Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases

Hiren Joshi, Lars Hansen, Yoshiki Narimatsu, Hudson Freeze, Bernard Henrissat, Eric Bennett, Hans Wandall, Henrik Clausen, Katrine Schjoldager

Glycobiology, 2018, vol. 28, no. 5, 284–294
doi: 10.1093/glycob/cwy015
Advance Access Publication Date: 22 March 2018
Original Article

OXFORD

Genetic Disorders of Glycosylation

Glycosyltransferase genes that cause monogenic congenital disorders of glycosylation are distinct from glycosyltransferase genes associated with complex diseases

Hiren J Joshi[1,2], Lars Hansen[1,2], Yoshiki Narimatsu[2], Hudson H Freeze[3], Bernard Henrissat[2,4], Eric Bennett[2], Hans H Wandall[2], Henrik Clausen[2], and Katrine T Schjoldager[2]

[2]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark, [3]Human Genetics Program, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037, USA, and [4]Architecture et Fonction des Macromolécules Biologiques, Centre National de la Recherche Scientifique (CNRS), Aix-Marseille University, F-13288 Marseille, France

[1]To whom correspondence should be addressed: Tel: +45-35-33-5504; e-mail: joshi@sund.ku.dk (H.J.J.); Tel +45-35-33-5499; e-mail: lah@sund.ku.dk (L.H.)

Received 16 November 2017; Revised 15 January 2018; Editorial decision 15 February 2018

Abstract

Glycosylation of proteins, lipids and proteoglycans in human cells involves at least 167 identified glycosyltransferases (GTfs), and these orchestrate the biosynthesis of diverse types of glycoconjugates and glycan structures. Mutations in this part of the genome—the GTf-genome—cause more than 58 rare, monogenic congenital disorders of glycosylation (CDGs). They are also statistically associated with a large number of complex phenotypes, diseases or predispositions to complex diseases based on Genome-Wide Association Studies (GWAS). CDGs are extremely rare and often with severe medical consequences. In contrast, GWAS are likely to identify more common genetic variations and generally involve less severe and distinct traits. We recently confirmed that structural defects in GTf genes are extremely rare, which seemed at odds with the large number of GWAS pointing to GTf-genes. To resolve this issue, we surveyed the GTf-genome for reported CDGs and GWAS candidates; we found little overlap between the two groups of genes. Moreover, GTf-genes implicated by CDG or GWAS appear to constitute different classes with respect to their: (i) predicted roles in glycosylation pathways; (ii) potential for partial redundancy by closely homologous genes; and (iii) transcriptional regulation as evaluated by RNAseq data. Our analysis suggest that more complex traits are caused by dysregulation rather than structural deficiency of GTfs, which suggests that some glycosylation reactions may be predicted to be under tight regulation for fine-tuning of important biological functions.

Key words: GALNT, gene regulation, glycogenome, glycosyltransferase, mutation

Introduction

Glycosylation of proteins, lipids and proteoglycans (glycoconjugates) in mammalian cells is directed by a large number of glycosyltransferases that each serve unique functions in building the diverse set of glycan structures produced—designated the glycome (Cummings 2009; Hansen et al. 2015). A subset of glycosyltransferases (GTfs) can be grouped into a set of enzymes that are involved in initiating glycoconjugates, usually catalyzing the first attachment of a monosaccharide to a protein backbone or lipids. An exception to this activity is found with the oligosaccharyltransferase complex, that initiates N-glycosylation of proteins by transferring a preformed oligosaccharide. The initiation step is followed by groups of enzymes that elongate, branch and cap growing oligosaccharides on glycoconjugates. The human genome contains over 214 GTf-genes (Hansen et al. 2015), of which 167 can be predicted to direct fairly specific biosynthetic steps in glycosylation pathways of proteins, lipids and proteoglycans as illustrated in Figure 1 (Narimatsu et al. 2018).

Glycans serve important functions in essentially all biological processes in a cell and organism (Varki 2017), and genetic deficiencies in GTf-genes and glycosylation capacities have been shown to cause/underlie over 58 rare congenital disorders of glycosylation (CDGs) (Table I) (for reviews see Freeze et al. 2014; Jaeken and Péanne 2017). Most of these CDGs have so far been found to be in the N-glycosylation pathway (Freeze et al. 2014), the O-Man glycosylation pathway directed by the *POMT1/T2* genes (Sheikh et al. 2017), and the proteoglycan glycosylation pathway (Mizumoto et al. 2014), although deficiencies in the GPI-anchor assembly (Freeze et al. 2015; Ng and Freeze 2015) and other types of O-glycosylation have also emerged (Jaeken and Péanne 2017). Most of the identified CDGs to date are caused by severe deficiencies in GTf functions as result of structural defects in the coding regions and/or splicing, and with global effects on the glycosylation capacity and resulting glycan structures. However, CDGs caused by more subtle glycosylation deficiencies are emerging with GTf genes that are members of large homologous gene families, where paralogs may provide partial compensation. The most illuminating examples are members of the large polypeptide GalNAc-transferase (*GALNT2* and *3*) gene family (GT27) (Kato et al. 2006; Khetarpal et al. 2016), the β4Gal-transferase family (GT7) (*B4GALT1*) (Hansske et al. 2002), and the large sialyltransferase family (GT29) (*ST3GAL3* and *5*) (Hu et al. 2011; Edvardson et al. 2013; Simpson et al. 2004), where deficiencies in individual genes are predicted to lead to subtle partial losses of glycosites or glycan structures that only affect a subset of glycoconjugates and/or specific cell types.

Deleterious single-nucleotide variants (SNVs) in GTf-genes are rarely found in whole exome sequencing (WES) of different populations, which we recently confirmed by analysis of deep WES data from 2000 Danes (Hansen et al. 2015). In this study, we only identified a few alleles with known disease-causing SNVs from *POMGNT1*, *ALG1*, *ALG6* and *ALG12*. Interestingly though, we did identify two less rare deleterious SNVs in 2 out of the 20 *GALNT* genes (*GALNT5* and *T14*) initiating O-GalNAc-glycosylation, although the role of these SNVs are still unknown.

However, Genome-Wide Association Studies (GWAS) increasingly associate a large number (~98) of the human GTf-genes with different complex traits or predisposition to diseases (Table I). While GWAS candidates clearly require validation, a cursory survey of the literature suggested that most of the implicated GTf-genes were different from the classical group of identified CDG genes. Many GWAS candidate GTf-genes are members of large homologous gene families (such as GT10, GT27, GT29 and GT31) with poorly characterized nonredundant functions (see Supplementary data Table SIV for references), and with potential for functional overlap by isoenzymes. This redundancy may cloud our understanding of the nonredundant functions of individual enzymes, rendering it extremely difficult to rationally predict how these genes may be implicated in the particular candidate GWA traits. We hypothesize that this partly explains the lack of attention to the potential significance of GTfs in more common complex traits or disease conditions. Recently, we therefore pursued one of the most substantiated GWAS candidate genes, *GALNT2*, with a GWAS predicted role in regulating high-density lipoprotein (HDL) and triacylglycerol (Kathiresan et al. 2008; Teslovich et al. 2010; Willer et al. 2013). *GALNT2* encodes one of a family of up to 20 polypeptide GalNAc-transferase isoenzymes controlling initiation of GalNAc-type (mucin-type) O-glycosylation (Bennett et al. 2012; Steentoft et al. 2013). We were able to confirm that loss of *GALNT2* in fact does cause reduced HDL in two very rare consanguineous cases of complete deficiency of this gene, as well as in several animal models (Khetarpal et al. 2016). In addition, we identified two *GALNT2* isoform-specific protein substrates, ANGPTL3 and PLTP, for which loss of glycosylation in both cases could affect HDL metabolism (Schjoldager et al. 2012; Khetarpal et al. 2016). Importantly, however, the GWAS signal for *GALNT2* and low HDL is located in the first large intron of the gene close to a liver-specific regulatory element, and several studies have demonstrated that the GWAS single nucleotide polymorphism (SNP) signal induced allele-specific transcription differences (Roman et al. 2015; Cavalli et al. 2016). Thus, *GALNT2* underlies a new rare CDG caused by complete loss-of-function due to biallelic structural deficiency in the gene, and simultaneously *GALNT2* is perhaps the first validated GWAS candidate GTf-gene. Importantly, the molecular basis for the latter appears to be organ-specific dysregulation of transcription rather than structural, and the phenotypic outcome remains altered HDL metabolism. These findings suggest that GWAS point to a different group of disorders of glycosylation caused by transcriptional dysregulation.

