

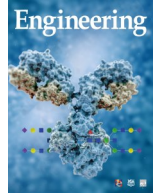


ELSEVIER

Contents lists available at ScienceDirect

Engineering

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



Research
Glycomedicine—Review

近十二年人类蛋白质N-糖基化的全基因组关联研究综述

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

关键词

糖组

聚糖

N-糖基化

基因组学

遗传学

全基因组关联研究

摘要

人类分泌的以及膜结合的大多数蛋白质都有共价连接的寡糖链或聚糖。糖基化影响着蛋白质的物理和化学性质及其生物功能。不难理解,蛋白质糖基化的改变与越来越多的人类疾病有关,而聚糖也越来越多地被视为潜在的治疗靶点、治疗药物的重要组成部分和生物标志物。尽管糖基化途径在生化方面被研究得很透彻,但人们对指导人类体内这些生化反应的细胞和组织特异性调节的基因网络知之甚少。由于缺乏对糖组变化的调控机制以及糖组与人类健康和疾病之间联系的详细了解,人类糖生物学的临床应用进展缓慢。人类遗传学和基因组学是用于提供人类体内糖生物学知识的两种亟待开发的工具,它们为剖析复杂性状的生物学特性提供了一种强大的数据驱动型不可知方法。本文总结了人类群体糖原组学的现状。在第1节中,简要概述了N-聚糖的结构组织;在第2节中,介绍了主要的血浆糖蛋白。接下来,在第3节中,总结、分类并归纳了当前N-糖基化全基因组关联研究(GWAS)的结果,这些研究提供了有关糖基化群体变异遗传调控的新知识。迄今为止,此类研究仅限于分析人体血浆N-糖组以及免疫球蛋白G和转铁蛋白的N-糖基化。虽然与不同组织和糖蛋白的大量糖组相比,这三种糖组只构成了相当有限的集合,但这三种糖组的研究确实可以提供有力的分析和归纳。然后,在第4节中讨论已确定位点中的基因,并在第5节中特别关注具有强支持作用的基因。最后,在第6节和第7节中介绍新发现的特殊案例,重点关注与人类蛋白质N-糖基化有关的基因变异的潜在作用机制和生物靶标。

© 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-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

N-聚糖是由5种单糖[即甘露糖、半乳糖、岩藻糖、唾液酸和N-乙酰葡萄糖胺(GlcNAc)]组成的不规则聚合物,通过与天冬酰胺(Asn)侧链的糖苷键连接,附着在多肽结构上形成糖蛋白[1]。糖基化影响蛋白质的物理和化学性质及其生物功能。N-聚糖蛋白的糖基部分种类繁多;一个N-聚糖蛋白可能有单个或多个N-糖基化位点,

而且不一定是糖基化的。N-聚糖在蛋白质内部和蛋白质之间都存在差异:同一蛋白质的不同分子上可能附着不同的N-聚糖,不同蛋白质上附着的N-聚糖种类繁多。人体血浆中约有一半的蛋白质都带有寡糖链(“聚糖”)[2]。蛋白质糖基化受遗传、表观遗传和环境因素控制[3–5]。尽管糖基化途径在生物化学方面已被研究得很透彻[6],但人们对调控这些生化反应的细胞和组织特异性调节的基因网络知之甚少。人们对蛋白质N-糖基化调控的兴趣不仅限于学术研究,因为人们越来越认识到N-糖基

* Corresponding author.

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

化在人类健康和疾病中的作用[7-8]。聚糖被认为是潜在的治疗靶点[9-11]、治疗药物的重要组成部分[12-14]和生物标志物[15-17]，使糖生物学成为未来临床应用的一个前景广阔的领域[18]。

与其他“-组学”(-omics)类似，生物体的“糖组”(glycome)由该生物体合成的所有聚糖结构组成[19]。不同类型的细胞会根据其分化状态和生理环境合成糖组子集。在此，使用“糖组”一词来定量表示细胞或组织在特定的时间、地点和环境条件下产生的聚糖集合。本文还使用“单个蛋白质糖组”(如“免疫球蛋白-G N-糖组”)一词来定量描述附着在特定蛋白质上的聚糖谱。

人类糖组学和基因组学的结合提供了一种强大的“数据驱动假设”方法[20]，可以通过不可知论的方式剖析人类体内复杂的糖生物学[21]。迄今为止，针对成千上万个样本进行的N-糖基化分析仅限于少数几种可用的组织和蛋白质。更具体地说，直到最近，人类N-糖基化的大型队列研究还仅限于免疫球蛋白G(IgG)的N-糖基化和血浆糖组(血浆N-糖组)研究，后者是指附着在血浆糖蛋白上的聚糖谱。已有数以万计的样本对这些糖组进行了表征[22]。最近，对其他糖蛋白，如转铁蛋白(TF)[23]、免疫球蛋白A(IgA)[24-25]和脂蛋白CIII[26]，也可以进行大规模的糖蛋白研究，但研究的样本数量仍然有限。

迄今为止，有关N-糖基化群体变异遗传调控的研究(需要对大量样本进行糖组和基因组测量)仅限于分析血浆糖组、IgG和TF N-糖基化。与不同组织和糖蛋白的大量糖组相比，这三种糖组组成的集合相当有限；不过，这三种糖组的研究可以提供有效的分析和归纳。

本文旨在补充和扩展参考文献[21,27]中对群体糖组学的总结和讨论。更具体地说，本文详细总结了自前几篇综述发表以来所取得的成果。本文总结了这些新的信息，并根据新的发现考虑了一些特殊的案例，重点关注与人类蛋白质N-糖基化有关的潜在作用机制和遗传变异的生物靶标。关于已经确定的遗传性和相关基因座的详细讨论，感兴趣的读者可参阅以前的综述和原文。

2. IgG、TF和总血浆蛋白质组的N-糖基化

IgG是全身体液免疫的关键因素。它是血液循环和组织中最常见的同种型。由于广泛的活性，IgG参与多种免疫反应，包括补体激活、吞噬和抗体依赖性细胞毒性(ADCC)[28-29]。IgG是人体血液中含量最高的糖蛋白，可能占血浆N-聚糖总量的近一半(表1[2,30-31])。IgG由两个生物学上不同的区域组成：抗原结合片段(Fab)

和可结晶片段(Fc)。这两个区域都可以糖基化。IgG的Fc在CH₂结构域的Asn297位置上有一个保守的N-糖基化位点。由于IgG是同源二聚体，因此两条重链中的每一条都可以在Asn297上进行糖基化。IgG连接的N-聚糖要么是高甘露糖型聚糖，要么是双触角型聚糖，通常以岩藻糖基化为核心，可能带有一个或两个半乳糖残基，并可能延伸出一个或两个唾液酸。研究表明，在病理和生理过程中，Fc连接的N-糖基化会发生变化，并强烈影响抗体的炎症特性[32-36]。相比之下，人们对Fab连接的N-糖基化知之甚少。已知15%~25%的IgG Fab上存在N-聚糖，它们会影响抗体对抗原的亲合力和热敏性，以及IgG的半衰期[37]。

TF是血浆糖蛋白，是铁代谢的关键部分，负责铁离子的输送。铁在人体生理学的许多重要生化过程中发挥着核心作用；人体细胞对作为氧化剂的铁的需求催生了一个复杂的系统，该系统严格调控铁的水平、组织分布和生物利用率[38]。人类TF有两个N-糖基化位点，分别位于N432和N630残基，该位点黏附最丰富的聚糖，其结构是不含岩藻糖的双触角型双唾液酸化二半乳糖基聚糖结构[39-40]。TF与纤维蛋白原一起，是最丰富的非免疫球蛋白人血浆糖蛋白之一(表1)。

毫无疑问，特定蛋白质的糖基化有其独特的调节机制。例如，如果编码突变扰乱了典型的N-糖基化氨基酸序列基团，就会导致该部位没有N-聚糖附着。众所周知，在经典模式之外的氨基酸序列改变的突变会导致突变蛋白质的N-糖基化谱发生变化[41-42]。这类突变可能通过三级结构的改变来显示其效果，从而改变糖基化位点的可及性。

除上述考虑外，还应注意的，糖基化机制是特定细胞中分泌蛋白和膜蛋白所共有的。这一机制的独特性，即特定酶的存在或不存在(或更一般地说，活性)在一个组织的细胞之间是共享的。因此，尽管独特的蛋白质和单个细胞的特异性是可以预期的，但在一个组织或一种细胞类型中，支配不同糖蛋白N-糖基化的一系列规则预计是相同的。

表1显示了人体血浆中主要糖蛋白的平均浓度和可能来源。很明显，大多数糖蛋白只来自两个组织：淋巴和肝脏。因此，可以合乎逻辑地假设，血浆蛋白质组的糖基化可能主要是由淋巴组织(主要是浆细胞)和肝脏(肝细胞)糖基化调节的特殊性所驱动的。根据这一假设，对于总的血浆、IgG和TF糖组的研究为什么会相互补充、相互促进这一问题就变得很清楚了(将在下文说明这一点)。

表1 人血浆糖蛋白的平均浓度和可能来源。推断可能的来源组织和细胞类型为已知在血液中分泌蛋白质并高表达相关蛋白质的组织(更多详细信息见附录 A 中的表 S1)

