Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Matthew Kudelka Date

Decoding the Cellular Glycome: Mucin-type O-glycans in Health and Disease

By

Matthew Robert Kudelka

Doctor of Philosophy

Graduate Division of Biological and Biomedical Science

Immunology and Molecular Pathogenesis

Richard D. Cummings, Ph.D.
Advisor

Andrew S. Neish, M.D.
Committee Member

Asma Nusrat, M.D.
Committee Member

Ifor Williams, M.D., Ph.D.
Committee Member

Periasamy Selvaraj, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Decoding the Cellular Glycome: Mucin-type O-glycans in Health and Disease

By

Matthew Robert Kudelka

B.A. (*magna cum laude*), Columbia University, 2010

Advisor: Richard D. Cummings, Ph.D.

An abstract of a dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Graduate Division of Biological and Biomedical Science

Immunology and Molecular Pathogenesis

2017

Abstract

Cells are coated with a glycocalyx rich in carbohydrates, or glycans, that facilitate interactions of cells with the environment. Mucin-type O-glycans are a major class of cell surface glycosylation and are critical for embryogenesis and altered in a range of diseases, such as inflammatory bowel disease (IBD) and cancer. O-glycan synthesis is initiated in the Golgi with addition of GalNAc to serine or threonine in a glycoprotein and further extended by various monsaccharides to form complex, but poorly defined, structures. We sought to define the normal O-glycome, evaluate how it changes in disease, and determine the consequences of those changes.

Traditional approaches for O-glycomics chemically release a finite number of glycans from a sample in an inefficient and degradative process. To address this, we developed a technology termed Cellular O-glycome Reporter/Amplification (CORA) that amplifies and sequences the O-glycome from cells, yielding ~100-1000× increased sensitivity over traditional approaches. We incubated cells with a peracetylated GalNAc-Bn chemical O-glycan precursor that is taken up into cultured cells and modified by endogenous glycosyltransferases utilizing native sugar donors to form Bn-O-glycans reprenting the native repertoire of cellular O-glycans, followed by secretion, purification, and analysis. We combined our next-generation glycomics across a range of cell types with computational models to predict the size of the human glycome and have begun to define the normal cellular O-glycome and how it changes in disease.

IBD arises from a disrupted immune response to the gut microbiota in genetically susceptible individuals. Recent evidence indicates that the gut microbiota is spatially disrupted in IBD with more prominent dysbiosis in the mucosa versus the luminal contents. O-glycans are the major class of glycosylation in the intestinal epithelia and overlying mucus layer and are truncated in IBD with loss of terminal epitopes and increased expression of simple sugars such as GalNAc (Tn), and Sialyl Tn (STn). Cosmc is a molecular chaperone on the X-chromosome required for normal extension of O-glycans beyond the simple Tn and STn antigens. A recent GWAS implicated *Cosmc* as a sex-specific risk gene in IBD. To explore the role of Cosmc and altered glycosylation in IBD, we deleted *Cosmc* in the intestinal epithelia and observed loss of the mucus layer, enhanced bacterial-epithelial, and ultimately spontaneous microbe-dependent inflammation. KO mice exhibited a loss of microbial diversity and dysbiosis in the colonic mucosa but not in the overlying lumen or small intestine resembling the spatial disruption of the gut microbiota in IBD. In contrast to male mice with loss of one allele of *Cosmc*, female mice with loss of one allele were completely protected from spontaneous colitis and partially protected from experimental colitis due to lateral migration of WT mucin glycocalyx over KO cells. These results validate Cosmc as a sex-specific risk gene that spatially regulates the microbiota and links IBD disease genetics, altered glycosylation, and spatial disruption of the gut microbiota. Collectively, these studies define the normal O-glycome, how O-glycans change in disease, and how these changes contribute to disease biology.

Decoding the Cellular Glycome: Mucin-type O-glycans in Health and Disease

By

Matthew Robert Kudelka

B.A. (*magna cum laude*), Columbia University, 2010

Advisor: Richard D. Cummings, Ph.D.

A dissertation submitted to the Faculty of the

James T. Laney School of Graduate Studies of Emory University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Graduate Division of Biological and Biomedical Science

Immunology and Molecular Pathogenesis

2017

Acknowledgements

More than any other endeavor science is a human affair. Everyday in lab my colleagues, mentors, and friends challenge me to think critically, question dogma, and devise elegant, incisive experiments. Outside the lab, family and friends inspire me to do beautiful science that matters and share that beauty with the world. For this, I am grateful to countless people, many of whom I regrettably do not have space to list.

First and foremost, I would like to thank my PI, Dr. Richard D. Cummings. It is a great a gift to mentor a student and one that Dr. Cummings embraces with his time, commitment, and generosity. No matter how busy, Dr. Cummings takes time to help plan experiments, discuss results, and entertain the crazy ideas of a young scientist, simultaneously providing concrete guidance and the freedom, flexibility, and inspiration for scientific creativity. More than anything else, Dr. Cummings spreads a passion and energy that makes doing science in his lab an extraordinary joy. I could not imagine a better Ph.D. mentor and, perhaps most importantly, I will leave the lab with an even greater love and commitment to science than when I began.

I have had numerous collaborators who have made this thesis possible. I thank my many coauthors who are too numerous to list and members of the Cummings lab, including R. Aryal, T. Ju, J. Zeng, S. Lehoux, Y. Matsumoto, X. Sun, R. Barnes, R. Kardish, N. Prasanphanich, T. McKitrick, L. Byrd-Leotis, M. Hanes, J. Heimburg-Molinaro, and Sandy Cummings for making the lab a home. I also thank Dr. Andy Neish for his

generosity and mentorship when my lab moved to Boston, Dr. Asma Nusrat for guidance and support, and Dr. Elliot Chaikof for welcoming me into the BIDMC community. Both the Emory MD/PhD and IMP programs have provided invaluable guidance, including Drs. Parkos, Ressler, Gross, Evavold, Boise, Mary Horton, and Emily Morran, and I also thank my thesis committee – past and present, including Drs. Richard Cummings, Andy Neish, Asma Nusrat, Ifor Williams, Fadlo Khuri, and Periasamy Selvaraj for countless insights and mentorship.

I am eternally grateful to my family and friends, including my mom, Margaret Roback; my dad, Andrzej Kudelka; my maternal grandparents, Robert and Frances Roback; my paternal grandparents, Leon and Halina Kudelka; my aunt and uncle Lori and Adam Miller; and my extended family. Together they have supported and loved me and taught me how to laugh, respect others, and bravely walk my own path. Lastly, I would like to thank Amanda Mener whose love, support, joy, and friendship has enriched my life during the past years.

Table of Contents

Figure and Table Index

Chapter 1.

Chapter 3.

Chapter 4.

Chapter 1

An Introduction: Mucin-type O-glycans in Cancer

(Adapted from Kudelka et al, Advances in Cancer Research, 2015.)

Abstract

Mucin-type O-glycans are a class of glycans initiated with N-acetylgalactosamine (GalNAc) α-linked primarily to Ser/Thr residues within glycoproteins and often extended or branched by sugars or saccharides. Most secretory and membrane-bound proteins receive this modification, which is important in regulating many biological processes. Alterations in mucin-type O-glycans have been described across tumor types and include expression of relatively small-sized, truncated O-glycans and altered terminal structures, both of which are associated with patient prognosis. New discoveries in the identity and expression of tumor associated O-glycans are providing new avenues for tumor detection and treatment. This review will describe mucin-type O-glycan biosynthesis, altered mucin-type O-glycans in primary tumors, including mechanisms for structural changes and contributions to the tumor phenotype, and clinical approaches to detect and target altered O-glycans for cancer treatment and management.

I. Introduction

Altered glycosylation is a hallmark of cancer that has helped to shape the management and understanding of cancer. Currently, several glycan-based biomarkers are in use worldwide and glycans have been established as key participants in tumorigenesis and progression. In the 1950s, glycopeptides isolated from transformed cells were found to be larger in size than those from their non-transformed counterparts(1-3). Around the same time, some plant lectins were found to exhibit enhanced binding to tumor cells, and in the 1970s and 1980s, researchers discovered that many of the anti-tumor monoclonal antibodies generated against tumors recognized glycans(4-6). These observations indicated that glycans are altered in cancer, setting the stage to investigate when, where, how, and what glycan structures are altered in cancer, which have led to new strategies that have fundamentally altered our view of cancer and approach to attack this deadly disease.

Glycans are present in all living organisms, required for life, and regulate a diversity of biological process. In mammals, glycans are constructed from a combination of 10 monosaccharides (Gal, Glc, Man, Fuc, Xyl, GalNAc, GlcNAc, GlcA, IdoA, and N-acetylneuraminic or sialic acids), which are attached via α or β glycosidic bond to form linear and/or branched structures (**Figure 1.1**).