In the present study, we performed a systematic survey of GTf-genes identified as CDG genes and those implicated as GWAS candidates. This showed that GTf-genes associated with CDGs mainly functioned in glycosylation reactions specific to a single glycosylation pathway, they were rarely part of close homologous families of putative isoenzymes, and they were not predicted to exhibit substantial organ-specific regulation. In striking contrast, GTf-genes implicated as GWAS candidates often functioned in multiple glycosylation pathways, and they were mainly members of isoenzyme families, and predicted to exhibit organ-specific regulation. Our analysis suggests the existence of two distinct types of congenital glycosylation defects—one based on well-established and rare structural gene defects (CDGs)—and one less well studied based on regulatory gene effects (GWAS). Although the many GWAS candidate genes will require validation, we predict that this group points to glycosylation features that require tight regulation to serve in fine-tuning and specific co-regulation of protein functions.

Results

A global view of the GTf-genome

An assembly of the human GTf-genome was based on the CAZy database annotation of GTfs (Lombard et al. 2014). Some 214

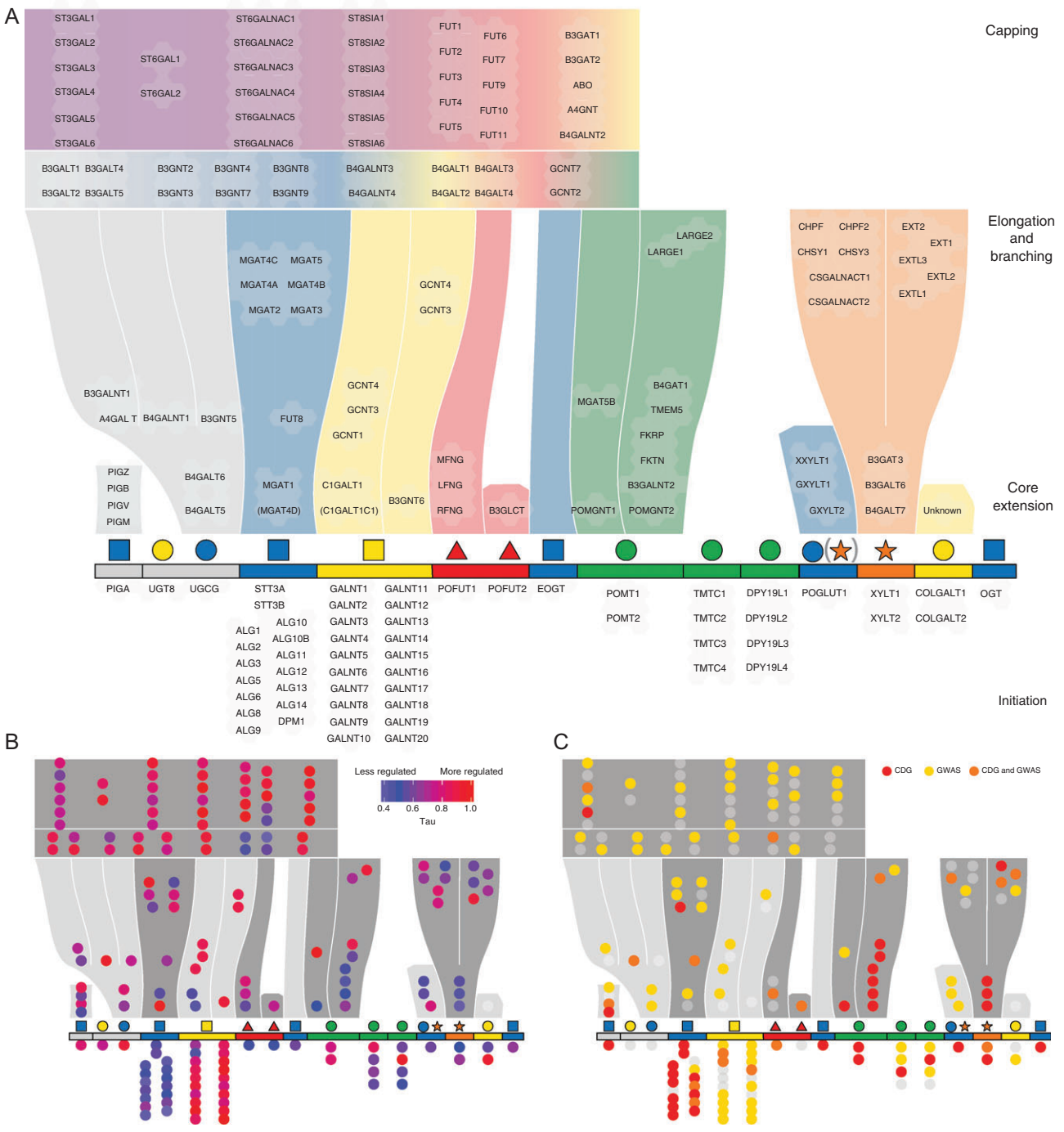

Fig. 1. (**A**) Graphic rainbow depiction of predicted roles of the 167 human GTfs involved in glycosylation. *Pathway-specific* GTf-genes (114) serving the 15 distinct glycosylation pathways are organized as vertical colored flows. *Pathway-nonspecific* GTf genes (53) serving elongation and capping are organized in horizontal layers at the top. The glycosylation pathways are: GPI-anchor, glycolipids (two pathways), N-linked glycans, O-GalNAc mucin-type, O-Fuc type (two pathways), O-GlcNAc type (EGF), O-Man type (POMT-directed), O-Man type (TMTC-directed), C-Man type, O-Glc type, O-Xyl type (proteoglycans), O-Gal type (collagen) and O-GlcNAc type (cytosolic). GTf-genes directing the initiation and pathway determining steps are indicated at the bottom. For N-glycosylation the *ALG1-14* GTf-genes required for the preformed oligosaccharide (Glc3Man9GlcNAc2) are indicated as well. Six glycosylation pathways are directed by isoenzyme families with partial functional redundancies including the *GALNT*s, *POFUT*s, *TMTC*s, *DPY19L*s, *XYLT*s and *COLGALT*s. Partial redundancy is also found for the oligosaccharyltransferase complex with the STT3A/B catalytic units (Ruiz-Canada et al. 2009). The assigned pathway-specific GTfs involved in core extension and elongation/branching (36) are mainly unique without apparent close isoenzymes with overlapping functions. Exceptions include *B4GALT5/T6* (LacCer glycolipid), *GCNT1/T2/T4* (O-GalNAc), *LFNG*/*MFNG*/*RFNG* (O-Fuc) and *GXYLT1*/*T2* (O-Glc). The assigned pathway-nonspecific GTf-genes involved in elongation and capping are mainly members of large isoenzyme families with considerable overlapping functions. While a number of studies have demonstrated preferences of, e.g. the many galactosyltransferase and sialyltransferase isoenzymes for certain glycoconjugates and pathways, we chose here to consider these as pathway-nonspecific because conclusive evidence is still missing (Tsuji et al. 1996, Amado et al. 1999; Togayachi et al. 2006). (**B**) Predicted transcriptional regulation of GTf-genes overlaid on the graphic depiction of GTf-genes in panel (A). A heatmap scale for τ-values based on RNAseq organ expression data represent high (red) to low (blue) degree of regulation. (**C**) Known CDG GTf-genes (red) and GWAS GTf-gene candidates (yellow) overlaid on the graphic depiction of GTf-genes in panel (A). Overlaps between CDG and GWAS candidate GTf-genes are shown in orange. Glycan symbols are drawn according to the SNFG format (Varki et al. 2015).

Table I. Summary of the 167 human GTf genes characteristic's found in Figure 1A

	Glycan biosynthetic steps in Figure 1A	Number of GTfs involved in each step[d]	Isoenzyme/ nonisoenzyme GTfs[e]	Regulated/not-regulated GTfs[f]	CDG-GTfs[g]	GWA-GTfs[h]
Pathway-specific GTfs[a]	Initiation step	58	36/22	26/32	27	25
	Core extension step	33	10/23	14/19	15	15
	Elongation and branching step	23	16/7	12/11	7	13
	Total	114	62/52	52/62	49	53
Pathway-nonspecific GTfs[a]	Elongation step	18	18/0	13/5	1	9
	Capping step	35	32/3	32/3	2	18
	Total	53	50/3	45/8	3	27
Total number of GTfs in Figure 1A[b]		167	112/55	97/70	52	80
Miscellaneous GTfs[c]		47	–	–	6	18
Total number of GTfs in the human genome		214	–	–	58	98

[a]Pathways-specific GTfs refers to Figure 1A lower part, pathway-nonspecific GTfs refers to Figure 1A upper part (elongation and capping). Details for the individual GTfs can be found in Supplementary data, Table SI.

