

Probing the O-glycoproteome of Gastric Cancer Cell Lines for Biomarker Discovery

Diana Campos^{1,2}, Daniela Freitas², Joana Gomes², Ana Magalhães², Catharina Steentoft¹, Catarina Gomes², Malene B. Vester-Christensen^{1*}, José Alexandre Ferreira^{3,4}, Lúcio L. Santos³, João Pinto de Sousa⁵, Ulla Mandel¹, Henrik Clausen¹, Sergey Y. Vakhrushev^{1#}, Celso A. Reis^{2,5,6#}

¹ Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and School of Dentistry, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

² IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

³ Experimental Pathology and Therapeutics Group, Portuguese Institute of Oncology, Rua Dr. António Bernardino de Almeida 4200-072 Porto, Portugal

⁴ QOPNA, Department of Chemistry of the University of Aveiro, Campus Universitário de Santiago 3810-193 Aveiro, Portugal

⁵ Faculty of Medicine of the University of Porto, Al. Prof. Hernâni Monteiro, 4200-319 Porto, Portugal

⁶ Institute of Biomedical Sciences Abel Salazar, ICBAS, Rua de Jorge Viterbo Ferreira n.228, 4050-313 Porto, Portugal

* Present address: Department of Mammalian Cell Technology, Novo Nordisk A/S, DK-2760 Måløv, Denmark

#To whom correspondence should be addressed: Sergey Y. Vakhrushev, Copenhagen Center for Glycomics, Department of Cellular and Molecular Medicine, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark. Tel.: 45-35327797; Fax: 45-35367980; E-mail: seva@sund.ku.dk

#Celso Reis, IPATIMUP, Institute of Molecular Pathology and Immunology of the University of Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal. Tel.: 351-225570700; Fax: 351-225570799; E-mail: celsor@ipatimup.pt

Originally published in *Molecular & Cellular Proteomics*. 2015 Jun;14(6):1616-29. doi: 10.1074/mcp.M114.046862. Epub 2015 Mar 26.

Running Title: O-Glycoproteome of Gastric Cancer Cell Lines

Key words: Glycobiomarkers, Cancer Serum Biomarker, Gastric Cancer

ABSTRACT

Circulating O-glycoproteins shed from cancer cells represent important serum biomarkers for diagnostic and prognostic purposes. We have recently shown that selective detection of cancer-associated aberrant glycoforms of circulating O-glycoprotein biomarkers can increase specificity of cancer biomarker assays. However, the current knowledge of secreted and circulating O-glycoproteins is limited. Here, we used the “SimpleCell” (SC) strategy to characterize the O-glycoproteome of two gastric cancer SC lines (AGS, MKN45) and as well as a wild type cell line (KATO III) naturally expressing truncated O-glycosylation. Overall we identified 499 O-glycoproteins and 1,236 O-glycosites in gastric cancer SCs, and a total 47 O-glycoproteins and 73 O-glycosites in the KATO III cell line. We next modified the glycoproteomic strategy to apply it to pools of sera from gastric cancer and healthy individuals to identify circulating O-glycoproteins with the STn glycoform. We identified 37 O-glycoproteins in the pool of cancer sera, and only 9 of these were also found in sera from healthy individuals. Two identified candidate O-glycoprotein biomarkers (CD44 and GalNAc-T5) circulating with the STn glycoform were further validated as being expressed in gastric cancer tissue. A proximity ligation assay was used to demonstrate that CD44 was expressed with the STn glycoform in gastric cancer tissues. The study provides a discovery strategy for aberrantly glycosylated O-glycoproteins and a set of O-glycoprotein candidates with biomarker potential in gastric cancer.

INTRODUCTION

Most broad proteomic studies for discovery of cancer biomarkers in serum have been designed to interrogate the proteome and not taking into account that cancer cells often produce aberrant glycoforms (1). Many cancer biomarkers currently used in the clinic are based on circulating O-glycoproteins that are detected in established serological assays (CA125, CA15-3, CEA, CA19.9) (2). In addition to be overexpressed in cancer, these proteins also carry aberrant glycans, which open for the opportunity to selectively detect aberrant glycoforms. An inherent problem with most cancer biomarker assays is that they often have poor specificity because the detected glycoprotein is found in elevated levels in non-malignant conditions (2, 3). We recently found that the specificity of the widely used CA125 biomarker assay can be increased by selectively detecting aberrant O-glycoforms of the MUC16 mucin probed in the CA125 assay (4). Thus, the truncated O-glycan STn (NeuAc α 2-6GalNAc α 1-O-Ser/Thr) (Fig. 1) was particularly suited for discrimination of MUC16 circulating in cancer patients in contrast to MUC16 circulating in benign conditions (4).

One of the most characteristic phenotypes of cancer cells is the expression of truncated O-glycans, and the structures T (Gal β 1-3GalNAc α 1-O-Ser/Thr), STn (NeuAc α 2-6GalNAc α 1-O-Ser/Thr), and Tn (GalNAc α 1-O-Ser/Thr) (Fig. 1) are considered pancarcinoma antigens (2, 5). These truncated O-glycans are essentially not produced in normal and benign cells, which suggests that circulating O-glycoproteins in normal and benign conditions should have more mature O-glycans, whereas O-glycoproteins shed from cancer cells are expected to display truncated glycan structures. Cancer cells produce, secrete and shed many different O-glycoproteins with truncated O-glycans, and provided these glycoproteins reach the circulation they may be detectable in serum. However, it is also known that non-sialylated glycoproteins are cleared from circulation through innate immune lectin receptors (6). In fact, we were previously unable to detect circulating T and Tn glycoforms of

MUC1 and MUC16, while the sialylated ST (NeuAc α 2-3Gal β 1-3[NeuAc α 2-6]+/-GalNAc α 1-O-Ser/Thr) and STn glycoforms were readily detectable (4, 7). Furthermore, two classical serological biomarker assays, CA19-9 (8) and CA72.4 (9-11), are based on the detection of sialylated O-glycans, and especially the latter that detects STn demonstrates that proteins expressing the STn glycoform circulate in serum of cancer patients. Interestingly, although CA72.4 has been used for decades, it is still largely unknown which O-glycoproteins carry STn and are detected by the CA72.4 assay (9, 10).

The truncated STn O-glycan has attracted much attention because it is highly expressed in most gastric (12), colorectal (13), ovarian (14), breast (15), pancreatic (16) and bladder (17) carcinomas, whereas expression of STn on normal tissues is highly restricted (11, 18). In addition, STn expression is associated with carcinoma aggressiveness and poor prognosis (15, 19). We have recently described the presence of a few STn bearing glycoproteins in serum from individuals with gastric cancer and gastric cancer precursor lesions (20). The biosynthetic and genetic mechanisms underlying the expression of this truncated O-glycan in cancer have remained poorly understood, and a number of mechanisms have been proposed which may not be mutually exclusive. One mechanism is the altered expression of the sialyltransferase ST6GalNAc-I, which is believed to be the main STn synthase (21, 22) (Fig. 1), and in fact overexpression of this enzyme in cell lines appears to override the normal O-glycan elongation machinery and result in expression of STn (22, 23). Another mechanism may be reduced core1 elongation that leads to accumulation of Tn, which serves as substrate for ST6GalNAc-I (22). The core1 synthase C1GALT1 is dependent on a private chaperone Cosmc, and several studies have reported that somatic mutations in COSMC gene (24), or hypermethylation of COSMC gene in cancer (25) lead to increased expression of Tn and STn. We have further shown that knockout (KO) of COSMC in a number of human cancer cell lines produce cells that express different levels of Tn and STn truncated O-glycans ranging from exclusive Tn to exclusive STn (26). A third potential mechanism offered recently may be related to cancer-associated relocation of the polypeptide GalNAc-transferases (GalNAc-Ts) that initiate O-glycosylation (Fig. 1) from Golgi to ER, which appear to induce expression of the Tn truncated O-glycans, although expression of STn has not been explored yet (27).

In the present study, we applied a glycoproteomics strategy to explore potential biomarker O-glycoproteins with the STn glycoform in gastric cancer. We first characterized the O-glycoproteome and including the secretome of two gastric cancer cell lines (AGS and MKN45) using our SimpleCell (SC) discovery platform where we identified a total of 499 O-glycoproteins (1,236 O-glycosites). This strategy involves genetic engineering of cell lines to produce homogenous truncated O-glycans (Tn and/or STn) by KO of COSMC, followed by Vicia Villosa lectin (VVA) enrichment of Tn glycoproteins and/or glycopeptides for sensitive identification of O-glycoproteins and O-glycosites by mass spectrometry (26, 28) (Fig. 1). We applied the same glycoproteomics workflow to a wild type (wt) gastric cancer cell line, KATO III, which naturally expresses Tn and STn O-glycans in a mixture with more complex structures, and identified a significantly smaller O-glycoproteome (total of 47 O-glycoproteins) compared with SimpleCells (total of 499 O-glycoproteins). We next modified the strategy to enrich for STn O-glycoproteins in pools of serum from cancer patients and normal controls using pretreatment with neuraminidase to remove sialic acid and expose Tn for VVA capture. This approach enabled us to isolate and identify 37 O-glycoproteins (49 O-glycosites) in gastric cancer serum. Finally, we confirmed that two of the identified serum O-glycoproteins (CD44 and GalNAc-T5) were expressed in gastric cancer tumours by immunohistology, and further used proximity ligation assay (PLA) to demonstrate that STn glycoforms of CD44 was expressed in cancer tissue. This study clearly demonstrates that cancer

patients have a variety of circulating O-glycoproteins with the STn glycoform, and supports the hypothesis that these glycoproteins originate from the cancer tissue. The identified secreted and circulating aberrant O-glycoproteins serve as a discovery set for biomarkers of gastric cancer.

EXPERIMENTAL METHODS

Generation and Characterization of AGS and MKN45 SimpleCells - We targeted the *COSMC* gene in two human gastric cell lines, AGS and MKN45, using zinc finger nuclease (ZFN) precise gene editing as previously described (26, 29, 30). Briefly, cells were transfected with 4 µg of compoZr® C1GalT1C1 DNA using an Amaxa™ Nucleofector™ according to the manufacturer's cell lines specific protocols (Lonza). Cells were screened on acetone fixed slides by immunocytochemistry using monoclonal antibodies (MAbs) to the Tn (5F4) and STn (3F1) O-glycans (See Supplemental Table S1 for MAbs used in study). The cells were then cloned by limited dilution and the final clones were further characterized by staining with monoclonal antibodies to either the C1GalT1 enzyme (5B6) or the T antigen (3C9) O-glycan structure with and without pretreatment with neuraminidase (Sigma). Slides were examined using a Zeiss Optical Microscope. Finally, we analyzed selected clones by PCR followed by sequencing to define the induced mutations in *COSMC*.