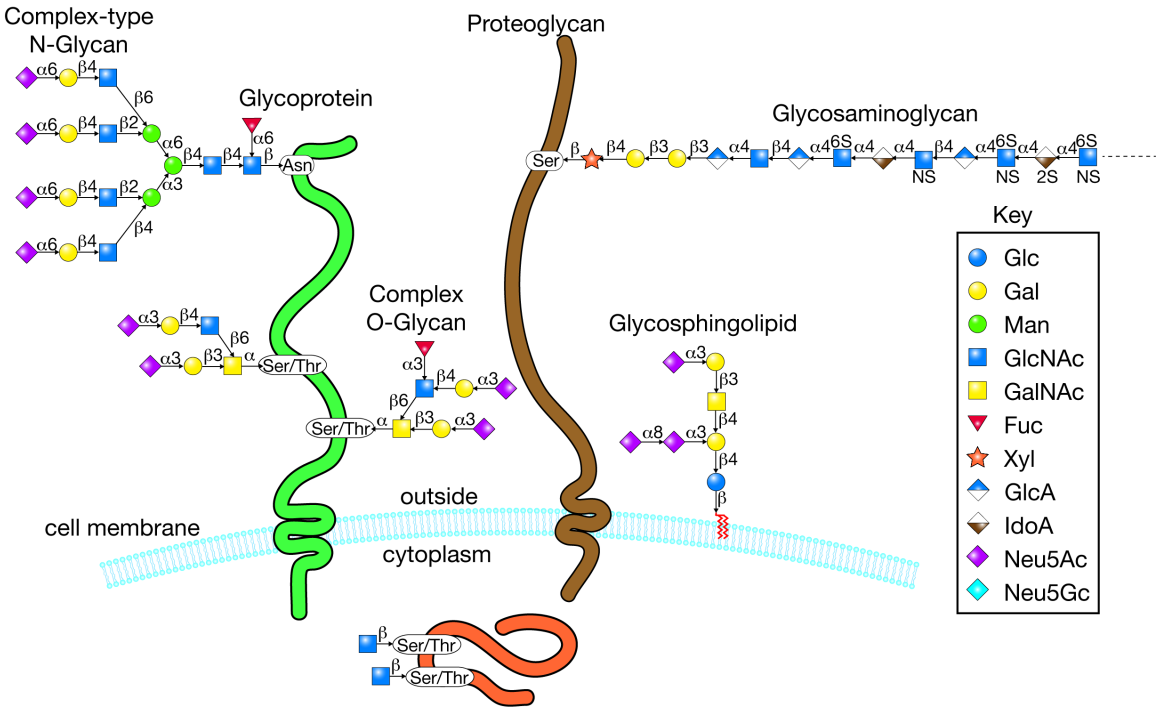


Figure 1.1. Diverse glycoconjugates of the cell. Human cells are covered with a dense assortment of glycoproteins, proteoglycans, and glycolipids, in addition to GPI-anchored glycoproteins (not shown here). The glycoproteins contain Asn-linked oligosaccharides (N-glycans) and Ser/Thr-linked oligosaccharides (O-glycans). The proteoglycans contain Ser-linked glycosaminoglycans, comprised of heparan sulfate, chondroitin sulfates, dermatan sulfate, and keratin sulfate. The glycolipids are largely glycosphingolipids, comprised of ceramide to which glucose is the linking sugar. In addition, O-linked GlcNAc is found in cytoplasmic, nuclear, and mitochondrial glycoproteins. The large repertoire of glycans in such glycoconjugates constitute the *glycome* of the cell and each cell type expresses its own relatively unique glycome, which is also subject to development and disease-specific changes. The symbols used to represent the monosaccharides are indicated.

Further structural diversity is obtained through modifications of the saccharides, e.g. phosphorylation, sulfation, and acetylation, as well as glycan linkage to various macromolecules. Such complex modifications occur in glycoproteins, glycolipids, GPI-anchored proteins, and in free glycans as found in milk and secretions. Glycoproteins can be broadly divided into two classes, N-glycans and O-glycans, although many types exist and 9 of the 20 amino acids can be modified with sugars. N-glycans are linked via an amide bond to asparagine in the Asn-X-Ser/Thr sequon where X is any amino acid except proline. O-glycans are linked most often to serine or threonine, and in some cases to tyrosine, and can be further subdivided into nuclear/cytoplasmic O-glycans, consisting of O-GlcNAc which functions in conjunction with phosphorylation to regulate signal transductions, and secreted or membrane-bound glycoproteins with O-glycans. The most common O-glycan in both membrane and secretory proteins is the mucin-type or GalNAc-type O-glycan initiated by GalNAcα1-linked to Ser/Thr of both mucin and non-mucin glycoproteins(7-11) (**Figure 1.1**). Unlike N-glycans, no conserved glycosite sequon has been identified for O-GalNAc-linked glycans(12-14). Other types of O-glycans include O-glucose, O-fucose, O-mannose, O-galactose, and O-xylose, the latter occurs in proteoglycans. In contrast to nuclear/cytoplasmic O-GlcNAc, which is dynamic, O-glycans in the secretory pathway are stable through the life of the glycoprotein, unless acted upon by glycosidases, such as sialidases (neuraminidases) derived from pathogens during infection. In addition to glycoproteins, glycolipids form a major component of cellular glycoconjugates and in mammals consist primarily of ceramide-linked glycans, forming what are called glycosphingolipids or GSLs (**Figure 1.1**), divided into the lacto, globo, and ganglio series.

Mucin-type O-glycans were first observed on mucins, but later shown to be ubiquitous. Eichenwald discovered that mucins contain carbohydrates in 1865 and Gottschalk and colleagues discovered that GalNAc links the carbohydrate to the mucin in the 1960s(15-19). Recently, glycoproteomics and prediction algorithms identified mucin-type O-glycans on ~83% of proteins entering the ER-Golgi secretory apparatus, including many non-mucin proteins(14). O-glycoproteins contain hundreds of O-glycans, as on MUC2, a dozen or so O-glycans, as on the LDL-Receptor, or a single O-glycan, as on erythropoietin and the transferrin receptor(20-25).

O-glycans regulate various physiological processes. Blockage of extensions of O-glycans in mice is embryonically lethal, while tissue specific deletion results in defects in platelets, endothelia, kidneys, GI tract, immune cells, and lipid metabolism, indicating O-glycans regulate these processes(26-35). Related defects have also been observed in humans, resulting in endocrine, immune, and developmental dysfunction, in addition to cancer. Non-malignant diseases include familial tumoral calcinosis (FTC), dyslipidemia, Wiskott-Aldrich Syndrome, Tn syndrome, and congenital heart disease(36-41).

Like glycans in general, O-glycans on glycoproteins use a variety of mechanisms to regulate biological processes. These are broadly categorized into direct and indirect effects(42). Direct effects involve direct interaction of a glycan epitope with a glycan binding protein (GBP). GBPs include soluble and cell surface proteins from self or microbes or parasites. Many classes of GBPs have been identified including lectins (C-

type, P-type, I-type, L-type, R-type, galectins, etc.), GAG-binding proteins, antibodies, and others(43). Indirect effects of protein glycosylation include effects on protein conformation, stability, recycling, solubility, proteolysis, immune surveillance, etc. A classic example is the LDL receptor, which requires mucin-type O-glycans for protein stability and activity(44-46).

Cancers express altered mucin-type O-glycans, in addition to altered N-glycans and glycolipids as described elsewhere(47-63). These tumor O-glycans comprise 1) oncofetal antigens, which are rare in normal adult tissue but expressed embryonically; 2) neo-antigens, which are novel structures not appreciably expressed either embryonically or in normal tissues; and 3) altered levels of normal antigens. Normal adult tissues do not express oncofetal or neo-antigens, making these ideal for targeted diagnostics and therapeutics; however, all three types of alterations are important in tumor biology and can be useful in clinical management.

Tumor O-glycans consist of both relatively small and very extended structures, including the truncated glycans Tn, sialyl Tn, and T, as well as the extended glycans ABO(H) and sialylated Lewis antigens on poly-N-acetyllactosamine (**Figure 1.2**). Tumors also express dysregulated post-glycosylational modifications, such as reduced sulfation and sialic acid acetylation. In tumors, truncated O-glycans tend to be tumor-specific, or only found in tumors but not normal cells, while altered terminal structures tend to be tumor-associated, with distinct changes noted in tumors but the structures themselves present in some normal tissues. Alterations in O-glycans terminal structures are also observed on N-

glycans and glycolipids, in contrast to truncated O-glycans found only on O-glycans. Despite these differences, both small and extended tumor O-glycans are present across carcinomas and contribute to the tumor phenotype.