[b]Figure 1A includes a total of 167 GTfs represented by the pathway-specific and pathway-nonspecific GTf-genes; the number of isoenzymes/nonisoenzymes, regulated/not-regulated CDG- and GWA-GTfs appear in this row.

[c]Miscellaneous GTf-genes include GTfs in the biosynthesis of hyaluronan and glycogen, GTfs involved in de-toxification or quality processes, and GTfs with unknown function. These are not included in the text but details can be found in Supplementary data, Table SII.

[d]Number of GTf-genes in the individual steps illustrated in Figure 1A.

[e]Isoenzymes include CAZy families or subfamilies of paralogous GTf-genes sharing sequence homology and overlapping enzymatic function; nonisoenzymes include GTfs without human genome paralogs.

[f]Regulated and not-regulated GTfs represent GTfs with $\tau > 0.71$ for regulated and $\tau \leq 0.71$ for not-regulated genes, for details see Materials and methods and Supplementary data, Table SVI.

[g]The CDG GTf-genes are listed in Supplementary data, Table SIII.

[h]The GWA GTf-genes are listed in Supplementary data, Table SIV.

human GTf genes representing 48 CAZy GT-families have been identified and their catalytic properties characterized to some detail or their properties predicted based on close sequence similarity to characterized genes (Hansen et al. 2015) (Supplementary data, Tables SI and SII). We recently developed a graphical overview in rainbow style of the 167 GTfs found in 44 out of the 48 GT-families, that we can assign or predict to be involved in defined roles in the biosynthesis of glycans, illustrating their predicted functions in glycosylation pathways (Figure 1) (Narimatsu et al. 2018). An elaboration upon the overview, providing more detailed assignment of GTfs to specific glycan structures and glycosylation pathways can be found in Figure 2. Currently, 15 distinct pathways of glycosylation can be defined including the glycophosphatidylinositol (GPI) anchor, glycosphingolipid (Gal-Cer and Glc-Cer), N-linked glycan, O-GalNAc mucin-type, O-Fuc type (*POFUT1* and *POFUT2* directed), O-GlcNAc (EGF-specific), O-Man type (POMT-directed), O-Man type (TMTC-directed), C-Man type, O-Glc type, O-Xyl type (proteoglycans), O-Gal type (collagen) and O-GlcNAc (cytosolic) pathways.

A group of 47 GTf-genes in 12 GT-families encode transferases involved in other processes such as synthesis of hyaluronan or glycogen, de-toxification of xenotoxins or endogenous compounds, as well as genes with unknown functions. These are listed in Table I and Supplementary data, Table SII as miscellaneous GTfs, and are not considered in the analyses performed in this study. The current assembly differs slightly from our previous GlyMAP report (Hansen et al. 2015): *MGAT4D, KDELC2* and *A3GALT2*, all not annotated in CAZy, and the novel gene *TMTC* gene family (*TMTC1-4*, CAZy GT105) recently identified as protein O-mannosyltransferases is now included (Larsen et al. 2017). Furthermore, a group of genes with proven GTf activity comprised of the ribitol-transferases *FKRP* and *FKTN* (Kanagawa et al. 2016; Praissman et al. 2016) and the Xyl-transferase *TMEM5* (Manya et al. 2016) are included in Figure 1 and Supplementary data, Table SI.

Assignment of GTf-genes to glycosylation pathways

As illustrated in Figures 1A and 2, GTf-genes can be classified according their predicted functions in the 15 glycosylation pathways recognized to date. A large proportion of the GTf-genome (114 genes) is predicted to encode for enzymes directing pathway-specific glycosylation. These are mainly involved in the initiation step and immediate core extension and/or branching of distinct types of glycoconjugates. In contrast, most genes controlling elongation and capping (53 genes) are predicted to be pathway-nonspecific and serve several different types of glycoconjugates (Figure 3A and Supplementary data, Figure S1A).

Classification of GTf-genes into isoenzyme families

A total of 112 out of the 167 GTf-genes can be classified as being members of paralogous gene families (Figure 3). Isoenzyme families are found both in pathway-specific (62/114) and pathway-nonspecific GTf-genes (50/53) (Table I and Supplementary data, Figure S1B). Among the pathway-specific GTf-genes are the genes involved in initiation of different types of O-glycosylation and C-mannosylation including *GALNT1-20* (GT27), *TMTC1-4* (GT105), *DPY19L1-4* (GT98), *XYLT1-2* (GT14) and *COLGALT1-2* (GT25). Among the pathway-nonspecific are the large gene families involved in elongation (GT7, GT10, GT29 and GT31) and capping (GT10 and GT29) (Figure 1A). Although a number of studies suggest that several isoenzymes in these GTf families may in fact function in specific glycosylation pathways, we have chosen to consider these as pathway-nonspecific because conclusive evidence is still missing. The most important examples are the sialyltransferases (GT29) specifically serving the glycolipid (*ST3GAL3/5*, *ST6GALNAC5/6* and *ST8SIA1/5*) and the O-glycan (*ST3GAL1* and *ST6GALNAC1/2*) pathways (Tsuji et al. 1996).