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	<i>IGHG1</i>	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)
	<i>IGHG2</i>	Bone marrow	Plasma cells		
	<i>IGHG3</i>	Bone marrow	Plasma cells		
	<i>IGHG4</i>	Bone marrow	Plasma cells		
Fibrinogen	<i>FGB</i>	Liver	Hepatocytes	3	2.0–4.5
	<i>FGA</i>	Liver	Hepatocytes		
	<i>FGG</i>	Liver	Hepatocytes		
IgA	<i>IGHA1</i>	Lymphoid tissues	Plasma cells	2.62 (IgA1: 2.358, IgA2: 0.262)	2.62
	<i>IGHA2</i>	Lymphoid tissues	Plasma cells		
TF	<i>TF</i>	Liver	Hepatocytes	2.5	2–3
IgM	<i>IGHM</i>	Spleen	Plasmoblasts (marginal zone B-cells)	1.47	0.5–2.0
Haptoglobin	<i>HP</i>	Liver	Hepatocytes	1.32	0.8–2.5
α -2-macrogobulin	<i>A2M</i>	Liver	Hepatocytes	1.2	1–2
α -1-antitrypsin	<i>SERPINA1</i>	Liver	Hepatocytes	1.1	1.1
Hemopexin	<i>HPX</i>	Liver	Hepatocytes	0.8	0.4–1.5
α -1-acid glycoprotein	<i>ORM1</i>	Liver	Hepatocytes	0.77	0.36–1.46
	<i>ORM2</i>	Liver	Hepatocytes		
Apolipoprotein B-100	<i>APOB</i>	Liver	Hepatocytes	0.5	0.5
α -2-heat stable-glycoprotein	<i>AHSG</i>	Liver	Hepatocytes	0.45	0.3–0.6
Ceruloplasmin	<i>CP</i>	Liver	Hepatocytes	0.355	0.15–0.96
Vitronectin	<i>VTN</i>	Liver	Hepatocytes	0.3	0.2–0.4
α -1B-glycoprotein	<i>A1BG</i>	Liver	Hepatocytes	0.22	0.22
β -2-glycoprotein 1	<i>APOH</i>	Liver	Hepatocytes	0.2	0.2
Antithrombin-III	<i>SERPINC1</i>	Liver	Hepatocytes	0.15	0.15
Histidine-rich glycoprotein	<i>HRG</i>	Liver	Hepatocytes	0.125	0.10–0.15
Apolipoprotein D	<i>APOD</i>	Brain	Glial cells	0.1	0.1
Apolipoprotein F	<i>APOF</i>	Liver	Hepatocytes	0.0845	0.073–0.096
Kininogen-1	<i>KNG1</i>	Liver	Hepatocytes	0.0725	0.055–0.090
Zinc- α -2-glycoprotein	<i>AZGP1</i>	Liver	Hepatocytes	0.05	0.05
		Breast	Glandular cells		
IgD	<i>IGHD</i>	Respiratory mucosa	IgD-secreting B-cells	0.035	< 0.003–0.400
IgE	<i>IGHE</i>	Thymus	Plasma cells	0.0003	0.0003

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

3. 蛋白质 N-糖基化的全基因组关联研究

由于高通量 N-糖组质谱分析方法的进步, 成千上万的样本的糖组得到了表征[43–44], 同时还出现了可用于测量常见 (小等位基因频率大于 1%) 遗传变异的经济实惠的精确 DNA 技术, 因此 GWAS 已被用于了解人类糖组变异的遗传决定因素。GWAS 是一种用于确定基因组区域

(基因位点) 中哪些区域的变异会导致所研究性状 (即人类数量性状和复杂疾病) 变异的方法。简而言之, 这种方法通过比较成千上万个体的基因组和 N-糖组, 可靠地发现遗传变异 [如单核苷酸多态性 (SNP)] 对特定聚糖丰度的影响。由于存在连锁不平衡 (LD), GWAS 通常会产一组相关的 SNP, 这些 SNP 与所关注的性状有显著的统计学关联。经典的 GWAS 设计分为 “发现” 阶段和

“复制”阶段。“发现”阶段是在全基因组显著水平上确定相关基因位点，“复制”阶段是在独立样本中证实已发现基因位点的关联性[45]。复制提高了GWAS研究结果的可靠性，有助于减少观察到的基因型与表型的关联是偶然发现或是人工设计和分析的产物的可能性[46]。因此，在针对特定位点启动复杂而昂贵的定向实验跟踪研究之前，复制是必不可少的一步。在最近的综述[21]中更详细地介绍了将GWAS方法应用于糖组学的情况。

确定与糖组相关的基因位点，为更好地了解糖生物学及其与人类健康和疾病的联系提供了起点。然而，接下来的工作并不简单，因为功能多态性和受影响的基因都不能通过GWAS被马上确定。在GWAS中确定的基因座可能包含数十个基因；此外，基因座中的未知功能元件可能会调控远相关区域之外的基因[47]。在接下来的章节中将讨论将关联区域与生物学功能联系起来所面临的挑战。下面将总结对IgG、TF和总血浆蛋白质组（血浆糖组）的N-糖基化研究的基因图谱结果。

迄今为止，已发表了5项针对多达8000人的IgG N-糖基化的GWAS [48–52]。在这些GWAS中发现的33个全基因组重要位点中，有29个在独立样本中得到了重复。第一项GWAS用液相色谱法测量IgG糖组，发现了4个糖基转移酶基因中或其附近的相关变异，以及另外5个位点（不含在蛋白质糖基化中起作用的基因）[48]。Shen等[49]开发并应用了多变量GWAS发现和复制方法，通过另外5个位点扩展了该列表，其中只有一个位点包含糖基转移酶基因。Wahl等[50]在一项使用液相色谱-电喷雾质谱联用法（LC-ESI-MS）测量IgG N-糖基化的研究中，又发现了一个不含糖基转移酶基因的位点。

上述研究仅限于约2000人的发现阶段样本量，而Klarić等[51]最近的一项研究则在8000多人的发现样本中研究了IgG糖组与基因组之间的关联。通过超高效液相色谱法（UHPLC）测量IgG糖组，结果在2400个样本中得到重复，并通过LC-ESI-MS测量IgG糖组，结果在约1800个样本中得到验证。总之，这项工作使全基因组范围内与IgG糖组显著相关的基因位点数量达到27个，其中22个在独立样本中得到了重复。

Shadrina等[52]使用基于多变量汇总的GWAS发现和复制方法[53]重新分析了Klarić等生成的汇总统计数据。由于采用了多变量方法并扩大了复制样本（ $N = 3147$ ），该分析产生了6个新的基因位点（其中5个得到了复制），并复制了两个以前发现的基因位点。

与IgG N-糖基化相关的29个复制位点中有5个含有糖基转移酶；这些酶负责核心（*FUT8*）和触角（*FUT6*）岩

藻糖基化、平分型GlcNAc（*MGAT3*）添加以及半乳糖基化（*B4GALT1*）和唾液酸化（*ST6GAL1*）等末端修饰。虽然其他24个基因位点包含的基因属于多个类别，但有一组基因涉及抗体产生细胞的生物学，如*RUNX3*、*RUNX1*、*IKZF3*、*IKZF1*和*TNFRSF13B*值得注意。其他值得注意的关联还包括与人类白细胞抗原（*HLA*）区域的关联，以及与包含编码免疫球蛋白重链（*IGH*）基因区域的关联，尽管该区域还包含一个次要候选基因[*TMEM121*，通过其表达定量性状位点（eQTL）的共定位而受到牵连]。下一节将详细讨论在受牵连的基因座中优先考虑的候选基因。

与对IgG N-糖基化的研究相比，整个人体血浆蛋白质组N-糖基化的遗传学研究在研究数量、样本大小以及由此确定的基因位点数量方面都有些滞后。迄今为止，利用基于液相色谱的技术对多达3500人的血浆N-糖基化进行测量的三项人类血浆N-糖基化GWAS共发现了16个基因位点[54–56]。无论是在原始研究中，还是在Sharapov等[57]的研究中，其中绝大多数位点（15个）在独立样本中得到了复制。这些研究使用了迄今为止最大的复制样本量（4800人）。除了一个基因位点（第22号染色体，靠近*KREMEN1*）外，该研究对所有基因位点都有很高的（>95%）预期复制能力。就后一个位点而言，相关插入/缺失多态性rs140053014的小等位基因频率在研究人群中很低（约为2%），这使得对该基因推算变得复杂，从而将可用样本数量和功率降低到约10%，并使复制结果具有不确定性。

15个复制位点中有11个与IgG N-糖基化遗传控制研究中发现的位点重叠。这些基因位点包含编码糖基转移酶的基因：*FUT8*、*FUT6*、*MGAT3*、*ST6GAL1*和*B4GALT1*；以及其他基因：*HLA*、*IGH*、*IKZF1*和*RUNX3*。鉴于免疫球蛋白占血浆糖蛋白组的主要部分（表1），这可能并不特别令人惊讶。在这种情况下，IgG N-糖基化的GWAS尽管样本量更大，但没有被发现的4个血浆糖组特异性位点却可能特别值得关注。其中三个位点含有编码糖基转移酶的基因：*MGAT5*，其产物添加GlcNAc，促进双触角聚糖的进一步分支；*ST3GAL4*，其产物促进末端糖基化；*B3GAT1*，编码葡萄糖醛酸基转移酶。此外，含有*HNF1A*基因的区域与血浆N-糖组相关，但与IgG N-糖组无关。该基因被证明是肝脏中岩藻糖基化的主调节因子[54]。参考文献[21]对血浆N-糖组的GWAS进行了更详细的综述。

