Cellular O-Glycome Reporter/Amplification to Explore O-Glycans of Living Cells

Matthew R. Kudelka[1], Aristotelis Antonopoulos[3], Yingchun Wang[1], Duc M. Duong[2], Xuezheng Song[1], Nicholas T. Seyfried[1,2], Anne Dell[3], Stuart M. Haslam[3], Richard D. Cummings[1] and Tongzhong Ju[1,*]

[1]Department of Biochemistry and [2]ENNCF Proteomics Core
Emory University School of Medicine, Atlanta, GA 30322

[3]Department of Life Sciences, Imperial College London, London, SW7 2AZ, UK.

Running Title: Cell O-Glycome Reporter/Amplification

*Correspondence to:
Tongzhong Ju, M.D./Ph.D.
Emory University School of Medicine
1510 Clifton Rd NE, Room 4029
Atlanta, GA 30322
Phone: 404-727-6166 (O)
Email: tju@emory.edu

Abstract

Protein O-glycosylation plays key roles in many biological processes, but the repertoire of O-glycans synthesized by cells is difficult to determine. Here we describe a new approach termed Cellular O-Glycome Reporter/Amplification (CORA), a sensitive method to amplify and profile mucin type O-glycans synthesized by living cells. Cells incubated with peracetylated benzyl-α-N-acetylgalactosamine (GalNAc-α-Benzyl) convert it to a large variety of modified O-glycan derivatives that are secreted from cells, allowing easy purification for analysis by HPLC and mass spectrometry (MS). CORA results in ~100-1000-fold increase in sensitivity over conventional O-glycan analyses and identifies a more complex repertoire of O-glycans in more than a dozen cell types analyzed. Furthermore, CORA coupled with computational modeling allows predictions on the diversity of the human O-glycome and offers new opportunities to identify novel glycan biomarkers for human diseases.

Introduction:

Protein glycosylation is a common post-translational modification in all animals that helps to create post-genomic diversity[1]. Although systems-level approaches to evaluate genomes and proteomes have revolutionized our understanding of physiology and disease, similar approaches for glycomics are lacking in most biological settings. A primary need in glycomics is simple and sensitive technologies to analyze all glycans synthesized by cells (the cellular glycome)[2]. This has been challenging due to the diversity and complexity of glycans, low abundance of certain glycan species, poor sensitivity of existing glycomics approaches, and lack of efficient and unbiased strategies to release glycans from complex samples. Current technologies to evaluate glycans in biological samples require their release from glycoconjugates by chemo-enzymatic methods, followed by analyses by mass spectrometry (MS) and other technologies[2]. Relatively large amounts of biological samples are often needed for detailed structural analyses, limiting the widespread application of glycomics. One way to at least partly overcome these challenges would be to "amplify" the glycome, similar to PCR-hybridization-based technologies for the genome[3]. Such amplification technology would allow analysis of microscale samples of biological material, facilitating clinical and biological discovery.

One of the most common types of protein glycosylation is mucin type O-glycosylation (R-GalNAcα1-O-Ser/Thr where R represents extended structures). O-glycans are present on >80% of proteins that traverse the secretory apparatus and are important in many normal and pathologic settings[4-7]. Nonetheless, little is known about how specific O-glycan structures regulate biology, largely due to a lack of effective technologies for O-glycomics. In contrast to N-glycans, which can be released enzymatically, O-glycans require chemical strategies – primarily alkaline β-elimination, which is inefficient, potentially biased, may result in O-glycan degradation via the peeling reaction, and requires extensive expertise not available in most laboratories[8].

To meet these challenges, here we describe a novel method for profiling and amplifying mucin type O-glycans from living cells, termed Cellular O-Glycome Reporter/Amplification (CORA). To assess the repertoire of glycosyltransferases and glycosylation reactions in the secretory pathway for the O-glycome, we developed a chemical O-glycan precursor, peracetylated Benzyl-α-GalNAc (Ac_3GalNAc-α-Benzyl), which is taken up by living cells, de-acetylated, modified by native glycosyltransferases in the presence of nucleotide sugar-donors in the secretory pathway, and then secreted into media (**Fig. 1**). Benzyl-α-GalNAc structurally mimics the precursor GalNAcα1-O-Ser/Thr (Tn antigen) in O-glycoproteins[9] and its peracetylation promotes transport across the plasma membrane[10]. Living cells incubated with Ac_3GalNAc-α-Benzyl secrete a diverse set of benzyl-α-O-glycans that represent the cellular O-glycome. Using CORA, we identified 76 O-glycans, including those observed by traditional alkaline β-elimination/MS. Importantly, CORA provided ~100-1000-fold enhanced sensitivity, much cleaner MS profiles of O-glycans, and revealed novel O-glycans in a variety of cancer and primary cells not seen by classic approaches. This simple, sensitive, and versatile method for amplifying and profiling the cellular glycome could be a transformational technology in the biomedical sciences and allow dynamic studies of O-glycosylation in living cells.

Results

Cells efficiently uptake Ac_3GalNAc-α-O-Bn and secrete Bn-O-glycans

Mucin type O-glycan biosynthesis begins with transfer of GalNAc to Ser/Thr residues in glycoproteins to generate GalNAcα1-O-Ser/Thr (Tn antigen), which T-synthase extends to the dominant Core 1 O-glycan Galβ3GalNAcα1-O-Ser/Thr. Bn-α-GalNAc structurally mimics GalNAcα1-O-Ser/Thr and has been used as an acceptor in *in vitro* assays of enzyme activities for generating Core 1 and 3 disaccharides[11, 12]. However, Core 3 GnT is only observed in primary GI tissues and not cell lines[11, 12]. Prior studies have shown that high concentrations of Bn-α-GalNAc (2~20mM) can inhibit extension of Core 1 O-glycans in cultured cells[13], apparently by acting as a surrogate acceptor for T-synthase and compromising extension of the Core 1 O-glycan on natural glycoprotein acceptors[14]. However, we hypothesized that Bn-α-GalNAc at low concentrations might not inhibit biosynthesis but could be useful as an O-glycan precursor to form free Bn-O-glycans representing the cellular O-glycome. We utilized the more highly hydrophobic derivative Ac_3GalNAc-Bn (**Supplementary Fig. 1a**) to enhance cellular uptake[10] and predicted that upon entry into cells Ac_3GalNAc-Bn would become activated by cytosolic esterases to regenerate Bn-GalNAc. Bn-GalNAc would then be transported into the secretory pathway, modified by glycosyltransferases, and secreted into media as biosynthetic Bn-O-glycans that could be easily purified and analyzed by MS. We termed this technology Cellular O-Glycome Reporter/Amplification or CORA and the workflow of the technology is described **Fig. 1**.

To test this, we cultured adherent (HEK293) and suspension (Molt-4) cells in complete media containing 50μM Ac_3GalNAc-Bn or vehicle (DMSO) for 3 days. Putative Bn-O-glycans from media were separated from larger material using a cut-off membrane, purified by C18 chromatography, permethylated, and analyzed by MALDI-TOF-MS. Here as elsewhere we only analyzed the permethylated non-sulfated/non-phosphorylated glycans. We observed clean MALDI-TOF profiles with Bn-O-glycan compositions corresponding to Core 1 and 2-based structures (**Fig. 2**) from cells cultured with Ac_3GalNAc-Bn but not vehicle, indicating efficient uptake and modification of the O-glycan precursor by glycosyltransferases *in vivo*.

To assess whether peracetylation of Bn-α-GalNAc enhanced uptake and subsequent sensitivity, we incubated breast cancer cells (MDA-MB-231) with 0-250μM of Ac_3GalNAc-Bn or Bn-GalNAc for 3 days. Bn-O-glycans with predicted sialylated Core 1 structure were observed at a concentration as low as 25μM Ac_3GalNAc-Bn, but such derivatives were only observed at the highest concentration of 250μM Bn-GalNAc (**Supplementary Fig. 2).** Thus, peracetylation improved the sensitivity, avoiding higher concentrations that might have side effects, as explored below.

To optimize conditions, we cultured MDA-MB-231 cells with 0-250μM of Ac_3GalNAc-Bn for 2-4 days. Bn-O-glycans were seen at all time points and concentrations, down to 5μM of Ac_3GalNAc-Bn for 2 days (**Supplementary Fig. 3**) and their profiles were stable over time. Remarkably, increasing Ac_3GalNAc-Bn concentration shifted abundance from disialylated to monosialylated Core 2 (**Supplementary Fig. 3d**), supporting observations that glycosyltransferase:substrate ratios drive glycan microheterogeneity[15]. This shift also suggests that a sialyltransferase and/or CMP-NeuAc required for generating disialyl Core 2 is limiting,

thus CORA should be performed at low concentrations of Ac_3GalNAc-Bn. For most subsequent studies, we incubated cells with 50μM Ac_3GalNAc-Bn for 3 days. At these conditions, Ac_3GalNAc-Bn was not toxic to cells (**Supplementary Fig. 4**) and did not alter cellular properties (**Supplementary Fig. 5**) or cell surface O- or N-glycosylation (**Supplementary Fig. 6, 7**).