O-glycans are altered at the earliest stages of cellular transformation, and genetically engineered mouse models recapitulating some of these alterations suggest that these alterations are important in cancer initiation(27, 64). Tumor O-glycans also correlate with cancer invasion and metastasis and can be engineered into cell lines, resulting in enhanced metastatic potential in xenotransplant studies. Altered O-glycans contribute to metastasis through various mechanisms, ranging from supporting tumor-endothelial interactions to survival in the blood via interaction with platelets and immune evasion(65-68).

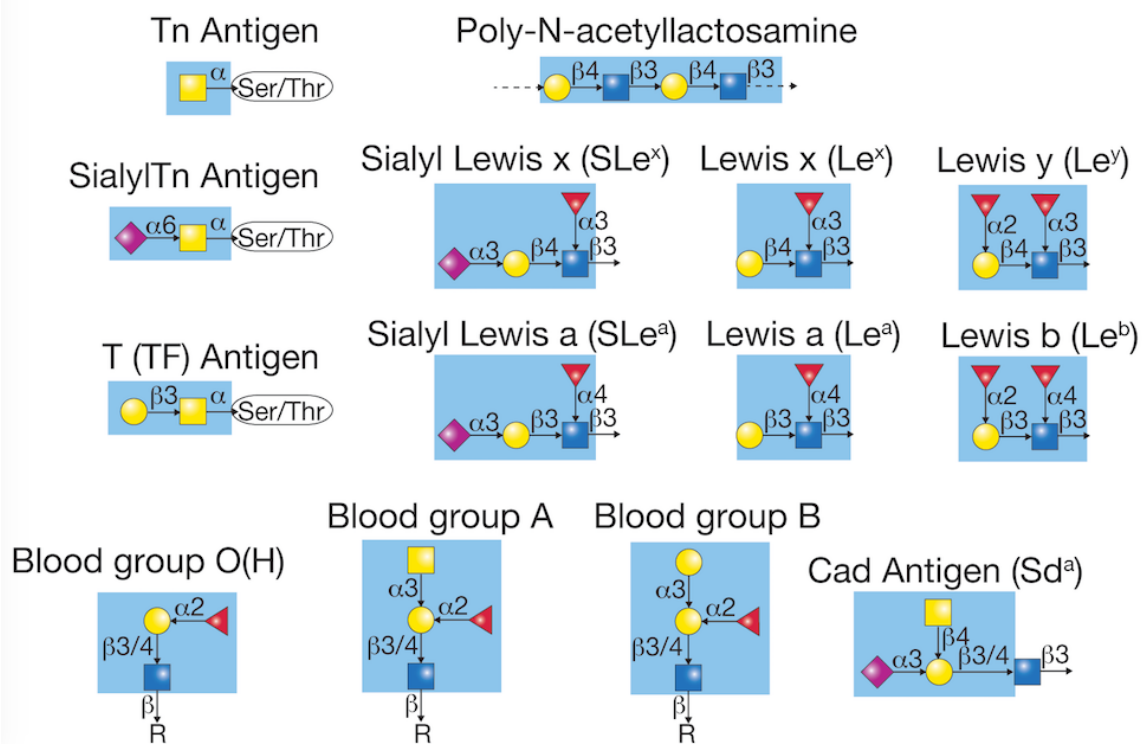


Figure 1.2. Tumor-associated carbohydrate antigens. The structures of many tumor-associated carbohydrate antigens are indicated. The colored box in each structure represents the known antigenic determinant recognized by antibodies.

Knowledge of altered O-glycan structures in cancer has led to the development of O-glycan based biomarkers, including glycan or glycoprotein targeted antibodies, such as CA15-3, CA125, CA19-9, and B72.3, as well as auto-antibody arrays and glycan based imaging. Glycan-targeted therapeutics have also been developed or are in development including passive immunotherapies, carbohydrate-based vaccines, and various strategies to block glycan-GBP interactions, such as sialyl Lewis x (sLe^x)-selectin interactions(66).

This review will introduce O-glycan biosynthesis, describe alterations observed in human tumors and possible mechanisms for these alterations, as well as how these alterations may contribute to tumor biology. Genetic and transcriptional alterations in genes contributing to O-glycosylation will also be discussed as well as tissue and serum biomarkers, imaging, and glycan targeted therapeutics. We will conclude with our perspectives and where we believe the greatest opportunities are for translating what we know about altered O-glycans in cancer to improve patient care.

II. O-glycan Biosynthesis

Overview

Mucin-type O-glycans consist of branched and linear arrangements of monosaccharides that are transferred by glycosyltransferases to glycoproteins on serine/threonine residues as they traverse the Golgi apparatus (**Figure 1.3**). The synthesis of mucin-type O-glycans is complex and depends on many factors. 1) Expression of glycosyltransferase genes: Glycosyltransferases are first synthesized and undergo transcriptional regulation, which depends on tissue-specific, environmental, and pathologic factors. 2) Localization of glycosyltransferases: After transcription, glycosyltransferases must be translated in the rough endoplasmic reticulum and transported to the appropriate location in the secretory apparatus. The localization and levels of enzyme in the Golgi are regulated by retrograde and anterograde vesicular cycling, post-translational modifications such as cytoplasmic tail phosphorylation, and also general Golgi regulation. 3) Golgi structure: The structure of Golgi stacks differs between cell types, under physiologic, environmental, and pharmacologic stress, and in different cellular state, such as proliferation or cytokinesis. These changes affect routes of protein export and glycosylation.

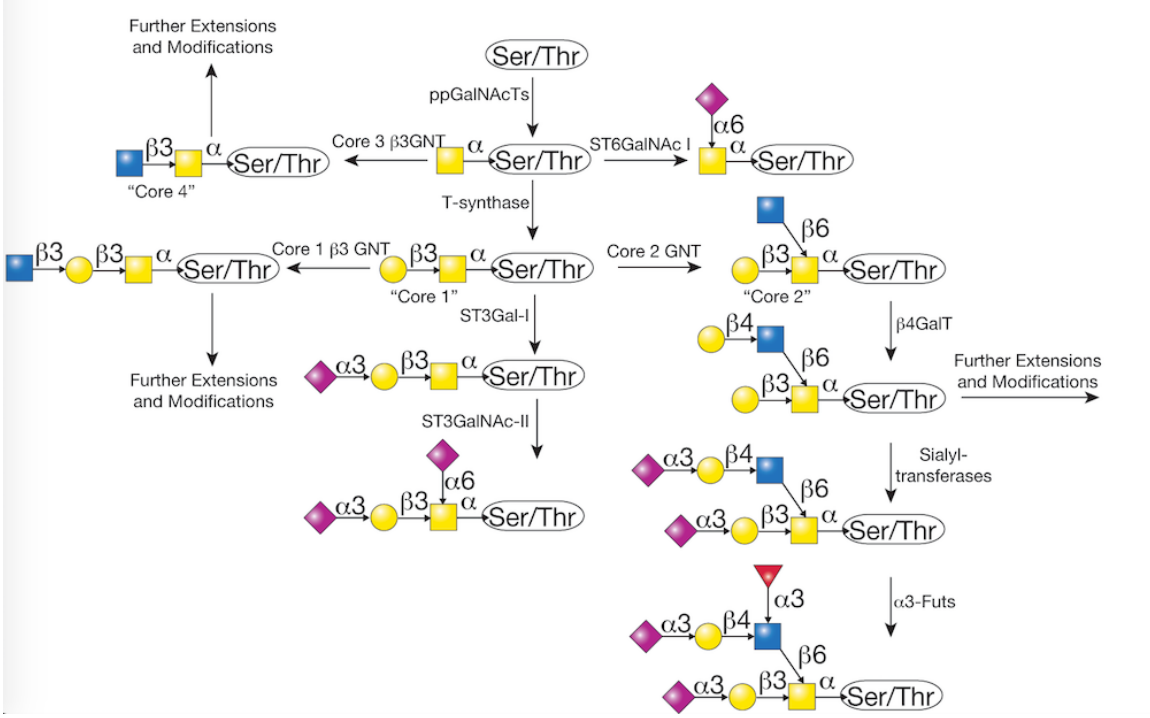


Figure 1.3. The biosynthesis of O-GalNAc-type O-glycans. O-glycan synthesis is initiated and completed in the Golgi apparatus. The ppGalNAcT family of enzymes adds N-acetylgalactosamine from the nucleotide sugar donor UDP-GalNAc to proteins entering the Golgi to form the Tn antigen. The Tn antigen is normally a precursor to a wide variety of other structures, deriving from modifications of the GalNAc residue, to generate core 1, core 2, and core 3 O-glycans. The key reaction is the addition of galactose from UDP-Gal by the enzyme termed T-synthase, which generates the common core 1 O-glycan. The core 1 and/or core 2 O-glycans are found in all human cells. Such glycans are extended by various glycosyltransferases using specific nucleotide sugar donors, e.g. UDP-Gal, UDP-GlcNAc, UDP-GalNAc, GDP-Fuc, CMP-Sialic acid, etc.

