Deconstruction of O-glycosylation—GalNAc-T isoforms direct distinct subsets of the O-glycoproteome

Katrine T Schjoldager*, Hiren J Joshi, Yun Kong, Christoffer K Goth, Sarah Louise King, Hans H Wandall, Eric P Bennett, Sergey Y Vakhrushev & Henrik Clausen**

Abstract

GalNAc-type O-glycosylation is found on most proteins trafficking through the secretory pathway in metazoan cells. The O-glycoproteome is regulated by up to 20 polypeptide GalNAc-Ts and the contributions and biological functions of individual GalNAc-Ts are poorly understood. Here, we used a zinc-finger nuclease (ZFN)-directed knockout strategy to probe the contributions of the major GalNAc-Ts (GalNAc-T1 and GalNAc-T2) in liver cells and explore how the GalNAc-T repertoire quantitatively affects the O-glycoproteome. We demonstrate that the majority of the O-glycoproteome is covered by redundancy, whereas distinct subsets of substrates are modified by non-redundant functions of GalNAc-T1 and GalNAc-T2. The non-redundant O-glycopeptide subsets and specific transcriptional responses for each isoform are related to different cellular processes; for the GalNAc-T2 isoform, these support a role in lipid metabolism. The results demonstrate that GalNAc-Ts have different non-redundant glycosylation functions, which may affect distinct cellular processes. The data serves as a comprehensive resource for unique GalNAc-T substrates. Our study provides a new view of the differential regulation of the O-glycoproteome, suggesting that the plurality of GalNAc-Ts arose to regulate distinct protein functions and cellular processes.

Keywords: apolipoproteins; dimethylation; GALNT; mass spectrometry

Introduction

Post-translational modifications greatly expand the size and functional space of the proteome, and glycosylation is one of the most abundant and diverse modifications. GalNAc-type O-glycosylation is uniquely geared for differential regulation of protein functions with up to 20 polypeptide GalNAc-transferases (GalNAc-Ts) in humans controlling the first decisive step in glycosylation and thereby determining both which proteins and where on these proteins O-glycans are attached [1]. GalNAc-Ts catalyze the addition of GalNAc residues to serine and threonine (and possibly tyrosine), and GalNAc-T isoenzymes have been shown to exhibit different, albeit partly overlapping, peptide substrate specificities and kinetic properties. GALNTs are differentially expressed and the repertoire in cells changes during cellular maturation and differentiation and in cancer [1–3]. It is therefore predicted that the GalNAc-T family of enzymes furnish cells with the capacity to differentially and perhaps dynamically regulate the O-glycoproteome and thereby modulate protein function [4], not dissimilar to the regulation of the phosphoproteome by the much larger family of kinases [5]. However, currently this is a hypothesis based largely on in vitro enzyme analyses.

In vitro analyses of GalNAc-T isoenzymes suggest that these have wide roles in glycosylation of proteins without appreciable systematic or specific roles in biological processes and pathways. However, deficiencies in GALNT genes appear to cause subtle diseases both in humans and in animal models [6]. For example, loss of GALNT3 causes the rare related diseases familial tumoral calcinosis (FTC) and hyperostosis syndrome (HHS) characterized by hyperphosphatemia [7], which is due to inactivation of fibroblast growth factor 23 (FGF23) by proprotein convertase (PC) processing as a result of lack of site-specific O-glycosylation by GalNAc-T3 in the PC processing site [8]. Strikingly, deficiency in the FGF23 gene itself produces largely the same disease phenotype [9], indicating that GalNAc-T3 despite its predicted broader role in protein glycosylation perhaps primarily serves to co-regulate phosphate homeostasis. Notably, this function is conserved in rodents [10,11], and GALNT3 has been associated with human bone mineral density and fracture risk [12]. Thus, it is possible that the seemingly broad and partly overlapping roles of the many GalNAc-Ts in protein glycosylation are deceptive and that instead the many GalNAc-Ts serve to fine-tune highly selective protein functions in limited cellular pathways.
A long-standing question then is why nature evolved the large family of isoenzymes with 20 members in mammals, 9 in C. elegans and 12 in Drosophila [1,3], and more specifically whether the evolutionary requirement for multiple GalNAc-Ts was driven by a need to accommodate efficient glycosylation of a great diversity of acceptor peptide sequences, or—and perhaps not mutually exclusively—if this was driven by a need for differential regulation and fine-tuning of distinct protein functions and biological pathways. Other types of protein O-glycosylation in metazoans are catalyzed by only one or two isoenzymes [1,13], although yeast O-mannosylation, that in many ways resembles GalNAc-type O-glycosylation, is also catalyzed by a large family of isoenzymes with at least six Dol-P-Man O-mannosyltransferases [14]. We and others have recently identified examples of non-redundant biological functions of site-specific O-GalNAc glycosylation in man and model organisms [4,15–17], and multiple Genome-Wide Association Studies (GWAS) point to different roles of GALNTs in human diseases [18]. However, it remains a challenge to characterize the non-redundant contributions of individual GalNAc-Ts within the O-glycoproteome of a cell, and within these to dissect the important glycosylation events that have biological consequences and ultimately could underlie disease. Thus, there is a need to develop quantitative strategies to probe GalNAc-T isomeric-specific O-glycosylation and the effects these may have on protein function, cellular processes, and ultimately whole organisms.