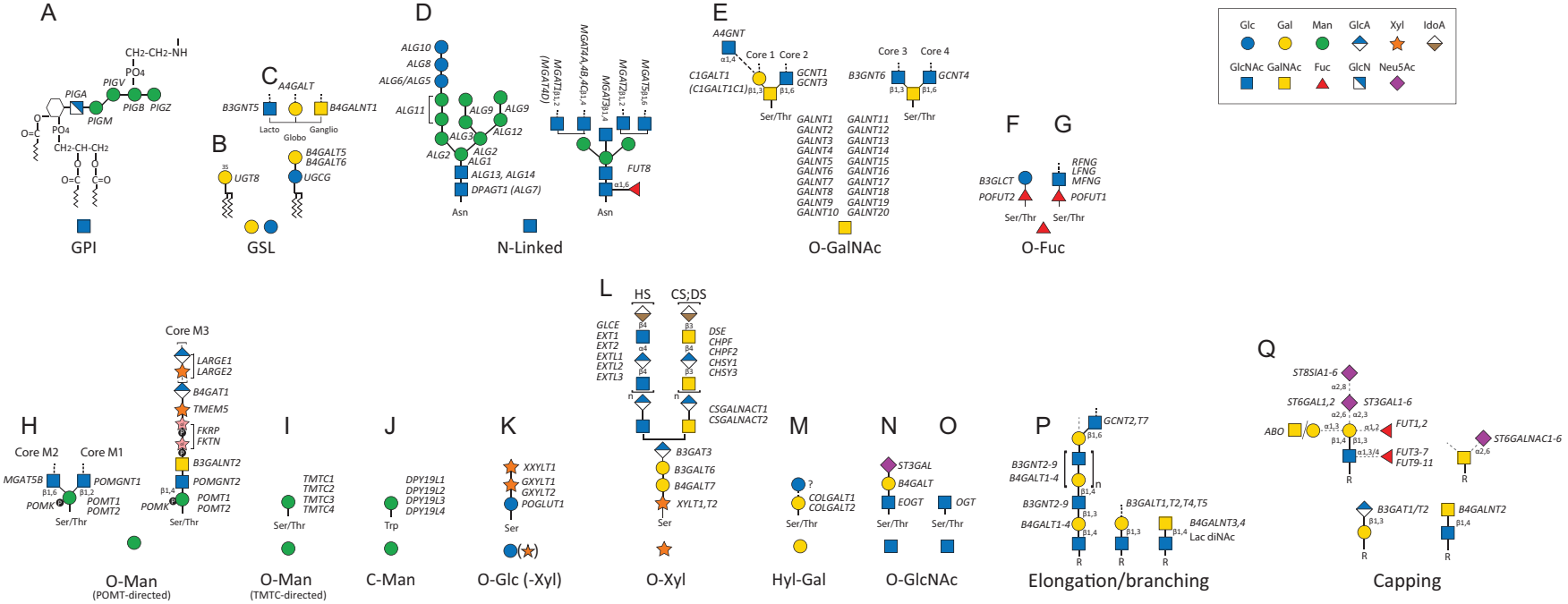

Fig. 2. Depiction of common glycan structures found on the 15 different types of glycosylation with (predicted) GTf-genes directing synthesis indicated. Glycan symbols are drawn according to the SNFG format (Varki et al. 2015). (**A**) GPI-anchors, (**B**) Galactocerebroside type glycosphingolipid, (**C**) Glycosphingolipid lacto-, globo and ganglio-type core structures, (**D**) N-linked glycans with oligosaccharide precursor structure and core complex type structure, (**E**) O-GalNAc mucin-type core 1–4 structures, (**F** and **G**) O-Fuc glycans with distinct *POFUT1* and *POFUT2* directed structures, (**H**) O-Man POMT1/2 directed core M1-M3 structures, (**I**) O-Man TMTC-directed simple monosaccharide glycan, (**J**) C-Man simple monosaccharide glycan, (**K**) O-Glc structure, which may exist as an O-Xyl glycan (Takeuchi et al. 2011), (**L**) O-Xyl linked glycosaminoglycan core structures for heparin/heparin sulfate (HS) and chondroitin/dermatin sulfate (CS/DS), (**M**) Hydroxyl lysine (HyL) O-Gal structure found on collagens, (**N**) O-GlcNAc directed by ER-located eOGT, (**O**) O-GlcNAc directed nucleocytoplasm located OGT, (**P**) common structures used in elongation and branching found in multiple glycosylation pathways and (**Q**) common structures used in capping found in multiple glycosylation pathways.

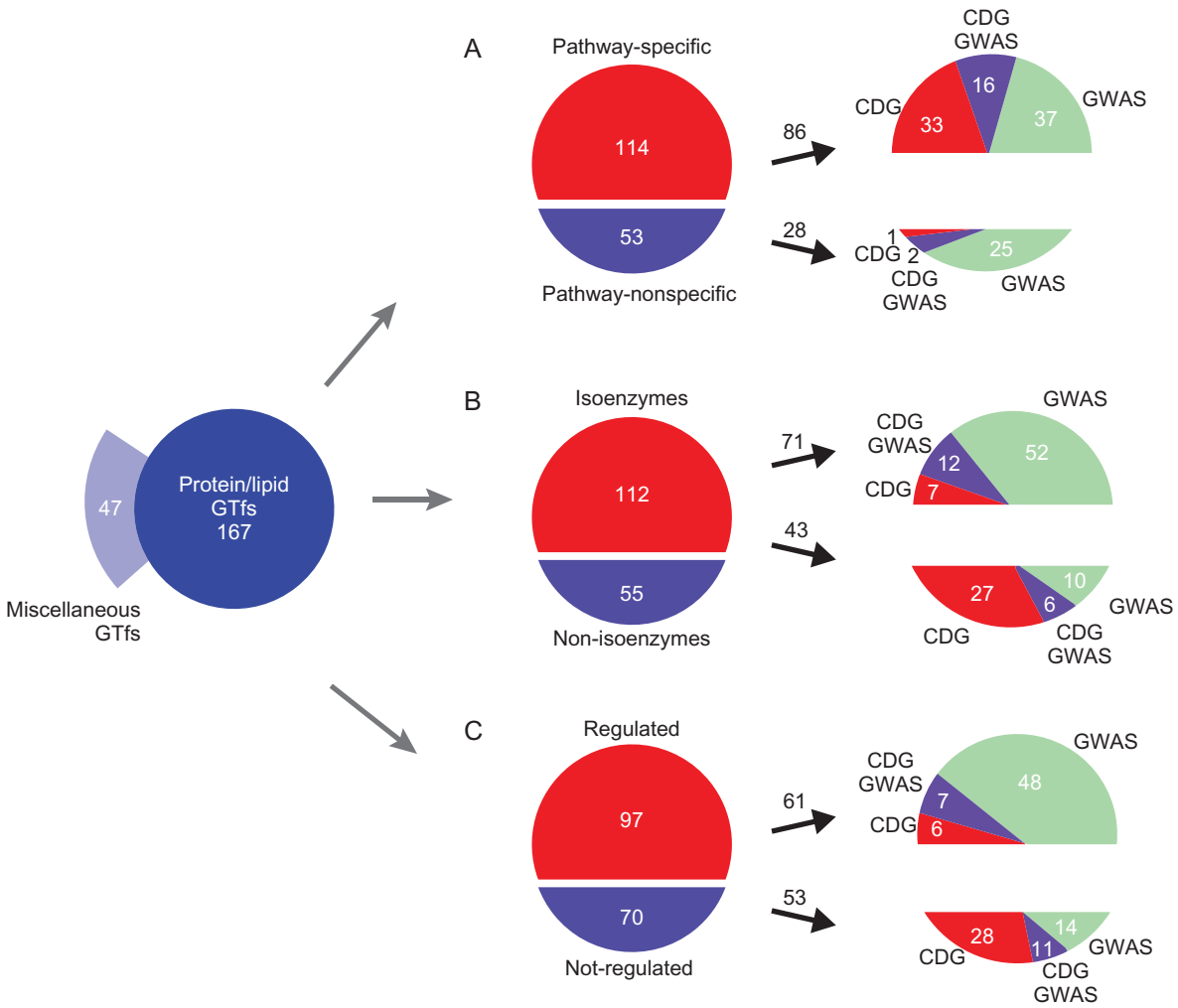

Fig. 3. Classification of the 167 human GTf-genes into (**A**) pathway-specific and pathway-nonspecific GTfs, (**B**) isoenzyme and nonisoenzyme GTfs, and (**C**) regulated and not-regulated GTfs as represented in Table I. (**A**) The pathway-specific GTfs refers to genes in the initiation, core extension, or branching and elongation steps (lower part of Figure 1A), and the pathway-nonspecific GTfs refer to genes in the elongation and capping steps (upper part Figure 1A). (**B**) The isoenzyme GTfs refer to subfamilies of paralogous GTf genes with sequence homology and overlapping function; the nonisoenzymes GTfs refers to genes without paralogs in the genome. (**C**) The regulated GTfs refer to genes ($\tau > 0.71$) and not-regulated GTfs ($\tau < 0.71$) refers to Figure 1B. For all three subdivisions of the 167 GTf-genes, (**A–C**) the distribution of CDG and GWA GTf-genes as illustrated in Figure 1C is shown in the right column of pie-charts. Detailed information for the individual GTf-genes are deposit in Supplementary data, Table SI.

Global analysis of transcriptional regulation of GTf-genes

Past studies of the expression of GTf-genes by traditional Northern blot analysis, qPCR and immunohistochemistry have clearly demonstrated that many of the GTf-genes are differentially regulated in tissues and disease conditions, but a large fraction has also been assigned more ubiquitous expression status. In order to probe the GTf-genome more globally, we turned to the rapidly increasing volume of RNAseq data, and selected the Human Protein Atlas (HPA) (Uhlén et al. 2015) RNAseq bodymap data as provided by the EMBL-EBI Expression Atlas (Petryszak et al. 2016). As a first estimate we grouped GTf-genes into regulated and not-regulated using a defined τ-value (Yanai et al. 2005; Kryuchkova-Mostacci and Robinson-Rechavi 2017) assigning a value between 0 and 1 for each GTf-gene (Supplementary data, Table SVI). A threshold value was chosen at $\tau > 0.71$ for regulated and $\tau \leq 0.71$ for not-regulated GTfs (Supplementary data, Figure S2). Using this cutoff we could classify 97/167 GTf-genes as regulated and 70/167 as being not-regulated (Table I). In Figure 1B we overlaid τ-values as a heatmap on the graphic view of glycosylation pathways. This revealed that GTf-genes responsible for pathway-specific glycosylation were more often classified as not-regulated compared to pathway-nonspecific genes (Supplementary data, Figure S1C). Moreover, among the pathway-specific GTf-genes those without paralogs such as those involved in the *N*-glycan precursor assembly and the immediate core extension of different types of O-glycosylation, including the O-Man core M3 assembly and the linker region for GAGs, were predominantly classified as not-regulated. In contrast, the pathway-specific GTf-genes with paralogs like the *GALNT*s directing O-GalNAc initiation, or those involved in decision of branching in glycolipid, *N*-glycan and Heparan and Chondroitin Sulfate glycosylation pathways were classified as regulated. Among the pathway-nonspecific GTf-genes the group of *B4GALT1-4* (GT7) genes stand out as the only group classified as not-regulated.