Synthesis of Core 1-based Bn-O-glycans in cells requires active T-synthase

The above results show that cells cultured with Ac_3GalNAc-Bn produce and secrete free Bn-O-glycans into media containing the Core 1-based structures requiring T-synthase. To confirm that Bn-GalNAc can only be utilized by glycosyltransferases involved in mucin-type O-glycan biosynthesis, but not other irrelevant or unknown pathway(s), we tested CORA in T-synthase-competent versus T-synthase-null cells. Biosynthesis of active T-synthase requires a specific molecular chaperone Cosmc and genetic or epigenetic disruption of Cosmc results in inactive T-synthase and consequent expression of Tn and its sialylated version, sialylTn antigens[16-18].

Thus, we performed CORA on cells with mutant *Cosmc* and their wildtype/rescued counterparts. Only cells with functional *Cosmc* and active T-synthase, but not cells with dysfunctional *Cosmc* and inactive T-synthase, secreted Bn-O-glycans when administered Ac_3GalNAc-Bn (**Fig. 3a-d**). Furthermore, cells given the isomer Ac_3GlcNAc-Bn (**Supplementary Fig. 1b)** did not secrete Bn-O-glycans (**Supplementary Fig. 8**), but rather no glycan or the simple trisaccharide Neu5Ac-Gal-GlcNAc. These demonstrate that CORA faithfully reports the O-glycome and that modification in cells requires active T-synthase.

Accuracy of CORA

We compared O-glycome profiles from CORA to β-elimination, available through the CFG (http://www.functionalglycomics.org). WEHI-3 and HL-60 cells were analyzed because they have complex O-glycomes with unique structures, such as Cad, and extended poly-N-acetyllactosamines [3Galβ4GlcNAcβ-]$_n$ that are challenging to detect by β-elimination of lysates but observed on purified glycoproteins[19, 20].

From **Figure 4** and **Supplementary Figures 9-10**, the MALDI-TOF-MS and MS/MS profiles show that WEHI-3 and HL-60 cells produced 11 and 40 glycan structures, respectively, including sialylated Core 1 and 2-based glycans for both cells and Cad antigen in WEHI-3 (**Fig. 4a, b**). CORA detected most of the compositions from β-elimination (12 of 13 for WEHI-3, 4 of 7 for HL-60) and many additional compositions (16 for WEHI-3, 6 for HL-60) (**Fig. 4c–e**), which were generally the most complex, including poly-N-acetyllactosamines with ~3 repeats and I antigen. Importantly, the 4 masses observed by β-elimination but not CORA (**Fig. 4c, d**) are Core 1 and 2-based glycans, lacking terminal sialylation, and therefore are most likely biosynthetic intermediates, such as from glycoproteins within an intermediate Golgi compartment prior to secretion.

To confirm our results, we repeated the experiments twice and obtained nearly identical O-glycome profiles (**Supplementary Fig. 11**) and also performed ESI-MS and obtained similar results to MALDI-MS and MS/MS (**Supplementary Fig. 12**). Together, these demonstrate that

CORA reflects the cellular O-glycome, which is relatively stable under optimal culture conditions.

Sensitivity of CORA

β-elimination often requires ≥10^7 cells and produces many unassignable peaks[21]. To determine how many cells are needed to get clean, interpretable profiles with CORA, we profiled four cell lines each seeded at 5x10^5, 10^5, or 2x10^4 cells. In **Figure 5a-d,** we obtained O-glycomes from all lines seeded at 5x10^5 and 10^5 cells and 3 of 4 lines seeded at 2x10^4 cells (**Fig. 5e)**. Significantly, profiles did not change with different cell numbers. The detection of Bn-O-glycans from 2x10^4 cells cultured for 3 days to a total of ~8x10^4 cells (assuming ~24 hour doubling times) represents a ~100-1000-fold increase in sensitivity compared to β-elimination.

Profiling the O-glycome of mouse and human primary cells

Because primary cells differ metabolically from cancer cells[22], we validated CORA on primary human and mouse cells. First we profiled immortalized murine pulmonary endothelial cells (mPECs) with or without *Cosmc* from Tie2-Cre^+;$Cosmc^{F/+}$ mice[23] to determine if, similar to other cell lines, Bn-O-glycan secretion requires functional *Cosmc*. mPECs with functional Cosmc, but not deleted *Cosmc*, secreted Bn-O-glycans (**Supplementary Fig. 13a, b**). Next we isolated wt Tn(-) mPECs from these same mice and performed CORA (**Supplementary Fig. 13c, d**). Glycan structures were similar in immortalized and primary mPECs, except for a glycan with the disialyl motif, only found in the primary cells. However, the ratio of O-glycans differed in these two mPECs, suggesting that transformation may disrupt glycan biosynthesis, as in human tumors[24].

Next, we evaluated primary human dermal fibroblasts and umbilical vein endothelial cells (HUVECs) (**Fig. 6, Supplementary Fig. 14-15**). HUVECs produced 43 O-glycans, including those containing poly-N-acetyllactosamine, Lewis structures, blood group antigens, and I antigen, as confirmed by MS/MS (**Fig. 6a, Supplementary Fig. 14a, 15**). Fibroblasts also produced at least 18 glycans (unique masses), including poly-N-acetyllactosamine, Lewis structures, and blood group antigens (**Fig. 6b, Supplementary Fig. 14b**). O-glycans are essential for endothelia but unexplored in fibroblasts; however, their remarkable diversity, described here for the first time, highlights their potential importance.

Ac3GalNAc-Bn treatment does not alter cellular properties

High concentrations (2~20 mM) of GalNAc-Bn can disrupt cell morphology and glycosylation[14]. To avoid this, we used a low concentration of 50μM Ac_3GalNAc-Bn and assessed toxicity by XTT after 3 days culture. XTT conversion to a formazan dye was similar in vehicle and Ac_3GalNAc-Bn treated cells (**Supplementary Fig. 4**), indicating that 50μM Ac_3GalNAc-Bn was not toxic to cells. After culture with Ac_3GalNAc-Bn, we also assessed gross morphology by light microscope and cellular granularity by side scatter on FACS. No changes were observed (**Supplementary Fig. 5**), demonstrating intact cell viability and morphology.

Low concentrations of Ac_3GalNAc-Bn do not significantly affect O- and N-glycosylation of glycoproteins in cells

To investigate whether low concentrations of Ac_3GalNAc-Bn alter protein O- or N-glycosylation in cells, we incubated cells with 50μM Ac_3GalNAc-Bn for 3 days and blotted with multiple lectins. Lectins allow a sensitive 'global' analysis of protein glycosylation (specific terminal sugar and linkage-specific determinants) in multiple glycoproteins, which is preferred over conventional MS analysis of released glycans that again might be limited in sensitivity and does not provide linkage information. We used PNA to assess O-glycosylation (**Supplementary Fig. 6**) and RCA-I, SNA, and MAL-I to assess N-glycosylation (**Supplementary Fig. 7**). No changes in lectin binding were observed after culture with 50μM Ac_3GalNAc-Bn, demonstrating no apparent change in O or N-glycosylation of cellular glycoproteins.

Ac_3GalNAc-Bn and Bn-O-glycans are stable in cell culture media

Peracetylation of GalNAc-Bn clearly enhanced production of Bn-O-glycans due to efficient cellular uptake of the compound. To evaluate the stability of GalNAc-Bn acetylation, we incubated 50μM Ac_3GalNAc-Bn in complete media for 3 days. Ac_3GalNAc-Bn was recovered without significant loss of the signal or appearance of GalNAc-Bn **(Supplementary Fig. 16)**, indicating that Ac_3GalNAc-Bn is stable in media and not appreciably deacetylated.

Cells might produce cell surface or secreted glycosidases that reshape their glycome[1] and perhaps modify secreted Bn-O-glycans. However, this is unlikely given that Bn-O-glycan profiles are stable across time (**Supplementary Fig. 3**) and closely resemble those from β-elimination/MS (**Fig. 4**). Nonetheless, we directly investigated this scenario using LS174T-Tn(-) and Tn(+) cells, with active or inactive T-synthase. We cultured LS174T-Tn(-) cells with Ac_3GalNAc-Bn and directly analyzed Bn-O-glycans or further incubated this media with LS174T-Tn(+) cells before analyzing Bn-O-glycans. Because LS174T-Tn(+) cells cannot synthesize O-glycans beyond Bn-GalNAc, any change in Bn-O-glycan structure would indicate glycosidases. Importantly, Bn-O-glycans from LS174T-Tn(-) were not altered after culture with LS174T-Tn(+) cells (**Supplementary Fig. 17**), showing that Bn-O-glycans are stable after secretion from cultured cells.