O-glycoproteomics has long remained a challenge, but recent advances with the “SimpleCell” (SC) strategy [19] using genetically engineered cell lines with simplified homogenous O-glycans have enabled proteome-wide discovery of O-glycosites. A first draft of the human O-glycoproteome is available [20], and here, we present a quantitative O-glycoproteomics strategy for sensitive mapping of the non-redundant contributions of the major GalNAc-Ts (GALNT1 and GALNT2) in the human hepatocellular carcinoma HepG2 liver cell. Interestingly, these non-redundant contributions involve proteins related to cellular processes inferred from mouse knockout studies (GalNAc-T1 or GWAS (GalNAc-T2) [21–23]. The study points to a new view of the regulation of protein O-glycosylation, suggesting that individual GalNAc-T isoforms actually serve more specialized roles in distinct cellular processes. The results provide a fruitful discovery platform for essential non-redundant functions of site-specific O-glycosylation against the background of which might be labeled as variational noise.

Results and Discussion

A quantitative differential O-glycoproteomics strategy to map GalNAc-T functions

We used the so-called SimpleCell strategy [19] to develop quantitative O-glycoproteome analyses. The strategy relies on knockout (KO) of the COSMC gene, a private chaperone for the C1GalT1 enzyme controlling most O-glycan extension in cells, which leaves the O-glycoproteome with rather homogenous Tn O-glycans that can be captured by Vicia villosa (VVA) lectin chromatography. As the GALNT2 gene has been identified as a candidate gene for dyslipidemia, we choose to use the human HepG2 cell line, which is a highly differentiated liver cell line and widely used as a model system for lipid and cholesterol metabolism [24,25]. Human liver and HepG2 cells express a limited number of GalNAc-T isoforms, mainly the GalNAc-T1 and GalNAc-T2 isoforms with only low levels of expression of GalNAc-T10 and GalNAc-T11 (Figs 1A and EV1A). We developed two clones each of a parent HepG2 SimpleCell line (HepG2SC) and isogenic lines without GalNAc-T1 (HepG2SCAT1) or without GalNAc-T2 (HepG2SCAT2) by ZFN-mediated knockout (KO) [4]. We also introduced GalNAc-T3 to HepG2 cells to probe the effects of expressing a dominant foreign GalNAc-T isoform not normally expressed in liver. In contrast, to GalNAc-T1 and GalNAc-T2, GalNAc-T3 is differentially expressed in cells and interestingly, markedly unregulated in cancer [26]. We used a previously described site-specific ZFN-mediated knockin (KI) strategy [4] to introduce stable expression of GalNAc-T3 in HepG2SC (HepG2SC-T3) and the expression level and Golgi topology estimated by immunocytochemistry was comparable to what is found in cell lines normally expressing this isoform (Fig EV1B). All clones were validated by DNA sequencing and characterized by immunocytchemistry (ICC) for loss or gain of gene products as well as O-glycophenotype change (Fig EV1B). The total collection of cell lines used in the study is summarized in Table EV1.