GTFs underlying CDGs and implicated by GWAS represent distinct groups

More than 100 CDGs have been identified to date (Freeze et al. 2014; and 2015; Jaeken and Péanne 2017), and half of these (52) are caused by partial or, rarely, as complete deficiencies in a GTf-gene assigned to roles in the glycosylation pathways depicted in Figure 1. Another six CDGs involve GTf-genes classified as miscellaneous (Table I and Supplementary data, Table SIII). We considered the 52 CDG GTf-genes and overlaid these on the graphic view of glycosylation pathways in Figures 1C and 3. Interestingly, the overlay pattern generated by CDG GTf-genes resembled the pattern produced by not-regulated classified GTf-genes (Figure 1B). A vast majority of the GTf-genes underlying CDGs (49/52) were classified as pathway-specific, and only three GTf-genes were classified as pathway-nonspecific (*ST3GAL3*, *ST3GAL5* and *B4GALT1*) (Figure 3A). However, ST3GAL5—which has been associated with infantile-onset epilepsy and a Salt and Pepper developmental regression syndrome—has been shown to selectively serve in the formation of ganglioside GM3 glycolipids (Berselli et al. 2006). Moreover, a smaller majority of the CDG GTf-genes (33/52) were classified as being nonisoenzymes (Figure 3B), but only a quarter (13/52) as not being regulated (Figure 3C).

We next surveyed the GWAS databases (primarily the GWAS Catalog (MacArthur et al. 2017)) and literature for GWAS candidates among the 167 GTf-genes with predicted roles in the glycosylation pathways. We applied the following criteria for including GWA GTf-genes in the analysis: (i) GWA-traits located in gene exons and introns and in intergenic regions within 50 kb of the 5′ and 3′ regions of GTf-genes; and (ii) GWA-SNPs located in adjacent genes within the 50 kb flanking regions was excluded. A total of 178 GWA traits were found for 80 GTf-genes, and these were reported in 181 studies and associated with 269 SNPs (Figure 4 and Supplementary data, Table SIV).

Overlaying the 80 GWAS candidate GTf-genes on the graphic view of glycosylation pathways (Figure 1C), revealed that a small majority (53/80) were classified as pathway-specific (Table I and Figure 3A). Interestingly, the overlay pattern generated by GWA GTf-genes resembled the pattern produced by GTf-genes classified as regulated (Figure 1C), which looked like the mirror image of that found for CDG GTf-genes. The majority of the GTf-genes implicated by GWAS (64/80) were classified as being members of isoenzyme families (Figure 3B), and similarly the majority (55/80) found to be classified as regulated (Figure 3C).

The comparison of GTf-genes underlying CDGs and implicated by GWAS, illustrated by the heatmap in Figure 1C, clearly indicates that these represent two quite distinct groups; however, there is a significant overlap for 18 GTf-genes that have been found to cause CDGs and are implicated by GWAS. The detailed analysis with regards to classifications into pathway specificity, isoenzyme nature and transcriptional regulation as summarized in Figure 3, further revealed that the main differences between the two groups primarily lies in the almost exclusive classification of CDGs to pathway-specific glycosylation, whereas GWA GTf-genes are mainly associated with isoenzyme families and dynamic transcriptional regulation.

GWAS implicate GTfs in diverse traits and phenotypes

The 178 GWA-traits associated with the 80 GTf-genes represented a large spectrum of predisposition traits spanning from common traits

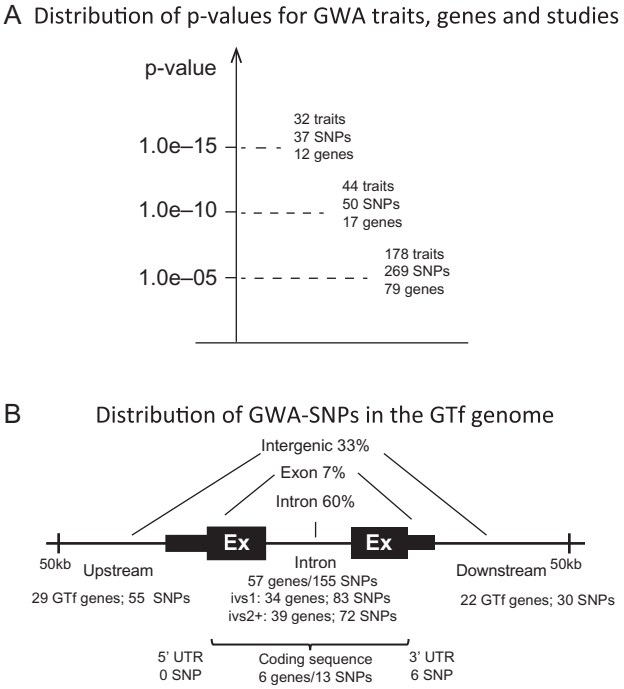

Fig. 4. Summary of GWAS candidate GTf-genes included in the study. (**A**) Graphic presentation of GWA-SNP p-values showing number of GTf-genes, traits and SNPs reported with intervals of $P < 10^{-5}$, $P < 10^{-10}$ and $P < 10^{-15}$. (**B**) Graphic distribution of GWA-SNPs with respect to gene structure.

such as low density lipoprotein and HDL cholesterol, triglycerides, obesity (body mass index), bone mineral density (BMD), over to neurological traits such as dementia, attention deficit hyperactivity disorder, and schizophrenia, and finally to more general phenotypes such as height and aging (Supplementary data, Table SIV). We found SNP signal *P*-values less than 10^{-10} for 44 traits in 17 GTf-genes, and SNP signal *P*-values less than 10^{-15} for 32 traits in 12 GTf-genes (Figure 4A). Among traits with the strongest SNP signals were several recurrent traits mapped in different studies and with multiple SNPs associating to the same GTf-gene. These include dyslipidemia traits for *GALNT2*, BMD trait for *GALNT3*, and body height for *COLGALT2* (Supplementary data, Table SIV).

We further inspected the genomic locations of SNPs and found that a majority of the SNPs (60%) were intronic and half of these were in the first intron of the GTf-genes, suggesting involvement of regulatory regions (Figure 4B). A third of the SNPs were intergenic and most of these located in the 5′ upstream 50 kb region. Very few SNPs were in the 3′ UTRs and none in 5′ UTRs. As expected, few SNPs were found in the coding regions of GTf-genes defining ABH blood groups (Supplementary data, Table SV).

Discussion

The presented global analysis of the 167 human GTf-genes assignable to distinct glycosylation pathways demonstrates that genes causing CDGs have distinct features from the group of genes implicated by GWAS. CDG genes are characterized by primarily serving pathway-specific roles in glycosylation, with a tendency to have unique enzymatic functions and low transcriptional regulation. In contrast, GWAS candidate genes are characterized by primarily being members of paralogous gene families, where isoenzymes may

provision for overlapping functions and differential transcriptional regulation. The characteristic features of GTf-genes causing rare CDGs were perhaps expected and consistent with their global effects on glycosylation and often severe outcome (Freeze et al. 2014; Jaeken and Péanne 2017). On the other hand, it may appear surprising that the many GTf-genes implicated by GWAS were found to be rather distinct from the CDG groups. However, we clearly do not have the complete picture and reports on new CDGs and GWA-traits continue to appear. Moreover, a number of factors may bias our current view including technical limitations in identifying subtle nonglobal changes in the glycome and lack of understanding of non-redundant functions of isoenzymes involved in glycosylation. Given the large number of GTf-genes implicated by GWAS there is a clear need to consider these and develop strategies for experimental validation, and the few examples validated to date seem to point to subtle regulatory roles of glycosylation in more common diseases and predispositions.