CORA promotes the discovery of novel glycans

MS/MS sequencing is needed for definitive determination of glycan structure. However, this is often not possible with β-elimination because of insufficient material. Using CORA, we performed MS/MS sequencing on HL-60, WEHI-3, and HUVEC cells and observed many diverse glycan structures (11 for HL-60, 40 for WEHI-3, 43 for HUVEC) as well as novel and unexpected O-glycans, including extended Core 1 (HL-60), Vim-2 (HL-60, HUVEC), disialic acid (WEHI-3), and I antigen (WEHI-3, HUVEC) (**Fig**. **4, 6, Supplementary Fig. 9, 10, 15**). Although this is the first report of extended Core 1 on HL-60 cells and of Vim-2 and disialic acid on O-glycans from myelocytic cells, these epitopes have been observed in related contexts or structures, indicating that they are biosynthetically reasonable products[25-27].

The I antigen replaces i antigen on RBCs after embryogenesis, but had not been identified on O-glycans from other cell lineages, except in secretions[28]. The discovery of I antigen on O-glycans from 2 distinct non-RBC cell lineages suggests that this may be a common, yet unappreciated structure. We have no explanation as to why I antigen was not previously observed on O-glycans from these cell lines, except that in general, O-glycans released by β-elimination are present in insufficient quantities to identify larger O-glycans in the high mass range and when larger O-glycans are identified, MS/MS usually is not feasible. As CORA is an amplification technology, it permits analyses of small numbers of cells since they continuously produce more Bn-O-glycans, thus amplifying their O-glycomes.

CORA provides information to evaluate the complexity of cellular O-glycome

The complete repertoire of O-glycans is not known. Here we analyzed 18 cell lines including primary cells and *Cosmc*-deficient cells and identified 57 unique O-glycan compositions (**Supplementary Table 1, 2**). Detailed MS/MS structural characterization on 3 of these cells (HUVEC, HL-60, WEHI-3) revealed 76 unique structures, derived from 48 unique compositions (#structures:#compositions = 1.6) (**Supplementary Table 3)**. From our 18 cell lines, we evaluated the frequency of cell lines expressing a given glycan composition (**Supplementary Fig. 18a, b, Supplementary Table 2**) and determined that a few outstanding cell lines encompassed most of the complexity of all 18 glycomes (**Supplementary Fig. 18c**). To estimate the size of the non-sulfated/phosphorylated cellular O-glycome, we developed a computational model to predict the number of unique O-glycan compositions (Y) we would observe on average from a given number of cell lines (X). Our data best fit the equation Y = 10^(0.5856*log(X) + 1.024) (**Supplementary Fig. 19**), which predicts that the estimated 200 cell types in the human body produce 235 unique compositions. Using the observation above that the number of glycan structures may be 1.6-fold more than the number of compositions, we could thus predict that there are 1.6x235=376 non-sulfated/phosphorylated structurally different O-glycans in the human body. This is the first estimate of the size of the human cellular O-glycome and provides a roadmap to comprehensively sequence the human cellular O-glycome.

Discussion

Here we show that CORA provides a sensitive new approach to both amplify and profile the mucin-type O-glycome in living cells. O-glycans are present on most secreted and cell surface proteins, but previous strategies for O-glycomics have limited their investigation. Compared to traditional methods for O-glycan analysis, CORA 1) amplifies the O-glycome, 2) bypasses glycan release, 3) avoids O-glycan degradation, 4) enhances sensitivity ~100-1000x, 5) increases signal-to-noise, 6) detects novel complex O-glycans, 7) provides a biosynthetic process for real-time studies of dynamic changes in the O-glycome of living cells, and 8) is technically simple with potential for wide-spread and high-throughput application.

It is interesting that CORA is effective at low concentrations (<250μM) and incubation times with no observed impact on cellular properties or glycosylation. Prior studies had shown that treatment of cells with high concentrations of Bn-α-GalNAc produced mainly small Galβ3GalNAcα1-O-benzyl and sialyl-Galβ3GalNAcα1-O-benzyl derivatives[29, 30]. However, the success of CORA using low concentrations of this new Ac_3GalNAc-Bn derivative as a precursor

indicates that it is readily taken up by live cells, converted to Bn-α-GalNAc, efficiently utilized by the T-synthase, and further accessed by a wide range of enzymes in the secretory pathway, including the most terminal types of glycan modifications and extensions.

This approach may allow assessment of the total diversity and repertoire of O-glycans in an animal O-glycome. Recent studies using transfected CHO cells engineered to express all major O-glycan core structures, along with chemical release techniques, identified ~70 different glycan structures[31]. The repertoire of O-glycans is likely much larger, as mucin-type O-glycan determinants have been estimated to be nearly 1,000, with probably <500 non-sulfated O-glycans[32]. In this regard, our computational modeling allowed us to predict the size of the non-sulfated/phosphorylated animal O-glycome to be in the range of ~376 unique glycan structures. What drives this diversity? Amplification of an entire O-glycome from a simple chemical precursor suggests that the biosynthetic machinery and not diverse protein substrates drives O-glycan heterogeneity.

O-glycans are altered in cancer[7], but limitations in O-glycomics have made it difficult to define these changes. The distinguishing features of CORA enable analysis of the O-glycome of commonly available tumor cell lines without the need for chemical cleavage techniques. Such approaches involving β-elimination and treatment with alkaline borohydride, as originally developed by Carlson in 1966[33], lead to peeling reactions at the reducing end and partial loss of glycan integrity[8]. Theoretically, CORA could also be applied to small numbers of fresh tumor cells excised from patients and be used to monitor changes in the O-glycome during tumor progression, cellular differentiation, maturation, and in the presence of various stimuli.

An advantage of CORA is that live cells are used to generate the O-glycome, but the cells afterward can also be analyzed by conventional techniques if desired for comparative analyses. Advances in cell and organoid culture, have enabled culture of many normal and diseased tissues[34-37], yet glycan release is often not sufficiently sensitive to analyze such precious specimens. CORA could address this challenge by amplifying the glycome. Although we have limited our analyses to mucin type O-glycans, this may be a general strategy for amplifying and profiling many classes of glycosylation, with appropriate precursors. While we used small numbers of cells for glycan analysis, large numbers of cells in continuous culture could also be used as glycan biosynthesis factories to prepare any and all natural O-glycans, even those difficult to synthesize chemically. Bn-O-glycans or their derivatives could be isolated to generate therapeutic glycans, test the roles of unusual glycans in cell recognition or growth, or for display in glycan microarrays. Amplifying the O-glycome by CORA offers a new paradigm in cellular glycomics that will enable new types of investigations in a wide range of basic and clinical settings to give new insights into O-glycans in physiology and disease.

Methods

Compounds

Benzyl-α-D-GalNAc was purchased from Sigma. Benzyl-α-D-GlcNAcAc$_3$ and 4MU-α-D-GalNAc were purchased from Carbosynth. Benzyl-α-D-GlcNAcAc$_3$ was generated by addition of 2:1

pyridine:acetic anhydride to Benzyl-α-D-GalNAc for 1 hr at 65°C in molar excess, dried by centrivap, resuspended in 1ml water, and lyophilized.

Cell culture

Molt-4, Jurkat (clone E6-1), LS174T, HL-60, WEHI-3, HUVEC, and human dermal fibroblasts were purchased from ATCC. Lox were a kind gift from Dr. Oystein Fodstad. MKN45, Colo205, MDA-MB-231, and MCF7 were a kind gift from Dr. Henrik Clausen. Tn(-) and Tn(+) LS174T were subcloned from a mixed population as previously described[17]. Lox and Jurkat were transfected with full length Cosmc or empty vector (pcDNA3.1+) and selected by G418. Tn(-) cells were further sorted by FACS. Molt-4, Jurkat, Lox, HL-60, MKN45, and Colo205 were cultured in RPMI (Corning) supplemented with 10% FBS and 2% P/S. LS174T, MDA-MB-231, and MCF7 were cultured in DMEM (Corning) supplemented with 10% FBS and 2% P/S. MCF7 were further supplemented with 0.01mg/ml insulin. WEHI-3 were cultured in Iscove's (Corning) supplemented with 10% FBS, 0.05mM 2ME, and 2% P/S. HUVEC and human dermal fibroblasts were cultured with endothelial cell growth kit-VEGF (ATCC) and fibroblast growth kit-low serum (ATCC) as instructed. All cells were cultured on plastic, except HUVECs, which were cultured on plastic pre-coated with 0.1% gelatin.