Ultimately glycosylation results in production of glycan structures from the cumulative enzymatic activity of many glycosyltransferases and perhaps host or foreign

glycosidases. Glycosyltransferases exhibit varying activities on different glycoprotein or glycopeptide substrates and sometimes occupy distinct or overlapping compartments in the Golgi, enabling competition between glycosyltransferases in glycan synthesis. Availability and levels of sugar donors impact glycosylation, and congenital disorders of glycosylation have been observed due to defects in glycosyltransferases as well as defects in sugar transporters(69).

In glycobiology, a non-template driven set of glycosyltransferase reactions, results in glycosylation microheterogeneity: one glycosite on one type of protein contains various structurally distinct glycans. Microheterogeneity has been observed in various systems, for example in the production of immunoglobulins for biopharmaceuticals, and is considered a principle of glycobiology. How this happens and what benefit microheterogeneity may confer to the cell is not completely clear. Nonetheless, protein conformation, structure, oligomerization, ratio of glycosyltransferase-to-substrate, and whether a protein is membrane bound or secreted affect glycosylation and heterogeneity. Although glycosylation is complex and incompletely understood, much is known about how O-glycans are synthesized to produce a variety of structures, some of which are altered in cancer. Here we outline key pathways, enzymes, and structures involved in mucin-type O-glycan biosynthesis as these are critical to informing our understanding of altered O-glycosylation in cancer.

Core structures 1 – 4

Mucin-type O-glycosylation initiates with transfer of GalNAc from UDP-GalNAc to Ser/Thr in a glycoprotein via an α-linkage to form GalNAcα1-Ser/Thr, which is also recognized as the Tn antigen(8-10). This reaction is catalyzed by a family of enzymes called polypeptide GalNAc-transferases (ppGalNAcTs), consisting of 20 members in humans (**Figure 1.3**). In contrast, *Drosophila* has 14 and *C. elegans* has 9 members(70). ppGalNAcTs are thought to initiate O-glycosylation in the cis Golgi, although some reports indicate that these enzymes may be variably distributed in the medial and trans Golgi, in addition to the cis Golgi(71, 72). ppGalNAcTs are unique among glycosyltransferases in that many contain a lectin domain, facilitating interaction not just with peptide but also with glycans on the peptide. This has led to the idea that there are two classes of ppGalNAcTs: initiator glycosyltransferases and glycopeptide glycosyltransferases(73). The first group transfers UDP-GalNAc to unglycosylated peptides while the second group utilizes glycosylated peptides. Notably, some ppGalNAcTs have both activities, so these groups are not mutually exclusive.

Each mammalian cell does not express all ppGalNAcTs, but rather, different tissues have unique expression patterns of particular family members(74). Similarly, different ppGalNAcTs are thought to modify different, though possibly overlapping, sets of glycoproteins and glycosites, although ppGalNAcTs can compensate to some degree for defects in other transferases(75, 76). These ideas are supported by evidence that deletion of individual ppGalNAcTs in mice result in viable mice with variable and sometimes subtle defects depending on the ppGalNAcT deleted(77). Similarly, findings are observed in humans in which defects in ppGalNAcT11 are associated with congenital heart disease

and defects in ppGalNAcT3 are associated with calcium/phosphate dysregulation(36, 40). SNPs, mutations, and altered transcription of different ppGalNAcTs have been implicated in cancer as discussed later.

Synthesis of Tn antigen is normally followed by transfer of Gal, GlcNAc, or GalNAc to the Tn antigen to form core O-glycan structures 1-8 (**Figure 1.3**). Core 5 – 8 are rare structures, whereas core 1 – 4 are common and will be discussed here. Core 1 or the T antigen is Galβ1-3-GalNAcα-Ser/Thr. This structure is synthesized by the T-synthase (Core 1 β3-galactosyltransferase, C1GalT1), which transfers Gal from UDP-Gal to Tn in the cis and medial Golgi. T-synthase is ubiquitously expressed in all cells and in mammals requires its unique molecular chaperone Cosmc (Core 1 β3 GalT Specific Molecular Chaperone or C1GalT1C1), which is also ubiquitously expressed(78-82). Cosmc is unique in the chaperone field in that it has a single specific client, and is unique in the glycobiology field in that it was the first and only chaperone identified for a glycosyltransferase (**Figure 1.4**). Interestingly, Cosmc shares sequence similarity to the T-synthase suggesting that it originally arose from a duplication and transposition of the *T-synthase* in an evolutionary ancestor. *C. elegans* and *Drosophila* T-synthase orthologs do not require Cosmc for proper folding, presumably due to the presence of N-glycans that were lost in mammalian T-synthase but may facilitate interaction with calnexin/calreticulin in the ER(83). Mammalian Cosmc interacts with unfolded T-synthase, perhaps co-translationally in the RER via a unique peptide region in T-synthase called the CBRT (Cosmc Binding Region of T-synthase)(84, 85). This results in proper folding of the T-synthase and production of an active enzyme, which is transported to the

Golgi, preventing non-productive aggregation, ubiquitination, retrotranslocation to the proteasome, and degradation(78). Loss of Cosmc or T-synthase activity results in expression of Tn in all cells described to date and various pathologies such as Tn syndrome (also known as permanent mixed-field polyagglutinability), possibly IgA nephropathy, and cancer as discussed below(9, 10, 33, 34, 86). Also, deletion of *Cosmc* or the *T-synthase* in a mouse results in embryonic lethality and expression of the Tn antigen and also bleeding dysfunction when deleted in platelet and endothelial cells(32-34). T antigen is normally sialylated or modified by GlcNAc transferases; however, it can be expressed in various pathologies, such as cancer, as well as on activated B cells during a germinal center reaction.