The current knowledge of CDGs caused by deficiencies in GTf-genes is still incomplete (Freeze et al. 2014) and structural defects in GTf-genes are extremely rare with allele frequencies of ≤0.001 (Hansen et al. 2015). While CDGs initially were discovered by, e.g. analysis of the N-glycosylation state of plasma transferrin or characteristic phenotypes associated with deficiencies in α-dystroglycan glycosylation and GAG biosynthesis (Freeze et al. 2014), more recent use of whole genome sequencing (WGS) or WES of patient groups and consanguineous families are leading to the discoveries of CDGs caused by GTf-genes that are members of isoenzyme families with partially overlapping functions and only subtle nonglobal imprints on the glycome (Freeze et al. 2014; and 2015). WGS and WES strategies are not limited by the technical difficulties with finding subtle changes in glycosylation perhaps restricted to certain organs or even specific glycoproteins, yet once a potential CDG is identified, it may be quite complex to identify the resulting glycosylation changes and the molecular mechanisms underlying phenotypic characteristics. Recent examples are found among the polypeptide GalNAc-transferase genes, where the phenotypic consequences of complete deficiency are subtle and distinct due to loss of O-glycosylation on one or only very few glycoproteins (Kato et al. 2006; Khetarpal et al. 2016). These findings are also reflected in the many knockout (KO) studies of GTf-genes in mice and other animal models (Lowe and Marth 2003; Stanley 2016). KO of GTf-genes that result in early global elimination of specific glycosylation pathways generally lead to embryonic lethality, although this is not the case with the GalNAc-transferase genes as there is substantial overlap in functions and only a limited number of proteins are affected (Schjoldager et al. 2015; Khetarpal et al. 2016). KO of pathway-specific GTf-genes acting later in the biosynthesis and without substantial redundancies in function generally produce postnatal severe phenotypes (Stanley 2016). In contrast, KO of pathway-nonspecific GTf-genes generally produce subtle or no apparent phenotypes, and the targeted genes usually have substantial functional redundancy from other isoenzymes.

Our analysis of GWAS implicated GTf-genes revealed, as predicted, that a significant number of these overlap with CDG GTf-genes (Figure 1C). Moreover, in several cases the GWA traits are consistent with the phenotypes associated with complete deficiency in the corresponding CDGs (7/18). The best examples may be the GalNAc-transferase genes *GALNT2* and *GALNT3*. Loss of *GALNT3* causes hyperphosphatemia and ectopic ossifications and the GWAS trait is bone mineral density (Topaz et al. 2004; Kato et al. 2006; Estrada et al. 2012; Kemp et al. 2014), while loss of *GALNT2* results in low HDL and the GWAS trait indeed includes low HDL (Willer et al. 2013; Khetarpal et al. 2016). In some cases, the GWA traits may be related to phenotypes found in murine KO models. Thus, GWA traits for *MGAT5* include type 2 diabetes and aging (Palmer et al. 2015; Pilling et al. 2016), and the *Mgat5* KO mice showed diminished glycemic response and increased insulin sensitivity (Johswich et al. 2014) and accelerated aging (Dennis et al. 2002). The GWA traits for *B3GNT2* include autoimmunity and chronic inflammation (Okada et al. 2012, 2014), and the *B3gnt2* KO mice showed hyperactivation of lymphocytes and enhanced initiation of immune responses by T cells and B cells (Togayachi et al. 2010). The GWA traits for *B3GAT1* include verbal declarative memory and schizophrenia (Debette et al. 2015; Goes et al. 2015), and *B3gat1* KO mice suffer from defects in higher brain functions including special memory, leaning phenotypes and impaired synaptic plasticity (Yamamoto et al. 2002). In most cases though, the murine KO phenotypes do not appear related to the identified GWA traits.

The phenotypic correlations between loss of GTf-gene functions in mice or man and GWAS candidate GTf-genes present one level of validation of the GWAS candidates, and highlight that GTf-genes are involved in highly specific and conserved glycosylation reactions that impact common diseases. How this occurs is still unclear, but as summarized in Figure 4 the location of the diagnostic SNPs for most GWA traits are in nonexon regions and predicted to affect transcriptional regulation. It is a general finding that GWA studies point to regulatory parts of candidate genes suggesting that in most cases the underlying mechanism for involvement in traits are allele-specific transcription consequences (Maurano et al. 2012; Lappalainen et al. 2013; Cavalli et al. 2016). As discussed above this is clearly the case for the diagnostic SNPs identified for the HDL trait in *GALNT2*. These are located in intron 1 close to a liver-specific regulatory element, and they were demonstrated to produce allele-specific transcript variations specifically in liver (Roman et al. 2015; Khetarpal et al. 2016). More examples are needed to confirm if this is a common finding, but to explore the functional consequence of such dysregulatory events, we will likely need to be able to monitor quantitative changes in the glycosylation of specific glycoproteins derived from specific organs or cell types.

The finding that most GWAS candidate GTf-genes appear to be members of larger isoenzyme gene families may not be surprising (Figure 3B). Evolution of isoenzymes is a way to diversify and fine-tune specific functions in addition to providing functional backup for essential functions. This clearly applies to the many GTf-gene families, where individual isoenzymes often are found to serve distinct functions, but with sufficient overlap in specificities to accommodate partial compensation. The crude analysis of transcriptional regulation in human organs provided by others (Uhlén et al. 2015; Petryszak et al. 2016) also pointed selectively to the GWAS candidate GTf-genes (Figure 3C), which suggests that differential and/or tight regulation of these enzymes are important. An example illustrating this is again the GalNAc-T3 isoenzyme, which co-regulates phosphate (Pi) homeostasis through specific nonredundant glycosylation of FGF23 (Kato et al. 2006). Expression of *GALNT3* as well as *FGF23* is tightly regulated (Chefetz et al. 2009; Ito et al. 2015), and nonredundant GalNAc-T3 O-glycosylation of FGF23 fine-tunes the biological function of FGF23 (Kato et al. 2006) essentially without major biological effects on the many other proteins glycosylated by this enzyme. It is still not clear how transcriptional regulation of GTfs-genes can serve in fine-tuning highly specific and tightly regulated processes.

In summary, our global analysis of GTf-genes and diseases draws attention to a group of glycosylation disorders caused by dysregulation rather than deficiency of GTfs. There are a number of challenges in exploring these, and new strategies including single cell transcriptomics and simple model systems for glycome analysis are required. However, the results presented suggest that at least some of the GWAS candidates are valid and play roles in more common general traits and health conditions.

Materials and methods

Assembly of the human GTf-genome

The initial human GTf-gene list was retrieved from the GlyMAP list of genes (Hansen et al. 2015), and updated using the recent Human Reference Genome (GRCh38/hg38, NCBI, https://www.ncbi.nlm.nih.gov) and CAZy database annotations (http://www.cazy.org/) (Lombard et al. 2014). GTf genes with known function, but not annotated in CAZy were curated from primary literature.

Classification of isoenzyme/nonisoenzyme and collation of protein annotation data

An initial grouping of isoenzymes was based upon the CAZy GT-family classification and was further refined by manual analysis and literature search. For each GTf, the donor nucleotide, protein function, subcellular localization and protein type was curated from CAZy, UniProt (http://www.uniprot.org/) and primary literature. Signal peptides and transmembrane spanning domains were annotated using the prediction servers SignalP (http://www.cbs.dtu.dk/services/SignalP/) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/), respectively.

Compilation of CDG data

The assembly of the CDGs caused by GTf-genes was built upon recent summarizing reviews (Freeze et al. 2014; 2015; Jaeken and Péanne 2017), supplemented with data from OMIM (https://omim.org/) and the Human Genome Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/).

Analysis of transcriptome information

The transcriptomic dataset for analysis was chosen from amongst the datasets available on Expression Atlas (Petryszak et al. 2016) (https://www.ebi.ac.uk/gxa/), selecting for a dataset that had a reasonable number of tissues included. Consequently, the transcriptomic data from the HPA (Uhlén et al. 2015) was chosen for further analysis. The RNAseq data comprises a map of gene expression in 32 different human tissues (Supplementary data, Table SVI) obtained from 122 healthy individuals. Since we only needed access to summarized data across different individuals, mean TPM (Transcripts Per Million) values per tissue were retrieved from the Expression Atlas in CSV format.