We first evaluated gross effects of KO/KI of the GALNT genes in HepG2 cells by SDS–PAGE Western blotting with the anti-Tn lectin VVA of secreted proteins from each cell line which revealed only subtle differences in labeled protein banding patterns (Fig 1B). For quantitative evaluation of changes in the O-glycoproteome, we used stable differential isoform dimethyl labeling [27] of tryptic digests of total cell lysates and VVA lectin-enriched culture medium (Fig 1C). Total digests from HepG2SC were labeled with light (L) reagent, and the digests from each isogenic HepG2SCAT1/HepG2SCAT2/HepG2SC-T3 mutant lines with medium (M) reagent. Thus, each analysis included a mixture of (L) labeled peptides from HepG2SC and (M) labeled peptides from one of the ΔT1/ΔT2 mutant cell lines, which were enriched for glycopeptides by VVA Lectin Weak Affinity Chromatography (LWAC) and further analyzed by LC-MS/MS. This approach ensured triplicate analysis of the HepG2SC glycoproteome and direct comparative and quantitative analysis of the effects of the altered cellular repertoire of GalNAc-Ts. We analyzed total cell lysates (TCL) and secretomes (SEC) separately as this has previously resulted in substantially different O-glycoproteomes [28,29].

The overall O-glycoproteome of HepG2SC

The complete datasets from the three differential O-glycoproteomes (HepG2SCΔT1, HepG2SC ΔT2, and HepG2SCΔT3) and two sources (TCL and SEC) are listed in Table EV2 and a summary shown in Fig 2 (all glycoproteome data are available at http://www. ebi.ac.uk/pride/archive/projects/PXD002770). We identified 631 O-glycopeptides (1,742 unambiguously assigned O-GalNAc sites) from TCL and SEC, which is a substantial expansion compared to the 74 distinct O-glycopeptides (219 sites) identified using a non-quantitative approach and analysis on a LTQ-Orbitrap XL hybrid spectrometer with lower sensitivity [4] (Fig 2A). More than 50% of the O-glycopeptides identified are novel compared to our previous analysis of 12 human SimpleCell lines from different organs [20] (Fig 2A), which may partly be attributed to use of a more sensitive mass spectrometer.
The GalNAc-T differential O-glycoproteomes in HepG2SC

Out of the total 1,742 unambiguously assigned glycosites, 1,603 (92%) were quantifiable (Fig 2B). Quantification takes place at the level of peptide or peptide fragment mass spectrum, and this poses a number of complications with peptides containing more than one O-glycosite. In particular, interpretation of the contribution of individual GalNAc-Ts and their substrate specificities is ambiguous because GalNAc-Ts often function successively to cooperate in glycosylation of clustered glycosites through lectin-mediated interactions with partially glycosylated substrates or through direct recognition of a substrate site with and adjacent GalNAc glycosite [1]. However, most of the identified O-glycosites were found as monoglycosylated peptides in the protease digest in agreement with our previous findings (Fig 2C) [20], and we therefore limited the analysis and exclusively considered monoglycosylated peptides in the comparative analysis of the three sample datasets (HepG2SC ΔT1/HepG2SC ΔT2/HepG2SC +T3). The distribution of total quantified and monoglycosylated peptides were similar among the three samples with the same relative number of unique glycopeptide identifications (Fig EV2C and D), which provides evidence that this
Figure 2. The overall HepG2 O-glycoproteome.
A Overall identified O-glycoproteins in all HepG2 cell lines (631 glycoproteins) with comparison to the HepG2 O-glycoproteome presented in Schjoldager et al [4] (74 glycoproteins) and the human O-glycoproteome identified in 12 SCs presented in Stensto et al [20] (662 glycoproteins).
B Relative numbers of all sites, quantified sites, and quantified single sites identified in this study.
C Distribution of glycosites per protein identified in this study.