Administration of compound

Ac_3GalNAc-Bn or Bn-GalNAc was dissolved to 50 or 100mM in DMSO and further diluted to 5 – 250μM in complete media with 5% FBS, except for HUVECs and human dermal fibroblasts which were incubated in ATCC pre-formulated media. Media with compound was administered 1 day (most cells) or 2 days (HUVEC, human dermal fibroblasts) after seeding. Cells were then incubated 2 – 4 days with compound before collecting media.

Glycan purification from media

Compound was added to complete media and collected after incubation with cells. Media was run over 10kDa centrifugal filter (Amicon – Ultra 4, Millipore) for ~30 minutes at 2465xg and flow through was collected. Bn-O-glycans were subsequently purified from flow through by Sep-Pak 3cc C18 cartridge (Waters) by gravity chromatography. The column was equilibrated with 2 x 2ml acetonitrile, then 4 x 2 ml 0.1% TFA. Media was applied and then column washed with 4 x 2ml 0.1% TFA. Bn-O-glycans were then eluted with 2 x 1.5 ml 50% acetonitrile/0.1% TFA. Eluent was divided into 3 fractions, centrivapped to remove organic solvents, and lyophilized.

Permethylation and glycan analysis

Dried samples were permethylated by standard procedures[38]. Two-hundred μl NaOH/DMSO slurry was added to samples followed by 200μl methyl iodide. Samples were shaken for 30 minutes and then spun down at 5000xg for 5 minutes. Supernatant was collected and chloroform extraction performed to isolate permethylated glycans. Five hundred μl chloroform and 500μl water were added to supernatant, mixed, and centrifuged 5000xg for 1 minute. Two more washes with 500μl water were performed before evaporating chloroform by centrivap for 30 minutes. Bn-O-glycans were then resuspended in 25 or 50 μl 50% methanol. 0.5μl matrix

(10mg/ml 2,5-dihydrobenzoic acid (Sigma), 50% acetonitrile, 0.1% TFA) and 0.5µl sample were spotted on an Anchorchip target plate, air dried, and analyzed by MALDI-TOF mass spectrometry using Ultraflex-II TOF-TOF system (Bruker Daltonics). Peak masses were identified and structures assigned by composition and knowledge of glycan biosynthetic pathways, or MS/MS where indicated. This procedure only detects non-sulfated/phosphorylated structures; however, these modifications have not been previously reported from cells in our study.

Cell and enzyme assays

T-synthase and mannosidase assay were performed as previously described[9]. Briefly, 4MU-GalNAc was incubated with cell lysate and reaction mix containing O-glycanase and UDP-Gal in triplicate for 45 minutes at 37°C to assay T-synthase. O-glycanase releases 4MU from 4MU-Core 1, which can be assessed in a fluorimeter. Reaction mix without UDP-Gal is used as control. To assess mannosidase activity, 4MU-mannoside was added to lysate in reaction mix in triplicate for 45 minutes at 37°C. Boiled lysate was used as control. Mannosidase releases 4MU from 4MU-mannoside. Fluorescence was converted to specific activities for T-synthase and mannosidase as described.

Supplementary Methods

Cell viability

XTT was performed per manufacturer instructions (ATCC) to assess cell viability after incubation with Ac_3GalNAc-Bn.

Endothelial cells

Primary ECs were obtained from EHC-Cosmc mice[23] and isolated from lung as previously described, with minor modifications[39, 40]. Briefly, murine lung was dissected, finely minced, and digested with 2 mg/ml type I collagenase (Worthington) at 37°C for 1 hr. Cell suspensions were filtered through a 70 µm strainer and then cultured in complete DMEM with endothelial cell growth supplement (Sigma). ECs were purified from culture with anti-PECAM-1 (CD31), and anti-CD102 (clone 3C4; BD) using Dynabeads®. Tn(+)-ECs were isolated with biotin-labeled anti-Tn antibody and Dynabeads® Biotin Binder (Invitrogen). Immortalized ECs were generated from primary ECs transfected with pPSVE1-B1a (from Jeffrey D. Esko, UCSD, La Jolla, California, USA), carrying SV40 viral T antigen.

Immunofluorescence

ECs were grown on 0.1% gelatin-coated chamber slides, fixed with 4% paraformaldehyde for 1 hr, and stained with anti-CD31 and anti-Tn antibodies. Staining was performed per the manufacturer's protocol (Zymed Laboratories) and visualized by Ix71 Olympus microscope with Deltavision Software Suite 6.0 (GE Healthcare).

Lectin blots

Biotin-labeled *Sambucus nigra* (SNA), *Ricinus communis agglutinin I* (RCA-I), and *Maackia amurensis I* (MAL-I) lectins were purchased from Vector and HRP-labeled *Peanut agglutinin* (PNA) lectin was purchased from Sigma. Cells were treated with 50μM Ac_3GalNAc-Bn or DMSO for 3 days and then harvested and lysed by vortexing cell pellet in lysis buffer (0.5% TritonX-100 in TBS supplemented with protease inhibitor tablet (Roche)) once per 5 minutes for 20 minutes over ice. Supernatant was collected after 16000xg centrifugation for 15 minutes at 4°C. Protein concentration was determined by BCA (Thermo Scientific) following manufacturer's instructions. Lysates were divided into 2 fractions, and 1 fraction was treated with 1μl neuraminidase (Roche)/100 μg lysate at 37°C overnight. Fetuin (Sigma) was used as control for lectin staining and enzymatic treatment. Lysates were then boiled in reducing SDS buffer, run on 4-20% Mini-PROTEAN-TGX gels (BIO-RAD) in Tris/Glycine/SDS running buffer, and transferred by Trans-Blot Turbo (BIO-RAD) semi-dry system to nitrocellulose. Membranes were rinsed and blocked with BSA and Tween-20 in essential buffer (25 mM Tris-HCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.15 M NaCl, pH 7.0) (SNA, RCA-I, MAL-I) or TBS (PNA) for 1 hr. Blocking buffer included 0.5% BSA/0.5% Tween-20 for SNA and RCA-I, 0.2% BSA/0.2% Tween-20 for MAL-I, and 5% BSA/.05% Tween-20 for PNA. PNA-HRP was incubated at 1:1000 for 1 hr at RT in blocking and biotin labeled SNA, RCA-I, and MAL-I were incubated in essential buffer at 1, 0.2, and 1 μg/ml, respectively for 1 hr at RT or overnight at 4°C. Blots were washed 15 minutes 3 times with TTBS (PNA) or wash buffer (essential buffer, 0.1% Tween-20, for SNA, RCA-I, MAL-I). ECL was immediately added to PNA-HRP and films were exposed. Biotin labeled lectins were then incubated for 1 hr with 1:5000 streptavidin-HRP at RT in blocking buffer, then washed 15 minutes, 3 times with wash buffer, 15 minutes once with essential buffer, and then incubated with ECL and exposed. Blots were stripped with 25mM glycine, 1% SDS, pH 2 at RT for 30 minutes and then rinsed with PBS for 10 minutes, 2 times before β-actin staining. Blot was reblocked with 5% milk/TTBS (0.05%) for 1 hr RT, washed 5 minutes with TTBS, and incubated with 1:3000 anti-β-actin (Santa Cruz) in block overnight. Blot was washed 5 minutes, 4 times with TTBS, and incubated with 1:3000 HRP labeled secondary at RT for 45 minutes. Blot was then washed 5 minutes, 4 times with TTBS, incubated with ECL, and exposed.

ACKNOWLEDGEMENT

The authors thank Drs. Jamie Heimburg-Molinaro for help in editing, David F. Smith and Junwei Zeng for critical reading of the manuscript. This work was supported by National Institutes of Health Grant U01CA168930 to TJ and RDC, P41GM103694 to RDC, Georgia Cancer Coalition (now Georgia Research Alliance, GRA) Award to TJ, and by Biotechnology and Biological Sciences Research Council grant BB/K016164/1 (AD and SMH for Core Support for Collaborative Research). AD is supported by a Wellcome Trust Senior Investigator Award.

CONTRIBUTIONS

M.R.K., R.D.C., and T.J. conceived of the project. M.R.K., Y.W., N.T.S., A.D., S.M.H., R.D.C., and T.J. designed experiments. M.R.K, A.A., Y.W., D.M.D., and X.S. performed experiments. M.R.K., A.A., Y.W., A.D., S.M.H., R.D.C, and T.J. analyzed the data. M.R.K., R.D.C, and T.J. wrote the manuscript. All authors edited the manuscript.