Serum Samples - Serum samples from gastric carcinoma patients were collected from The Portuguese Institute of Oncology (IPO-Porto, Portugal), and control samples from individuals undergoing hernia surgery from the Hospital de São João in Porto, Portugal. All samples were collected with informed consent and use of samples was approved by the local Ethical committees (CHSJ and IPO). A total of 29 individual samples (approximately 0.6g ml each, from a sample taken at time of surgery) from both intestinal and unclassified subtype, according to Lauren's classification (31), with different disease stages, were used for the gastric carcinoma serum pool. Individuals were selected for blood group B and O to avoid potential cross-reactivity of VVA lectin with blood group A, although glycan array data from the Consortium for Functional Glycomics show no such cross-reactivity (<http://www.functionalglycomics.org>). Although some O-glycoproteins may potentially be lost in coagulation, most comparative proteomic studies of serum and plasma have found similar glycoprotein composition (24, 32). The age of carcinoma patients varied from 49 to 81 with an average value of 65 years old, and included 17 males and 12 females. Age of control individuals varied from 21 to 83 years old with an average of 69 and included 20 males and 5 females.

Sample Preparation and Lectin Affinity Chromatography - Total cell lysates (TCL) and culture medium (SEC) were processed as previously described (28, 33). In brief, spent culture medium (~80 mL harvested from two 175 ml T-flasks seeded at 5×10^5 cells and cultured for three days) was cleared, dialyzed, and subjected to neuraminidase treatment (10 U *Clostridium perfringens* neuraminidase Type VI (Sigma)) before loaded on a 0.3 ml *Vicia villosa* agglutinin (VVA) agarose (Vector laboratories) column. The column was washed with 10 column volumes (CV) of 0.4 M Glucose in LAC A buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 M Urea, 1 mM CaCl₂, MgCl₂, MnCl₂, and ZnCl₂) followed by 1 ml 50 mM AmBic. The glycoproteins were then eluted by 4x 500 µl 0.05% RapiGest with heating to 90°C for 10 min as described previously (28). Fractions were pooled and directly reduced with 5 mM DTT for 40 min at 60 °C, alkylated with 10 mM iodoacetamide in dark for 45 min, and digested with trypsin (25 µg).

Cell pellets were obtained by removing media, washing with phosphate buffer solution (PBS) and, followed by scraping of the cells in PBS, lysed in RapiGest and sonicated. The cleared lysate was reduced, alkylated with iodoacetamide, and then digested overnight with trypsin.

For AGS SC we also used chymotrypsin and Glu-C in addition to trypsin. Thus, for this cell line three data sets for TCL and three for SEC were generated.

Serum pools (20 mL) were diluted in PBS to 100 mL and neuraminidase treated (10 U), and loaded twice on a 0.3 mL VVA agarose column. The glycoproteins were eluted and reduced and alkylated as described above, and digested with chymotrypsin (25 µg) overnight at 37 °C.

All digests were finally subjected to lectin weak affinity chromatography (LWAC) on VVA agarose as previously described (26, 28). Briefly, the digestion was quenched with 1% trifluoroacetic acid (TFA), and the resulting peptides were desalted with C18 stage tips and dried in a Speed Vac. The digest was then neuraminidase treated (New England Biolabs), diluted in 2 mL LAC A buffer and injected onto a pre-equilibrated 2.6 m long VVA agarose (Vector Laboratories) column, similar to the system described previously (28, 34, 35). The flow was set to 100 µL min⁻¹ and 500 µL fractions were collected. The column was washed with 0.4 M glucose in LAC A buffer and eluted with 2 x 2 mL 0.2 M GalNAc and 1x 2 mL 0.4 M GalNAc.

Analysis of KATO III further included isolation of T-glycopeptides. The flowthrough fractions (2 mL) of the VVA LWAC step for TCL were concentrated on C18 and dried as before and further re-enriched for T glycoforms by PNA LWAC as previously described (26).

MS approach and Data Analysis - EASY-nLC 1000 UHPLC (Thermo Scientific) interfaced via nanoSpray Flex ion source to an LTQ-Orbitrap Velos Pro spectrometer (Thermo Scientific) was used for analysis. The nLC was operated in a single analytical column set up using PicoFrit Emitters (New Objectives, 75 µm inner diameter) packed in-house with Reprosil-Pure-AQ C18 phase (Dr. Maisch, 1.9-µm particle size, 19-21 cm column length) and gradient elution times of 120 min were employed.

A precursor MS₁ scan (*m/z* 350–1,700) of intact peptides was acquired in the Orbitrap at a nominal resolution setting of 30,000, followed by Orbitrap HCD-MS₂ and ETD-MS₂ (*m/z* of 100–2,000) of the five most abundant multiply charged precursors in the MS₁ spectrum; a minimum MS₁ signal threshold of 50,000 was used for triggering data-dependent fragmentation events; MS₂ spectra were acquired at a resolution of 7,500 for HCD MS₂ and 15,000 for ETD MS₂. Activation times were 30 and 200 ms for HCD and ETD fragmentation, respectively; isolation width was 4 mass units, and usually 1 microscan was collected for each spectrum. Automatic gain control targets were 1,000,000 ions for Orbitrap MS₁ and 100,000 for MS₂ scans, and the automatic gain control for fluoranthene ion used for ETD was 300,000. Supplemental activation (20 %) of the charge-reduced species was used in the ETD analysis to improve fragmentation. Dynamic exclusion for 60 s was used to prevent repeated analysis of the same components. Polysiloxane ions at *m/z* 445.12003 were used as a lock mass in all runs.

Data analysis - Data processing was performed using Proteome Discoverer 1.4 software (Thermo Scientific) as previously described with small changes (28). Due to the high speed of data processing Sequest HT node was used instead of Sequest. All spectra were initially searched with the full

cleavage specificity, filtered according to the confidence level (medium, low and unassigned) and further searched with the semi-specific enzymatic cleavage. In all cases the precursor mass tolerance was set to 6 ppm and fragment ion mass tolerance to 50 mmu. Carbamidomethylation on cysteine residues was used as a fixed modification. Methionine oxidation and HexNAc attachment to serine, threonine and tyrosine were used as variable modifications for ETD MS₂. All HCD MS₂ were pre-processed as described (28) and searched under the same conditions mentioned above using only methionine oxidation as variable modification. All spectra were searched against a concatenated forward/reverse human-specific database (UniProt, January 2013, containing 20,232 canonical entries) with the maximum of two missed cleavage sites. In addition, another 251 common contaminants were included in the search using a target false discovery rate (FDR) of 1 %. FDR was calculated using target decoy PSM validator node, a part of the Proteome Discoverer workflow. The resulting list was filtered to include only peptides with glycosylation as a modification. This resulted in a final glycoprotein list identified by at least one unique glycopeptide. ETD MS₂ data were used for unambiguous site assignment. HCD MS₂ data were used for unambiguous site assignment only if the number of GalNAc residues was equal to the number of potential sites on the peptide.

Immunocytochemistry - Cells were set on multiwall glass slides and fixed in ice cold acetone for 10 minutes before immunofluorescence as previously described (36). Cells were then incubated with MAbs (undiluted hybridoma supernatants) (Supplemental Table S1) overnight at 4°C. To assess the expression of ST antigen, cells were pretreated with neuraminidase diluted in 0.1M sodium acetate buffer (pH 5.5) to a final concentration of 0.1 U/ml, for 1 hour at 37°C. FITC-conjugated rabbit anti-mouse immunoglobulin (Dako, Denmark) diluted 1:100 in 0.1% bovine serum albumin/PBS (BSA/PBS) solution was used as a secondary antibody during 45 minutes at room temperature. Cells were visualized and pictures were acquired using a Zeiss Optical Microscope.

Immunohistochemistry - Tissue samples were obtained from University Hospital São João and their use approved by the local Ethical committee (CHSJ). Expression of CD44v6 (clone MA54, Invitrogen, 1:50) and STn (Mab 3F1, 1:200) were evaluated in 24 cases of human gastric carcinomas (11 Intestinal sub-type and 13 unclassifiable of Lauren's classification (31)). Paraffin sections were dewaxed and rehydrated. Antigen retrieval was carried out by microwave treatment in sodium citrate buffer (10 mM, pH 6.0) for 20 min. Treatment with 3% hydrogen peroxide (H₂O₂) in methanol for 10 minutes was performed to block endogenous peroxidase. Tissue sections were incubated for 20 minutes with normal rabbit serum in PBS with 10% BSA, before incubation with MAbs to CD44v6 or STn for 1 hour diluted in PBS with 5% BSA. After three washing steps with PBS, sections for immunohistochemistry were incubated with biotin-labelled secondary rabbit anti-mouse antibody (Dako, Glostrup, Denmark) diluted 1:200 in PBS with 5% BSA for 30 min and with ABC kit (Vector Labs, CA, USA) for 30 min. Sections were stained for 2-3 minutes with 0.05% 3,3'-diaminobenzidinetetrahydrochloride (DAB) containing 0.01% H₂O₂. Sections were counterstained with Mayers' haematoxylin solution, dehydrated and mounted. Slides were examined using a Zeiss Optical Microscope.

To evaluate the expression of GalNAc-T₅ we produced a novel MAb 5F11 as previously described (37). Briefly, human GalNAc-T₅ was recombinant expressed in insect cells as a secreted HIS-tagged construct, purified by Ni-chromatography and used for immunization of mice. Hybridomas were selected and characterized as previously described (37). The specificity of MAb 5F11 for GalNAc-T₅ was evaluated by immunocytology with SF9 cells expressing different human GalNAc-T isoforms

(GalNAc-T5, T7, T10, T13) and human cell lines (K562, K562 GalNAc-T5 KO, colo205, Hek293, LST174, HL60, A431). MAb 5F11 was used in immunofluorescence labeling of frozen sections of human gastric carcinomas (n=8, from intestinal, diffuse and unclassified subtype, according to Lauren's classification (31)) and adjacent gastric mucosa (n=4) and intestinal metaplasia (n=4). Serial sections selected for immunofluorescence were stained by haematoxylin and eosin. Immunofluorescence labeling with MAb 5F11 was performed as previously described (21).