To identify candidate GalNAc-T isoform-specific substrate glycosites, we analyzed histograms based on glycopeptide M/L ratios for each paired sample set (Fig 3A–C). Loss of function of a GalNAc-T isoform is expected to result in loss of glycopeptides with specific acceptor sites for that isoform and appearance of L singlets (high M/L ratios), while de novo introduction of a GalNAc-T is expected to produce M singlets (high M/L ratios). However, glycopeptides with, for example, two glycosylation sites of which only one is isoform-specific can potentially produce a glycopeptide with only one site, which would appear as a M singlet based on the criteria of our analysis, thus explaining the non-Gaussian distribution with both light and medium shoulders and singlets (see Fig EV3 and Appendix Supplementary Methods).

We selected a conservative M/L cutoff ratio of 1:10 for selection of candidate isoform-specific glycopeptides (< 1:10 for GalNAc-T1 or GalNAc-T2 KO and > 10:1 for GalNAc-T3 KO) and included L or M singlets in this analysis (Fig 3 and Table EV2 sheet 2–7). Using these criteria, we identified a total of 75 glycopeptides (44 from TCL and 36 from SEC) that were essentially only found in HepG2SC and not in Delta1 isogenic cells with this isoform (Fig 3A). The same analysis with AT2 isogenic cells resulted in identification of 81 glycopeptides (32 from TCL and 57 from SEC) that were present only in HepG2SC+T3 (Fig 3C). The majority of identified GalNAc-T isoform-specific candidate sites were found on glycoproteins with only one O-glycosite identified (Fig EV3C and Table EV2). The quantitative O-glycroteomics strategy thus enabled us to monitor ~1,000 O-glycosites in isogenic cell pairs with and without a single GalNAc-T isoform, and for the first time, we can estimate the size and nature of the non-redundant O-glycroteome in a cell line expressing the two dominant GALNT1 and GALNT2 isoforms. About 10% of the identified glycosites are substantially affected (more than 10-fold) by loss or gain of individual GalNAc-T isoforms, while the majority are unaffected or partially affected. We provided validation of a fraction of the isoform-specific candidate O-glycosites using in vitro enzyme analysis. We previously showed that in vitro enzyme assays with short peptides correlate well with in vivo activity and specificity in cells [20,30], and here, we expanded the panel of peptides to include representative subsets of those identified here. For comparison, we also selected a subset of glycosites that were found to be representative of redundant functions of GalNAc-T3 and predicted not to be isoform-specific substrates. We tested recombinant soluble versions of the three isoforms GalNAc-T1, GalNAc-T2, and GalNAc-T3, as well as two other isoforms (T4 and T11) weakly expressed in liver and HepG2 cells (Fig 1A). We also analyzed the non-redundant O-glycosites for prominent sequence motifs but did not identify such (Fig EV4C), although GalNAc-T2 clearly shows preference for Pro in position –1 and –3, and GalNAc-T3 tolerates more charged residues both N- and C-terminal to glycosylation sites as previously demonstrated for synthetic random peptide libraries [31]. The results confirmed that the majority of the candidate sites represent GalNAc-T isoform-specific acceptor substrate glycosites and thus provide further validation for the differential glycroteomics strategy (Fig EV4A and B and Table EV3). Clearly further validation and testing in other cell systems are required for arriving at a definitive isoform-specific dataset.

Comparison of the identified candidate glycosites for the three GalNAc-T isoforms (Fig 3D–F) revealed that these were largely different for each GalNAc-T isoform, which provides strong support for the validity of the strategy and the identified candidates. Among proteins with T1-specific sites (Table EV2), we found several proteins involved in basement membrane (BM)/extracellular matrix (ECM) organization (e.g., HSPG2, FN1, VCAN, DAG1, SCD4, FRAS1, FGA, MIA3, VWA1, MATN3), which is in agreement with and confirms the proposed association of murine Galnt1 and BM/ECM formation [32,33]. Furthermore, we found a T1 isoform-specific glycosite on Osteopontin (Ser280), which correlates well with and confirms a previous report demonstrating reduced O-glycosylation of Osteopontin in Galnt1−/− KO mice [34]. In agreement with our previous studies [4], we found ApoC-III (Thr94) among the T2-specific sites. Additionally, among the T2-specific sites identified in our previous non-quantitative study [4] for which quantitative data were obtained in this study, > 80% were identified as T2-specific candidates. Interestingly, a total of 13 of the 67 proteins with identified T2 candidate-specific O-glycosites (Table EV2) are
Figure 3. Differential O-glycoproteomes in HepG2Sc.