Figure Legends

Figure 1. Overview of Cellular O-glycome Reporter/Amplification (CORA). Cells are incubated with chemical O-glycan precursor or primer (Benzyl-α-D-GalNAc), which is peracetylated (Ac_3GalNAc-Bn) to facilitate passive transport across the plasma membrane. Cytosolic esterases generate Bn-GalNAc, which is taken up in the Golgi and modified by native glycosyltransferases during anterograde transport. Elaborated Bn-O-glycans are secreted into the media, purified, and analyzed by MS (reported here), HPLC, or printed on glycan microarray for interrogation by GBPs.

Figure 2. Bn-O-glycans produced by HEK293 and Molt-4 cells. 50μM Ac_3GalNAc-Bn or DMSO was administered to (**a**) HEK293 or (**b**) Molt-4 cells and incubated for 3 days. Bn-O-glycans were purified from media, and one-third of eluent was permethylated and analyzed by MALDI-MS (composition). Spectra for Ac_3GalNAc-Bn and DMSO are off-set, but scaled to same absolute intensity for each cell. Representative profiles are shown (n = 2).

Figure 3. The chaperone Cosmc and active T-synthase are required for production of Core 1- and 2-based Bn-O-glycans. (**a**) Loss of Cosmc results in loss of T-synthase activity in LS174T colorectal cells, LOX melanoma cells, and Jurkat T cell leukemia. Cells with and without Cosmc were subcloned from a mixed population (LS174T) or cells with defective Cosmc were rescued with Cosmc or empty vector (LOX, Jurkat). Mannosidase was assayed as control. T-synthase and Mannosidase activities were determined in triplicate (n = 2), and a representative experiment is shown, mean ± SD of triplicates. (**b – d**) Only cells with functional Cosmc secrete Bn-O-glycans when administered Ac_3GalNAc-Bn. Only major glycans are annotated (composition) for LS174T (**b**) for clarity; highly fucosylated minor species as shown in **Supplementary Fig. 17b** were also observed. Spectra for each graph are off-set, but scaled to same absolute intensity for each cell. Representative profiles are shown (n = 2).

Figure 4. Accuracy of CORA for profiling the O-glycome. 50μM Ac_3GalNAc-Bn was incubated with (**a**) HL-60 and (**b**) WEHI-3 cells for 3 days and Bn-O-glycans were purified, permethylated, and analyzed by MALDI-TOF-MS. Note: Here as in all studies we only analyzed the non-sulfated/non-phosphorylated glycans. Bn-O-glycans observed by CORA versus alkaline β-elimination (CFG) of (**c**) HL-60 and (**d**) WEHI-3 cells were compared. Compositions are depicted for β-elimination and structures are depicted for CORA. Blue boxes indicate glycan compositions observed in both CORA and β-elimination. O-glycan MS profiles by β-elimination were previously deposited at the Consortium for Functional Glycomics (CFG, http://www.functionalglycomics.org). CORA amplified the O-glycome and therefore generated sufficient quantities of Bn-O-glycans for high quality MALDI-TOF/TOF-MS/MS analysis. In contrast, only MALDI-TOF-MS and compositional analysis was previously performed for CFG data, presumably due to insufficient material for MS/MS. (e) The number of glycans (by composition) observed in CORA, β-elimination (CFG), or both are indicated for each cell line. All spectra of O-glycans were derived from the 75% acetonitrile fraction (see Methods). All molecular ions are $[M+Na]^+$. Putative structures are based on composition, tandem MS, and biosynthetic knowledge.

Figure 5. Sensitivity of CORA. (**a**) MKN45, (**b**) Colo205, (**c**) MDA-MB-231, and (**d**) MCF-7 cells were seeded at $5x10^5$, 10^5, and $2x10^4$ cells/well in 6 well, 12 well, and 48 well flasks. 50μM Ac_3GalNAc-Bn was added and Bn-O-glycans were purified and permethylated after 3 days. $1/150^{th}$ (6 well, 12 well) or $1/50^{th}$ (48 well) of total glycans were analyzed by MALDI-MS (composition). (**e**) O-glycomes were observed from all cells seeded at $5x10^5$ and 10^5 cells/well and 3 of 4 cells seeded at $2x10^4$ cells/well. Spectra are off-set for each seeding density and scaled relative to maximum intensity. Representative profiles are shown (n = 2).

Figure 6. MALDI-TOF-MS and MS/MS profiling of the O-glycome of primary cells. CORA was used to profile the O-glycome of (**a**) HUVECs and (**b**) primary human dermal fibroblasts. $12.5x10^4$ (**a, b**), $5x10^4$ (**b**), and $2.5x10^4$ (**b**) cells were seeded in T25, 6 well, or 12 well flasks, respectively. 50μM Ac_3GalNAc-Bn was added after 2 days, and Bn-O-glycans were purified, permethylated, and analyzed by MALDI after 3 more days. MALDI-TOF/TOF-MS/MS (structure) analysis was performed for HUVECs and MS (composition) was performed for fibroblasts. Spectra are off-set for each seeding density and scaled relative to maximum intensity.

Supplementary Fig. 1. MALDI profiles of compounds used in this study. (**a**) Bn-GalNAc, Ac_3GalNAc-Bn, and (**b**) Ac_3GlcNAc-Bn.

Supplementary Fig. 2. Bn-GalNAc peracetylation increases CORA sensitivity. 0 – 250μM of (**a**) Ac_3GalNAc-Bn or (**b**) Bn-GalNAc was incubated with MDA-MB-231 cells for 3 days before collecting, purifying, permethylating, and analyzing Bn-O-glycans by MALDI. Bn-O-glycans were observed at a lower initial concentration for peracetylated than unmodified Bn-GalNAc. Spectra are off-set for each concentration and scaled relative to maximum intensity. Representative profiles are shown (n = 2).

Supplementary Fig. 3. Optimizing concentration of Ac_3GalNAc-Bn and incubation time. MDA-MB-231 breast cancer cells were incubated with 0 – 250μM Ac_3GalNAc-Bn for (**a**) 2, (**b**) 3, or (**c**) 4 days. Profiles did not change across time. (**d**) Increasing concentration resulted in a shift in relative intensities of disialylated Core 2 (glycan #6) to monosialylated Core 2 (glycan #4). (**a – d**) Glycans are labeled 1 – 6 and indicated in (**e**). Spectra are off-set for each concentration and scaled relative to maximum intensity. Representative profiles and graphs are shown (n = 2).

Supplementary Fig. 4. CORA does not alter cell viability. (**a**) LS174T-Tn(-) and (**b**) MDA-MB-231 cells were seeded in 96 well plates in triplicate. 50μM Ac_3GalNAc-Bn or DMSO vehicle was added to cells for 3 days and XTT assay was performed. One-third and $1/9^{th}$ dilutions of cells were plated as a positive control for reduced cell numbers. No difference was observed between Ac_3GalNAc-Bn and DMSO for either cell line. Representative experiments are shown (n = 2), mean ± SD of triplicates.

Supplementary Fig. 5. CORA does not alter cell morphology or granularity. HEK293 cells were incubated for 3 days with (**a, e**) no treatment, (**b, f**) vehicle, (**c, g**) 50μM Ac3GalNAc-Bn, or (**d, h**) 10% DMSO as a positive control. (**a – d**) Bright field microscopy was used to assess changes in cell morphology and (**e – h**) side scatter was determined by FACS to assess cell granularity. No changes were observed upon treatment with 50μM Ac3GalNAc-Bn (n = 2).

Supplementary Fig. 6. CORA does not alter cell surface O-glycosylation. (**a**) Colo205, (**b**) MDA-MB-231, (**c**) MCF7, and (**d**) MKN45 cells were incubated with 50μM Ac_3GalNAc-Bn or DMSO for 3 days. Cells were then collected, lysed, and one fraction was treated with neuraminidase before blotting with PNA, which recognizes Core 1 O-glycan. (**a – d**) Actin was used as loading control and (**e**) fetuin was used as a control for neuraminidase treatment and lectin staining. Without neuraminidase treatment, Ac_3GalNAc-Bn did not reveal significant PNA binding, indicating that 50μM Ac_3GalNAc-Bn did not inhibit addition of sialic acid to Core 1 structure. With neuraminidase treatment, Ac_3GalNAc-Bn did not reduce PNA binding, indicating that 50μM Ac_3GalNAc-Bn did not block Core 1 O-glycan synthesis. Representative blots are shown (n = 2).