In situ Proximity Ligation Assay - *In situ* Proximity Ligation Assay (PLA) was performed in paraffin sections from gastric carcinoma cases for the detection of colocalization of STn and CD44v6. PLA was performed adapting the procedure previously described (38). Briefly, Duolink *In Situ* Probemaker was used to conjugate the PLA oligonucleotide arms directly to primary antibodies according to the manufacturer instructions (MAb MA54 to CD44v6 and MAb 3F1 for STn). Duolink II reagents from Olink Bioscience were used according to the manufacture instructions. Paraffin sections were dewaxed, rehydrated, antigen retrieval was performed as described previously (38), and endogenous peroxidase blocked before sections were incubated for 30 minutes at 37°C with the Blocking Solution (Olink Bioscience). Conjugated MABs to CD44v6 (1:50) and STn (1:200) (labeled with PLA probes) were diluted in PBS with 5% BSA and with 1:20 of assay reagent (Olink Bioscience) and slides incubated for 1 hour at room temperature. Slides were then washed with a filtered solution of 0.01 M Tris, 0.15 M NaCl and 0.05% Tween 20 (Wash buffer A). For the ligation step, a solution with the ligation stock, consisting of two oligonucleotides, (1:5) and the ligase (1:40) both diluted in pure water, was added to slides for 30 minutes at 37°C in order to hybridize to the two PLA probes. After washing with wash buffer A, the amplification solution, consisting of nucleotides and fluorescently labeled oligonucleotides, (1:5) together with polymerase (1:80) both diluted in pure water, were added to the samples. The oligonucleotide arm of one of the PLA probes acted as a primer for a rolling-circle amplification (RCA) reaction using the ligated circle as a template, generating a repeated sequence product. The fluorescently labeled oligonucleotides are then hybridized to the RCA product. The slides were then washed with 0.2 M Tris and 0.1 M NaCl (Wash buffer B) and then with 0.01x wash buffer B. To visualize the nuclei cells were incubated DAPI (Sigma-Aldrich, 0.4 mg/mL). Samples were examined under a Zeiss Imager.Z1 Axio fluorescence microscope (Zeiss, Welwyn Garden City, UK). Images were acquired using a Zeiss Axio cam MRm and the AxioVision Rel. 4.8 software.

RESULTS

Characterization of the O-glycoproteome of gastric cancer AGS and MKN45 SCs - In order to identify secreted or shed O-glycoproteins from gastric cancer cells, we first established gastric cancer cell lines displaying simplified homogeneous O-glycoproteomes. We achieved this by targeting the Cosmc chaperone in two different gastric cancer cell lines (AGS and MKN45) to produce SCs (26). The COSMC knockout clones were initially screened and selected by immunocytology with MAb 5F4 detecting the Tn structure. AGS and MKN45 wt cells produce little if any detectable Tn and STn antigens as evaluated by immunocytology with specific MABs (Fig. 2). Detailed characterization showed that both AGS and MKN45 SCs expressed a mixture of Tn and STn antigens. We confirmed loss of the core1 synthase C1GALT1 and elongated O-glycans by immunocytology with MABs to C1GalT1 and T, respectively. COSMC KO was verified by target specific PCR followed by sequencing. The selected AGS SC clone had a small +1bp insertion and wt alleles were not detectable. With the MKN45 SC clone we were not able to identify mutated alleles using primers spanning 600 bp around the ZFN target site, and no wt allele was found. It is likely that the ZFN

mediated deletion is large and thus not amplified with the used PCR strategy, as previously described for other cell lines (28). We therefore PCR amplified the genes MCTS1 and GLUD2 flanking COSMC and verified their presence in both MKN45 wt and SCs, suggesting that the deletion introduced only affected COSMC (not shown). Both MKN45 and AGS cell lines expressed very little if any ST6GalNAc-I levels detectable by immunocytochemistry with MAb 2C3 (Fig. 2).

Interestingly, AGS SC showed weak reactivity with MAb 3C9 detecting the core1 T O-glycan despite KO of COSMC and loss of the C1GalT1 enzyme as evaluated by immunocytochemistry (Fig. 2). It is currently unclear why AGS SC appear to produce detectable T structures in contrast to other SCs characterized previously (28), but one possibility may be that the relatively high level of expression of C1GalT1 in AGS wt cells (Fig. 2), may result in folding and expression of a minor amount C1GalT1 despite the lack of COSMC in AGS SC.

The analysis of both TCL and SEC of AGS and MKN45 identified a total of 499 O-glycoproteins and 1,236 O-glycosites as listed in Supplemental Table S2. The TCL analysis identified 252 O-glycoproteins in AGS SC and 216 in MKN45 SC, whereas the SEC analysis identified 135 O-glycoproteins in AGS SC and 303 in MKN45 SC (Fig. 3). Although there was considerable overlap between the set of O-glycoproteins identified in TCL and SEC for each cell line, a majority of O-glycoproteins identified were only found in one of the sources. Of the total glycoproteins and glycosites, 178 O-glycoproteins (36%) and 310 O-glycosites (25%) were shared among the two cell lines (Fig. 3, Supplemental Fig. S1).

We compared the O-glycoproteome from the two gastric cancer SC lines with the O-glycoproteome derived from 12 human cancer cell lines from different organs of which none were generated from gastric tissues (designated "Other SCs") (28) (Fig. 4). This revealed that we identified a new subset of O-glycoproteins and O-glycosites in the gastric lines that had not been previously found. Specifically, out of the total 499 O-glycoproteins identified in the two gastric SimpleCells, 324 (65%) overlapped with the O-glycoproteins (663) identified in other SimpleCells. Among the novel 175 O-glycoproteins identified only in gastric cells, there were proteins typically expressed in gastric tissues, such as MUC5AC, a gel-forming mucin highly expressed in the gastric epithelium (39), and gastrin, a peptide hormone produced by G cells in the stomach (40). Comparing the O-glycosylation sites showed that 733 out of the 1,236 (59%) O-glycosites identified in the gastric cell lines overlapped with previously identified O-glycosites (Fig. 4). The finding of new O-glycoproteins and new glycosites in previously identified O-glycoproteins in the gastric cell lines compared to our previous dataset from 12 human cell lines, may reflect cell specific differences in proteomes and in particular repertoire of GalNAc-Ts. We previously found that analysis of different cell lines continuously produced a subset of novel O-glycoprotein identifications without apparent saturation (28), however, it should also be considered that in the present analysis we have used more advanced and sensitive mass spectrometry instrumentation.

Characterization of the aberrant O-glycoproteome of KATO III wt gastric cancer cells - KATO III, in contrast to AGS and MKN45, express Tn/STn glycans as evaluated by immunocytochemistry, although it also produces more complex O-glycans as evidenced by detection of the core1 (T) glycoform ST (after neuraminidase treatment) by anti-T MAbs (Fig. 2) (36). Moreover, expression of the C1GalT1 enzyme is readily detectable by immunocytochemistry (Fig. 2). We wanted to evaluate the performance of the VVA LWAC and MS workflow developed for SimpleCells in enrichment and identification of the subfraction of the O-glycoproteome carrying the aberrant Tn/STn O-glycan structures as a

prelude to analysis of serum. Remarkably, the application of the VVA glycoproteomics workflow with KATO III wt using the same amount of cells/media (0.5 ml packed cells, 80 ml media) provided a relatively small data set (47 O-glycoproteins, 73 O-glycosites) compared to those obtained with SimpleCells (Fig. 5). However, almost a third of the identified O-glycoproteins (17) were not previously found (Supplemental Table S2) (28).

In addition to Tn-glycopeptides we also identified 49 O-glycoproteins (38 unambiguous O-glycosites) with T O-glycans using PNA LWAC after the VVA LWAC step from KATO III TCL (Fig. 5), which is in agreement with the presence of T and ST O-glycans (Fig. 2). Twelve of the T O-glycoproteins overlapped with Tn O-glycoproteins indicating that these proteins carried both Tn and T O-glycosites. Although PNA LWAC appears to have lower sensitivity than VVA (33), we identified a similar number of O-glycoproteins as with VVA.

Characterization of the STn O-glycoproteome in serum- We then modified the SimpleCell enrichment strategy to capture and analyze O-glycoproteins with the STn glycoform in serum pools from healthy and gastric carcinoma patients (Fig. 6A). We used neuraminidase pretreatment of whole serum to uncover STn glycoproteins and enable VVA LWAC enrichment. We initially demonstrated that the desialylation and enrichment step resulted in accumulation of proteins reactive with Tn antibodies and lectins by ELISA, furthermore showing higher Tn reactivity in samples from cancer patients compared with healthy control serum (not shown). These preliminary results also demonstrated that a large amount of sera would be needed for the LWAC and MS strategy, since semi-quantification of Tn reactivity in VVA enriched O-glycoprotein samples obtained from serum versus that obtained from the spent medium of a SimpleCell, suggested that 1 ml of serum generated the equivalent Tn reactivity as 1 ml culture medium from a SimpleCell conditioned media.

Application of 20 ml of pooled gastric cancer serum resulted in identification of a total of 37 O-glycoproteins and 49 O-glycosites (Fig. 6B; Supplemental Table S2). Applying the same strategy to a 20 ml normal serum pool in contrast only resulted in identification of 9 O-glycoproteins (15 O-glycosites) and all of these overlapped with those identified in the cancer serum pool (Fig. 6B; Table 1). Some proteins had two or more glycosites and for 35% only a single site was identified.

We also searched MS spectra for HexHexNAc representing the T structure, which would be expected to be widely found on most serum glycoproteins after pretreatment with neuraminidase, although this should not be specifically captured by the VVA lectin enrichment steps unless the structure coincided on proteins/peptides with Tn structures. We did find 11 HexHexNAc sites distributed in 6 O-glycoproteins, and all of these were found on glycopeptides in combination with Tn O-glycosites.

A total of 24 glycoproteins found in the serum analyses were not found in the gastric SimpleCell glycoproteome (Fig. 6C), and of these 14 were not found in any of the SimpleCell glycoproteomes reported previously (Fig. 6E). These 14 were mainly complement proteins, coagulation factors and immunoglobulin chains, and all abundant normal proteins in serum. The O-glycoproteins identified in healthy individuals were predominantly abundant serum proteins (Table 1). The relative amounts of a large number of proteins in serum are fairly well established and proteomics strategies often relate their sensitivities to these levels (41, 42). Of the serum O-glycoproteins identified where reliable quantification exists, we found that many are placed among the classical plasma protein

group that is mainly produced and secreted by the liver. However, we did find glycoproteins such as ADAMTS, Syndecan-1, fractalkine, and Dickkopf related protein1 amongst a lower abundance set of proteins. Moreover, we identified CD44 and GalNAc-T5 selectively in the cancer serum pool, and both are found in the gastric cancer cell lines as well as in many other SCs (26, 28).