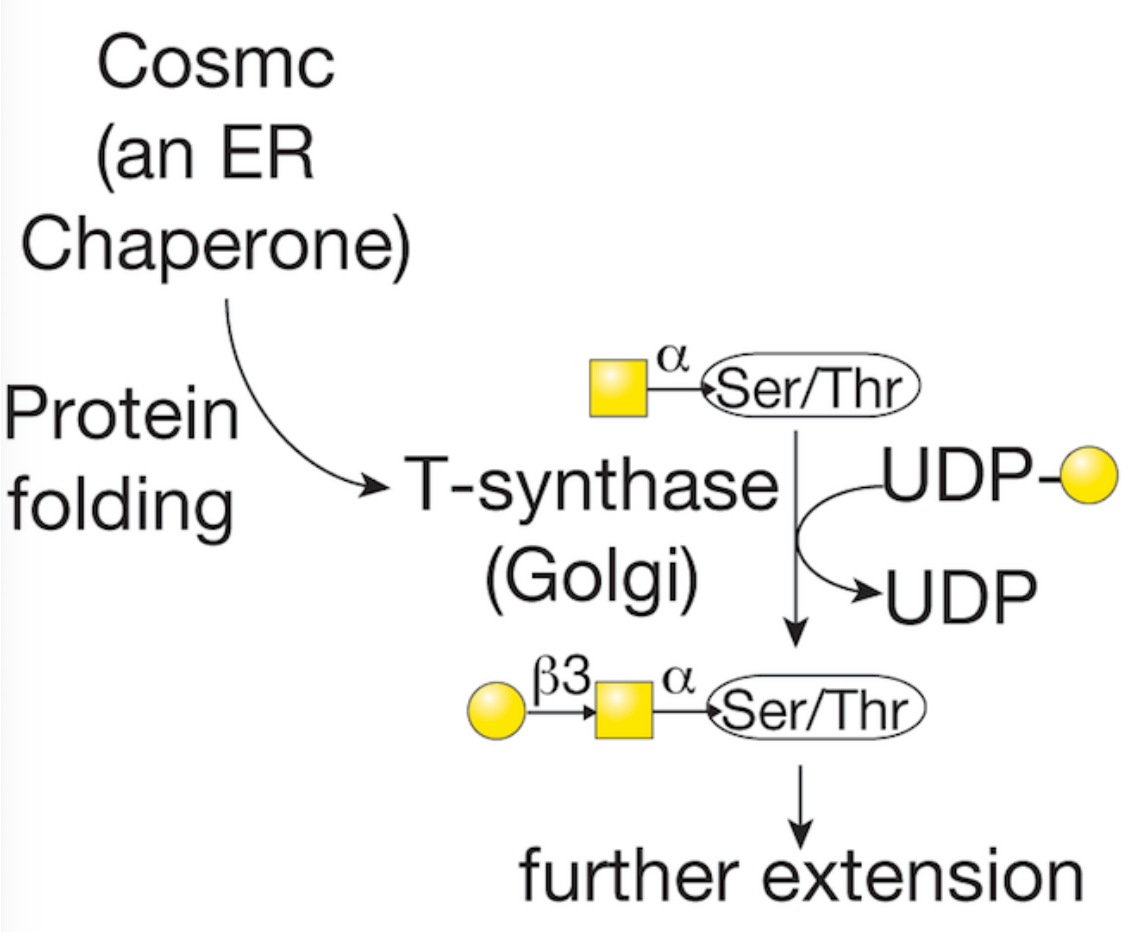


Figure 1.4. The role of Cosmc in core 1 O-glycan synthesis. The T-synthase is the enzyme that generates core 1 O-glycan, also termed the T antigen. However, the formation of active T-synthase (the core 1 β3galactosyltransferase or C1GalT1), which is a Golgi enzyme, requires its correct folding in the endoplasmic reticulum by the specific molecular chaperone Cosmc (Core 1 β3 GalT Specific Molecular Chaperone). Cosmc is encoded by a gene on the X-chromosome, and acquired alterations in expression of Cosmc, either by genetic mutation, epigenetic silencing, or other mechanisms, can lead to expression of the Tn antigen.

Core 1 or the T antigen can be further converted to core 2 by one of three Core 2 GlcNAc Transferases (C2GnT1-3), which transfer GlcNAc from UDP-GlcNAc via a β1-6 linkage to form GlcNAcβ1-6(Galβ1-3)GalNAcα1-Ser/Thr(87-91)(**Figure 1.3**). C2GnT1 and 3 only modify core 1 to form core 2 structures whereas C2GnT2 can also modify core 3 to form core 4 structure, as described below. Hence, C2GnT2 is also called C2/4GnT. C2GnT1 is ubiquitously expressed, C2GnT2 or C2/4GnT is restricted to GI-tract, pancreas, and kidney, and C2GnT3 is restricted to thymus and T cells(92). Presumably, distinct tissue distribution and activities, in the case of C2/4GnT, facilitate tissue-specific regulation and co-regulation of different core structures, such as core 2 and 4. C2GnTs are related to other β6GnTs, including the I-GnTs involved in formation of the I blood group structure and GnTV involved in β1-6 branching of N-glycans.

Unlike T-synthase, which appears to be constitutively transcribed and expressed, core 2 appears to be more sensitive to cellular state and differentiation. Activation of mature T cells upregulates C2GnT1, resulting in increased core 2-based structures. In contrast, resting mature T cells contain primarily core 1-based structures(93). Transcriptional regulation of C2GnTs is complex with multiple transcripts and promoters per enzyme. C2GnT1, for example, uses alternative promoters to produce 5 different mRNAs(94, 95). In addition to transcriptional regulation, enzymatic competition regulates synthesis of core 2 based structures.

C2GnTs functionally co-localize with ST3Gal-I which transfers N-acetylneuraminic acid (Sialic acid) via α2-3 linkage to Gal in core 1 to form sialyl core 1 or sialyl T. Formation of sialyl T by ST3Gal-I inhibits transfer of GlcNAc by C2GnTs. Although only activated T cells normally express core 2, deletion of ST3Gal-I results in elevated expression of core 2 in naïve and activated T cells suggesting that ST3Gal-I activity normally outcompetes C2GnT for substrate in naïve T cells(30).

Core 2 forms a platform for polyLacNAc $(\text{-3Gal}\beta\text{1-4GlcNAc}\beta\text{1-})_n$, which functions as a ligand for several galectins and a substrate to form blood group antigens and various Lewis antigens. In addition to galectins, polyLacNAc-containing glycans interact with other lectins, such as selectins. Hence, regulation of core 2 is critical to the regulation of structures attached to core 2. Core 2 is elevated in immunopathologies, such as Wiskott-Aldrich Syndrome and HIV, and expression of C2GnTs is elevated in many cancers and decreased in others, both correlating with progression of disease(37, 96, 97).

Core 1 and core 2-based structures are ubiquitously expressed. In contrast, core 3 and 4 are primarily expressed in the GI tract. Core 3 structure is synthesized by Core 3 N-acetylglucosaminyltransferase (C3GnT; β3GnT6) by transferring GlcNAc from UDP-GlcNAc to the Tn antigen in β1-3 linkage to form GlcNAcβ1-3GalNAcα1-Ser/Thr, which can then be further modified by the C2/4GnT branching enzyme which transfers an additional GlcNAc via β1-6 linkage to form GlcNAcβ1-3(GlcNAcβ1-6)GalNAcα1-Ser/Thr or core 4. Interestingly, although *C3GnT* is expressed in stomach > small intestine ~ colon in humans, core 3 is most often observed in the colon and appears less abundant in the stomach and not present in tissues outside of the GI tract(98). This suggests that either transcript level does not completely correlate with the activity, that C3GnT may compete with other glycosyltransferases, such as T-synthase, for its substrate the Tn antigen, or that C3GnT may have unique acceptor specificities while T-synthase has broad substrates. In support of this idea, core 1 and 2 structures predominate in the stomach. Core 3 and 4–based structures are found on mucins in intestines and may be important in maintaining the mucus barrier and preventing pathological interactions between bacteria and luminal epithelial cells. Accordingly, deletion of *C3GnT* in mice increases susceptibility to DSS-induced colitis(27). Core 3 may play a role in suppressing tumor development as discussed below.

Extended O-glycans

Although O-glycan structures are typically smaller in size than N-glycans, core 1–4 structures are often extended to form various structures including polyLacNAc chains, Lewis antigens, and various blood group antigens including well known ABO blood groups, as well as less well known Cad (Sd^a) antigens (**Figure 1.3**). Some of these terminal structures are also found on other glycoconjugates such as N-glycans and glycolipids. In some cases, terminal structures can confer biological activity whether on an O-glycan, N-glycan, or glycolipid, for example in sialyl Lewis x (sLe^x) mediated sperm-egg interactions; however, in some cases the class of glycan presenting a terminal structure is biologically important(99). For example, P-selectin requires a very specific glycopeptide epitope to engage its glycoprotein partner, PSGL1. This epitope includes sLe^x on a core 2 residue with nearby sulfated tyrosine(100, 101). Deletion of O-glycans abrogates this binding(28, 102). O-glycans likely share some glycosyltransferase machinery, such as β4GalT, with other classes of glycoconjugates to extend their O-glycans; however, O-glycan specific extensions are also observed. In addition to glycosyltransferases, monosaccharide modifications, such as acetylation and sulfation, are critical to synthesize glycan-binding epitopes, whether for endogenous lectins or monoclonal antibodies generated to recognize glycans or glycoconjugates.

Extended Core 1

Core 1 is most often sialylated by ST3Gal-I and/or ST6GalNAc I-IV to form mono or disialyl core 1 and branched to form core 2. However, other modifications of core 1 are sometimes observed. Core 1 is classically defined as a type 3 chain (Galβ1-3GalNAc-R)

and can serve as a platform for blood group antigens, such as H, A, and B antigens, as well as for O-glycan specific modifications such as the Cad (Sd^a) antigen, which is also found on extended core 2, 3, and 4 structures (**Figure 1.3**). Furthermore, core 1 can be elongated or extended by Core 1 β3-N-acetylglucoaminyltransferase (Core 1 GnT) by transferring GlcNAc from UDP-GlcNAc to form extended core 1, GlcNAcβ1-3(Galβ1-3)GalNAcα1-Ser/Thr(35). This can be further modified by other glycosyltransferases to form Sulfated sLe^x structures on extended core 1, which is expressed by activated endothelial for inflammatory leukocyte homing and recognized by monoclonal antibody MECA-79(35, 103, 104).

Extended Core 2

Extended core 2 are quite common and mediated by alternating activity of β4GalTs and β3GnTs, which form polyLacNAc chains based on type 2, repeats $(3Galβ1\text{-}4GlcNAcβ1\text{-})_n$ (**Figure 1.3**). These structures can be expressed as linear chains, also called i antigen, branched by β1-6GnT-I to form branched structures, and/or modified by fucosyltransferases, sialyl transferases, sulfotransferases, etc. to form various blood group antigens as well as Lewis, sialyl Lewis and sulfo sialyl Lewis structures. PolyLacNAc are also substrates for a class of animal lectins called galectins, which are important in immunity, cell turnover, and growth-factor activity(105). In addition to expression on O-glycans, polyLacNAc are also found on N-glycans as the Galβ4Ts and β3GnTs responsible for synthesizing i antigen can function on O-glycans, N-glycans, and glycolipids(106-110).