Supplementary Fig. 7. CORA does not alter cell surface N-glycosylation. (**a, f, k, p**) Colo205, (**b, g, l, q**) MDA-MB-231, (**c, h, m, r**) MCF7, and (**d, i, n, s**) MKN45 cells were incubated with 50μM Ac_3GalNAc-Bn or DMSO for 3 days. Cells were then collected, lysed, and one fraction was treated with neuraminidase, before running a denaturing, reducing gel. Lysates were then blotted for (**a – e**) RCA-I, which recognizes Galβ4GlcNAc of N-glycans, (**f – j**) SNA, which recognizes α6-linked sialic acids, (**k – o**) MAL-I, which recognizes α3-linked sialic acids, and (**p – s**) actin as loading control. (**e, j, o**) Fetuin was used as a control for neuraminidase treatment and lectin staining. Staining with RCA-I, SNA, and MAL-I with or without neuraminidase treatment did not differ between Ac_3GalNAc-Bn or DMSO incubated cells, indicating that 50μM Ac_3GalNAc-Bn did not disrupt addition of sialic acid to N-glycan termini or biosynthesis of N-glycans. Representative blots are shown (n = 2).

Supplementary Fig. 8. Evaluation of potential Bn-glycans produced from cells incubated with Ac_3GlcNAc-Bn. 50μM Ac_3GlcNAcBn or vehicle was incubated with (**a**) HUVEC, (**b**) primary human dermal fibroblasts, (**c**) MKN45, and (**d**) LS174T-Tn(-) and Tn(+) cells. Potential Bn-glycans were purified, permethylated, and analyzed by MALDI.

Supplementary Fig. 9. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from HL-60 cells (from Figure 4a): (**a**) *m/z* 2202, (**b**) *m/z* 2389, (**c**) *m/z* 2563. All molecular ions are $[M+Na]^+$. Horizontal blue arrows indicate the losses indicated from the molecular ion. Vertical blue lines indicate the corresponding peak ion. Letters “m” and “M” in bold characters suggest minor and major abundances respectively.

Supplementary Fig. 10. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from WEHI-3 cells (from Figure 4b): (**a**) *m/z* 1853, (**b**) *m/z* 2460, (**c**) *m/z* 2507, (**d**) *m/z* 2956, (**e**) *m/z* 3113, (**f**) *m/z* 3358. All molecular ions are $[M+Na]^+$. Horizontal blue arrows indicate the losses indicated from the molecular ion. Vertical blue lines indicate the corresponding peak ion. Letters “m” and “M” in bold characters suggest minor and major abundances respectively.

Supplementary Fig. 11. Reproducibility of CORA. (**a, b**) Ac_3GalNAc-Bn was added to HL-60 cells and Bn-O-glycans were purified, permethylated and analyzed. MALDI-MS profiles from two independent experiments are shown and nearly identical. MALDI-MS/MS experiments described in **Fig 4a**. were independently performed but used same material as those used for (**a**) here. Ac_3GalNAc-Bn and DMSO are off-set, but scaled to same absolute intensity for each experiment.

Supplementary Fig 12. Analysis of HUVEC, MKN45, and WEHI-3 O-glycans by ESI-MS. 50μM Ac_3GalNAcBn was incubated with (**a**) HUVEC, (**b**) MKN45, and (**c**) WEHI-3 cells for 3 days and Bn-O-glycans were purified, permethylated, and analyzed by LTQ Orbitrap Hybrid MS. Glycans indicated were not detected in vehicle only control (not shown).

Supplementary Fig. 13. O-glycome of murine pulmonary endothelial cells. Tn(-) and Tn(+) murine pulmonary endothelial cells (mPECs), with or without Cosmc, respectively, were obtained from mosaic EHC-Cosmc$^{+/-}$ mice (immortalized) or EHC-Cosmc$^{+/y \text{ or } -/y}$ mice (primary). Tn expression and CD31 (an endothelial marker) was analyzed by immunofluorescence for (**a**) immortalized and (**b**) primary cells. Ac_3GalNAc or DMSO was added to (**a, b**) immortalized or (**c, d**) primary cells and Bn-O-glycans were purified after 3 days, permethylated, and analyzed by MALDI. Similar O-glycan structures were observed in immortalized and primary cells; however, relative ratios of O-glycans differed. Spectra are off-set, but scaled to same absolute intensity for each cell type (primary, immortalized). Representative profiles are shown (n = 2).

Supplementary Fig. 14. Profiling the O-glycome of Primary Cells. Ac_3GalNAc-Bn or DMSO was added to (**a**) HUVECs or (**b**) primary human dermal fibroblasts seeded at $12.5x10^4$, $5x10^4$, or $2.5x10^4$ cells/well. Glycan structures and masses are omitted for clarity, but similar to those indicated in **Figure 6**. Spectra are off-set, but scaled relative to absolute intensity for cells seeded at a given density and relative to maximum intensity across cell densities for a given cell type. Representative profiles are shown (n = 2).

Supplementary Fig. 15. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from HUVEC cells (from Figure 6): (**a**) *m/z* 1578, (**b**) *m/z* 2027, (**c**) *m/z* 2303, (**d**) *m/z* 2389, (**e**) *m/z* 2651, (**f**) *m/z* 2838. All molecular ions are $[M+Na]^+$. Horizontal blue arrows indicate the losses indicated from the molecular ion. Vertical blue lines indicate the corresponding peak ion. Letters "m" and "M" in bold characters suggest minor and major abundances respectively.

Supplementary Fig 16. Stability of acetylation in complete media. (**a**) 50μM Bn-GalNAc, Ac_3GalNAc-Bn, or vehicle was added to DMEM with 5% FBS and incubated at 37°C, 5% CO_2 for 3 days. Media was directly analyzed by MALDI (n = 2).

Supplementary Fig. 17. Stability of O-glycans in media with cells. (**a**) LS174T-Tn(-) cells were seeded overnight. Ac_3GalNAc-Bn was added. After 3 days media was collected and stored for further analysis or mixed 1:1 with fresh media and incubated for 3 days with LS174T-Tn(+). (**b**) Bn-O-glycans were then analyzed to assess cell or media-associated glycosidase or non-enzymatic degradation. Profiles were nearly identical for 3 and 6 day incubations, indicating that Bn-O-glycans are stable and not degraded by cells or media. Spectra are off-set and scaled relative to maximum intensity. Representative profiles are shown (n = 2).

Supplementary Fig. 18. Glycan frequency across cell lines evaluated in CORA. Histograms were generated based on data from **Supplementary Table 2** indicating (**a, b**) number of cells lines expressing a given glycan (mass listed on x-axis). (**b**) Glycans were reordered from glycans expressed in greatest number of cells to glycans expressed in least number of cells. Eight unique glycans were each observed in >5 cell lines, 17 glycans in 2 – 5 cell lines, and 32

glycans in at most 1 cell line. (**c**) Cells were arranged in descending order from cells expressing most glycans to cells expressing fewest glycans. Then number of new glycans observed with evaluation of each additional cell line was determined. We have only considered non-sulfated and non-phosphorylated O-glycans in this study, in part because these modifications have not been previously reported in our 18 cell lines.

Supplementary Fig. 19. Computational model to estimate size of non-sulfated human cellular O-glycome. Data from CORA was input into a custom computational model generated in MATLAB that determines # of unique glycan masses (Y) observed given X number of randomly selected cells. (**a**) 256 iterations were plotted for each X (1 to 18) and Y was reported +/-SD. Nonlinear regression was performed to generate a best-fit line using Prism. Visually satisfactory models (Semi-log, log-log, quadratic) were compared by F test or AIC to determine best-fit. 95% prediction band was also plotted. SE of slope and Yintercept as well as R^2 are reported. This equation has no upper-limit suggesting that evaluation of more cells would on average identify more glycans. As described we use this equation to predict the size of the non-sulfated/phosphorylated human cellular O-glycome to be 376. This represents a lower limit as there are probably >200 cell types in the human body, a given cell type may express different glycans in different environments, and we have only considered the non-sulfated and non-phosphorylated O-glycans.

Supplementary Table 1. Cell lines and primary cells.

Supplementary Table 2. Glycan masses and compositions.

Supplementary Table 3. Glycan MS/MS structures.