最近一项关于TF N-糖基化的GWAS [23]，尽管样本量相对较小（发现样本量 $N = 948$ ；复制样本量 $N = 952$ ），却带来了重要的概念性进展。在全基因组显著性水平的8个基因位点中，有一个位点（包含TF本身）在以前的

N-糖基化研究中没有被报道过。与TF N-糖基化相关的5个基因位点含有编码糖基转移酶的基因，这些基因已经参与控制总血浆（*B3GAT1*、*MGAT5*和*ST3GAL4*）或IgG和总血浆（*FUT8*和*FUT6*）的N-糖基化。其他共享位点包括*HNF1A*（TF和总血浆N-糖基化共享）和*NXPE2/4*（TF和IgG N-糖基化共享）。在对发现样本和复制样本进行荟萃分析后，又有两个基因位点达到了全基因组意义。其中一个重要的基因位点包含一个编码叉头转录因子家族成员的基因*FOXJ1*，另一个区域位于基因*MSR1*附近。

Landini等[23]还专门研究了与TF和IgG N-糖基化都相关的*FUT8*和*FUT6*区域的TF和IgG N-糖基化共定位关联模式。他们的结论是，这些位点与TF和IgG N-糖基化的关联很可能是由不同的致病变异体造成的。参考文献[23,51]中报道的与人类蛋白质糖基化变化相关的6个基因位点包含一个以上的独立相关变体（图1中用星号标出）。

总之，9项已发表的GWAS确定了34个与TF、IgG或总血浆蛋白质组N-糖基化变异相关的重复位点。

从方法论的角度看，根据上述结果，可以得出结论：只要有足够大的复制样本，GWAS发现的糖基因位点就有

很高的复制率。例如，就血浆N-糖组而言，在全基因组意义上报道的16个位点中，有15个位点得到了重复，有一个位点的重复没有结果[57]；而在33个IgG N-糖基化位点中，迄今为止有29个位点得到了重复[52]。这些结果凸显了GWAS结果的普遍稳健性，以及研究蛋白质N-糖基化遗传调控的群体所采用的高标准。

至少在单一蛋白IgG N-糖基化研究中，还可以得出结论：基因分析的多元方法显示出更强的能力。对IgG N-糖基化的研究经验表明，应用多变量分析相当于将样本量增加20%~50% [49,52]。这种分析能力的显著提高可能是由于在不同的聚糖之间观察到了很大的相关性，这种相关性不仅很大，而且具有生物学意义[58]，这可能为多变量分析提供了一个理想的案例[59]。

图1概述了已确定的基因位点。从图中可以明显看出，IgG N-糖基化研究中建立的基因位点占主导地位，血浆-糖组特异性基因位点的缺失也很明显。前者可能只是统计上的假象，而后者则可能反映了真正的生物学。

事实上，IgG N-糖基化研究的样本量已达8000个，而血浆糖组和TF N-糖基化研究的最大样本量分别为3500个和1900个。根据经验可知，对于一个复杂的性状，发现

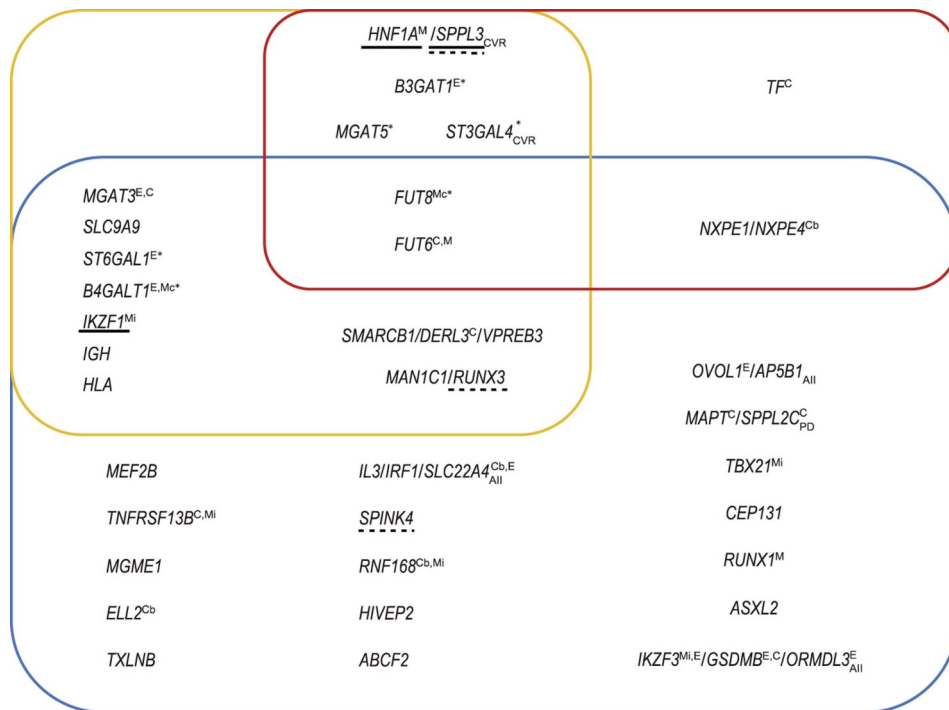


图1. 与N-糖基化密切相关的34个位点概述。基因位点以所提出的候选基因命名。左上方（黄色矩形框）：与血浆糖组相关的位点；右上方（红色矩形框）：与TF N-糖基化相关的位点；下方（蓝色矩形框）：与IgG N-糖基化相关的位点。蓝色的基因名称为糖基转移酶基因，绿色的基因名称为转录调节基因，黑色的基因名称为其他相关基因；黄色矩形框表示机制尚未被假设，所标示的基因为最近的基因。带下划线的基因具有实验支持的因果关系[实线：经典检测；虚线：基于HEK-293F细胞的瞬态系统，其中稳定整合了聚类规则间隔短回文重复（CRISPR）/失活的CRISPR相关核酸内切酶9（dCas9）]。指数越低，表示该基因位点与以下基因位点共定位，即All：自身免疫或炎症性疾病；CVR：心血管疾病或其风险因素；PD：帕金森病。上标表示有证据表明该基因位点的变异与该基因的功能有关。C：基因中的编码变异；Cb：预测为良性的编码变异；E：与该基因的eQTL共同定位；M：已知相关的单基因表型；Mc：先天糖基化紊乱的单基因表型；Mi：免疫缺陷的单基因表型。标有*的基因座包含一个以上独立相关的遗传变异。

的基因位点数量与样本大小大致成正比[60–61]。因此，可以预计，IgG N-糖基化（29个基因位点）的基因位点数量大约是血浆糖组（15个基因位点）的两倍，是TF N-糖基化（8个基因位点）的4倍，事实的确如此。此外，可以推测，对于所有三个研究的糖组结果，每个固定样本大小的基因位点数量大致相同，这可能反映了它们的遗传效应大小分布之间的相似性，以及遗传和环境变异所起的相对作用。

迄今为止，任何已知的血浆总N-糖基化（复制）位点都可以作用于IgG和（或）TF的N-糖基化，这一观察结果很可能反映了基本的生物学原理。这一观察结果证实了一个假设，即血浆总糖组或许可被视为由两个主要贡献者，即淋巴组织（主要是浆细胞）和肝脏（肝细胞），分泌到血液中的蛋白质糖组叠加，并主要由这些组织中的N-糖基化调节的特殊性所驱动。然而，其他组织的糖基化机制也可能对血浆蛋白的糖基化产生影响[62]。

4. 已确定位点的候选基因

尽管复制将一个基因位点确定为进一步研究的可靠起点，但无论是GWAS还是复制都无法回答基因座的变异导致性状变异的潜在作用机制。LD的存在限制了用经典统计遗传方法绘制功能多态性图谱的精确性。通过GWAS，可以建立一个多态性列表，其中包含的致病多态性的概率很高；但是，对于一个基因座来说，这样一个列表可能包括数十个甚至数百个候选多态性。如果基因座的效应足够大，且该区域的LD不过度，则只有超大规模的研究才能将关联映射到每个基因座的几个SNP上[63]。然而，即使有可能将列表限制在单个致病变异上，这仍然很少能直接回答该多态性的生物学作用机制是什么。也许可以明确地确定受影响的基因靶点，但只有当变异被可靠地预测为改变了该基因编码的蛋白质序列时才有可能。与此同时，复杂性状所涉及的大多数遗传变异都是调节性的[64]。虽然通常情况下，与关联性最强的变异最接近的基因确实是该变异的靶基因[65]，但要将调控变异与受调控的基因或基因集联系起来通常并不简单。例如，受相关SNP干扰的功能元件可能会影响相隔数百万碱基对的靶标，而且可能只在特定组织中发挥作用[47]。更离奇的是，一个功能基因可能通过反式调控机制作用于位于不同染色体上的基因[66–67]。