Extended Core 3, 4

Core 3, 4 and extended structures are less well detailed, in part because core 3 structures are restricted to the GI tract in humans, but by enzyme activity are reduced in GI-cancers and generally not observed in cancer cell lines(98, 111). Further, although core 3-based structures are thought to be a major component of colonic glycans, based on studies of purified or partially purified mucins from the GI tract, core 3 is minimally expressed in the mouse GI tract(112). Evaluating enzyme activity as a supplement or correlate to structural data is difficult because C3GnT is an extremely unstable enzyme(113). Nonetheless, a few studies have evaluated mucins from GI tract and observed core 3, core 4, and extended core 3 and 4 structures in human colonic mucins(114). Extended core 3 structures are most often observed with one of the most abundant structures being Siaα2-6 core 3 with sialic acid on the GalNAc, extended by β1-3/4Gal and with variable extension of a few type 1 or type 2 chains and presence of fucosylation, sialylation, and sulfation. Additionally, branching off Galβ1-3 core 3 has been observed as well as core 4 structures, core 5 structures (GalNAcα1-3GalNAc), Cad/Sd^{a} antigen, blood group determinants, and Lewis structures(21, 114, 115).

ABO blood group antigens

Blood group antigens are observed on O-glycoproteins, N-glycoproteins, and glycolipids, both on red blood cells and various other cells of the body. Blood group antigens are

synthesized on type 1, 2, 3, or 4 structures. Type 1 and 2 structures are Galβ1-3GlcNAc-R and Galβ1-4GlcNAc-R, respectively (**Figure 1.3**). Both are present on O- and N-glycoproteins as well as glycolipids. Type 3 and 4 structures are both Galβ13GalNAc-R, however the R group for type 3 and 4 differ. R for type 3 is Ser/Thr of an O-glycopeptide and R for type 4 is a glycolipid moiety. Type 2 structures are ubiquitous while type 1 structures are found in the GI tract. Type 1 and 2 can both be found in polymers of (Type 1)n and (Type 2)n, with the latter forming polyLacNAc chains, also called i blood group. In addition to forming a linear chain, i blood group can also be branched by various β1-6GnTs to form I blood group. I blood group predominates after embryonic development, increasing through adulthood(116).

Synthesis of blood group antigens requires at least 2 steps. The first is synthesis of H antigen, the structure corresponding to O blood type. The second is synthesis of either A or B structure. The H antigen is generated by addition of fucose in α1,2 linkage to a terminal galactose on a type 1 – 4 chain. Two genetic loci encode the H transferase. The H loci is functional in red blood cells and the secretor loci is function in GI epithelia, getting its name for the secreted blood group antigens produced from secreted glycoconjugates(117). These transferases are also important in synthesizing some Lewis antigen as discussed below.

After synthesis of the H structure, the A and B transferases, which differ by 4 amino acids, utilize the H structure to synthesize A and B structures on Type 1 – 4 chains. The A transferase transfers GalNAc from UDP-GalNAc via α3 linkage to the terminal Gal of

the H structure, while the B transferase transfers Gal from UDP-Gal via α3 linkage also to the terminal Gal of the H structure. Individuals carrying mutated A/B transferases encode neither functional A or B transferase, making them O-blood group; only one functional A or B transferase, making them AA/AO or BB/BO; or both functional transferases, making them AB+. More rarely, individuals can be H-, Se-, or H-/Se-, making them unable to synthesize AB/H or other blood group structures such as Lewis antigens.

Susceptibility or protection from various diseases, such as certain infections, has been associated with the presence of different blood group antigens. Pathogens contain GBPs that may recognize cells of an individual with one blood type but not another. Alternatively, individuals with a given blood type cannot mount an adaptive immune response to pathogens expressing the same blood group or blood group-like structures. Galectins appear to be able to fill this immunologic gap by recognizing and killing ABO expressing bacteria(118). In addition to infections, AB/H structures and changes in these structures are observed in cancers and contribute to the tumor phenotype, as discussed later.

Lewis antigens

Lewis antigens are synthesized primarily by endodermal epithelia, such as GI epithelia, but are found in endodermal epithelia and RBCs due to transfer of glycolipids to RBCs(117). Lewis structures are found on type 1 and 2 chains of O-glycans, N-glycans,

and glycolipids. Type 1 chains contain Lewis$^{a/b}$ while type 2 chains contain Lewis$^{x/y}$ (**Figure 1.3**). The Lewis locus encodes the fucosyltransferases responsible for synthesizing the Lewis antigens(119). These transferases exhibit similar expression to the secretor loci.

The Lewis transferase is an α3/4FucT which transfers fucose from GDP-Fuc to GlcNAc in a type 1 or 2 chain. An α3 linkage is formed when transferred to a type 2 chain and an α4 linkage is formed when attached to a type 1 chain due to prior occupancy of the Gal on the 4 or 3 position of the GlcNAc, respectively. Addition of the fucose forms the Le^{a} (type 1) or Le^{x} (type 2) structure. Transfer of fucose to the terminal galactose in α1-2 to form the H antigen prior to action of the α3/4FucT is responsible for forming the Le^{b} (type 1) and Le^{y} (type 2) antigens. Formation of the H antigen uses the same α1-2FucT responsible for synthesizing the H precursor to A and B blood groups(120). In summary, Lewis antigens are synthesized by addition of α3/4fucose to an unsubstituted type 1 or 2 chain to form Le$^{a/x}$ antigens or to an H type 1 or 2 chain to form Le$^{b/y}$ antigens.

Lewis antigens can also be sialylated and/or sulfated to form sialyl and sulfo Lewis antigens. Sialylation most often occurs at the 3 position of the terminal galactose of the type 1 or 2 chain to form SLe$^{a/x}$. Sulfation can also occur at the 3 position of the terminal galactose, denoted 3' (' indicates terminal galactose whereas no ' indicates modifications of the subterminal GlcNAc), the 6 position of the terminal galactose, denoted 6', or the 6 position of the subterminal GlcNAc, denoted 6. Sialylation and sulfation on the terminal galactose are compatible, resulting in the possibility of structures such as 6,6' bis sulfo-

Sialyl Le^x(120). Sulfo, sialyl, and sulfo sialyl Lewis antigens are important in physiological processes such as inflammation, in particular because of their role in leukocyte rolling and as selectin ligands. These antigens also play an important role in cancer, which additionally express dimeric Lewis antigens such as sialyl-dimeric Lewis x(121). Regulation of these structures is complex and involves coordinated synthesis and activity of multiple enzymes and careful regulation at both the genetic/transcriptional level as well as in the secretory apparatus.

Sialic Acids

Sialic acids (SA) are an important component of O-glycans as well as N-glycans and glycolipids. Over 50 different sialic acids have been observed. Neu5Ac is the most common in humans, while Neu5Gc is common in lower mammals but normally absent in humans due to a mutation in the synthase. Interestingly, Neu5Gc is observed in pathologic conditions in humans, such as cancer, presumably due to dietary uptake(122, 123). Sialic acid can be acetylated, methylated, etc. and contributes to glycan binding epitopes, such as sialyl Lewis antigens.

Approximately 20 sialyltransferases mediate transfer of CMP-SA to glycoconjugates in mammals. These transfer sialic acids in α2-3 and α2-6 linkage to Gal, α2-6 linkage to GalNAc, and α2-8 to other sialic acids as observed in polysialic acid on N-glycans in N-CAM and on O-glycans in neuropilin-2. Four families of sialyltransferases catalyze these reactions including ST3Gal-I-VI, ST6Gal-I,II, ST6GalNAc-I-VI, and STSia-I-VI(124).

Within a sialyltransferase family, enzymes are further divided based on properties of the acceptor, in particular the glycan structure, e.g. type I (Galβ1-3GlcNAc) or type II chains (Galβ1-4GlcNAc), and the class of glycoconjugate, e.g. O-glycans, N-glycans, and glycolipids.

Many sialyltransferases transfer sialic acids to O-glycans. These include ST3Gal-I,III-V; ST6Gal-II; and ST6GalNAc-I-IV. ST3Gal-I makes Siaα2-3 core 1, while ST3Gal-III-V forms Siaα2-3Galβ1-3/4GlcNAc- (i.e. on type 1 or type 2 chains), which is found on extended core 2 chains. ST6Gal-I-II makes Siaα2-6Galβ1-4GlcNAc- (i.e. on type 2 chains), which is also found on extended core 2 chains. ST6GalNAc-I-IV all modify GalNAc of core 1 to form sialyl or disialyl T but differ in substrate preference based on whether core 1 is unsubstituted or whether it is monosialylated by ST3Gal-I. ST6GalNAc-I modifies unsubstituted acceptors and is the only ST6GalNAc that synthesizes STn, ST6GalNAc-II modifies unsubstituted or monosialylated acceptors, and ST6GalNAc-III-IV modify monosialylated acceptors. ST6GalNAc-IV also modifies glycolipids.