References

1. Ohtsubo, K. & Marth, J.D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855-867 (2006).
2. Cummings, R.D. & Pierce, J.M. The challenge and promise of glycomics. *Chemistry & biology* **21**, 1-15 (2014).
3. Shriver, Z., Raguram, S. & Sasisekharan, R. Glycomics: a pathway to a class of new and improved therapeutics. *Nature reviews. Drug discovery* **3**, 863-873 (2004).
4. Tian, E. & Ten Hagen, K.G. Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconj J* **26**, 325-334 (2009).
5. Ju, T. et al. Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *Proteomics. Clinical applications* (2013).
6. Ju, T., Aryal, R.P., Kudelka, M.R., Wang, Y. & Cummings, R.D. The Cosmc connection to the Tn antigen in cancer. *Cancer biomarkers : section A of Disease markers* **14**, 63-81 (2014).
7. Kudelka, M.R., Ju, T., Heimburg-Molinaro, J. & Cummings, R.D. Simple sugars to complex disease--mucin-type O-glycans in cancer. *Advances in cancer research* **126**, 53-135 (2015).
8. Furukawa, J., Fujitani, N. & Shinohara, Y. Recent advances in cellular glycomic analyses. *Biomolecules* **3**, 198-225 (2013).

9. Ju, T. et al. A novel fluorescent assay for T-synthase activity. *Glycobiology* **21**, 352-362 (2011).
10. Fuster, M.M., Brown, J.R., Wang, L. & Esko, J.D. A disaccharide precursor of sialyl Lewis X inhibits metastatic potential of tumor cells. *Cancer Res* **63**, 2775-2781 (2003).
11. Brockhausen, I. et al. Control of O-glycan synthesis: specificity and inhibition of O-glycan core 1 UDP-galactose:N-acetylgalactosamine-alpha-R beta 3-galactosyltransferase from rat liver. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **70**, 99-108 (1992).
12. Vavasseur, F., Yang, J.M., Dole, K., Paulsen, H. & Brockhausen, I. Synthesis of O-glycan core 3: characterization of UDP-GlcNAc: GalNAc-R beta 3-N-acetyl-glucosaminyltransferase activity from colonic mucosal tissues and lack of the activity in human cancer cell lines. *Glycobiology* **5**, 351-357 (1995).
13. Zheng, Z., Cummings, R.D., Pummill, P.E. & Kincade, P.W. Growth as a solid tumor or reduced glucose concentrations in culture reversibly induce CD44-mediated hyaluronan recognition by Chinese hamster ovary cells. *J Clin Invest* **100**, 1217-1229 (1997).
14. Kuan, S.F., Byrd, J.C., Basbaum, C. & Kim, Y.S. Inhibition of mucin glycosylation by aryl-N-acetyl-alpha-galactosaminides in human colon cancer cells. *J Biol Chem* **264**, 19271-19277 (1989).
15. Wilson, J.R. & Zimmerman, E.F. Yolk sac: site of developmental microheterogeneity of mouse alpha-fetoprotein. *Developmental biology* **54**, 187-199 (1976).
16. Ju, T. & Cummings, R.D. A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Natl Acad Sci U S A* **99**, 16613-16618 (2002).
17. Ju, T. et al. Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res* **68**, 1636-1646 (2008).
18. Mi, R. et al. Epigenetic silencing of the chaperone Cosmc in human leukocytes expressing tn antigen. *J Biol Chem* **287**, 41523-41533 (2012).
19. Wilkins, P.P., McEver, R.P. & Cummings, R.D. Structures of the O-glycans on P-selectin glycoprotein ligand-1 from HL-60 cells. *J Biol Chem* **271**, 18732-18742 (1996).
20. Kawar, Z.S., Johnson, T.K., Natunen, S., Lowe, J.B. & Cummings, R.D. PSGL-1 from the murine leukocytic cell line WEHI-3 is enriched for core 2-based O-glycans with sialyl Lewis x antigen. *Glycobiology* **18**, 441-446 (2008).
21. Morelle, W. & Michalski, J.C. Analysis of protein glycosylation by mass spectrometry. *Nature protocols* **2**, 1585-1602 (2007).
22. Hsu, P.P. & Sabatini, D.M. Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703-707 (2008).
23. Wang, Y. et al. Platelet biogenesis and functions require correct protein O-glycosylation. *Proc Natl Acad Sci U S A* **109**, 16143-16148 (2012).
24. Dube, D.H. & Bertozzi, C.R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nature reviews. Drug discovery* **4**, 477-488 (2005).
25. Babu, P. et al. Structural characterisation of neutrophil glycans by ultra sensitive mass spectrometric glycomics methodology. *Glycoconj J* **26**, 975-986 (2009).
26. Macher, B.A., Buehler, J., Scudder, P., Knapp, W. & Feizi, T. A novel carbohydrate, differentiation antigen on fucogangliosides of human myeloid cells recognized by monoclonal antibody VIM-2. *J Biol Chem* **263**, 10186-10191 (1988).
27. Sato, C. et al. Frequent occurrence of pre-existing alpha 2-->8-linked disialic and oligosialic acids with chain lengths up to 7 Sia residues in mammalian brain glycoproteins. Prevalence revealed by highly sensitive chemical methods and anti-di-, oligo-, and poly-Sia antibodies specific for defined chain lengths. *J Biol Chem* **275**, 15422-15431 (2000).
28. Karlsson, N.G. & Thomsson, K.A. Salivary MUC7 is a major carrier of blood group I type O-linked oligosaccharides serving as the scaffold for sialyl Lewis x. *Glycobiology* **19**, 288-300 (2009).

29. Delannoy, P. et al. Benzyl-N-acetyl-alpha-D-galactosaminide inhibits the sialylation and the secretion of mucins by a mucin secreting HT-29 cell subpopulation. *Glycoconj J* **13**, 717-726 (1996).
30. Zanetta, J.P. et al. Massive in vitro synthesis of tagged oligosaccharides in 1-benzyl-2-acetamido-2-deoxy-alpha-D-galactopyranoside treated HT-29 cells. *Glycobiology* **10**, 565-575 (2000).
31. Liu, J., Jin, C., Cherian, R.M., Karlsson, N.G. & Holgersson, J. O-glycan repertoires on a mucin-type reporter protein expressed in CHO cell pools transiently transfected with O-glycan core enzyme cDNAs. *Journal of biotechnology* **199**, 77-89 (2015).
32. Cummings, R.D. The repertoire of glycan determinants in the human glycome. *Molecular bioSystems* **5**, 1087-1104 (2009).
33. Carlson, D.M. Oligosaccharides isolated from pig submaxillary mucin. *J Biol Chem* **241**, 2984-2986 (1966).
34. Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262-265 (2009).
35. Gao, D. et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176-187 (2014).
36. Park, I.H. et al. Disease-specific induced pluripotent stem cells. *Cell* **134**, 877-886 (2008).
37. Dimos, J.T. et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218-1221 (2008).
38. Jang-Lee, J. et al. Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies. *Methods in enzymology* **415**, 59-86 (2006).
39. Dong, Q.G. et al. A general strategy for isolation of endothelial cells from murine tissues. Characterization of two endothelial cell lines from the murine lung and subcutaneous sponge implants. *Arteriosclerosis, thrombosis, and vascular biology* **17**, 1599-1604 (1997).
40. Hartwell, D.W. et al. Role of P-selectin cytoplasmic domain in granular targeting in vivo and in early inflammatory responses. *The Journal of cell biology* **143**, 1129-1141 (1998).

Kudelka, *et al*, Table S1

Cell #	Unique Cell #	Cell name	Species	Cell type	Cancer/ Immortalized (I) or Primary (P)	Analysis	Number of unique glycan masses (MALDI only)	Number of unique glycan structures (MALDI-MS/MS)	Figure (Main unless S for Supplementary, ESI in italics)
1	1	HEK-293	Human	Embryonic kidney	I	MALDI-MS	12	N/A	2A
2	2	Molt-4	Human	T cell luekemia	I	MALDI-MS	7	N/A	2B
3	3	LS174T-Tn(-)	Human	Colon cancer	I	MALDI-MS	12	N/A	3B, S17B
4	3	LS174T-Tn(+)	Human	Colon cancer	I	MALDI-MS	0	N/A	3B
5	4	Lox-empty	Human	Melanoma	I	MALDI-MS	0	N/A	3C
6	4	Lox-Cosmc	Human	Melanoma	I	MALDI-MS	1	N/A	3C
7	5	Jurkat-empty	Human	T cell leukemia	I	MALDI-MS	0	N/A	3D
8	5	Jurkat-Cosmc	Human	T cell leukemia	I	MALDI-MS	2	N/A	3D
9	6	HL-60	Human	Promyelocytic leukemia	I	MALDI-MS/MS	10	11	4A, S5, S7
10	7	WEHI-3	Murine	Myelomonocytic leukemia	I	MALDI-MS/MS, ESI-MS	28	40	4B, S6, *S8*
11	8	MKN45	Human	Gastric cancer	I	MALDI-MS, ESI-MS	11	N/A	5A, *S8*
12	9	COLO205	Human	Colon cancer	I	MALDI-MS	14	N/A	5B
13	10	MDAMB231	Human	Breast cancer	I	MALDI-MS	6	N/A	5C, S2, S3
14	11	MCF7	Human	Breast cancer	I	MALDI-MS	5	N/A	5D
15	12	HUVEC	Human	Endothelial cell	P	MALDI-MS/MS, ESI-MS	29	43	6A, S11, S10A, *S8*
16	13	Human dermal fibroblasts	Human	Fibroblast	P	MALDI-MS	18	N/A	6B, S10B
17	14	Immortalized mPEC	Murine	Endothelial cell	I	MALDI-MS	5	N/A	S9B
18	15	Primary mPEC	Murine	Endothelial cell	P	MALDI-MS	6	N/A	S9D
Number of unique compositions or structures across all samples							57	76	

