: 10.3389/fimmu.2020.00617
[19] Jones MB, Oswald DM, Joshi S, Whiteheart SW, Orlando R, Cobb BA. B-cell-independent sialylation of IgG. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(26), 7207-7212. DOI: 10.1073/pnas.1523968113
[20] Shan Q, Zeng Z, Xing S, Li F, Hartwig SM, Gullicksrud JA, et al. The transcription factor Runx3 guards cytotoxic CD8+ effector T cells against deviation towards follicular helper T cell lineage. Nature Immunology, 2017, 18(8), 931-939. DOI: 10.1038/ni.3773
[21] Paneque A, Fortus H, Zheng J, Werlen G, Jacinto E. The hexosamine biosynthesis pathway: regulation and function. Genes (Basel), 2023, 14(4), 933. DOI: 10.3390/genes14040933
[22] Lau KS, Partridge EA, Grigorian A, Silvescu CI, Reinhold VN, Demetriou M, et al. Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. Cell, 2007, 129(1), 123-134. DOI: 10.1016/j.cell.2007.01.049
[23] John LB, Ward AC. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. Molecular Immunology, 2011, 48(9-10), 1272-1278. DOI: 10.1016/j.molimm.2011.03.006
[24] Ferreira R, Ohneda K, Yamamoto M, Philipsen S. GATA1 function, a paradigm for transcription factors in hematopoiesis. Molecular and Cellular Biology, 2005, 25(4), 1215-1227. DOI: 10.1128/MCB.25.4.1215-1227.2005
[25] Dicken J, Mildner A, Leshkowitz D, Touw IP, Hantisteanu S, Goldenberg D, et al. Transcriptional reprogramming of CD11b+Esam(hi) dendritic cell identity and function by loss of Runx3. PLoS One, 2013, 8(10), e77490. DOI: 10.1371/journal.pone.0077490
[26] Krištić J, Lauc G. The importance of IgG glycosylation—What did we learn after analyzing over 100,000 individuals. Immunological Reviews, 2024, 328(1), 143-170. DOI: 10.1111/imr.13407
[27] Stanley P. Genetics of glycosylation in mammalian development and disease. Nature Reviews Genetics. 2024, 25(10), 715-729. DOI: 10.1038/s41576-024-00725-x
[28] Nimmerjahn F. Role of antibody glycosylation in health, disease, and therapy. Handbook of Experimental Pharmacology, 2025, 288:189-209. DOI: 10.1007/164_2025_744
[29] Huang HW, Zeng YF, Shivatare VS, Tseng TH, Wong CH. Cell-based glycoengineering for production of homogeneous and specific glycoform-enriched antibodies with improved effector functions. Proceedings of the National Academy of Sciences of the United States of America, 2025, 122(8), e2423853122. DOI: 10.1073/pnas.2423853122
[30] Peng X, Li Y, Sun X, Ren G, Li H, Wang X, et al. Therapeutic effect of galactosyltransferase- and sialyltransferase-encoding mRNA in rheumatoid arthritis. Molecular Therapy, 2025, 33(9), 4143-4155. DOI: 10.1016/j.ymthe.2025.06.018
[31] Wu Y, Wang M, Hu C, Zhang S, Zhao J, Wang Q, et al. IgG glycosylation profiling of systemic lupus erythematosus using lectin microarray. Lupus Science & Medicine, 2025, 12(1), e001413. DOI: 10.1136/lupus-2024-001413
[32] Azzam T, Bournazos S, Gunduz H, et al. Asymmetrically glycosylated IgG1 antibodies are universal and drive human disease. Nature Communications, 2026, 17:383. DOI: 10.1038/s41467-025-67070-3
[33] Serafin B, Henry O, de Crescenzo G. On the use of glycoengineering methods for the generation of antibodies with enhanced effector functions. Glycobiology, 2026, cwag019. DOI: 10.1093/glycob/cwag019
[34] Yan R, Wang C, Zhao X, Wang Y, Liu M, Shi X, et al. Immunoglobulin G N-glycosylation remodeling in viral infections: Immunomodulatory roles and potential biomarker implications. Frontiers in Immunology, 2026, 17, 1758338. DOI: 10.3389/fimmu.2026.1758338

Supplementary Material

The supplementary files listed below accompany this article and replace the supplementary tables included in the previous version.

Supplementary File 1 (Supplementary Tables S1-S8). S1, query-gene provenance ledger (51 genes, 29 loci; source study, locus coordinates, lead variant and P value, mapping strategy and library coverage); S2, gene-set libraries, versions, universes, SHA-256 checksums and multiplicity families; S3, FDR-significant and suggestive multi-locus terms with single-gene/single-locus classification; S4, leave-one-locus-out robustness of all FDR-significant terms; S5, Enrichr-convention sensitivity analysis; S6, PubChem and HMDB identifiers for glycosylation-relevant metabolites and nucleotide-sugar donors; S7, evidence scores and priority for all 51 genes; S8, negative-control summary.

Supplementary Data S1 (code). Analysis pipeline, leave-one-locus-out script, evidence-scoring script, figure script and README with exact commands, software versions and random seed.

Supplementary Data S2 (Excel workbook). Query-gene ledger, per-library coverage, complete tested-term output (2,271 tests) for the primary analysis, the Enrichr-convention sensitivity analysis and the leave-glycosyltransferases-out analysis, leave-one-locus-out results, per-iteration negative-control results and network edge list.