要将GWAS的统计结果转化为可操作的生物学知识，关键是要确定可解释关联的致病多态性和基因。有针对性的功能实验是确定作用机制的黄金标准；然而，鉴于一个

基因座可能存在多种假说，*in silico*功能调查在确定可能机制的优先次序方面发挥着重要作用。*in silico*方法利用了DNA变化的潜在后果[68–69]和不同基因组元素的功能作用[70–71]方面积累的大量知识，以及在组学研究（包括研究组学性状的遗传控制）中积累的数据[72–76]。后者的数据在假设特定关联的潜在分子遗传机制时尤其有用。基于摘要统计的共定位分析方法[74,77–78]可用于解决以下问题：两个性状与同一遗传区域的关联是否可能是由于同一遗传变异的多效性作用，或者是两个性状与两个不同的（尽管很接近的）多态性关联的结果。整合相关信息的（在线）数据库和工具为基于摘要统计的*in silico*研究提供了便利[79–83]。

在本节中，根据提出的候选基因的功能，对34个与蛋白质N-糖基化可复制相关的位点进行了分类。有6个基因位点无法提出候选基因；在这种情况下，该区域以最接近最相关多态性的基因命名（图1中的基因名称为黄色）。对于5个基因位点，已知有多个具有不同功能的候选基因，这些基因被分为几类。需要注意的是，原则上，对于复杂性状而言，与一个位点相关的基因不一定是单一的致病基因，因为一个位点上的多个候选基因并不一定相互排斥。

8个基因座包含糖基转移酶（图1中蓝色为基因名称），作为各自基因位点的主要候选基因。研究发现，在8个糖基转移酶中，*MGAT3*、*ST6GAL1*和*B4GALT1*等区域的遗传变异与IgG和总血浆N-糖基化的变化有关；*B3GAT1*、*MGAT5*和*ST3GAL4*附近的遗传变异与TF和总血浆N-糖基化有关；*FUT8*和*FUT6*附近的遗传变异与所研究的所有三种糖蛋白有关。大多数候选糖基转移酶在肝细胞和浆细胞中都有转录[图2（a）[84]]，但也有几个例外。

*MGAT3*优先在浆细胞中表达，而*MGAT5*则在肝细胞中表达[图2（a）]。这与以下事实是一致的：IgG连接的聚糖是双触角的，可选择性地呈现平分型GlcNAc（*MGAT3*编码的酶促进了这一反应），而由肝脏产生的糖蛋白连接的聚糖通常是多链的（*MGAT5*在聚糖分支中发挥作用）。 α -2,3唾液酸转移酶编码的*ST3GAL4*在肝脏中的表达量较高，而在浆细胞中的表达量极低，这与该基因座与浆细胞和TF有牵连，但与IgG N-糖基化GWAS没有牵连这一事实是一致的；而IgG的 α -2,6连接的唾液酸化作用几乎是众所周知的事实[85]。与浆细胞相比，半乳糖基转移酶基因*B4GALT1*在肝脏中的表达更强，这似乎与半乳糖存在于IgG和肝脏糖蛋白N-聚糖的观察结果不一致，而且该基因座只与IgG N-糖基化有关。但需要注意的是，

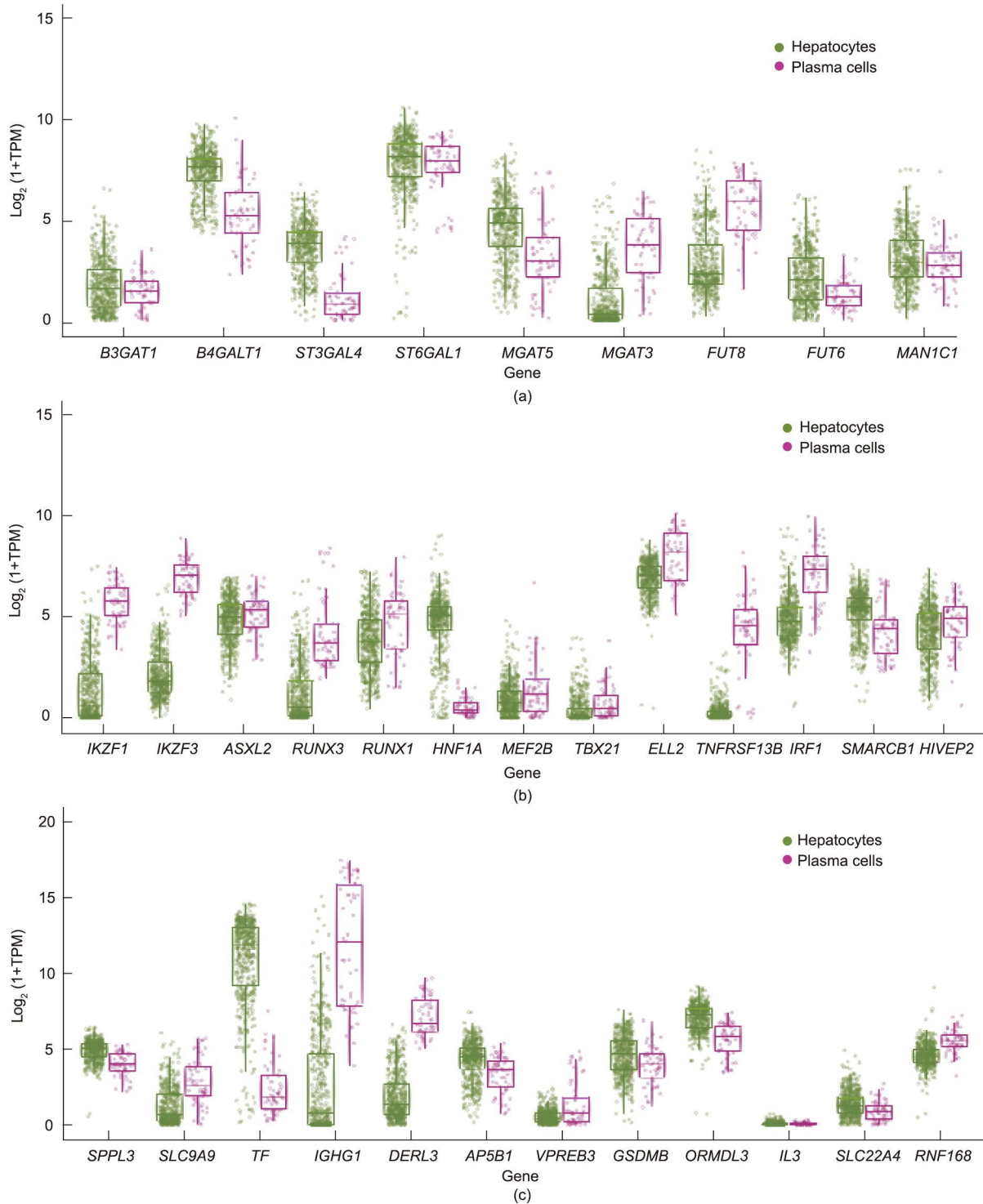


图2. 候选基因在肝细胞（绿色）和骨髓浆细胞（粉红色）中的表达。候选基因按功能分组：(a) 糖基转移酶和糖苷酶；(b) 转录调节因子；(c) 其他。TPM：每百万转录本。经许可转载自参考文献[84]，其方法和编码与参考文献[23]中生成图5的方法和编码相同。

B4GALT1 在浆细胞中的表达虽然低于肝细胞，但仍然相当高，即高于 *MGAT3* 在浆细胞中的表达水平。其次，应该考虑基因调控的组织特异性可能会导致在一个组织中存在糖组学数量性状基因座 (QTL)，而在另一个组织中却不存在。后者也可以解释为什么 *ST6GAL1* 附近的区域似

乎只与 IgG N-糖蛋白的变异有关，尽管它在肝细胞和浆细胞中的表达量同样高，并且编码一种唾液酸转移酶（促进附着在肝脏分泌蛋白质和免疫球蛋白上聚糖中 α -2,6 键合）。

除了糖基转移酶外，另一个基因座（包括 *RUNX3* 候

选基因) 包含编码糖苷酶的 *MAN1C1* 基因。该基因的表达组织特异性较低, 其编码的蛋白质参与了高尔基体从高甘露糖结构向混合聚糖结构过渡的早期阶段。鉴于参考文献 [51] 中的功能网络结果以及相关的糖型, 该基因座的致病基因更有可能是 *RUNX3*。

一大批优先候选基因编码 (组织特异性) 转录调节因子 (图 1 中的基因名称为绿色)。就血浆和 TF N-糖基化而言, *HNF1A* (与另一个候选基因 *SPPL3* 位置相近) 就是这样一个基因。在 IgG N-糖基化方面, 有 8 个位点含有候选基因, 这些基因编码参与 B 细胞分化和运转的转录因子。这些基因包括 *RUNX1* [86]、*RUNX3*、*IKZF1* [87] 和 *IKZF3* [88], 它们在 B 细胞成熟和分化中起核心作用; 基因 *MEF2B* 在约 15% 的滤泡淋巴瘤和弥漫性大 B 细胞淋巴瘤中发生突变, 在生殖中心发育中发挥作用 [89]; 基因 *ASXL2* 的产物的靶基因与 *RUNX1* 的靶基因高度重叠, 是正常造血所必需的 [90]; 转录延伸因子 *ELL2* 指导浆细胞中免疫球蛋白的分泌 [91]。 *TNFRSF13B* 基因位于最近确定的一个基因座上 [52], 它编码肿瘤坏死因子受体超家族中淋巴细胞特异性的成员, 称为跨膜激活剂及钙调亲环素配体相互作用 (TACI) 蛋白。通过该蛋白发出的信号可激活多种转录因子, 包括活化 T 细胞核因子 (NFAT)、激活蛋白 (AP) -1 和核因子- κ B (NF- κ B) [92]。TACI 在体液免疫中起着至关重要的作用, 它与其他信号网络协同作用, 促进 B 细胞分化为浆细胞并产生免疫球蛋白 [93–94]。