Kudelka, *et al*, Table. S2

MALDI		Composition (+Bn)				Observed Mass (Da) by Cell																	Glycan Co-expression by frequency	
Glycan #	Theoretical mass [M+Na] (Da)	Hex	HexN Ac	Fuc	Neu5 Ac	HEK-293	Molt-4	LS174T-Tn(-)	LS174T-Tn(+)	Lox-empty	Lox-Cosmc	Jurkat-empty	Jurkat-Cosmc	HL-60	WEHI-3	MKN45	COLO-205	MDAMB-231	MCF7	HUVEC	Human dermal fibro-blasts	Immor-talized mPEC	Primary mPEC	# cell lines expressing glycan	Theoretical mass [M+Na] (Da)	# cell lines expressing glycan
1	594.4	1	1	0	0																593.9			1	1316.8	15
2	768.5	1	1	1	0															768				1	955.6	12
3	955.6	1	1	0	1	955.8	955.6	955.6					955.4	955	955	955.8	955.9	955.8	955.9	955	955.5			12	1766.1	11
4	1043.7	2	2	0	0	1043.9									1043					1043	1043.5			4	1404.9	11
5	1129.7	1	1	1	1															1129				1	1940.2	10
6	1200.8	1	2	0	1										1200						1200.5			2	1579.0	10
7	1217.8	2	2	1	0	1218.0		1217.8						1217			1218.2			1217	1217.6			6	2389.4	7
8	1247.8	3	2	0	0										1247									1	1217.8	6
9	1316.8	1	1	0	2	1317.1	1316.7	1316.8			1316.7		1316.5	1316	1316	1317.1	1317.3	1317.1	1317.3	1316	1316.7	1316.7	1317.1	15	2028.2	5
10	1391.9	2	2	2	0			1391.9												1391	1390.4			3	1854.1	5
11	1404.9	2	2	0	1	1405.1	1404.8	1404.9						1404	1404	1405.2	1405.3	1405.1	1405.3	1404	1404.9			11	1667.0	5
12	1445.9	1	3	0	1										1445									1	2215.3	4
13	1492.9	3	3	0	0															1492	1492.8			2	1841.1	4
14	1562.0	1	2	0	2										1561									1	1043.7	4
15	1566.0	2	2	3	0			1566.1																1	2058.2	3
16	1579.0	2	2	1	1	1579.3	1578.8	1579						1578		1579.3	1579.4	1579.3	1580.3	1578	1579.3			10	1609.0	3
17	1609.0	3	2	0	1										1608							1609.5	1609.3	3	1391.9	3
18	1650.0	2	3	0	1										1649									1	3113.8	2
19	1667.0	3	3	1	0			1667								1667.4	1667.5			1666	1666.8			5	3025.8	2
20	1678.0	1	1	0	3																		1678.4	1	2563.5	2
21	1753.1	2	2	2	1			1753.2																1	2303.4	2
22	1766.1	2	2	0	2	1766.4	1765.9							1765	1765	1767.4		1766.4	1767.5	1765	1766	1766.6	1766.3	11	2202.3	2
23	1783.1	3	2	1	1										1782									1	1901.2	2
24	1841.1	3	3	2	0			1841.2									1842.6			1841	1840.1			4	1492.9	2
25	1854.1	3	3	0	1	1854.4									1853	1855.5				1853	1854.3			5	1200.8	2
26	1901.2	5	3	0	0																	1901.9	1902.3	2	3475.0	1
27	1940.2	2	2	1	2	1941.5	1939.9	1940.7						1939	1939	1941.6	1941.6	1940.6		1939	1940.5			10	3359.0	1
28	2011.2	2	3	0	2										2010									1	3287.9	1
29	2015.2	3	3	3	0			2015.3																1	2956.7	1
30	2028.2	3	3	1	1	2029.6										2028.6	2029.7			2027	2028.4			5	2926.7	1
31	2058.2	4	3	0	1										2058							2059.4	2059.4	3	2868.7	1
32	2202.3	3	3	2	1									2202			2203.8							2	2838.7	1
33	2215.3	3	3	0	2	2216.7										2216.8				2215	2215.4			4	2752.6	1
34	2256.3	2	4	0	2										2256									1	2711.6	1
35	2303.4	4	4	0	1										2303					2303				2	2705.6	1
36	2290.4	4	4	2	0												2291.9							1	2664.6	1
37	2389.4	3	3	1	2	2390.7	2390.1							2389		2390.9	2390.9			2389	2389.3			7	2651.6	1
38	2460.5	3	4	0	2										2460									1	2617.5	1
39	2464.5	4	4	3	0												2466.0							1	2507.5	1
40	2477.5	4	4	1	1															2477				1	2477.5	1
41	2507.5	5	4	0	1										2507									1	2464.5	1
42	2563.5	3	3	2	2									2563			2564.9							2	2460.5	1
43	2617.5	2	4	0	3										2617									1	2290.4	1
44	2651.6	4	4	2	1															2651				1	2256.3	1
45	2664.6	4	4	0	2															2664				1	2015.2	1
46	2705.6	3	4	0	2										2705									1	2011.2	1
47	2711.6	6	4	0	1										2711									1	1783.1	1
48	2752.6	5	5	0	1															2752				1	1753.1	1
49	2838.7	4	4	1	2															2838				1	1678.0	1
50	2868.7	5	4	0	2										2868									1	1650.0	1
51	2926.7	5	5	1	1															2926				1	1566.0	1
52	2956.7	6	5	0	1										2956									1	1562.0	1
53	3025.8	4	4	0	3										3025					3025				2	1445.9	1
54	3113.8	5	5	0	2										3113					3113				2	1247.8	1
55	3287.9	5	5	1	2															3287				1	1129.7	1
56	3359.0	5	6	0	2										3358									1	768.5	1
57	3475.0	5	5	0	3															3474				1	594.4	1

structure #	mass #	structure	mass	HL-60	WEHI-3	HUVEC
1	1		768			X
2	2		955	X	X	X
3	2		955		X	X
4	3		1043		X	X
5	4		1129			X
6	5		1200		X	
7	6		1217	X		X
8	6		1217			X
9	7		1247		X	
10	8		1316	X	X	X
11	9		1391			X
12	10		1404	X		
13	10		1404		X	X
14	10		1404		X	X
15	11		1445		X	
16	12		1492			X
17	13		1561		X	
18	14		1578	X		X
19	14		1578			X
20	15		1608		X	
21	16		1649		X	
22	17		1666			X
23	17		1666			X
24	18		1765	X	X	X
25	19		1782		X	
26	20		1841			X
27	21		1853		X	
28	21		1853		X	X
29	21		1853		X	X
30	22		1939	X	X	X
31	23		2010		X	
32	23		2010		X	
33	24		2027			X
34	24		2027			X
35	25		2058		X	
36	26		2202	X		
37	26		2202	X		
38	27		2215			X

structure #	mass #	structure	mass	HL-60	WEHI-3	HUVEC
39	28		2256		X	
40	29		2303		X	
41	29		2303			X
42	29		2303			X
43	29		2303			X
44	30		2389	X		X
45	31		2460		X	
46	31		2460		X	
47	32		2477			X
48	32		2477			X
49	32		2477			X
50	33		2507		X	
51	33		2507		X	
52	33		2507		X	
53	34		2563	X		
54	35		2617		X	
55	35		2617		X	
56	36		2651			X
57	36		2651			X
58	37		2664			X
59	37		2664			X
60	38		2705		X	
61	38		2705		X	
62	39		2752			X
63	40		2838			X
64	41		2868		X	
65	42		2926			X
66	42		2926			X
67	43		2956		X	
68	43		2956		X	
69	43		2956		X	
70	44		3025		X	X
71	45		3113		X	
72	45		3113		X	
73	45		3113			X
74	46		3287			X
75	47		3358		X	
76	48		3474			X