A–C Histograms showing the distribution of M/L quantitation ratios of monoglycosylated peptides identified in (A) HepG2Sc ΔT1, (B) HepG2Sc ΔT2, and (C) HepG2Sc+T3. Glycopeptides with M/L ratio less than −1 are colored red, and glycopeptides with a M/L ratio greater than +1 are colored green.

D–F Venn diagrams showing the distribution of candidates for isoform-specific sites among HepG2Sc ΔT1, HepG2Sc ΔT2, and HepG2Sc+T3 applying a log_{10} (±1) cutoff (D) (excluding sites identified in both TCL and SEC for each isoform), and (E) TCL alone and (F) SEC alone.

De novo introduction of GalNAc-T3 resulted in a large expansion of the O-glycoproteome of HepG2 comparable to the loss observed with knockout of GalNAc-T1 and GalNAc-T2 (Fig 3). GalNAc-T3 has unique substrate specificity and accepts charged residues in near proximity of acceptor sites (Table EV2 and Fig EV4). Proteins rely on limited proteolytic activity upon activation and such cleavage sites are often composed of basic residues. Interestingly, several of the T3-specific sites were located in close proximity to such activation sites either on the active protein or in the propeptide (PROS1, FV, ASHG, LOX, F2RL1, and RARRES2). A number of the identified candidates for T3-specific glycosylation have previously been found in human cell lines endogenously expressing GalNAc-T3, for example, Pro-SAAS (Thr53) and PAR-2 (Thr49) [20]. The analysis shows that de novo introduction of a GalNAc-T can have dramatic effects on the O-glycoproteome of a cell and may serve to help understand the functional consequences of the aberrant expression of GalNAc-T3 found in many cancers [35–37].

HepG2 and liver express minor levels of other well-characterized GalNAc-T isoforms (GalNAc-T4, GalNAc-T10, GalNAc-T11) (Figs 1 and EV1A), and their contribution was not addressed in the present study. GalNAc-T4 and GalNAc-T10 are mainly functioning as so-called follow-up enzymes using partially GalNAc-glycosylated substrates to complete glycosylation [1], and the analysis of these functions in a quantitative way is not simple with the current strategy where we have limited the analysis to single O-glycosites. GalNAc-T11 is unique in controlling O-glycosylation of the low-density-lipoprotein (LDL) receptor in the LDL binding region [38], and here, we identified one of these O-glycosites (TI08), and the glycosylation of this site was as predicted not affected by loss or gain of the three other GALNT genes studied (Table EV2). We recently found that GalNAc-T11-mediated O-glycosylation is required for LDL binding and uptake (unpublished observations), and it is our hypothesis that this isoform serves a selective role in these functions. However, at this stage, the finding that loss or gain of GalNAc-T1, GalNAc-T2, and GalNAc-T3 isoforms in HepG2 does not affect LDLR O-glycosylation provides additional validation for the strategy.

Our study provides the first comprehensive and quantitative cell-based evidence that the O-glycoproteome of a cell is differentially regulated by its repertoire of GalNAc-T isoenzymes expressed. The expression of individual GalNAc-T isoenzymes thus affects a limited subset of the O-glycoproteome, while the majority of the O-glycoproteome is covered by redundancy among the GalNAc-Ts expressed in a cell. This provides strong evidence that cells have the potential for differential and dynamic regulation of O-glycosylation of distinct subsets of proteins by the non-redundant functions of individual GalNAc-Ts. Moreover, the results suggest that the non-redundant contributions to the O-glycoproteome of the three GalNAc-T isoenzymes studied can be categorized to different biological processes. In particular, we found strong support for the putative roles in lipid metabolism for GalNAc-T2 and in basement membrane/extracellular matrix composition for GalNAc-T1, where each isoenzyme controlled O-glycosites in multiple proteins with pivotal roles in these pathways [22,23,32].