许多编码转录因子的候选者在浆细胞和肝细胞中的表达不同 [图 2 (b)]。在浆细胞中表达较强的一组基因包括 *IKZF1*、*IKZF3*、*RUNX3*、*TNFRSF13B* 和 (可能的) *IRF1*。 *HNF1A* 基因在浆细胞中的表达量极少, 但在肝细胞中的表达量很高。这种表达模式与 GWAS 的结果一致, 但也可能部分受其影响: 事实上, 第一组基因是在 IgG N-糖基化的 GWAS 中发现的, 而含有 *HNF1A* 的基因座只与血浆和 TF N-糖基化有关。

有三个基因位点含有的基因与 B 细胞免疫有关。这些基因包括 IgG 重链基因座 (14q32.33) [毫不奇怪, *IGHG1* 在浆细胞中表达很强, 但在肝细胞中却没有表达; 图 2 (c)]; *VPREB3* 基因 (靠近 *DERL3* 和 *SMARCB1*) 编码一种蛋白质, 该蛋白质很可能参与前 B 细胞受体的运输和组装 [95]; 以及最近提出的基因 *RNF168* [52], 该基因编码一种 E3 泛素连接酶蛋白, 参与 DNA 双链断裂修复和免疫球蛋白类开关重组 [96]。

5 个基因位点含有涉及 (非 B 细胞) 免疫功能的基因: *HLA* 区域, 该区域含有多个免疫相关基因; 一个基因位点含有编码 T-box 转录因子 (TBX21) 的基因, TBX21

的小鼠同源物可调节干扰素 γ (IFN γ) 的表达并控制 T 辅助细胞 1 (Th1) 系的形成 [97]; 基因 *HIVEP2* [98–99]; 以及 *IRF1*, 该基因编码干扰素调节因子 1。包括 *IRF1* 在内的区域还包括编码白细胞介素 3 的基因 *IL3*, 白细胞介素 3 是一种细胞因子, 在造血细胞系和淋巴细胞系的分化和增殖中具有重要功能; 编码 OCTN1 的 *SLC22A4*, OCTN1 是麦角硫因 (巯基组氨酸三甲基内盐) 的主要转运体, 麦角硫因是一种具有抗氧化和抗炎特性的氨基酸 [100]。与 B 系特异性转录因子 *IKZF3* 位于同一基因座的还有 *GSDMB*, 该基因编码的蛋白家族成员在细胞焦亡过程中发挥作用 [101–102]; 以及 *ORMDL3* 基因, 该基因的产物已被证明可通过控制抗原结合时内质网中钙离子的流入来调节 T 淋巴细胞的活化 [103]。

4 个基因位点含有涉及内质网、高尔基体或核内体功能的基因。这些基因如下: *SLC9A9* 基因, 该基因通过调节核内体的 pH 值来影响蛋白质糖基化; *DERL3* 基因 (靠近 *VPREB3* 和 *SMARCB1*), 该基因参与内质网中错误折叠糖蛋白的降解, 并优先在浆细胞中表达 [图 2 (c)]; *SPPL3* 基因 (靠近 *HNF1A*), 该基因编码一种膜内蛋白酶 (可裂解糖苷酶和糖基转移酶) [104]; 以及最近确定的含有 *AP5B1* 的基因座 [52], 该基因在骨髓中表达丰富, 编码适配蛋白复合物 AP-5 的一个亚基 [105], 可能参与促进晚期核内体到高尔基体的检索 [106]。

在同一功能组中, Frkatovic 等 [27] 提出了候选基因 *SPPL2C* (靠近 *MAPT*)。该基因编码一种膜内裂解天冬氨酸蛋白酶, 会影响囊泡转运, 并可能导致货物蛋白滞留在内质网中 [107], 从而可能通过聚糖修饰酶的错误定位而影响蛋白质糖基化等细胞变化过程。同一基因家族的另一个成员 *SPPL3* 的表达具有较低的组织特异性, 已知它能调节细胞的 N-糖基化 [104, 108]。然而, *PPL2C* 的表达富集于睾丸, 在其他组织中的表达量非常低。这一组的另一个可能候选者是由 *SPINK4* 基因编码的丝氨酸蛋白酶抑制剂。然而, 该抑制剂在肠道中的表达量较高, 在淋巴组织中的表达量有限; 其可能的作用机制仍不清楚, 实验验证也不具有完全的确定性 [109]。鉴于有关 *SPPL2C* 和 *SPINK4* 基因可能的作用的信息有些相互矛盾, 图 1 中用黄色标出了这两个基因。

有两个基因不属于上述任何一类。一个是 *SMARCB1* (靠近 *VPREB3* 和 *DERL3*), 它也被称为 *BAF47*。该基因编码三磷酸腺苷 (ATP) 依赖性染色质重塑复合物的核心亚基, 在基因转录调控中发挥关键作用 [110]。最后, 对于 TF N-糖基化, *TF* 基因本身就是参考文献 [23] 确定的基因座中的候选基因。根据已知的生物学原理, *TF* 基因在肝

细胞中强烈表达，而在浆细胞中却没有表达[图2 (c)]。

总之，即使对候选基因的描述略显肤浅，也可以清楚地看出，蛋白质N-糖基化的种群变异是由影响①参与N-聚糖的岩藻糖基化、分支和末端修饰的糖基转移酶，②参与产生相关糖蛋白的细胞（肝细胞和浆细胞）的分化、成熟和调节的组织特异性调节因子，③参与分泌糖蛋白的产生和转运的细胞机制（也许更具推测性），以及④所研究的糖蛋白（TF和IgG）本身的产生和结构的遗传变异来调节的。

5. 得到有力支持的候选基因

对于上一节讨论的基因是否确实是决定N-糖基化与已确定基因位点遗传变异之间联系的基因，人们的信心差别很大。例如，对于含有糖基转移酶的基因座（也许除了*B3GAT1*），证据是相当完整和令人信服的。众所周知，这些糖基转移酶参与蛋白质的N-糖基化，与基因座相关的N-糖基化变化与候选基因编码的产物的已知酶活性一致。少数基因对N-糖基化的影响已在细胞系和（或）实验动物模型中得到验证。有些由基因的突变导致的单基因表型包括N-糖基化（或可能与N-糖基化的变化有关）。然而，有些基因是基于已知参与同一类过程（如免疫系统功能）而提出的。对于这些候选基因，可信度较低。此外，即使已确定的基因座中存在优秀的候选基因，除非能提出并验证相关变异与该基因功能的关联机制，否则对该基因座作用机制的了解仍是不完整的。

在本节中，选择性地总结了从单基因表型和（或）实验研究中得到有力支持的候选基因的证据；讨论了潜在的作用机制，并强调了可能存在的 inconsistence 之处。关键事实和参考文献见表2 [23,51–52,54,56,108,111]。

如前所述，除了一个（*B3GAT1*）含糖基转移酶的基因座外，其他所有与基因座相关的N-糖基化变化都与候选基因编码产物的已知酶活性一致。对于半数糖基转移酶（*MGAT3*、*ST6GAL1*、*B4GALT1*和*B3GAT1*）来说，遗传变异转化为功能的可能机制是转录调控，这一点从糖组学GWAS与相应的eQTL之间的共定位中可以看出[51–52, 54,107]（图1和表2）。

对于由*FUT6*编码的一种糖基转移酶，编码变异可能解释了所观察到的与血浆和TF N-糖基化的关联[23,56]。事实上，在欧洲血统中，关联性最强的多态性rs7255720的C等位基因（频率为4%）[56]与rs17855739-T处于完美的LD ($r^2=1$)，后者编码*FUT6*催化结构域中的E247K取代失活，导致同型结合时（血浆）岩藻糖基转移酶6

（*FUT6*）缺乏[112]。在这种表型中，患者肝脏中产生的血浆糖蛋白缺乏 $\alpha3$ -岩藻糖基化[113]。否则，*FUT6*的有害等位基因可能会在某些人群中达到很高的频率，但不会产生明显的临床后果。

最近，Puan等[114]证实，*FUT6*缺乏会通过选择性削弱嗜碱性粒细胞的sialyl-Lewis^x表达而损害其功能；*FUT6*缺乏的嗜碱性粒细胞在E选择素上的滚动功能严重受损。研究人员还证明，*FUT6*空等位基因携带者会表现出血液中嗜碱性粒细胞数量升高，对昆虫叮咬的瘙痒敏感性降低，因此，*FUT6*缺乏会抑制外周嗜碱性粒细胞介导的过敏反应。IgE滴度降低和嗜酸性粒细胞数量减少的观察结果也支持这一结论。

含有*FUT6*和*FUT8*的基因座与IgG和TF的N-糖基化都有牵连，表明这些基因在不同组织的蛋白质的N-糖基化中起着重要作用[23,51]。然而，Landini等[23]的研究得出一个具有生物学意义的重要结论：一个“共同”基因座对两种糖蛋白的作用机制的细节可能并不一致，这反映在与每种糖蛋白的N-糖基化相关的不同致病体上。这一发现表明，尽管一些酶在多个组织中参与N-糖基化，但它们在组织中的遗传调控可能不同。