CD44 and GalNAc-T5 are expressed in gastric cancer - The O-glycoproteomics strategy identified CD44 and GalNAc-T5 as potential secreted biomarker O-glycoproteins with aberrant STn O-glycans. Both were found in the secretomes of MKN45 SC and AGS SC lines, KATO III wt cells, and in the gastric cancer serum STn-glycoproteome, but not in the healthy control serum pool. The single O-glycosite identified in CD44 was ambiguously located within the peptide IPVTSAKTGSF584 placed in the v9 exon (Supplemental Fig. S2). All four possible Ser/Thr glycosylation sites in this peptide have been identified as being glycosylated (28), and in the present study these four sites were also identified in the MKN45 SC and AGS SC. The majority of the O-glycosites identified in CD44 in the gastric cancer cell lines were located in exons v5, v6, and v9.

We first used immunohistology to probe expression of CD44 and STn in gastric cancer tissue. We used a MAb directed to the CD44v6 splice-variant, which is known to be highly expressed in gastric cancer (43). Although the peptide found in serum was not located in exon v6, CD44v6 is described as cancer specific, and v9 exon should also be included in the CD44v6-containing isoform (43). Immunohistology showed that CD44v6 was expressed in all 20 gastric cancer tissue samples evaluated. The expression was heterogeneous among the cases ranging between low (< 25 % of positive cells), moderate (25-50% of positive cells) and high (> 50% of positive cells) and with variability in intensity (Supplemental Table S3; Fig. 7). The STn O-glycan was also expressed in all samples tested with varying intensity and regional heterogeneity (Supplemental Table S3). There appeared to be substantial overlap in staining patterns for CD44 and STn. To further confirm co-localization we used PLA with MAbs to CD44v6 and STn on 20 of the same cases of gastric carcinoma, and found that the majority (85%) of the gastric carcinoma cases were positive for PLA signal (Fig. 7; Supplemental Table S3). The majority of the cases (65%) showed a strong positive signal, whereas 20% showed moderate positive PLA signal. The 15% of the cases that were PLA negative for CD44v6/STn did appear to display distinct staining patterns without clear co-localization of CD44v6 and STn by immunocytology using the individual MAbs in contrast to the remaining cases. The results support that CD44v6 is aberrantly glycosylated in gastric cancer and carries the STn glycoform.

In serum of cancer patients we identified one T429 glycosite from GalNAc-T5, and this site was also identified in the gastric cancer SimpleCells. The long stem region of approximately 500 residues in GalNAc-T5 has a large number of O-glycosites reported (30 sites) (28), and most of these (19 sites) and an additional 7 novel ones were also identified in the gastric SCs (Supplemental Fig. S3). The T429 glycosite is positioned in the stem region most proximal to the catalytic domain of the enzyme. Glycosyltransferases are often proteolytically cleaved in the stem region and released as soluble secreted active enzymes (44). GalNAc-T5 has the longest stem region of all known mammalian glycosyltransferases with almost 500 residues and the cleavage site for release of the soluble catalytic domain is unknown. We produced a new MAb (5F11) to human GalNAc-T5 and confirmed expression of this enzyme in the gastric cell lines by immunocytology (Supplemental Fig. S4). We also evaluated the expression of GalNAc-T5 in gastric carcinomas with adjacent normal appearing gastric mucosa. This was done on frozen sections as our MAbs developed to GalNAc-Ts are generated to recognize the native active enzymes (37). GalNAc-T5 was highly expressed in all

(n=8) gastric carcinomas tested (Fig. 8), and staining of cancer cells showed a strong diffuse cytoplasmic labeling including in the invasive tumor areas. GalNAc-T5 was also expressed in the epithelial cells of the adjacent normal gastric mucosa, but here with the typical supranuclear staining pattern found for Golgi-resident glycosyltransferases (37). The diffuse staining found in cancer cells resemble the staining pattern of other glycosyltransferases including GalNAc-Ts in cancer cells, where the Golgi complex is disorganized (45, 46). We also evaluated the expression of GalNAc-T5 in intestinal metaplasia displayed in the mucosa adjacent to gastric carcinomas in four of the cases, and observed strong distinct supranuclear Golgi-like staining in Goblet cells of metaplastic glands in all four cases (Fig. 8D).

DISCUSSION

It is our hypothesis that cancer cells produce and shed/secreted O-glycoproteins with truncated cancer-associated O-glycans that can be detected in circulation, and ultimately that such O-glycoproteins may serve as biomarkers with higher specificity. To address this we first characterized the O-glycoproteomes of two gastric cancer cell lines, and identified a large set of shed/secreted O-glycoproteins of which a major fraction was novel (28). We next developed a strategy for selective isolation and identification of circulating O-glycoproteins with truncated STn O-glycans based on our O-glycoproteomics strategy (26), and we were able to confirm that circulating truncated O-glycoproteins are selectively found in cancer patients in an explorative study with pools of sera. Although there was only partial overlap with the shed'ome of the gastric cancer cell lines and the identified serum O-glycoproteins, we confirmed that two interesting candidates, CD44 and GalNAc-T5, found in both gastric cancer cells and in gastric cancer serum, were expressed in gastric primary tumors. Furthermore, applying the PLA technology we could confirm expression of the CD44v6 splice variant with the STn glycoform in tumour tissues. While the design of our studies was mainly of developmental and explorative nature, the results support the hypothesis and warrant further efforts to identify and develop biomarker assays based on aberrant O-glycoproteins expressed by cancer cells.

The SC O-glycoproteomics strategy enables proteome-wide discovery of O-glycosites with high sensitivity (26). We previously applied this strategy to 12 human cancer cell lines derived from different organs to probe the human O-glycoproteome and found that each cell line expressed a minor subset of unique O-glycoproteins (26, 28). This suggests that the O-glycoproteomes of cells is differentially regulated by its proteome and its repertoire of GalNAc-Ts as predicted (28). Here we found that the O-glycoproteomes of gastric cancer cell lines also produced new subsets of O-glycoproteins with some being characteristic for gastric tissues such as MUC5AC (39) and gastrin (40). Moreover, we identified proteins playing pivotal role in the gastric carcinoma neoplastic transformation, such as E-Cadherin (47, 48) and c-Met (hepatocyte growth factor receptor), which is highly expressed in MKN45 (49). C-Met overexpression and increased activation has been associated with gastric cancer progression (50, 51). Recently, we demonstrated that increased cellular expression of sialylated Lewis antigens can modulate c-Met activation (52).

The sensitivity of the O-glycoproteome strategy is largely based on selective enrichment of O-glycopeptides and nLC-MS sequencing (53). The VVA lectin is quite efficient in selective enrichment both with O-glycoproteins and O-glycopeptides carrying the Tn O-glycans (54). This enabled us to develop a workflow for enrichment of STn O-glycoproteins in serum using initial desialylation and capture of Tn O-glycoproteins and subsequent enrichment of O-glycopeptides (Fig. 6A). Several O-

glycoproteome studies of serum have been undertaken in the past using PNA (35, 55), and these have provided insight into the general presence of O-glycoproteins and O-glycosites in serum. Our interest was to isolate the small subfraction of circulating O-glycoproteins with aberrant STn O-glycosylation, which are predicted to exist from studies with the CA72-4 assay (56, 57) and lectin arrays (58-61). High levels of STn are detectable in sera of patients with different cancer types (62-64) and this is associated with lower survival (65, 66). We confirmed that STn O-glycoproteins are selectively found in gastric cancer patients. Due to the amount of serum needed for our glycoproteomics workflow we had to use pools of sera, and it is clearly necessary to perform studies of individual sera in future.

We only identified a limited set of 37 O-glycoproteins in serum, which compared to our discovery rate in SimpleCells is quite low. However, applying the same workflow to wt KATO III gastric cancer cells, which express Tn and STn without COSMC knockout, we also only identified a small set of O-glycoproteins (Fig. 5). While this clearly illustrates the sensitivity of the SimpleCell strategy, it also shows that cancer cells expressing detectable Tn and STn phenotypes by immunocytology, may not provide the same level of sensitivity. The difference may reflect more inherent difficulties with analysis of heterogeneous glycoproteins and/or low quantities. In this respect we did identify O-glycoproteins in serum with estimated concentrations at 5-20 pg/mL (Dickkopf-related protein 1, GalNAc-T5, ST6GalNAc-I), that represents nearly a 109 fold scale of the higher abundant proteins.

Out of the O-glycoproteins identified in both SimpleCells and cancer serum, we focused on CD44 and GalNAc-T5. CD44 is the major cell surface receptor for hyaluronate (67), that serves as a gastric stem cell marker and has been implicated in important cellular functions (47, 68). The CD44 gene encodes several protein isoforms due to alternative splicing (69). The variant isoforms of CD44 (CD44v) seem restricted to subpopulations with stem cell potential, and they have been proposed to be involved in cancer development. The CD44v6 in particular has been suggested to play a major role in cancer progression due to its ability to bind hepatocyte growth factor (HGF), osteopontin (OPN), and other major cytokines produced in the tumor microenvironment (70). The glycosite identified in CD44 from gastric cancer serum was located within the peptide IPVTSAKTGSF584, which includes the sites also identified in the gastric cancer SimpleCells. This peptide sequence is not located in the exon v6 (Supplemental Fig. S2), but exon v9 has been suggested to be in the CD44v6-containing isoform expressed in gastric cancer (43). We also found O-glycosites in the v6 specific sequence in AGS and MKN45 SCs. CD44v6 containing isoforms have been shown to be highly expressed in premalignant and malignant lesions of the stomach, pinpointing its potential as biomarker for early transformation of the gastric mucosa (43). We confirmed that gastric cancer tissue express the CD44v6 variant and carries STn O-glycans by in situ PLA. It is therefore likely that CD44v6/STn in serum originate from cancer cells and may represent a useful biomarker.