Monosaccharide modifications

Monosaccharides in glycoconjugates can be sulfated, phosphorylated, acyl/deacylated, epimerized, and methylated by postglycosylational modifications(125, 126). All but methylation are observed in humans. These modifications increase the structural and functional diversity of glycans and are altered in disease. Sulfation, as discussed

previously, generates sulfo Lewis and sulfo sialyl Lewis antigens, as well as GAGs. Acylation/deacylation utilizes common substituents such as acetyl or less common substituents, such as ferrulate and lactyl, through the action of human enzymes and sometimes microbial/parasitic enzymes. In colon cancer, a loss of sialic acid O-acetylation is observed. Normally ~50% of colonic mucin sialic acids are O-acetylated. Glycan phosphorylation is best characterized for the lysosomal sorting signal Man-6-P on N-glycans and O-Mannose glycans on α–dystroglycan (α-DG) as well as some microbial/parasitic organisms, but probably modulate other glycan biology. Epimerization alters the stereochemistry of monosaccharides and converts glucuronic acid to iduronic acid in the synthesis of GAGs. Post-glycosylational modifications are not well studied and often difficult to assay, but observations to date indicate they are diverse and important in multiple classes of glycans in humans and other animals and organisms.

III. Altered O-glycan structures observed in cancer

Overview

Alterations in O-glycan structures were arguably first observed in the 1940s and 1950s with expression of immature blood group structures in gastric carcinoma(127). Later, purification and characterization of specificities of various lectins as well as generation of monoclonal antibodies led to the identification of truncated and shortened O-glycans, such as Tn, STn, and T antigens, as well as identification and confirmation of altered terminal O-glycan structures, such as Lewis blood group and AB/H structures(128-133).

Recent studies investigating glycan binding specificities of many of these reagents through glycan microarrays have allowed improved interpretation of these early studies. Further evidence for altered O-glycans in cancer derive from immunologic studies evaluating autoantibody signatures and cellular immunity through glycopeptide arrays and delayed type hypersensitivity reactions (DTHR), while advances in physical methods, such as mass spectrometry, gas chromatography, and NMR have revealed structural features of these altered O-glycans. The initial discovery of altered O-glycans in cancer led researchers to investigate the clinical applicability of these discoveries - including sensitivity and specificities, tissue localization, clinical stage of expression, and whether these structures correlate with survival/progression and/or contribute to the tumor phenotype. This section will focus on structural alterations observed in primary human tumors, including histology and mechanistic insights into structural alterations as well as potential contributions to the tumor phenotype.

a) Methods to identify altered O-glycosylation in cancer

There are three general approaches to identify alterations in glycosylation, including O-glycosylation, in human tumors. The first method uses affinity probes, the next method uses physical methods, such as mass spectrometry, and the third method involves indirect immunologic approaches, evaluating immunologic responses to altered glycosylation. There are advantages and limitations to all of these approaches.

Antibodies against carbohydrates have been generated through a variety of approaches. With the advent of monoclonal antibodies, many researchers began immunizing mice against tumor and tumor cell extracts and screening against tumor cells and/or histologic specimens. Although the initial goal was to develop anti-tumor antibodies and not necessarily anti-carbohydrate antibodies, many of the antibodies generated were against glycan or glycopeptide epitopes including O-glycoproteins or O-glycans, such as CA15-3 (MUC1), CA-125 (MUC16), B72.3/Tag-72 (STn), and CA19-9 (sLe^a)(128, 134-136). More recently, investigators have taken targeted approaches to generate anti-O-glycan tumor antibodies, immunizing mice with tumor cells, microorganisms, or glycoproteins containing defined tumor glycans and screening antibodies against histologic specimens, tumor cell lines, defined glycoproteins, and/or glycopeptide microarrays. Not all glycan structures are equally immunogenic, biasing the production of antibodies. Glycan determinates recognized by antibodies and other GBPs contain 2–6 monosaccharides, limiting generation of monoclonal antibodies against single monosaccharides, such as GalNAc(137). However, antibodies against monosaccharide clusters or monosaccharide peptide epitopes have been developed, for example to the Tn antigen(138).

Lectins have also been used as affinity reagents in conjunction with monoclonal antibodies. Lectins form multimeric units, facilitating detection of low affinity interactions through enhanced avidity. Altered binding of plant lectins to tumor cells, such as of wheat germ agglutinin, provided some of the earliest evidence that glycans are altered in cancer(4, 5). Lectins differ from monoclonal antibodies in that they tend to be more poly-reactive, recognizing many related glycan structures with a gradient of

affinities, in contrast to monoclonal antibodies which tend to be more specific. Despite these limitations for lectins, careful use of GBP inhibitors and multiple GBPs can provide important structural information. In addition to classic use of lectins and monoclonal antibodies, other affinity reagents may provide additional information, including VLR-Fcs generated from lampreys, which use unique binding domains to generate glycan reactivity(139, 140).

b) Truncated O-glycans

Overview

Tn, STn, and T antigens are biosynthetically related carbohydrate structures that are highly expressed in carcinomas, but not present in normal tissues or cells (**Table 1.1, 1.3**) (**Figure 1.3**). Various reagents have been used to assess these structures, including antibodies to all three structures and lectins that bind Tn and T antigens. Some of these reagents have been validated across platforms, including defined tissues for immunohistochemistry, hapten inhibition, column chromatography, and glycopeptide and glycan microarrays, whereas other reagents are less well-defined. For example BaGs6 and TKH2 monoclonal antibodies (mAbs), recognizing Tn and STn, are highly specific, whereas lectins, such as HPA, and other mAbs, cross-react with normal structure, such as blood group A, or are poorly characterized(8, 141).

Table 1.1. Truncated O-glycans in cancer

Antigen	Tissue (unless serum is noted)	% Tumor positive	% Normal positive	Stage of expression (pre-malig, primary, met)	Notes	Citation
Breast						
Tn	Breast	14/15 (93%)	1/5	In situ, grossly malig	By lysate absorption to anti-serum, lectin	144
Tn	Breast	48/50 (96%)		Primary, metastatic	Adsorption	145
STn	Breast	13/21 (62%)		Metastatic	B72.3	128
STn	Breast	19/41 (46%)	2/13 (15%); 2 positives benign, 9 total benign, rest normal non-cancer	Primary	B72.3	128
STn	Breast	37/44 (84%)	6/20 (30%); benign lesions, weak staining in positives			172
T	Breast	15/15 (100%)	2/5	In situ, grossly malig	By lysate absorption to anti-serum, lectin	144
T	Breast	Present, unspecified	Present, unspecified	Differentiated, undifferentiated	PNA binding tissue section	193
T	Breast	47/52 (90%)	2/21 (10%); 2 positives pre-malig		Adsorption	145
Tn	Colorectal	72% (n = 29) cancer; 35% (n = 25) Transitional mucosa	0% (n = 22)		ETn1.01; all tumor grades positive	146
Tn	Colorectal	72% (n = 29) cancer; 67% (n = 25) Transitional mucosa	14% (n = 22)		VVA; all tumor grades positive	146
Tn	Colorectal	81% (n = 29) cancer; 61% (n = 25) Transitional mucosa	14% (n = 22)		CU-1; all tumor grades positive	146
Tn	Colorectal polyps	103/103 (100%)		79 adenomatous; 24 hyperplastic	VVA	147
Tn	Colorectal	44/52 (85%) 5/20 (25%) 21/22 (95%)	0/17 (0%)	Primary Transitional Liver met	BaGS-6 Tec-02	148
STn	Colorectal	44/52 (85%)	0/17 (0%)	Primary	TKH-2	148