有人提出了含有*FUT8*的基因座的可能机制。众所周知，在浆细胞和肝细胞中，*FUT8*的表达分别受Ikaro家族锌指1（IKZF1）[51]和肝细胞核因子1- α （HNF1A）[54]转录因子的调控。*FUT8*区域的许多变体都是独立相关的，可以推测一些变体可能会影响HNF1A的结合，从而改变肝脏分泌的糖蛋白的岩藻糖基化，而另一些变体可能会影响IKZF1的结合，从而改变IgG的岩藻糖基化。因此，在这种情况下，最合理的生物学解释也许是，不同的致病体可能会影响不同组织中不同转录因子的结合，从而以组织特异性的方式调节组织特异性蛋白（即肝脏分泌的TF和浆细胞分泌的IgG）的糖基化。虽然这一解释符合转录调控的一般知识（即参考文献[74]），但这一假设的验证还需要进一步的*in silico*研究和实验功能验证。

就*FUT6*基因座而言，其机制至少应与上述*FUT8*基因座的机制有部分不同。在血浆糖基化和TF N-糖基化GWAS [23,56–57]中，关联性最强的变异与导致*FUT6*缺陷的失活E247K多态性（上文有更详细的描述）存在强LD。然而，在IgG N-糖基化GWAS中，关联性最强的SNP（rs35222081、rs874232和rs12986368）却截然不同[23,51–52]，它们与编码多态性存在较弱的LD ($r^2 < 0.1$)。失活*FUT6*基因突变与IgG N-糖基化不相关，可能说明在IgG N-糖基化的情况下，该区域的另一个基因负责相关性。

表2 候选基因与 GWAS 所涉及区域的映射证据摘要。表中仅列出了单基因表型和/或实验强烈支持的在 N-糖基化中发挥作用的基因。不耐受功能缺失(LoF)的概率(pLI)和观察到的(o)/预期的(e)是基因组聚合数据库(gnomAD)[111]的约束指标,pLI 值接近 1 和 o/e 值接近 0 表示约束更强。空白单元格表示在撰写文章时缺乏证据

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							
<i>MGAT5</i>	1.00	0.07 (0.03–0.18)	—	—	—	—	—
<i>ST3GAL4</i>	0.04	0.31 (0.17–0.61)	Yes, eQTL colocalization [56]	—	—	—	—
<i>FUT6</i>	—	—	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	—
<i>FUT8</i>	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	—
<i>MGAT3</i>	0.15	0.28 (0.14–0.63)	Yes, eQTL colocalization found in Ref. [51]	—	—	—	—
<i>ST6GAL1</i>	0.23	0.25 (0.12–0.57)	Yes, eQTL colocalization found in Ref. [51]	—	—	—	—
<i>B4GALT1</i>	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							
<i>HNF1A</i>	0.96	0.14 (0.07–0.33)	—	600496	Maturity-onset diabetes of the young, type 3 (MODY-3); autosomal recessive	Yes, directly	[54]
<i>IKZF1</i>	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]
<i>IKZF3</i>	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	—
<i>TNFRSF13B</i>	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	—
<i>TNFRSF13B</i>	0	1.73 (1.14–1.96)	Yes, eQTL colocalization found in Ref. [51]	609529	Immunoglobulin A deficiency, type 2 (IGAD2)	Perhaps, through IgA deficiency	—
<i>RUNX1</i>	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	—
<i>RUNX3</i>	0.84	0.13 (0.05–0.42)	—	—	—	—	[108]
<i>RNF168</i>	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.

B3GAT1 基因与血浆糖组和 TF N-糖基化的变异有关 [23,56], 关于该基因的证据有些矛盾。一方面, 支持该基因的证据很有力: 候选基因编码一种糖基转移酶, 其作用已通过共定位分析得到加强, 该分析表明, 同一功能变异可能解释了与血浆糖组以及 *B3GAT1* 在肝脏中的表达有关的关联 (见附录 A 中的图 S1)。然而, *B3GAT1* 是葡萄糖醛酸基转移酶基因家族的成员, 编码葡萄糖醛酸基转移酶, 它参与碳水化合物表位人类自然杀伤细胞抗原-1 (HNK-1) 的生物合成, 在末端 N-乙酰半乳糖胺 (LacNAc) 二糖上添加葡萄糖醛酸 (GlcA), 形成 HNK-1 表位前体 [115–116]。HNK-1 表位在人类淋巴细胞 (包括自然杀伤细胞) 的亚群上进行表达。该表位尚未被报道存在于血浆蛋白中。虽然 Huffman 等 [55] 报道了与 *B3GAT1* 区域相关的 DG13 血浆聚糖峰中的一些聚糖上存在 GlcA, 但这些结果尚未得到证实; 目前还不清楚这一观察结果是否能完全解释所观察到的关联。

位于 12q24 的基因座包含两个强大的候选基因 *SPPL3* 和 *HNF1A*。*SPPL3* 基因编码一种膜内天冬氨酰蛋白酶, 可裂解糖苷酶和糖基转移酶; 当该基因被抑制时, 会导致高糖基化 [104]。目前还不知道该基因的单基因表型。就 *HNF1A* 而言, Lauc 等 [54] 证实, 该基因及其下游靶标 *HNF4A* 可调控关键 FUT 和岩藻糖生物合成基因的表达, 表明 *HNF1A* 是驱动肝细胞中这些基因表达的必要条件和充分条件。研究人员认为, *HNF1A* 相互调节核心岩藻糖基化与双触角岩藻糖基化, *HNF1A* 的表达增加会抑制核心岩藻糖基化 (*FUT8*), 同时激活参与双触角岩藻糖基化的 FUT (如 *FUT6*)。与这一作用机制相一致的是, *HNF1A* 的罕见突变会导致成熟期发病的青年糖尿病 3 型 (MODY-3), 其特点是血浆三触角和四触角 N-聚糖的触角岩藻糖基化水平升高 [16,117], 血浆中 α 1-3,4 岩藻糖基化水平也普遍升高 [118]。12q24 中的基因变异与候选基因功能的关联机制尚未被提出。

两个基因位点包含属于 Ikaros 锌指转录因子家族的基因 *IKZF1* 和 *IKZF3*, 它们与 IgG-糖组相关 [51], 而 *IKZF3* 基因座也与总的血浆糖组相关 [56]。据报道, 这两个基因都存在导致单基因免疫缺陷的突变, 其特征是 B 细胞功能异常。*IKZF1* 基因突变导致常见变异性免疫缺陷-13, 这是一种常染色体显性原发性免疫缺陷病, 与低丙种球蛋白血症和 B 细胞数量减少有关 [119]。*IKZF3* 基因突变导致免疫缺陷-84 (IMD84), 这是一种常染色体显性原发性免疫缺陷病, 与 B 细胞水平低和早期 B 细胞发育受损有关 [120]。

Klarić 等 [51] 利用糖组学数据的多维性, 建立了一个

连接相关基因位点的网络。在这个网络中, 位置紧密的节点是具有相似相关聚糖谱的基因位点, 表明这些基因位点中的致病基因具有功能上的相似性。网络中存在的糖基转移酶可以通过一种有趣的方法进行功能推断。如果两个相连的基因位点含有编码糖基转移酶和转录因子的基因, 那么可以认为转录因子可能会调节糖基转移酶的表达。特别是, 从参考文献 [51] 中推导出的网络中可以看出, 含有 *IKZF1* 和 *IKZF3* 的基因位点可能会调控 *FUT8* (一种负责对附着 IgG 的 N-聚糖进行核心岩藻糖基化的酶)。此外, 研究人员还发现, 在 *IKZF3* 基因座中, 相同的遗传变异可能会同时影响 IgG 的岩藻糖基化和 *IKZF3* 在外周血中的表达。研究表明, 在淋巴母细胞系 MATAT-6 中敲除 *IKZF1* 会导致 *IKZF3* 表达减少、*FUT8* 表达增加、IgG 核心岩藻糖基化增加 [51], 这证实了上述假设。

在两个位点中, 属于 Runt 相关家族的基因, 即 *RUNX1* 和 *RUNX3*, 被强调为可能与 IgG N-糖基化有关的候选基因 [51]。这些基因编码转录因子, 在造血细胞分化中发挥重要作用 [121–122]。Mijakovac 等 [109] 提出, IgG 半乳糖基化和转录因子基因 *RUNX1* 和 *RUNX3* 受雌二醇调控。随后, 在 HEK-293FS 衍生细胞系 (该细胞系经过优化, 可分泌大量原生 IgG 抗体) 中使用成簇规律间隔短回文重复序列 (CRISPR) / 失活的 CRISPR 相关核酸内切酶 9 (dCas9) 分子工具进行分析, 结果证明, *RUNX3* 的上调会导致 IgG 半乳糖化水平的下降。尽管 *RUNX3* 在调控 IgG 糖组中的作用得到了充分支持, 但 *RUNX3* 与遗传相关的确切机制仍不清楚。