The non-redundant functions identified for each GalNAc-T are based on complete loss or gain of function, and hence display the extreme effects. It is therefore still unknown how quantitative changes in expression of individual GalNAc-Ts will affect the O-glycoproteome, and we are currently approaching this with reintroduction of inducible GalNAc-Ts in cell line systems. Within the non-redundant subsets, there will be a broad range of acceptor substrates for which the kinetic properties of each particular GalNAc-T isoform will clearly be very different as suggested by extensive in vitro enzyme analyzes (Fig EV4) [30], and it may be expected that only a fraction of the non-redundant O-glycosites identified, in fact, will be substantially affected by minor quantitative changes in expression of individual GalNAc-Ts. Thus, physiological regulation of individual GalNAc-Ts in cells may have even more subtle and specific consequences for the O-glycoproteome and a major challenge will be to identify O-glycosites that are truly regulatable by quantitative changes in GalNAc-T expression. More broadly, the challenge is to identify the few non-redundant O-glycosites that are essential for health, and that when not glycosylated, due to deficiencies in or dysregulation of GALNT genes, cause disease or susceptibility to disease.

Transcriptional analysis of isogeneic HepG2WT cell lines with different GalNAc-T repertoires

To evaluate potential transcriptional compensation of GalNAc-T genes in response to loss-of-function mutations, we generated a panel of isogenic HepG2WT cell lines with the same KO and KI of GALNT Ts as analyzed for the O-glycoproteomics studies (Fig EV1B and Table EV1). We used wild-type HepG2 cells for these experiments such that potential effects in the transcriptomes could be related solely to loss or gain of the GalNAc-T isoforms. First, we used immunocytology to compare expression of 12 GalNAc-T isoforms for which we have developed monoclonal antibodies to, and this demonstrated that the protein expression of these was essentially unaltered by loss or gain of GalNAc-T1, GalNAc-T2, and GalNAc-T3 in HepG2 SC and WT cells (Fig EV1 and data not shown). The transcriptomes of two independent clones of each HepG2WTAT1/HepG2WTAT2/HepG2WT+T3 KOs were analyzed by RNA-sequencing (RNAsq, data are available at www.ebi.ac.uk/ arrayexpress with accession number E-MTAB-3844), and although we saw minor differences in GALNT1 and GALNT2 transcript levels, we found no evidence of compensatory changes in transcript levels in any members of the large GALNT family (Fig 4) or other known glycosyltransferase genes (for an updated list of glycosyltransferase genes, see [39]), suggesting that they serve independent functions without cross talk.

To further explore the transcriptome data, we used a stringent and conservative statistical analysis for differential RNAseq transcriptomic data [40], and comparison of the transcriptomes of HepG2WT and HepG2WTAT1 showed that 908 genes were significantly upregulated and 791 genes downregulated (Fig EV5 and Table EV4). Considerably fewer transcripts were affected in HepG2WTAT2 with 254 genes upregulated and 210 downregulated (Fig EV5 and Table EV5). Remarkably, the de novo introduction of the GalNAc-T3 isoform in HepG2 that resulted in a marked expansion of the O-glycoproteome (Table EV2) had little effect on the transcriptome (Fig EV5 and Table EV6) as only 14 genes were
significantly upregulated and 53 genes downregulated in HepG2WT-T3. This demonstrates for the first time that loss of endogenous GalNAc-T isoforms has major impact on cells, while in contrast aberrant expression of GalNAc-T3 in the context of a liver cell only appears to have little effect.

Despite a limited number of transcripts being affected by de novo introduction of GalNAc-T3 (Fig EV5 and Table EV6), it may be interesting to note that matrix metalloproteinase 14 (MMP14 also known as MT1-MMP) was downregulated more than 50-fold by expression of GalNAc-T3 (HepG2WT-T3). MMP14 promotes cellular migration and invasion in vivo, and a number of specific substrates have been reported [41]. Specifically, we identified a non-redundant glycosite (T263) in close proximity to the MMP14 cleavage site in matrix metalloproteinase 11 (MMP11 also known as Stromelysin-3 [42]) (\(253^\text{VQLH}1^\text{YGQP}W\text{PPTY}TSRT}^{268}\)), cleavage site arrow and glycosite underlined). MMP11 is inactivated by the MMP14 processing in this site [42]. MMP11 is upregulated in cancer and suggested to play a role in invasive growth [43–45], and it is possible that the upregulation of GalNAc-T3 (and perhaps its paralog GalNAc-T6 with similar specificity [46] often found in cancer [47,48] may help block MMP14-mediated inactivation of MMP11 and thus promote invasive growth. Clearly, this is speculation and further studies are needed to address the functions of these GalNAc-T isoforms, but the example demonstrates the potential connectivity of site-specific O-glycosylation and relevant cellular processes.