GalNAc-T5 is a poorly studied member of the GalNAc-T family (71), which is expressed in kidney and stomach (72, 73). RNA-seq data obtained in various human tissues by the Illumina Human Body Map Project (HBM) (www.illumina.com; ArrayExpress ID: E-MTAB-513) indicates that GalNAc-T5 is expressed primarily in the lung in man and according to the Human Protein Atlas (www.proteinatlas.org) GalNAc-T5 is also primarily expressed in many tissues from the digestive tract as stomach, colon and rectum, while the original Northern blot analysis performed when GalNAc-T5 was cloned showed expression in sublingual gland, stomach, small intestine and colon of the rat tissues (71). Interestingly, we have found GalNAc-T5 derived O-glycosites in most human cancer cell lines tested with our SC O-glycoproteomics strategy (28), suggesting that this GalNAc-T

isoform is more widely expressed. Using a new MAb we found that GalNAc-T5 is expressed in the three gastric cancer cell lines studied and in K562, colo205, LST174, but only weakly or not detectable in other cancer cell lines like HL60, OVCAR3 and HeLa. We also found that GalNAc-T5 is strongly expressed in normal gastric epithelium and in gastric cancer (Fig. 8). The expression of GalNAc-T5 in normal gastric tissues was limited to a perinuclear Golgi-like localization, as previously found for other GalNAc-Ts. In contrast, GalNAc-T5 was expressed throughout the cytoplasm of cells in gastric cancer tissues. A recent study also reported that GalNAc-T5 was expressed in normal gastric epithelium and gastric cancer using a polyclonal commercial antibody generated to a recombinant peptide fragment from the stem region (73). This study suggested that GalNAc-T5 expression was decreased or lost in advanced TNM stages of gastric cancer and that GalNAc-T5 could serve as a prognostic biomarker. However, the decreased reactivity of the polyclonal antibody in advanced stages was not correlated with decreased mRNA levels for the enzyme. Interestingly, this apparent discrepancy could be related to our finding that the stem region of GalNAc-T5 is O-glycosylated. The polyclonal antibody used to detect GalNAc-T5 was raised to a fairly short non-glycosylated peptide and potential enhanced O-glycosylation could block reactivity. The MAb generated here was produced to the active catalytic domain of GalNAc-T5 to ensure reactivity with the native enzyme, and further studies with this MAb may help clarify whether GalNAc-T5 may serve as a biomarker. Nevertheless, the present study provides compelling evidence that GalNAc-T5 is shed from gastric tumors with STn O-glycans and may have potential as a biomarker.

In summary, we have developed a fruitful discovery strategy for aberrant O-glycoproteins in serum and identified a number of O-glycoproteins that may serve as serum biomarkers for gastric cancer. We also provided compelling evidence that at least some of the aberrant O-glycoproteins detected in cancer serum are derived from cancer cells, and the strategy should be applicable to many cancers known to express STn.

ACKNOWLEDGEMENTS

This work was supported by The Danish Research Councils, The Mizutani Foundation, The Danish National Research Foundation (DNRF107) and Fundação para a Ciência e a Tecnologia (FCT) and COMPETE (Programa Operacional Temático Factores de Competitividade, participado pelo fundo comunitário europeu FEDER) in the framework of the projects: PTDC/BBB-EBI/0786/2012; EXPL/CTM-BIO/0762/2013. Grants were received from FCT (SFRH/BD/73717/2010 to DC), (SFRH/BPD/75871/2011 to AM), (SFRH/BPD/96510/2013 to CG) and (SFRH/BPD/66288/2009 to JAF). IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education, and is partially supported by FCT.

REFERENCES

1. Surinova, S., Schiess, R., Huttenhain, R., Cerciello, F., Wollscheid, B., and Aebersold, R. (2011) On the development of plasma protein biomarkers. *J Proteome Res* 10, 5-16
2. Reis, C. A., Osorio, H., Silva, L., Gomes, C., and David, L. (2010) Alterations in glycosylation as biomarkers for cancer detection. *J Clin Pathol* 63, 322-329

3. Drake, P. M., Cho, W., Li, B., Prakobphol, A., Johansen, E., Anderson, N. L., Regnier, F. E., Gibson, B. W., and Fisher, S. J. (2010) Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin Chem* 56, 223-236
4. Chen, K., Gentry-Maharaj, A., Burnell, M., Steentoft, C., Marcos-Silva, L., Mandel, U., Jacobs, I., Dawnay, A., Menon, U., and Blixt, O. (2013) Microarray Glycoprofiling of CA125 improves differential diagnosis of ovarian cancer. *J Proteome Res* 12, 1408-1418
5. Springer, G. F. (1984) T and Tn, general carcinoma autoantigens. *Science* 224, 1198-1206
6. Wahrenbrock, M. G., and Varki, A. (2006) Multiple hepatic receptors cooperate to eliminate secretory mucins aberrantly entering the bloodstream: are circulating cancer mucins the "tip of the iceberg"? *Cancer Res* 66, 2433-2441
7. Wandall, H. H., Blixt, O., Tarp, M. A., Pedersen, J. W., Bennett, E. P., Mandel, U., Ragupathi, G., Livingston, P. O., Hollingsworth, M. A., Taylor-Papadimitriou, J., Burchell, J., and Clausen, H. (2010) Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer Res* 70, 1306-1313
8. Magnani, J. L., Nilsson, B., Brockhaus, M., Zopf, D., Steplewski, Z., Koprowski, H., and Ginsburg, V. (1982) A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. *J Biol Chem* 257, 14365-14369
9. Guadagni, F., Roselli, M., Amato, T., Cosimelli, M., Mannella, E., Tedesco, M., Grassi, A., Casale, V., Cavaliere, F., Greiner, J. W., and et al. (1991) Clinical evaluation of serum tumor-associated glycoprotein-72 as a novel tumor marker for colorectal cancer patients. *J Surg Oncol Suppl* 2, 16-20
10. Guadagni, F., Roselli, M., Cosimelli, M., Mannella, E., Tedesco, M., Cavaliere, F., Grassi, A., Abbolito, M. R., Greiner, J. W., and Schlom, J. (1993) TAG-72 (CA 72-4 assay) as a complementary serum tumor antigen to carcinoembryonic antigen in monitoring patients with colorectal cancer. *Cancer* 72, 2098-2106
11. Kjeldsen, T., Clausen, H., Hirohashi, S., Ogawa, T., Iijima, H., and Hakomori, S. (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2----6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. *Cancer Res* 48, 2214-2220
12. Baldus, S. E., and Hanisch, F. G. (2000) Biochemistry and pathological importance of mucin-associated antigens in gastrointestinal neoplasia. *Adv Cancer Res* 79, 201-248
13. Itzkowitz, S. H., Bloom, E. J., Kokal, W. A., Modin, G., Hakomori, S., and Kim, Y. S. (1990) Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. *Cancer* 66, 1960-1966
14. Kobayashi, H., Terao, T., and Kawashima, Y. (1992) Serum sialyl Tn as an independent predictor of poor prognosis in patients with epithelial ovarian cancer. *J Clin Oncol* 10, 95-101
15. Leivonen, M., Nordling, S., Lundin, J., von Boguslawski, K., and Haglund, C. (2001) STn and prognosis in breast cancer. *Oncology* 61, 299-305
16. Kim, G. E., Bae, H. I., Park, H. U., Kuan, S. F., Crawley, S. C., Ho, J. J., and Kim, Y. S. (2002) Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology* 123, 1052-1060
17. Ferreira, J. A., Videira, P. A., Lima, L., Pereira, S., Silva, M., Carrascal, M., Severino, P. F., Fernandes, E., Almeida, A., Costa, C., Vitorino, R., Amaro, T., Oliveira, M. J., Reis, C. A., Dall'Olio, F., Amado, F., and Santos, L. L. (2013) Overexpression of tumour-associated carbohydrate antigen sialyl-Tn in advanced bladder tumours. *Mol Oncol* 7, 719-731
18. Cao, Y., Stosiek, P., Springer, G. F., and Karsten, U. (1996) Thomsen-Friedenreich-related carbohydrate antigens in normal adult human tissues: a systematic and comparative study. *Histochem Cell Biol* 106, 197-207

19. Miles, D. W., Happerfield, L. C., Smith, P., Gillibrand, R., Bobrow, L. G., Gregory, W. M., and Rubens, R. D. (1994) Expression of sialyl-Tn predicts the effect of adjuvant chemotherapy in node-positive breast cancer. *Br J Cancer* 70, 1272-1275
20. Gomes, C., Almeida, A., Ferreira, J. A., Silva, L., Santos-Sousa, H., Pinto-de-Sousa, J., Santos, L. L., Amado, F., Schwientek, T., Levery, S. B., Mandel, U., Clausen, H., David, L., Reis, C. A., and Osorio, H. (2013) Glycoproteomic analysis of serum from patients with gastric precancerous lesions. *J Proteome Res* 12, 1454-1466
21. Marcos, N. T., Bennett, E. P., Gomes, J., Magalhaes, A., Gomes, C., David, L., Dar, I., Jeanneau, C., DeFrees, S., Krusturup, D., Vogel, L. K., Kure, E. H., Burchell, J., Taylor-Papadimitriou, J., Clausen, H., Mandel, U., and Reis, C. A. (2011) ST6GalNAC-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci* 3, 1443-1455
22. Marcos, N. T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., Kihlberg, J., Clausen, H., and Reis, C. A. (2004) Role of the human ST6GalNAC-I and ST6GalNAC-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Cancer Res* 64, 7050-7057
23. Sewell, R., Backstrom, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Gatgens, J., Clausen, H., Hansson, G. C., Burchell, J., and Taylor-Papadimitriou, J. (2006) The ST6GalNAC-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J Biol Chem* 281, 3586-3594
24. Ju, T., Lanneau, G. S., Gautam, T., Wang, Y., Xia, B., Stowell, S. R., Willard, M. T., Wang, W., Xia, J. Y., Zuna, R. E., Laszik, Z., Benbrook, D. M., Hanigan, M. H., and Cummings, R. D. (2008) Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer Res* 68, 1636-1646
25. Radhakrishnan, P., Dabelsteen, S., Madsen, F. B., Francavilla, C., Kopp, K. L., Steentoft, C., Vakhrushev, S. Y., Olsen, J. V., Hansen, L., Bennett, E. P., Woetmann, A., Yin, G., Chen, L., Song, H., Bak, M., Hlady, R. A., Peters, S. L., Opavsky, R., Thode, C., Qvortrup, K., Schjoldager, K. T., Clausen, H., Hollingsworth, M. A., and Wandall, H. H. (2014) Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc Natl Acad Sci U S A* 111, E4066-4075
26. Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., and Clausen, H. (2011) Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. *Nat Methods* 8, 977-982
27. Gill, D. J., Tham, K. M., Chia, J., Wang, S. C., Steentoft, C., Clausen, H., Bard-Chapeau, E. A., and Bard, F. A. (2013) Initiation of GalNac-type O-glycosylation in the endoplasmic reticulum promotes cancer cell invasiveness. *Proc Natl Acad Sci U S A* 110, E3152-3161
28. Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Bennett, E. P., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., and Clausen, H. (2013) Precision mapping of the human O-GalNac glycoproteome through SimpleCell technology. *EMBO J* 32, 1478-1488
29. Schjoldager, K. T., Vakhrushev, S. Y., Kong, Y., Steentoft, C., Nudelman, A. S., Pedersen, N. B., Wandall, H. H., Mandel, U., Bennett, E. P., Levery, S. B., and Clausen, H. (2012) Probing isoform-specific functions of polypeptide GalNac-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci U S A* 109, 9893-9898
30. Steentoft, C., Bennett, E. P., Schjoldager, K. T., Vakhrushev, S. Y., Wandall, H. H., and Clausen, H. (2014) Precision genome editing: a small revolution for glycobiology. *Glycobiology* 24, 663-680
31. Lauren, P. (1965) The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification. *Acta Pathol Microbiol Scand* 64, 31-49