Shadrina 等 [52] 通过对 IgG 糖组进行多变量遗传关联分析, 发现了一个携带 *TNFRSF13B* 基因的基因座。*TNFRSF13B* 编码 TACI 蛋白, 它是肿瘤坏死因子受体超家族的淋巴细胞特异性成员。罕见的 *TNFRSF13B* 突变可增加哮喘症状的风险 [123]。此外, 据报道, 该基因的几个常见变异与循环 IgG 水平有关 [124–125], 而 *TNFRSF13B* 的编码变异与常见变异性免疫缺陷和选择性 IgA 缺乏有关 [126]。

在参考文献 [52] 中发现并复制的另一个基因座包含 *RNF168* 基因, 该基因编码一种 E3 泛素连接酶蛋白, 该蛋白参与 DNA 双链断裂修复和免疫球蛋白类开关重组 [96]。以前曾有报道称, *RNF168* 基因座的常见遗传变异与 IgM 水平有关 [125], *RNF168* 基因的罕见突变会导致常染色体隐性遗传疾病、免疫缺陷、畸形特征、学习困难 (RIDDLE) 综合征, 其特征包括免疫球蛋白生成减少 [127]。

另一个在参考文献 [23] 中被发现和复制的有力候选基

因是 *TF*。如上所述, *TF* 本身是一种具有两个 N-糖基化位点的 N-糖蛋白。*TF* 基因座的遗传变异与碳水化合物缺乏的 *TF* 水平[41]以及 *TF* 连接的 N-聚糖丰度有关[23]。*TF* 存在一个顺式蛋白质定量性状位点 (pQTL), 可解释约 40% 的人群血清中 *TF* 水平的变异[128]。rs8177240 [128] 标记的 pQTL 与哨点 *TF* N-糖基化 SNP rs6785596 [23] 的 LD 值较低 ($r^2 < 0.1$)。另一方面, 哨点 N-糖基化 SNP 与一个与碳水化合物缺乏 *TF* 水平相关的 SNP rs1799899 存在相当强的 LD ($r^2 < 0.57$) [41]。后一个 SNP 编码 *TF* 中的 G277S 氨基酸变化。可以推测, 这种氨基酸替代可能会改变蛋白质的三级结构和 N-糖基化位点的可及性。对于 Landini 等 [23] 报道的哨点 SNP 与 Kutalik 等 [41] 报道的 SNP 之间的 LD 是否能完全解释 *TF* N-糖基化与 *TF* 区域的关联, 目前尚无定论。

总之, 在与 N-糖基化变异相关的 34 个基因位点中, 约有一半可根据强候选基因的存在来推测其生物学作用机制。对于其中大约一半的基因位点, 还可以假设遗传变异与靶基因之间的关联机制。对于作用机制难以猜测的大量基因位点可以成为人类糖生物学知识的重要来源。

6. N-糖基化和复杂疾病的共同遗传决定因素——特定基因位点的多义性

一些与蛋白质糖基化相关的基因座也显示与复杂的人类疾病相关[51–52,56]。一个基因座与两个性状的关联可能是由于一个遗传变异对两个性状的多效性作用, 也可能是由于两个性状的不同遗传变异的独立效应。需要注意的是, 如果两个变异体之间存在完美的 LD, 后一种假设与前一种假设在统计上就没有区别了。鉴于这一基本限制, 人们开发了区分这两种假说的方法[77–78]。这些方法比较了与同一区域相关的两个性状的区域关联模式。如果存在足够的相似性, 则拒绝多效性假说。相似性则表明可能存在多效性。

如果一个基因位点对聚糖水平和疾病具有多效性作用, 则可将聚糖的选择作为疾病的生物标志物进行研究 [129]。在这一逻辑的基础上, 人们发现血浆蛋白的双触角岩藻糖基化水平升高是最常见的 MODY-3 的功能性生物标志物之一, 而 MODY-3 是由 *HNF1A* 基因突变引起的。在第一项血浆糖组 GWAS 中, 含有 *HNF1A* 基因的基因座与人类血浆 N-糖组的变化有关。在细胞系实验中, 该基因被证实改变了核心岩藻糖基化和双触角岩藻糖基化之间的平衡[54]。MODY-3 经常被误诊为 1 型糖尿病 (T1D)

或 2 型糖尿病 (T2D) [130]。已成功地利用双触角岩藻糖基化生物标志物将 MODY-3 患者与 T1D 和 T2D 患者区分开来[118,131] (图 3)。

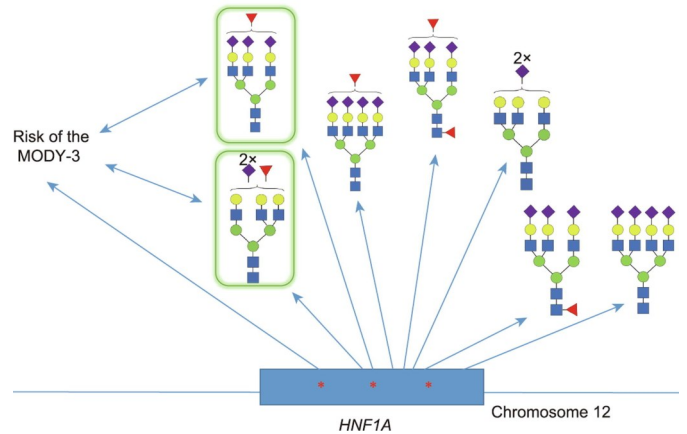


图 3. 通过 GWAS [56] 获得的聚糖性状与含基因 *HNF1A* 的位点之间的显著关联。方框内的聚糖是对 MODY-3 进行分类时表现最好的单一聚糖性状。该图根据参考文献[118,131]中报道的结果和数据绘制。

Klarić 等 [51] 证实了 IgG N-糖组相关基因座 17q12-21 (包含基因 *ORMDL3*、*GSDMB* 和 *IKZF3*) 对一系列自身免疫性和炎症性疾病的多效性作用。值得注意的是, 17q12-21 是最著名的哮喘相关基因位点之一 [132]。哮喘风险等位基因与无岩藻糖基化、半乳糖基化和单半乳糖基化聚糖水平降低以及核心岩藻糖基化聚糖水平升高相关 [51]。该位点还与自身免疫性疾病和炎症性疾病 (如类风湿性关节炎、克罗恩病、溃疡性结肠炎和原发性胆汁性肝硬化) 的发病风险存在多重相关性, 这一点已通过共聚焦分析得到证实。自身免疫性疾病风险等位基因与无岩藻糖基化、半乳糖基化和单半乳糖基化聚糖水平升高以及岩藻糖基化聚糖水平降低相关 [51]。与这些观察结果一致的是, 尽管在该基因位点上, 相同的 SNP 与哮喘和自身免疫性疾病相关, 但与自身免疫性疾病相关的风险等位基因和与哮喘风险相关的等位基因正好相反 [132]。

此外, Shadrina 等 [52] 证实了另一个包含基因 *OVAL1* 和 *AP5B1* 的 IgG-糖组相关基因座对哮喘的多效性作用。该基因座还被发现对皮肤病和过敏性疾病具有多效性作用 [133]。尽管很想推测 IgG N-糖基化与哮喘之间的关联或共同的生物学机制, 但一概而论可能有些为时过早。一些观察结果, 包括两个基因座与不同的聚糖谱和哮喘亚型的关联, 表明决定包括 *ORMDL3*、*GSDMB* 和 *IKZF3* 的基因座与包括 *OVAL1* 和 *AP5B1* 的基因座对哮喘和 IgG N-糖组的多效性作用的机制可能是不同的 (更详细的讨论见参考文献 [52])。

可以推测, 同一等位基因与自身免疫性疾病风险的关

联以及 *ORMDL3-GSDMB-IKZF3* 基因座上无岩藻糖基化半乳糖基化聚糖丰度的增加在生物学上是合理的，因为携带此类聚糖的 IgG 被认为具有促炎性[33]，并具有很强的诱导 ADCC 的能力[33]。与这一推测相一致的是，在 *IRF1-LC22A4* 基因座上，克罗恩病风险等位基因已被证实与无岩藻糖基化半乳糖基化聚糖水平的增加和岩藻糖基化水平的降低有关[51]。Klarić 等[51]的共定位分析也表明，在含有 *IRF7* 基因的基因座上，影响聚糖的变异与克罗恩病风险之间可能存在基因多效性，而在含有 *MAPT* 基因的基因座上，影响聚糖的变异与帕金森病风险之间也可能存在基因多效性。

总之，一些与 IgG N-糖基化相关的基因位点可能会对自身免疫性疾病和炎症性疾病的发病风险产生多重影响，尤其是炎症性肠道疾病和哮喘。此外，在 *MAPT* 基因座上还观察到与帕金森病的共定位。与 TF N-糖基化相关的 *HNF1A* 基因位点对 C 反应蛋白 (CRP)、低密度脂蛋白 (LDL) 和总胆固醇水平具有多效性作用[23]。