Finally, to further explore the marked differential gene expression observed in cells without GALNT1 or GALNT2, we analyzed the differential transcriptomes for HepG2WTΔT1 and ΔT2 by GO enrichment analysis to evaluate potential global effects. We found significant (hypergeometric test; \(P\)-value \(< 0.05\)) enrichment of distinct combinations of biological process terms for each isoform (Fig EV6), suggesting that individual GalNAc-Ts affect the transcriptome of cells in isoform-specific manner and that complete loss of function of a GalNAc-T gene induces a selective transcriptional response. Clearly, further studies are needed to validate these initial findings and probe into cause and effects. Interestingly, however, GO terms significantly enriched for among T2-specific upregulated transcripts were mainly related to lipid metabolism biological processes, which is in line with the isoform-specific glycosylation functions identified here and the genetic studies associating GALNT2 polymorphisms with dyslipidemia [22,23].

An increasing number of GALNT genes have been identified as candidate genes for human diseases (for review see [18]). Apart from the strong evidence for GALNT2 and dyslipidemia, studies have pointed to associations of GALNT4 with coronary artery disease [49], GALNT7 with Alzheimer’s disease [50], GALNT10 with obesity and high body mass index [51], GALNT13 with elevated tricuspid regurgitation jet in sickle cell disease patients [52], and GALNT14 with hepatocellular carcinoma [53]. The strategy presented here produced a list of candidate glycoproteins and glycosites specifically controlled by the GalNAc-T2 isoform in a liver cell line that may aid in dissection of molecular mechanisms underlying the association with dyslipidemia. More generally, the quantitative O-glycoproteomics strategy should be applicable to any GalNAc-T isoform in discovery and dissection of non-redundant biological functions in appropriate cell line systems and eventually in vivo in model organisms or patients. Although perhaps premature, the quantitative transcriptomics of appropriate isogenic cell models may also support discovery of disease-causing functions.

In conclusion, the repertoire of GalNAc-Ts of a cell directs its O-glycoproteome, and individual GalNAc-Ts serve limited non-redundant functions that may affect distinct subsets of metabolic processes. The developed isogenic cell models open for wider discovery of functions of the large GalNAc-T family and deciphering of molecular mechanisms causing disease.

Figure 4. RNAseq transcriptome analysis of isogenic HepG2 cells.
Transcript levels in RPKM (reads per kilobase transcriptome per million mapped reads) of GalNAc-Ts in two individual clones of isogenic HepG2 lines as indicated. Apart from minor changes in the GALNT1 and GALNT2 transcript levels, all transcript levels were similar. Bars represent mean values, and error bars represent SD.
Materials and Methods

ZFN gene targeting

HepG2ΔT1 and HepG2ΔT2 in WT and SC background were generated as previously described [4]. HepG2+T3 cells were generated by stable integration of GalNAc-T3 as previously described [4] (see Appendix Supplementary Methods). Cells were maintained in Dulbecco’s modified Eagle’s medium supplemented with 10% FBS, 1% glutamine and 1% NEAA.

For ICC characterization, cells were grown on sterile teflon-coated slides (Clearcell, Histolab) and briefly fixed with ice-cold acetone. Cells were then sequentially incubated with MAb to GalNAc-T1 (4D8), GalNAc-T2 (4C4), GalNAc-T3 (2D10), or Tn O-glycans (SF4) overnight and FITC-conjugated rabbit anti-mouse immunoglobulin (Dako) for 45 min and mounted with ProLong Gold antifade reagent (DAPI) (Invitrogen). Fluorescence microscopy was performed using a Zeiss Axioskop 2 plus with an AxioCam MR3.