32. Adamczyk, B., Struwe, W. B., Ercan, A., Nigrovic, P. A., and Rudd, P. M. (2013) Characterization of fibrinogen glycosylation and its importance for serum/plasma N-glycome analysis. *J Proteome Res* 12, 444-454
33. Yang, Z., Halim, A., Narimatsu, Y., Joshi, H. J., Steentoft, C., Schjoldager, K. T., Schulz, M. A., Sealover, N. R., Kayser, K. J., Bennett, E. P., Levery, S. B., Vakhrushev, S. Y., and Clausen, H. (2014) The GalNAc-type O-glycoproteome of CHO cells characterized by the SimpleCell strategy. *Mol Cell Proteomics* 13, 3224-35
34. Chalkley, R. J., Thalhammer, A., Schoepfer, R., and Burlingame, A. L. (2009) Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A* 106, 8894-8899
35. Darula, Z., and Medzihradsky, K. F. (2009) Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics* 8, 2515-2526
36. Marcos, N. T., Cruz, A., Silva, F., Almeida, R., David, L., Mandel, U., Clausen, H., Von Mensdorff-Pouilly, S., and Reis, C. A. (2003) Polypeptide GalNAc-transferases, ST6GalNAc-transferase I, and ST3Gal-transferase I expression in gastric carcinoma cell lines. *J Histochem Cytochem* 51, 761-771
37. Mandel, U., Hassan, H., Therkildsen, M. H., Rygaard, J., Jakobsen, M. H., Juhl, B. R., Dabelsteen, E., and Clausen, H. (1999) Expression of polypeptide GalNAc-transferases in stratified epithelia and squamous cell carcinomas: immunohistological evaluation using monoclonal antibodies to three members of the GalNAc-transferase family. *Glycobiology* 9, 43-52
38. Conze, T., Carvalho, A. S., Landegren, U., Almeida, R., Reis, C. A., David, L., and Soderberg, O. (2010) MUC2 mucin is a major carrier of the cancer-associated sialyl-Tn antigen in intestinal metaplasia and gastric carcinomas. *Glycobiology* 20, 199-206
39. Reis, C. A., David, L., Nielsen, P. A., Clausen, H., Mirgorodskaya, K., Roepstorff, P., and Sobrinho-Simoes, M. (1997) Immunohistochemical study of MUC5AC expression in human gastric carcinomas using a novel monoclonal antibody. *Int J Cancer* 74, 112-121
40. Larsson, L. I. (2000) Developmental biology of gastrin and somatostatin cells in the antropyloric mucosa of the stomach. *Microsc Res Tech* 48, 272-281
41. Anderson, N. L., and Anderson, N. G. (2002) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1, 845-867
42. States, D. J., Omenn, G. S., Blackwell, T. W., Fermin, D., Eng, J., Speicher, D. W., and Hanash, S. M. (2006) Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nat Biotechnol* 24, 333-338
43. da Cunha, C. B., Oliveira, C., Wen, X., Gomes, B., Sousa, S., Suriano, G., Grellier, M., Huntsman, D. G., Carneiro, F., Granja, P. L., and Seruca, R. (2010) De novo expression of CD44 variants in sporadic and hereditary gastric cancer. *Lab Invest* 90, 1604-1614
44. Paulson, J. C., and Colley, K. J. (1989) Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. *J Biol Chem* 264, 17615-17618
45. Gill, D. J., Chia, J., Senewiratne, J., and Bard, F. (2010) Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *J Cell Biol* 189, 843-858
46. Gill, D. J., Clausen, H., and Bard, F. (2011) Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends Cell Biol* 21, 149-158
47. Cancer Genome Atlas Research, N. (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202-209

48. Pinho, S. S., Carvalho, S., Marcos-Pinto, R., Magalhaes, A., Oliveira, C., Gu, J., Dinis-Ribeiro, M., Carneiro, F., Seruca, R., and Reis, C. A. (2013) Gastric cancer: adding glycosylation to the equation. *Trends Mol Med* 19, 664-676
49. Smolen, G. A., Sordella, R., Muir, B., Mohapatra, G., Barmettler, A., Archibald, H., Kim, W. J., Okimoto, R. A., Bell, D. W., Sgroi, D. C., Christensen, J. G., Settleman, J., and Haber, D. A. (2006) Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752. *Proc Natl Acad Sci U S A* 103, 2316-2321
50. Dua, R., Zhang, J., Parry, G., and Penuel, E. (2011) Detection of hepatocyte growth factor (HGF) ligand-c-MET receptor activation in formalin-fixed paraffin embedded specimens by a novel proximity assay. *PLoS One* 6, e15932
51. Lee, J., Seo, J. W., Jun, H. J., Ki, C. S., Park, S. H., Park, Y. S., Lim, H. Y., Choi, M. G., Bae, J. M., Sohn, T. S., Noh, J. H., Kim, S., Jang, H. L., Kim, J. Y., Kim, K. M., Kang, W. K., and Park, J. O. (2011) Impact of MET amplification on gastric cancer: possible roles as a novel prognostic marker and a potential therapeutic target. *Oncol Rep* 25, 1517-1524
52. Gomes, C., Osorio, H., Pinto, M. T., Campos, D., Oliveira, M. J., and Reis, C. A. (2013) Expression of ST₃GAL₄ leads to SLe(x) expression and induces c-Met activation and an invasive phenotype in gastric carcinoma cells. *PLoS One* 8, e66737
53. Lavery, S. B., Steentoft, C., Halim, A., Narimatsu, Y., Clausen, H., and Vakhrushev, S. Y. (2014) Advances in mass spectrometry driven O-glycoproteomics. *Biochim Biophys Acta pii: S0304-4165(14)00328-6*
54. Steentoft, C., Bennett, E. P., and Clausen, H. (2013) Glycoengineering of human cell lines using zinc finger nuclease gene targeting: SimpleCells with homogeneous GalNAc O-glycosylation allow isolation of the O-glycoproteome by one-step lectin affinity chromatography. *Methods Mol Biol* 1022, 387-402
55. Darula, Z., Sherman, J., and Medzihradsky, K. F. (2012) How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol Cell Proteomics* 11, O111 016774
56. Kim, D. H., Oh, S. J., Oh, C. A., Choi, M. G., Noh, J. H., Sohn, T. S., Bae, J. M., and Kim, S. (2011) The relationships between perioperative CEA, CA 19-9, and CA 72-4 and recurrence in gastric cancer patients after curative radical gastrectomy. *J Surg Oncol* 104, 585-591
57. Marrelli, D., Roviello, F., De Stefano, A., Farnetani, M., Garosi, L., Messano, A., and Pinto, E. (1999) Prognostic significance of CEA, CA 19-9 and CA 72-4 preoperative serum levels in gastric carcinoma. *Oncology* 57, 55-62
58. Gbormittah, F. O., Haab, B. B., Partyka, K., Garcia-Ott, C., Hancapie, M., and Hancock, W. S. (2014) Characterization of glycoproteins in pancreatic cyst fluid using a high-performance multiple lectin affinity chromatography platform. *J Proteome Res* 13, 289-299
59. Haab, B. B., Partyka, K., and Cao, Z. (2013) Using antibody arrays to measure protein abundance and glycosylation: considerations for optimal performance. *Curr Protoc Protein Sci* 73, Unit 27.6
60. Li, C., Simeone, D. M., Brenner, D. E., Anderson, M. A., Shedden, K. A., Ruffin, M. T., and Lubman, D. M. (2009) Pancreatic cancer serum detection using a lectin/glyco-antibody array method. *J Proteome Res* 8, 483-492
61. Zhao, J., Simeone, D. M., Heidt, D., Anderson, M. A., and Lubman, D. M. (2006) Comparative serum glycoproteomics using lectin selected sialic acid glycoproteins with mass spectrometric analysis: application to pancreatic cancer serum. *J Proteome Res* 5, 1792-1802
62. Motoo, Y., Kawakami, H., Watanabe, H., Satomura, Y., Ohta, H., Okai, T., Makino, H., Toya, D., and Sawabu, N. (1991) Serum sialyl-Tn antigen levels in patients with digestive cancers. *Oncology* 48, 321-326

63. Nanashima, A., Yamaguchi, H., Nakagoe, T., Matsuo, S., Sumida, Y., Tsuji, T., Sawai, T., Yamaguchi, E., Yasutake, T., and Ayabe, H. (1999) High serum concentrations of sialyl Tn antigen in carcinomas of the biliary tract and pancreas. *J Hepatobiliary Pancreat Surg* 6, 391-395
64. Takahashi, I., Maehara, Y., Kusumoto, T., Yoshida, M., Kakeji, Y., Kusumoto, H., Furusawa, M., and Sugimachi, K. (1993) Predictive value of preoperative serum sialyl Tn antigen levels in prognosis of patients with gastric cancer. *Cancer* 72, 1836-1840
65. Nakagoe, T., Sawai, T., Tsuji, T., Jibiki, M., Nanashima, A., Yamaguchi, H., Kurosaki, N., Yasutake, T., Ayabe, H., and Tagawa, Y. (2000) Prognostic value of circulating sialyl Tn antigen in colorectal cancer patients. *Anticancer Res* 20, 3863-3869
66. Nakagoe, T., Sawai, T., Tsuji, T., Jibiki, M., Nanashima, A., Yamaguchi, H., Yasutake, T., Ayabe, H., Arisawa, K., and Ishikawa, H. (2001) Pre-operative serum levels of sialyl Tn antigen predict liver metastasis and poor prognosis in patients with gastric cancer. *Eur J Surg Oncol* 27, 731-739
67. Aruffo, A., Stamenkovic, I., Melnick, M., Underhill, C. B., and Seed, B. (1990) CD44 is the principal cell surface receptor for hyaluronate. *Cell* 61, 1303-1313
68. Qiao, X. T., and Gumucio, D. L. (2011) Current molecular markers for gastric progenitor cells and gastric cancer stem cells. *J Gastroenterol* 46, 855-865
69. Sreaton, G. R., Bell, M. V., Jackson, D. G., Cornelis, F. B., Gerth, U., and Bell, J. I. (1992) Genomic structure of DNA encoding the lymphocyte homing receptor CD44 reveals at least 12 alternatively spliced exons. *Proc Natl Acad Sci U S A* 89, 12160-12164
70. Orian-Rousseau, V. (2010) CD44, a therapeutic target for metastasising tumours. *Eur J Cancer* 46, 1271-1277
71. Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., Van Wuyckhuysse, B., and Tabak, L. A. (1998) Cloning and expression of a novel, tissue specifically expressed member of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family. *J Biol Chem* 273, 27749-27754
72. Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., and Tabak, L. A. (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22, 736-756
73. He, H., Shen, Z., Zhang, H., Wang, X., Tang, Z., Xu, J., and Sun, Y. (2014) Clinical significance of polypeptide N-acetylgalactosaminyl transferase-5 (GalNAc-T5) expression in patients with gastric cancer. *Br J Cancer* 110, 2021-2029

Table 1. Summary of the O-glycoproteins found in serum and gastric cancer cell lines.