7. N-糖基化与复杂疾病的共同遗传决定因素——孟德尔随机化(MR)分析的结果

尽管在许多复杂的人类疾病中都观察到了蛋白质糖基化的变化[29]，但人们对糖基化与复杂疾病之间的因果关系知之甚少。在流行病学中，随机对照试验是检验风险因素与疾病之间因果关系的实验设计黄金标准。然而，在蛋白质糖基化的情况下，应用这种设计具有挑战性。在这里，遗传学提供了一种 MR 方法[80,134]，该方法利用与风险因素相关的遗传变异作为工具变量，模拟随机对照试验。如果遗传变异改变了风险因素的水平，并且这种改变对疾病结果产生了影响，那么该变异就提供了风险因素对疾病结果产生因果影响的证据。在这种情况下，可以建议将该风险因子作为治疗目标，从而改变疾病风险。反向因果检验也很重要，因为在某些情况下，疾病的表现对分子表型[如心血管风险和高密度脂蛋白 (HDL)]有因果影响[135]。在这种情况下，分子表型可能被认为是疾病的潜在生物标志物。值得一提的是，虽然 MR 背后的思想相对简单，但这里面有几个假设，违反这些假设可能会导致假阳性结果；因此，应谨慎对待 MR 的结果[134]。

迄今为止，MR 方法已被用于研究 IgG 糖组特征，12 种与 IgG N-糖组相关的常见疾病之间的因果关系[136–137]，以及总的血浆 N-糖组、IgG N-糖组与前列腺癌之间

的因果关系[138]。系统性红斑狼疮 (SLE) 对总 IgG N-糖组中具有平分型 GlcNAc 的 N-聚糖的丰度显示出因果效应[137]，表明这种聚糖特征是系统性红斑狼疮的生物标志物。分析结果相当可靠，因为使用了大量 (36 个) 与系统性红斑狼疮相关的基因变异、大样本量和一系列方法得出结论。附着在 IgG 上的平分型 N-聚糖的丰度与促炎状态有关[139]，在系统性红斑狼疮患者中也会升高[140]，但其机制尚不清楚。在系统性红斑狼疮患者中，由于 N-聚糖的分段和核心岩藻糖基化是相互竞争的过程，因此平分型聚糖水平的升高可能是 IgG 聚糖岩藻糖基化水平降低的结果[141]，尽管尚未观察到 IgG 岩藻糖基化与系统性红斑狼疮之间存在因果关系。在东亚血统的样本中，IgG 链接的高甘露糖型聚糖峰 5 (GP5; Man5) 和平分型的 GP13 显示出与 T2D 结果的因果关系[136]，这与已知的这些聚糖特征与 T2D 之间的相关性一致[142]。

同样的方法也被用于检测血浆总的 N-糖组和 IgG N-糖组特征对前列腺癌的因果效应[138]，结果表明，高甘露糖型聚糖 (Man9) 水平降低、唾液酸化二聚糖与唾液酸化三聚糖比例增加、四分支重半乳糖基化聚糖和唾液酸化聚糖 (附着于 α -1-酸蛋白——一种急性期蛋白) 水平增加会导致前列腺癌风险降低。然而，后者与所观察到的这种聚糖水平升高及低度炎症水平升高之间存在矛盾，因此需要在今后开展工作，以更好地了解两者之间的关系[143]。应该注意的是，所获得的结果是由一个或两个工具变量驱动的，表明可能存在微弱的工具偏倚[144]。需要更强大的 GWAS 来证实或反驳推测的关系。

总之，IgG 的 N-糖基化可作为系统性红斑狼疮等自身免疫性疾病的生物标志物，而血浆 N-糖组结果可作为内分泌疾病和癌症的生物标志物，这一假设得到了遗传学的支持。这一假设符合并证实了 (前瞻性) 流行病学研究的结果[29,145]。

8. 结论

在旨在发现人类健康和疾病机理的研究中，蛋白质糖基化是对其他层组学表型的补充。人类蛋白质的糖基化与许多人类疾病和病理状态有关，这使得糖基化结果成为生物标志物和药物靶点的丰富来源。与蛋白质组和代谢组相比，基于糖基的生物标记物和药物靶点的例子相对较少。部分原因是对人体内糖蛋白的调节机制了解不够。虽然人们对糖类生物合成的生化基础研究得很透彻，但对这一网络是如何被调控的却知之甚少，尤其是在以细胞、组织和环境特异性方式方面。

高通量糖组学、遗传学和基因组学的结合提供了一种“数据驱动假设”的方法，以不可知论的方式发现蛋白质糖基化的调控[20–21]。目前，对人类血浆、IgG和TF N-糖基化的遗传学研究已确定了与34个位点的遗传变异的关联性。通过对这些基因位点的候选基因进行分析，可以推断出这种遗传变异会影响参与N-聚糖的岩藻糖基化、分支和末端修饰的糖基转移酶，以及参与产生相关糖蛋白的细胞（肝细胞和浆细胞）的分化、成熟和调节的组织特异性调节因子。影响分泌糖蛋白的产生和转运的细胞机制以及所研究的糖蛋白（TF和IgG）本身的产生和结构的变异也变得越来越明显。在与N-糖基化变异相关的34个基因位点中，约有一半可根据强候选基因的存在来推测其生物学作用机制；此外，还可利用约一半的基因位点推测出遗传变异与目标基因的关联机制。对于大量难以假设作用机制的基因位点，可能会为人类糖生物学提供新的知识来源。有关作用机制的假设必须以功能性湿性实验室实验为目标。

为了建立和理解N-糖基化变异与人类健康和疾病的关联机制，人们付出了巨大的努力。研究表明，一些与IgG N-糖基化相关的基因位点可能会对自身免疫性疾病和炎症性疾病的风险产生多效性作用，而血浆N-糖组基因位点可能会对内源性心血管疾病风险因素产生多效性作用。这些位点的候选基因因为研究N-糖基化和复杂疾病的共同病因提供了一个起点。遗传学为以下假设提供了支持：IgG的N-糖基化可作为自身免疫性疾病的生物标志物，而血浆N-糖组结果可为内分泌疾病和癌症提供生物标志物。这些发现证实了流行病学研究的结果，这些研究已将部分N-糖组结果确定为心血管、代谢、自身免疫性疾病和癌症的风险因素和（或）生物标志物[29, 145]。不过，应该指出的是，糖原组学研究不仅支持而且还提出了人类疾病的新型生物标志物[16,48,137]。其中一些生物标志物已显示出很高的鉴别力，并正处于临床实践中[118]。

9. 展望

糖遗传学和糖原组学发展迅速。预计人类蛋白质糖基化遗传控制的进一步研究至少有三个方向。首先，与其他复杂性状的研究相比，人类N-糖组学遗传分析所用的样本量仍然相对较小，这限制了定量遗传学方法的应用。期待未来随着样本量的增加，GWAS的功能会更加强大。对于复杂的人类性状，预计样本量增加一倍将使发现的遗传位点数量增加大约一倍[60–61]。样本量的增加将会提高

对遗传效应的估计水平，从而为MR、共定位、遗传相关性等定量遗传学分析提供更可靠的输入，从而更好地理解糖组在疾病病因学中的作用。

其次，预计以单个蛋白质糖基化为重点的研究将会兴起。这些研究材料的可用性将取决于高通量糖组学技术的进步[22]。利用液相色谱-串联质谱（LC-MS/MS）等仪器将传统的糖组学研究扩大到大规模糖蛋白分析，对于推进功能性糖组学的发展非常重要，因为功能性糖组学可以高度详细地阐明蛋白质上附着的N-聚糖的位点、蛋白质和细胞特异性。因此，合乎逻辑的假设是，对特定GWAS基因座的作用机制的深入了解，至少部分取决于蛋白质组学和糖蛋白组学的发展。这些研究不仅能更好地解释所观察到的遗传关联，更深入地剖析组织特异性调控网络，而且还能促进临床应用。

最后，多组学方法的进步将补充和加强高通量糖组学和糖基化遗传学的进展。特别值得关注的是测量单细胞转录组的技术[146]。这种技术可以分解复杂的细胞混合物中的转录，从而使糖基化与特定——可能是稀有——细胞类型保持一致提供可能。测量的多组学样本的可用性将对人们了解人类糖生物学产生协同效应，并将改善糖组与细胞、组织和生物体的系统生物学的整合。多组学方法还将有助于发现新的生物标记物、药物靶点和治疗方法，并提供更好的证据。

因此，对蛋白质N-糖基化遗传控制的持续研究应能进一步加深人们对蛋白质糖基化生物学知识的了解，这将有助于开发新的方法来对抗糖蛋白依赖性自身免疫、癌症和其他疾病。

Acknowledgements

We are grateful to Anna Soplenskova 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

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, Wührer 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, Gloyd 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, Surolija 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, Rispen 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, Wührer 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 FcγRIII 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, Hafkenschheid L, Rispen 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] Kotalik 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, Klaric 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, Muẑinić 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 HNF1 α 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, Umičević 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, Hormozdiani F, et al.; ConsortiumGTEx. 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] Stattalo 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. FATHMMXF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics* 2018;34(3):511–3.
- [70] Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores 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, Bilenyk 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.; ConsortiumMUTHER. 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; ConsortiumGTEx, LDACC—Analysis Working Group, Statistical Methods groups—Analysis Working Group, groupseGTEx, Common FundNIH, 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, AnalystsLead, LDACC, Program ManagementNIH, CollectionBiospecimen, 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, Charleatoux 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, Styrismisdottir 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 Selektcii* 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, Elkhali 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 JY, 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.; ConsortiumgnomAD. 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 α 3-fucosylation of human α 1-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] Juszcak 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, Juszcak A, Štambuk T, et al. Development of an exoglycosidase plate-based assay for detecting α 1-3, 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, Mitsui 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, Hurler 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 randomization 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); β 1, 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, Genetics Collaboration CRP CHD. 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.