As a heuristic for evidence of tissue-specific gene regulation, we chose to use the dimensionless Tau (τ) value (Yanai et al. 2005), a metric that performs generally well with respect to measuring tissue-specific expression (Kryuchkova-Mostacci and Robinson-Rechavi 2017). τ-values were calculated in R for each of the GTf-genes across the 32 tissue samples. In order to choose a cutoff τ-value to distinguish between regulated and nonregulated, we modeled the τ-value distribution for all GTF-genes as a two component mixed Gaussian model. The model comprises a component with mean at τ 0.55 (not-regulated) and another with mean at τ 0.86 (regulated). The intersection of the component curves at τ 0.71 can be used as a cutoff point. τ values greater than this cutoff represent regulated, and lower than this value represent not-regulated (Supplementary data, Figure S2). This is clearly a crude estimation, but given that current RNAseq data is derived from complex tissues and not distinct cell types, we do not believe more detailed analysis is warranted. Adapted from the HPA, levels for TPM < 1 represent no detection of reads, TPM∈[1–10] represent low expression, TPM∈[10–50] represent medium expression and TPM > 50 high expression.

Compilation of GWAS data

The GWAS data analysis was based on data from the GWAS Catalog (download October 2017, https://www.ebi.ac.uk/gwas/), which is a manually curated literature-driven collection of published GWA Studies representing SNP-traits with a statistical significance of P-values $< 1 \times 10^{-5}$ (MacArthur et al. 2017). From the initial dataset retrieved from the GWAS Catalog, the data was filtered setting inclusion criteria for GWA-SNPs to be localization in GTf-gene exons, introns or flanking intergenic 50 kb regions. SNPs located within adjacent genes in the 50 kb flanking regions were excluded from further analysis. SNPs were analyzed for genomic localization in exons, UTRs, introns and intergenic regions using the longest RefSeq transcript. The GWAS catalog disease/trait associations terms were adapted for the GWA-GTf analysis and are cataloged in Supplementary data Table SIV including GWA GTf-genes, SNP id, P-value intervals and GWAS references.

Internet resources

CAZy, the Carbohydrate-Active enZYmes Database (http://www.cazy.org/)
CBS Prediction Servers (http://www.cbs.dtu.dk/services/)
EMBL-EBI Expression Atlas (https://www.ebi.ac.uk/gxa/home)
Ensembl (http://www.ensembl.org/)
NCBI (https://www.ncbi.nlm.nih.gov)
GWAS Catalog (https://www.ebi.ac.uk/gwas/)
HGMD, Human Genome Mutation Database Cardiff (http://www.hgmd.cf.ac.uk/)
Human Protein Atlas, HPA (https://www.proteinatlas.org/)
OMIM (https://omim.org/)
SignalP (http://www.cbs.dtu.dk/services/SignalP/)
TMHMM (http://www.cbs.dtu.dk/services/TMHMM/)
UniProt (http://www.uniprot.org/)

Supplementary data

Supplementary data is available at *GLYCOBIOLOGY* online.

Funding

This work was supported by Læge Sofus Carl Emil Friis og hustru Olga Doris Friis' Legat, The Lundbeck Foundation, The Danish Research Council Sapere Aude, Research Talent Grant to K.T.S., The Novo Nordisk Foundation, The Danish National Research Foundation (DNRF107), The Rocket Fund and National Institutes of Health grant (R01 DK099551) to H.H.F., and The Novo Nordisk Foundation to B.H.

Abbreviations

BMD, bone mineral density; CDGs, congenital disorders of glycosylation; HDL, high-density lipoprotein; HPA, Human Protein Atlas; GPI, glycophosphatidylinositol; GTFs, glycosyltransferases; GWAS, Genome-Wide Association Studies; KO, knockout; SNP, single nucleotide polymorphism; SNVs, single-nucleotide variants; WES, whole exome sequencing; WGS, whole genome sequencing.

References

Amado M, Almeida R, Schwientek T, Clausen H. 1999. Identification and characterization of large galactosyltransferase gene families: Galactosyltransferases for all functions. *Biochim Biophys Acta*. 1473: 35–53.

Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. 2012. Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*. 22:736–756.

Berselli P, Zava S, Sottocornola E, Milani S, Berra B, Colombo I. 2006. Human GM3 synthase: A new mRNA variant encodes an NH2-terminal extended form of the protein. *Biochim Biophys Acta*. 1759:348–358.

Cavalli M, Pan G, Nord H, Wallén Arzt E, Wallerman O, Wadelius C. 2016. Allele-specific transcription factor binding in liver and cervix cells unveils many likely drivers of GWAS signals. *Genomics*. 107:248–254.

Chefetz I, Kohno K, Izumi H, Uitto J, Richard G, Sprecher E. 2009. GALNT3, a gene associated with hyperphosphatemic familial tumoral calcinosis, is transcriptionally regulated by extracellular phosphate and modulates matrix metalloproteinase activity. *Biochim Biophys Acta*. 1792:61–67.

Cummings RD. 2009. The repertoire of glycan determinants in the human glycome. *Mol BioSyst*. 5:1087–1104.

Debette S, Ibrahim Verbaas CA, Bressler J, Schuur M, Smith A, Bis JC, Davies G, Wolf C, Gudnason V, Chibnik LB et al. 2015. Genome-wide studies of verbal declarative memory in nondemented older people: The Cohorts for Heart and Aging Research in Genomic Epidemiology consortium. *Biol Psychiatry*. 77:749–763.

Dennis JW, Pawling J, Cheung P, Partridge E, Demetriou M. 2002. UDP-N-acetylglucosamine:alpha-6-D-mannoside beta1,6 N-acetylglucosaminyltransferase V (Mgat5) deficient mice. *Biochim Biophys Acta*. 1573:414–422.

Edvardson S, Baumann AM, Muhlenhoff M, Stephan O, Kuss AW, Shaag A, He L, Zenvirt S, Tanzi R, Gerardy-Schahn R et al. 2013. West syndrome caused by ST3Gal-III deficiency. *Epilepsia*. 54:e24–e27.

Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, Oei L, Albagha OM, Amin N, Kemp JP et al. 2012. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet*. 44:491–501.

Freeze HH, Chong JX, Bamshad MJ, Ng BG. 2014. Solving glycosylation disorders: Fundamental approaches reveal complicated pathways. *Am J Hum Genet*. 94:161–175.

Freeze HH, Eklund EA, Ng BG, Patterson MC. 2015. Neurological aspects of human glycosylation disorders. *Annu Rev Neurosci*. 38:105–125.

Goes FS, McGrath J, Avramopoulos D, Wolyniec P, Pirooznia M, Ruczinski I, Nestadt G, Kenny EE, Vacic V, Peters I et al. 2015. Genome-wide association study of schizophrenia in Ashkenazi Jews. *Am J Med Genet B Neuropsychiatr Genet*. 168:649–659.

Hansen L, Lind-Thomsen A, Joshi HJ, Pedersen NB, Have CT, Kong Y, Wang S, Sparso T, Grarup N, Vester-Christensen MB et al. 2015. A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology*. 25:211–224.

Hansske B, Thiel C, Lubke T, Hasilik M, Honing S, Peters V, Heidemann PH, Hoffmann GF, Berger EG, von Figura EG et al. 2002. Deficiency of UDP-galactose:N-acetylglucosamine beta-1,4-galactosyltransferase I causes the congenital disorder of glycosylation type IId. *J Clin Invest*. 109:725–733.

Hu H, Eggers K, Chen W, Garshasbi M, Motazacker MM, Wrogemann K, Kahrizi K, Tzschach A, Hosseini M, Bahman I et al. 2011. ST3GAL3 mutations impair the development of higher cognitive functions. *Am J Hum Genet*. 89:407–414.

Ito N, Wijenayaka AR, Prideaux M, Kogawa M, Ormsby RT, Evdokiou A, Bonewald LF, Findlay DM, Atkins GJ. 2015. Regulation of FGF23 expression in IDG-SW3 osteocytes and human bone by pro-inflammatory stimuli. *Mol Cell Endocrinol*. 399:208–218.

Jaeken J, Péanne R. 2017. What is new in CDG? *J Inherit Metab Dis*. 40: 569–586.

Johswich A, Longuet C, Pawling J, Abdel Rahman A, Ryczko M, Drucker DJ, Dennis JW. 2014. N-glycan remodeling on glucagon receptor is an effector of nutrient sensing by the hexosamine biosynthesis pathway. *J Biol Chem*. 289:15927–15941.

Kanagawa M, Kobayashi K, Tajiri M, Manya H, Kuga A, Yamaguchi Y, Akasaka-Manya K, Furukawa J, Mizuno M, Kawakami H et al. 2016. Identification of a post-translational modification with ribitol-phosphate and its defect in muscular dystrophy. *Cell Rep*. 14:2209–2223.

Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS et al. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet*. 40:189–197.

Kato K, Jeanneau C, Tarp MA, Benet-Pagès A, Lorenz-Depiereux B, Bennett EP, Mandel U, Strom TM, Clausen H. 2006. Polypeptide GalNAc-transferase T3 and familial tumoral calcinosis. Secretion of fibroblast growth factor 23 requires O-glycosylation. *J Biol Chem*. 281:18370–18377.

Kemp JP, Medina-Gomez C, Estrada K St, Pourcain B, Heppe DH, Warrington NM, Oei L, Ring SM, Kruithof CJ, Timpson NJ et al. 2014. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet*. 10:e1004423.

Khetarpal SA, Schjoldager KT, Christoffersen C, Raghavan A, Edmondson AC, Reutter HM, Ahmed B, Ouazzani R, Peloso GM, Vitali C et al. 2016. Loss of function of GALNT2 lowers high-density lipoproteins in humans, nonhuman primates, and rodents. *Cell Metab*. 24:234–245.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Brief Bioinform*. 18:205–214.

Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PA, Monlong J, Rivas MA, Gonzàlez-Porta M, Kurbatova N, Griebel T, Ferreira PG et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 501:506–511.

Larsen ISB, Narimatsu Y, Joshi HJ, Siukstaite L, Harrison OJ, Brasch J, Goodman KM, Hansen L, Shapiro L, Honig B et al. 2017. Discovery of an O-mannosylation pathway selectively serving cadherins and protocadherins. *Proc Natl Acad Sci USA*. 114:11163–11168.

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 42:D490–D495.

Lowe JB, Marth JD. 2003. A genetic approach to Mammalian glycan function. *Annu Rev Biochem*. 72:643–691.

MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J et al. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 45:D896–D901.

Manya H, Yamaguchi Y, Kanagawa M, Kobayashi K, Tajiri M, Akasaka-Manya K, Kawakami H, Mizuno M, Wada Y, Toda T et al. 2016. The muscular dystrophy gene TMEM5 encodes a Ribitol β1,4-Xylosyltransferase required for the functional glycosylation of dystroglycan. *J Biol Chem*. 291:24618–24627.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 337:1190–1195.

Mizumoto S, Yamada S, Sugahara K. 2014. Human genetic disorders and knockout mice deficient in glycosaminoglycan. *BioMed Res Int*. 2014: 495764.

Narimatsu Y, Joshi HJ, Zhang Y, Gomes C, Chen Y-H, Lorenzetti F, Furukawa S, Schjoldager K, Hansen L, Clausen H et al. 2018. A validated

gRNA library for CRISPR/Cas9 targeting of the human glycosyltransferases. *Glycobiology*. 28:296–306.

Ng BG, Freeze HH. 2015. Human genetic disorders involving glycosylphosphatidylinositol (GPI) anchors and glycosphingolipids (GSL). *J Inherit Metab Dis*. 38:171–178.

Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Kawaguchi T, Stahl EA, Kurreeman FA, Nishida N et al. 2012. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat Genet*. 44:511–516.

Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, Yoshida S et al. 2014. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 506:376–381.

Palmer ND, Goodarzi MO, Langefeld CD, Wang N, Guo X, Taylor KD, Fingerlin TE, Norris JM, Buchanan TA, Xiang AH et al. 2015. Genetic variants associated with quantitative glucose homeostasis traits translate to Type 2 Diabetes in Mexican Americans: The GUARDIAN (Genetics Underlying Diabetes in Hispanics) Consortium. *Diabetes*. 64:1853–1866.

Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, Füllgrabe A, Fuentes AM, Jupp S, Koskinen S et al. 2016. Expression Atlas update—An integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res*. 44:D746–D752.

Pilling LC, Atkins JL, Bowman K, Jones SE, Tyrrell J, Beaumont RN, Ruth KS, Tuke MA, Yaghootkar H, Wood AR et al. 2016. Human longevity is influenced by many genetic variants: Evidence from 75,000 UK Biobank participants. *Aging (Albany NY)*. 8:547–560.

Praissman JL, Willer T, Sheikh MO, Toi A, Chitayat D, Lin YY, Lee H, Stalnaker SH, Wang S, Prabhakar PK et al. 2016. The functional O-mannose glycan on α-dystroglycan contains a phospho-ribitol primed for matriglycan addition. *eLife*. 5:e14473.

Roman TS, Marvelle AF, Fogarty MP, Vadlamudi S, Gonzalez AJ, Buchkovich ML, Huyghe JR, Fuchsberger C, Jackson AU, Wu Y et al. 2015. Multiple hepatic regulatory variants at the GALNT2 GWAS locus associated with high-density lipoprotein cholesterol. *Am J Hum Genet*. 97:801–815.

Ruiz-Canada C, Kelleher DJ, Gilmore R. 2009. Cotranslational and posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell*. 136:272–283.

Schjoldager KT, Joshi HJ, Kong Y, Goth CK, King SL, Wandall HH, Bennett EP, Vakhrushev SY, Clausen H. 2015. Deconstruction of O-glycosylation—GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Rep*. 16:1713–1722.

Schjoldager KT-BG, Vakhrushev SY, Kong Y, Steentoft C, Nudelman AS, Pedersen NB, Wandall HH, Mandel U, Bennett EP, Levery SB et al. 2012. Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci USA*. 109:9893–9898.

Sheikh MO, Halmo SM, Wells L. 2017. Recent advancements in understanding mammalian O-mannosylation. *Glycobiology*. 27:806–819.

Simpson MA, Cross H, Proukakis C, Priestman DA, Neville DCA, Reinkensmeier G, Wang H, Wiznitzer M, Gurtz K, Verganelaki A et al. 2004. Infantile-onset symptomatic epilepsy syndrome caused by a homozygous loss-of-function mutation of GM3 synthase. *Nat Genet*. 36:1225–1229.

Stanley P. 2016. What have we learned from glycosyltransferase knockouts in mice? *J Mol Biol*. 428:3166–3182.

Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Schjoldager KTG, Lavrsen K, Dabelsteen S, Pedersen NB, Marcos-Silva L et al. 2013. Precision mapping of the human GalNAc-type O-glycoproteome—First generation view from 12 human SimpleCells. *EMBO J*. 32:1478–1488.

Takeuchi H, Fernandez-Valdivia RC, Caswell DS, Nita-Lazar A, Rana NA, Garner TP, Weldeghiorghis TK, Macnaughtan MA, Jafar-Nejad H, Haltiwanger RS. 2011. Rumi functions as both a protein O-glucosyltransferase and a protein O-xylosyltransferase. *Proc Natl Acad Sci USA*. 108:16600–16605.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 466:707–713.

Togayachi A, Kozono Y, Kuno A, Ohkura T, Sato T, Hirabayashi J, Ikehara Y, Narimatsu H. 2010. Beta3GnT2 (B3GNT2), a major polylactosamine synthase: Analysis of B3GNT2-deficient mice. *Methods Enzymol*. 479:185–204.

Togayachi A, Sato T, Narimatsu H. 2006. Comprehensive enzymatic characterization of glycosyltransferases with a beta3GT or beta4GT motif. *Methods Enzymol*. 416:91–102.

Topaz O, Shurman DL, Bergman R, Indelman M, Ratajczak P, Mizrachi M, Khamaysi Z, Behar D, Petronius D, Friedman V et al. 2004. Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nat Genet*. 36:579–581.

Tsuji S, Datta AK, Paulson JC. 1996. Systematic nomenclature for sialyltransferases. *Glycobiology*. 6:v–vii.

Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A et al. 2015. Tissue-based map of the human proteome. *Science*. 347:1260419.

Varki A. 2017. Biological roles of glycans. *Glycobiology*. 27:3–49.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T et al. 2015. Symbol nomenclature for graphical representations of glycans. *Glycobiology*. 25:1323–1324.

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S et al. 2013. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 45:1274–1283.

Yamamoto S, Oka S, Inoue M, Shimuta M, Manabe T, Takahashi H, Miyamoto M, Asano M, Sakagami J, Sudo K et al. 2002. Mice deficient in nervous system-specific carbohydrate epitope HNK-1 exhibit impaired synaptic plasticity and spatial learning. *J Biol Chem*. 277:27227–27231.

Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, Bar-Even A, Horn-Saban S, Safran M, Domany E et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 21:650–659.