Accession	Gene Name	UniProt Name	Human serum	SC gastric cancer cell lines				KATO III WT	Other organs SCs (28)			
			Gas	AGS		MKN45		TC	SE	TC	SE	
				TC	SE	TC	SE					L
O6U	ADAMT	ADAMTS-like protein 4	+				+			+	+	
O9N	ST6GA	Alpha-N-acetylgalactosaminide	+	+			+				+	
P026	APOA1	Apolipoprotein A-I	+							+	+	
P026	APOA2	Apolipoprotein A-II	+							+	+	
P067	APOA4	Apolipoprotein A-IV	+									
P041	APOB	Apolipoprotein B-100	+							+	+	
P026	APOC3	Apolipoprotein C-III	+	+							+	
P040	C4BPA	C4b-binding protein alpha	+									
P160	CD44	CD44 antigen	+		+	+	+	+		+	+	
P004	F8	Coagulation factor VIII	+									
P051	F13B	Coagulation factor XIII B chain	+									
P066	C2	Complement C2	+	+								
P010	C3	Complement C3	+									
O94	DKK1	Dickkopf-related protein 1	+			+	+	+		+	+	
O9G	LACRT	Extracellular glycoprotein	+				+		+		+	
O9U	EMCN	Endomucin	+									
P026	FGA	Fibrinogen alpha chain	+		+	+				+	+	
P027	FN1	Fibronectin	+				+			+	+	
P784	CX3CL1	Fractalkine	+				+			+	+	
P055	SERPIN	Heparin cofactor 2	+	+								
P018	IGKC	Ig kappa chain C region	+	+								
P0C	IGLC3	Ig lambda-3 chain C regions	+									
P198	ITIH2	Inter-alpha-trypsin inhibitor	+								+	
O14	ITIH4	Inter-alpha-trypsin inhibitor	+									
P010	KNG1	Kininogen-1	+									
O9N	MXRA5	Matrix-remodeling-associated	+							+	+	
P145	NID1	Nidogen-1	+		+	+	+	+		+	+	
P051	SERPIN	Plasma serine protease inhibitor	+							+	+	
O7Z	GALNT	Polypeptide N-	+		+	+	+	+	+	+	+	
O9U	HEG1	Protein HEG homolog 1	+	+	+	+	+			+	+	
O92	PRG4	Proteoglycan 4	+	+								
A1L	SSC5D	Soluble scavenger receptor	+	+							+	
O6U	SBSN	Suprabasin	+								+	
P188	SDC1	Syndecan-1	+				+		+	+	+	
P079	THBS1	Thrombospondin-1	+			+		+		+	+	
O86	C14orf3	Uncharacterized protein	+	+						+	+	
P042	VWF	von Willebrand factor	+									

FIGURES

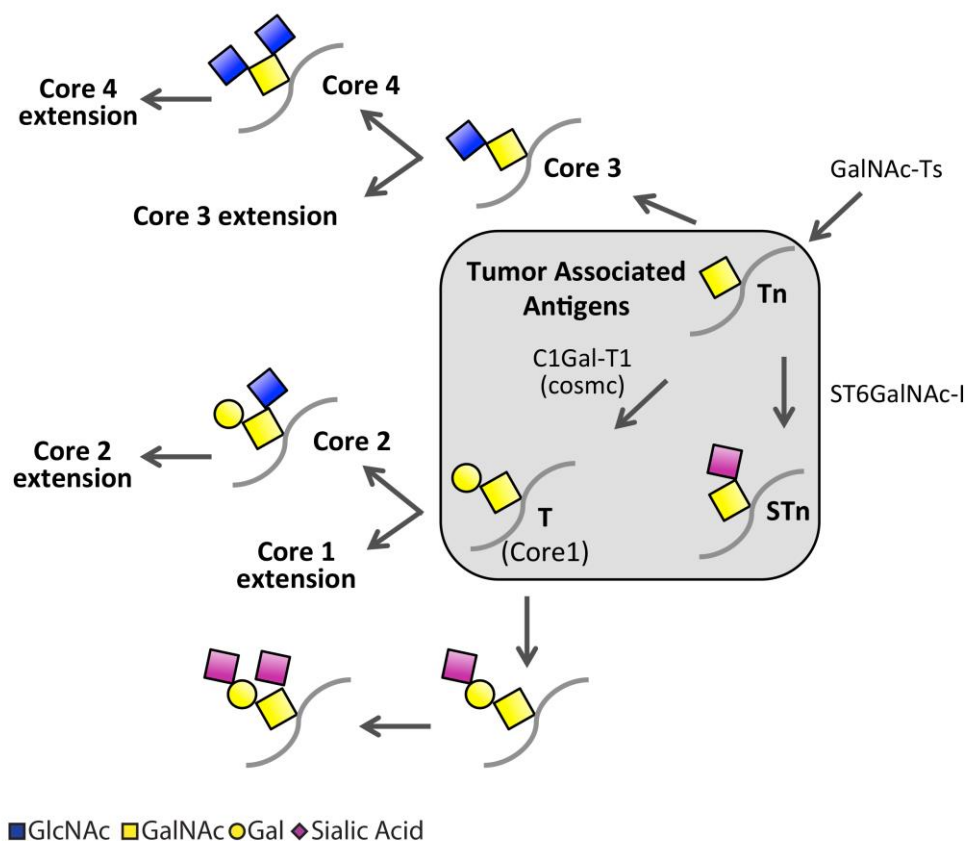


Fig. 1. Schematic depiction of the initial biosynthetic pathways of O-linked protein glycosylation. Overview of the O-linked protein glycosylation. O-GalNAc glycosylation is initiated by up to 20 different GalNAc-transferases. The addition of GalNAc to serines or threonines (or tyrosines) forms the Tn structure that can be sialylated by ST6GalNAc-I or further elongated to form up to 4 core structures. The core structures can be further elongated.

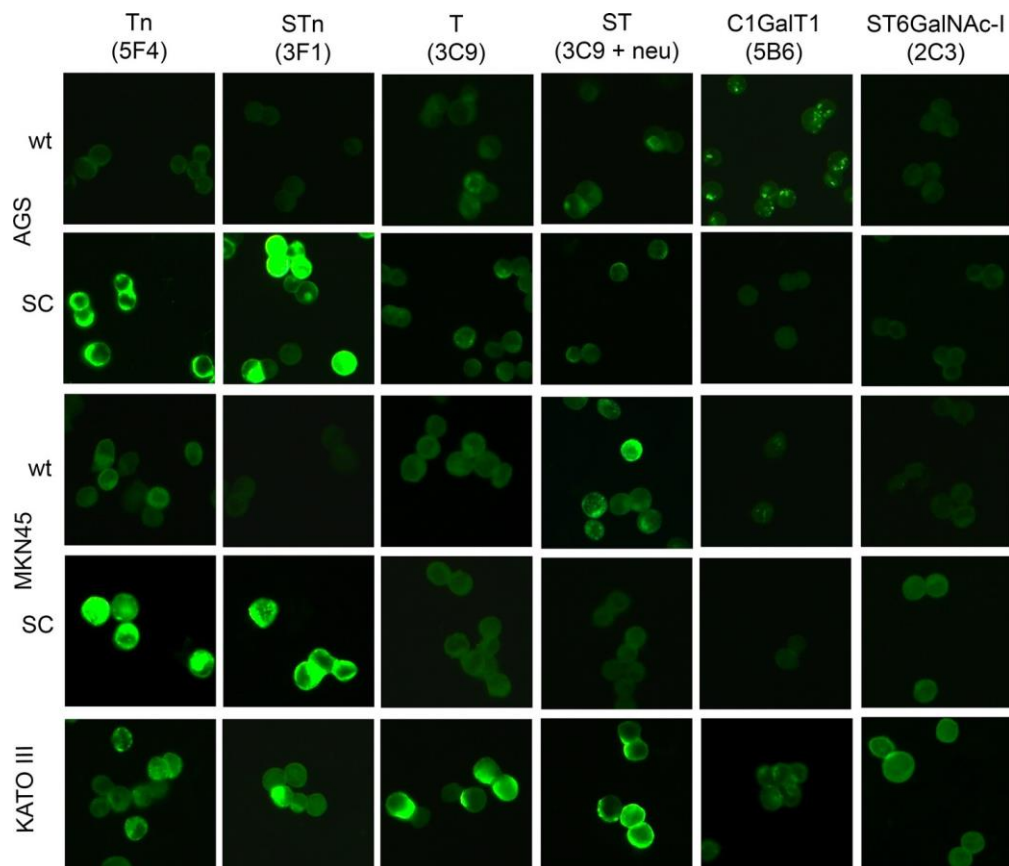


Fig. 2. Immunofluorescence staining of engineered SC cell lines. Staining of AGS and MKN45 SC lines and corresponding wt cells as well as KATO III wt cell line with MAbs to Tn (5F4), STn (3F1), T (3C9) with and without pretreatment with neuraminidase (Neu), C1GalT1 (5B6), GalNAc-T5 (5F11), and ST6GalNAc-I (2C3).