Dimethyl labeling and lectin weak affinity chromatography

Samples of 100 ml cell culture supernatant or 0.5 ml packed cells were processed [4] (see Appendix Supplementary Methods). For quantitative differential glycoproteomics, tryptic digests of isogenic pairs of SCs were labeled with light (L) and medium (M) isotopomer dimethyl labels [27]. Cleared acidified digests were loaded in equal amounts (peptide concentration) onto equilibrated SepPak C18 cartridges (Waters) followed by 3 equal amounts (peptide concentration) onto equilibrated SepPak mermic dimethyl labels [27]. Cleared acidified digests were loaded in quantitative differential glycoproteomics, tryptic digests of isogenic pairs of SCs were labeled with light (L) and medium (M) isotopomer dimethyl labels [27]. Cleared acidified digests were loaded in equal amounts (peptide concentration) onto equilibrated SepPak C18 cartridges (Waters) followed by 3 equal amounts (peptide concentration) onto equilibrated SepPak mermic dimethyl labels [27]. Cleared acidified digests were loaded in quantitative differential glycoproteomics, tryptic digests of isogenic pairs of SCs were labeled with light (L) and medium (M) isotopomer dimethyl labels [27]. Cleared acidified digests were loaded in equal amounts (peptide concentration) onto equilibrated SepPak C18 cartridges (Waters) followed by 3 equal amounts (peptide concentration) onto equilibrated SepPak mermic dimethyl labels [27]. Cleared acidified digests were loaded in quantitative differential glycoproteomics, tryptic digests of isogenic pairs of SCs were labeled with light (L) and medium (M) isotopomer dimethyl labels [27]. Cleared acidified digests were loaded in quantitative differential glycoproteomics, tryptic digests of isogenic pairs of SCs were labeled with light (L) and medium (M) isotopomer dimethyl labels [27].

Mass spectrometry

LWAC elution fractions were fractionated by isoelectric focusing according to [28] and analyzed by EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via nanoSpray Flex ion source to an LTQ-Orbitrap Velos Pro spectrometer (Thermo Scientific) as previously described [29]. Data were processed using Proteome Discoverer 1.4 software (Thermo Scientific) (see Appendix Supplementary Methods).

In vitro glycosylation

Peptides were designed and synthesized with acceptor residues in center position of a 20-mer peptide (NeoBioScience). GalNAc-transferase assays, monitored by MALDI-TOF mass spectrometry, were performed as previously described [30].

RNA transcriptomic analysis

Total RNA was extracted from exponentially growing cells using RNAeasy® kit (Qiagen). RNA integrity and quality were determined using Bioanalyzer instrumentation (Agilent Technologies). RNAseq was performed by Beijing Genetics Institute (BGI). Briefly, library construction was performed using Illumina Truseq RNA Sample Preparation Kit and subjected to next-generation sequencing using Illumina HiSeq 2000 System (Illumina, USA). See Appendix Supplementary Methods for more details.

Bioinformatic analysis

The aligned reads from the BGI analysis were analyzed combining DESeq [54] and EdgeR [55] methods to determine the differentially expressed transcripts (see Appendix Supplementary Methods). GO enrichment analyses of glycoproteomic and transcriptomic data were performed in R using GO and GOStats [56].

Data availability

The mass spectrometry glycoproteomics data have been deposited to the ProteomeXchange Consortium [57] via the PRIDE partner repository with the dataset identifier PXD002770 and RNAseq data can be found at www.ebi.ac.uk/arrayexpress with accession number E-MTAB-3844.

Expanded View for this article is available online.

Acknowledgements

We thank James Paulson, Ramon Hurtado Guerrero, and Frederic Bard for critical reading of the manuscript. This work was supported by The Danish Research Councils (Sapere Aude Research Talent Grant to KTS), The Mizutani Foundation, Kirsten og Freddy Johansen Fonden, A.P. Møller og Hustru Chastine Mc-Kinney Møllers Fond til Almene Formaal, The Novo Nordisk Foundation, a program of excellence from the University of Copenhagen (CDO2016), and The Danish National Research Foundation (DNRF107).

Author contributions

KTS, HJJ, SYV, and HC designed and performed experiments, analyzed data, and wrote the paper. HHW and EPB analyzed data. YK, CKG, and SLK performed experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

References


Deconstruction of protein O-glycosylation

Katrine T Schjoldager et al

EMBO reports

Published online: November 13, 2015