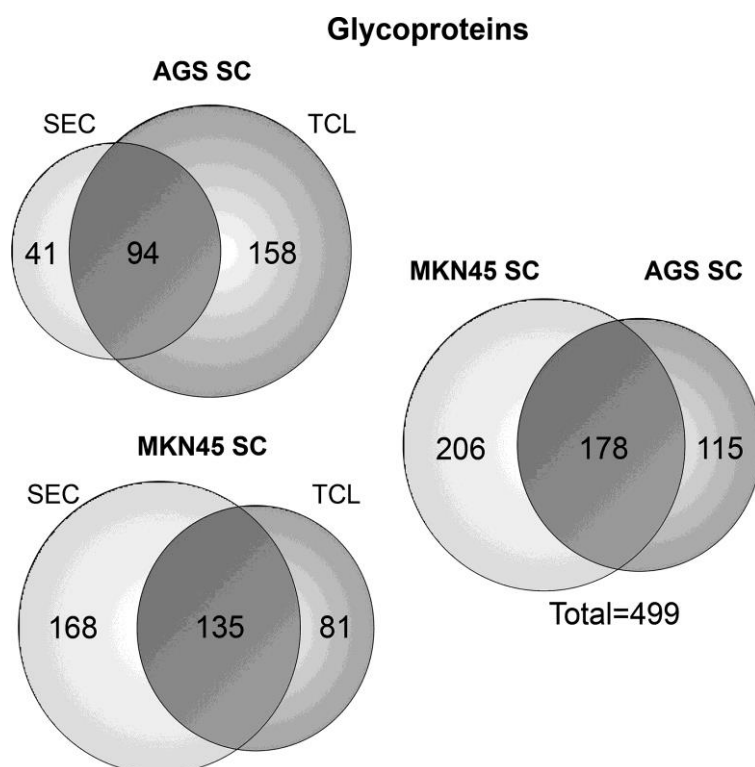


Fig. 3. Summary of O-glycoproteins identified in AGS SC and MKN₄₅ SC. Venn diagrams illustrating comparative analysis of glycoproteins identified in TCL and SEC.

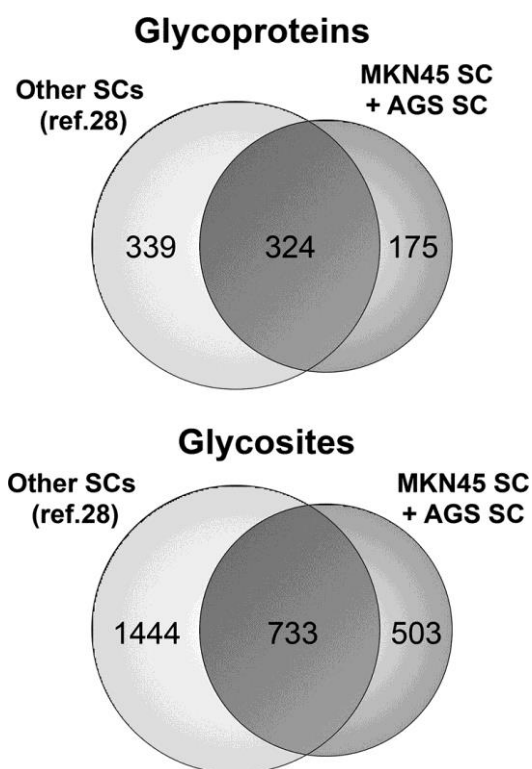


Fig. 4. Comparative analysis of O-glycoproteins identified from gastric cancer SCs and SCs from other organs. Illustration of both AGS SC and MKN₄₅ SC O-glycoproteomes compared with previously published glycoproteomes on 12 other organs SimpleCells (28) at the glycoprotein and glycosites level.

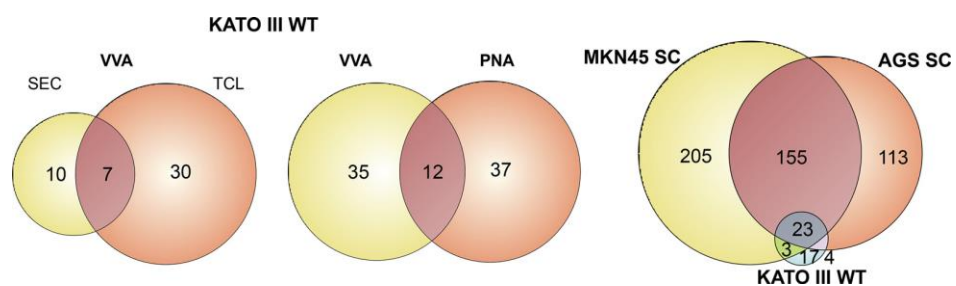


Fig. 5. O-glycoproteins identified in the gastric carcinoma KATO III wt cell line. Comparative analysis of VVA LWAC and PNA LWAC identified O-glycoproteins.

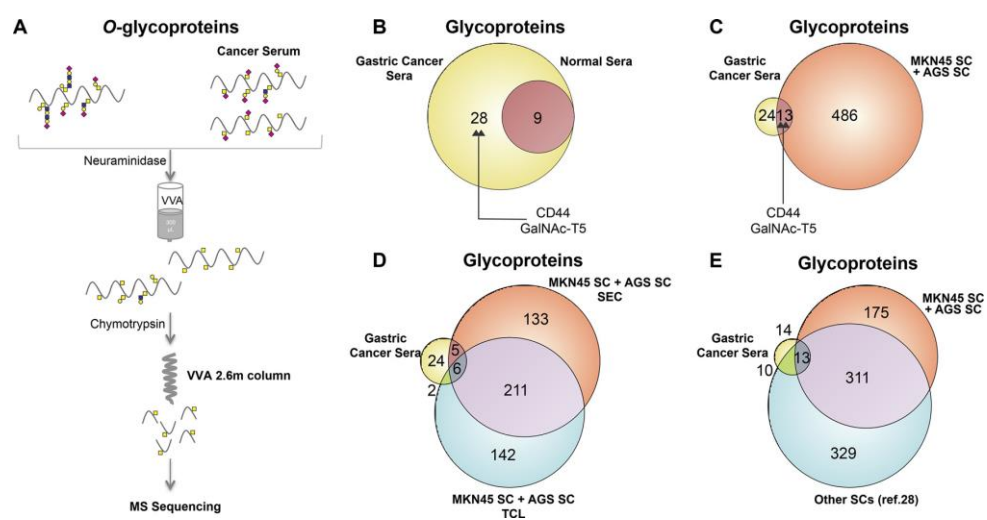


Fig. 6. The O-glycoproteome of the human gastric cancer serum. A, Schematic depiction of the serum O-glycoproteome strategy. B, Comparative analysis of glycoproteins determined in serum pools from gastric carcinoma and healthy control groups. C, Comparison of O-glycoproteins identified in serum and gastric cancer SCs. D, Comparison of O-glycoproteins identified in SEC and TCL from gastric SCs with those identified in gastric cancer sera. E, Comparison of the O-glycoproteins identified from gastric cell lines and gastric cancer sera with previously published O-glycoproteomes annotated in the GlycoDomain Viewer (28).

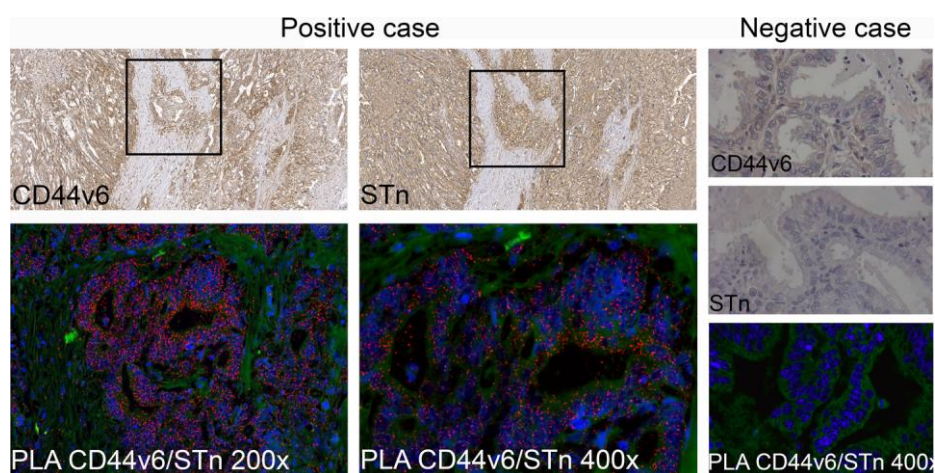


Fig. 7. Proximity Ligation Assay for detection of CD44 and Sialyl-Tn in human gastric carcinoma. Immunohistochemistry for CD44v6 and STn antigen demonstrating the expression in a gastric carcinoma. The vast majority of carcinoma cells are immunoreactive (100x). The squares in the picture indicate the area selected to be depicted in the Proximity ligation assay (PLA). PLA for the co-detection of CD44v6 protein and the STn antigen with microscope amplification of 200x and 400x. The PLA positive signals (red dots) result from the detection of proximity between the two antigens. PLA positive signals are restricted to the areas were both antigens are expressed. Nuclei are stained with DAPI (blue) and green corresponds to tissue auto-fluorescence.

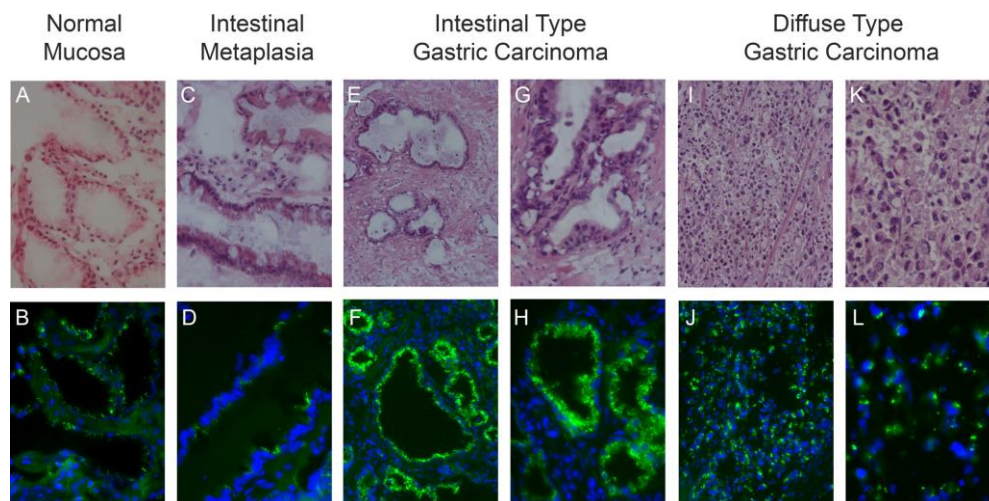


Fig. 8. Expression of GalNAc-T5 in human gastric tissues. The immunofluorescence staining of GalNAc-T5 showed that this enzyme is expressed in human normal gastric mucosa (B) and in intestinal metaplasia (D) with an expression limited to the cytoplasm and perinuclear region of the cells. In gastric carcinoma tissues, both intestinal (F and H) and diffuse (J and L) type also showed staining in cytoplasm, including in depolarized tumor cells. HE was performed in serial sections (A,C,E,G,I and K). Images are represented with a amplification of 200x (E,F and I,J) and 400x (A,B,C,D,G,H and K,L).