		11/20 (55%) 21/22 (95%)		Transitional Liver met	B72.3	
STn	Colorectal	40/60 (67%) 29/46 (63%)	7/46 (15%)	Carcinomas Transitional mucosa	HBSTn-1	173
STn	Colorectal	4/4			B72.3	128
STn	Colorectal	51/54 (94%)	5/27 (19%)		Of benign, highest is 20% of cells reactive in Crohn's sample	172
STn	Colorectal	96% (n = 29) tumor; 38% (n = 25) Transitional mucosa	0% (n = 22)		TKH2; all tumor grades positive	146
STn	Colorectal	93% (n = 29); 38% (n = 25) Transitional Mucosa	0% (n = 22)		B72.3; all tumor grades positive	146
STn	Colorectal	87.5% (112/128)			TKH2; STn(+) worse prognosis	174
STn	Colon and Serum	27.8%			RIA; 45U/ml cutoff	181
STn	Colorectal polyps	29% (7/24)		Hyperplastic	TKH2	147
STn	Colorectal polyps	56%		Adenomatous	TKH2	147
T	Colorectal	71% (n = 29) Tumor; 47% Transitional mucosa (n = 25)	0% (n = 22)		AH9-16; well/moderately differentiated positive; reduced in poorly differentiated	146
T	Colorectal	8 ng/ug (n = 11)	3.3 ng/ug (n = 5, UC) 1.5 ng/ug (n = 9 normal)		Units: ng TF/ug protein; Use O-glycanase to release and analyze by HPAEC	150
T	Colorectal	31/52 (60%) 0/20 (0%) 20/22 (91%)	0/17 (0%)	Primary Transitional Liver met	TF-α/β A78-G/A7 PNA HB-T1 TF-α HH8 BM22 TF- β A-68-B/A11	148
STn, T, Tn					15/24 Tn/STn/T(+)	146

					6/24 Tn/STn(+) 2/24 STn/T(+) 1/24 Tn/STn/T(-)	
Gastric						
Tn	Gastric	80/87 (91.9%)	0/58 (0%); intracellular staining noted in all			149
Tn	Gastric	96/163 (59%)			HPA	150
STn	Gastric	3/4				172
STn	Gastric and Serum	28.1%			RIA; 45U/ml cutoff	181
STn	Gastric	69/87 (19.3%)	0/58 (0%); 8 positive for intracellular staining		Discrepancy in manuscript, frequency not match % tumors positive	149
STn	Gastric	53/85 (62.3%)			TKH2	175
STn	Gastric	21/31 (68%)			TKH2; correlate with outcome	176
STn	Gastric	186/340 (54.7%)			TKH2; International study: Japan, Brazil, USA, Chile; cancer beyond stage I (advanced) express more frequently than stage I	177
T	Gastric	18/87 (20.7%)	0/58 (0%)			149
Pancreas						
Tn	Pancreas	36/36 (100%) IDC 5/5 IPT	0/45 (0%)		CU-1, 91S8 (similar staining reported) 100% Tn/STn(+); Localization: 100% of cyto v. 47% luminal surface, 31% luminal contents positive	151
Tn	Pancreas	3/6 (50%); adenoma 2/7 (29%); hyperplastic duct		Benign	All Tn+STn- (adenoma) Tn/STn(+) (hyperplastic ducts)	151
STn	Pancreas	3/3				172
STn	Pancreas and Serum	40.0%			RIA; 45U/ml cutoff	181

STn	Pancreas	36/36 (100%); IDC 5/5; IPT	0/45 (0%)		TKH2	151
STn	Pancreas	3/6 (50%); adenoma 2/7 (29%); hyperplastic duct		Benign	All Tn-/STn+ (adenoma) Tn/STn(+) (hyperplastic duct)	151
STn	Pancreas	77% (n = 64)	2% (n = 58)	Infiltrating pancreatic ductal adenocarcinoma	TKH2, increase in advanced cancer, reduced in PanINs (PanIN 3 = 67% n = 9 but <10 % PanIN < 3)	178
T	Pancreas	29/36 (81%)	36% PDs 71% ACs 53% ICs PD panc duct AC Acinar cells IC Islet cells n total = 45		PNA	151
Bladder						
Tn	Bladder	27/34 (77%)	0/10 (0%)	Primary	BaGs2	152
STn	Bladder	1/34 (3%)	1/10 (10%)	Primary	TKH2	152
Respiratory						
Tn	Lung	84/93 (90%)			HPA	153
STn	Lung	26/27 (96%)				172
T	Respiratory	4/5			Adsorption	145
Ovarian						
STn	Ovary	40/40 (100%)			B72.3	172
STn	Ovary	61/82 (74%)				180
Endometrial						
STn	Endometrial	36/43 (84%)	13/32; proliferative phase & menopausal negative; secretory phase, especially late were positive)		TKH2	179
Cervical						
Tn	Uterine cervix	50/111 (45%)			VVA	154
Tn	Uterine cervix	24/29 (82.8%) 17/29 (58.5%)		Met Primary	VVA; more Tn in met than primary, but no difference in T antigen in primary v. met	155
Salivary gland						
Tn	Salivary gland	12/13 (92%) 7/9 (78%) 7/18 (39%)		Mucoepidermoid carcinoma Adenocarcinoma	Tn: 1E3, HB-Tn-1; similar results for both; Tn &	156

		6/14 (43%) 1/6 (17%)		Carcinoma in pleomorphic adenoma Adenoid cystic carcinoma Acinic cell carcinoma	STn: 1C12; similar results to Tn	
STn	Salivary gland	12/13 (92%) 7/9 (78%) 7/18 (39%) 6/14 (43%) 1/6 (17%)		Mucoepidermoid carcinoma Adenocarcinoma Carcinoma in pleomorphic adenoma Adenoid cystic carcinoma Acinic cell carcinoma	STn: TKH2, HBSTn1; similar results for both; Tn & STn: 1C12; similar results to STn	156
GI unspecified						
STn	Digestive (unspecified) and Serum		4.1%		RIA; 45U/ml cutoff	181
T	GI-tract	5/5			Adsorption	145
Esophagus						
STn	Esophagus and Serum	0%			RIA; 45U/ml cutoff	181
Liver						
STn	Liver and Serum	7.1%			RIA; 45U/ml cutoff	181
Biliary tract						
STn	Bile/Pancreas (Serum)	8/15 (53%) 5/9 (56%)	2/14 (14%)	Bile duct Pancreas	<45U/ml (normal); ≥45U/ml (high)	182
STn	Biliary Tract and Serum	25.0%			RIA; 45U/ml cutoff	181
Melanoma						
Tn	Melanoma	0/2			Adsorption	145
STn	Melanoma	0/2				172
T	Melanoma	0/4			Adsorption	145
Non-epithelial solid tumor						
STn	Osteogenic sarcoma	0/1				172
STn	Glioblastoma multiforme	0/1				172
Blood cells						
Tn	Blood cells (bone marrow aspirates)	5/725 (<1%)	0/35 (0%)		FBT3	157
STn	Lymphoma	0/4				172
STn	Leukemia	0/1				172
STn	Thymoma	0/1				172
Various						
Tn	Various	7/8 (88%)			Adsorption	145
STn	Various normal, from non-cancer		0/25 (0%)		B72.3	128
STn	33 normal organ		minimal			172

	and tissue type					
T	Various non-breast benign		0/7 (0%)		Adsorption	145
T	Various non-breast, healthy		0/11 (0%)		Adsorption	145

Frequency of positive staining reported as number positive samples out of all samples unless indicated otherwise; % reported in parentheses for n>5; literature search performed for tumor antigen; abbreviations – IDC: invasive ductal carcinoma; IPT: intraductal papillary tumor; malig: malignant; met: metastasis; RIA: radioimmunoassay, PNA: Peanut agglutinin; VVA: Vicia Villosa agglutinin; HPAEC: high performance anion exchange chromatography; HPA: Helix Pomatia Agglutinin; PanIN: pancreatic intraepithelial lesion; cyto: cytosoplasm, (+): positive, (-): negative.

Table 1.3. Frequency of altered Tn, STn, T, SLe^a, and SLe^x in cancer

Tissue	**Tn**	**STn**	**T**	**SLe^a**	**SLe^x**
Breast	> 90 (1/5)	60 – 85 (< 30)	90 – 100 (10 – 40)	5	25
Colon	70 – 100 (0 – 15)	65 – 100 (0 – 20)	60 – 90 (0)	55 – 80 (0)	40 - 90
Gastric	60 – 90 (0)	55 – 75 (0)	20 (0)	80 – 90 (37)	> 90
Pancreas	100 (0)	80 – 100 (< 2)	80	55 (70)	> 90
Bladder	80 (0)	3 (10)			
Respiratory	90	95	4/5		
Lung				40	60

Esophagus				0	50
Cervical	45 - 80				
Ovarian		75 - 100			50
Endometrial		84			
Salivary	20 - 90	20 - 90			
Liver				10	
Melanoma	0/2	0/2	0/4		
Blood	< 1	0/6			
Mesothelioma				10	

All numbers are % positive samples out of all samples analyzed unless reported as fraction (for small n). Numbers in parentheses are normal tissues. Data adapted from Table 1 and 2 and presented as range. Literature search performed for antigens and left blank if not found using search criteria.

As discussed previously, Tn is synthesized by a family of ppGalNAcTs and normally extended to form core O-glycan structures 1–4. Alterations in biosynthetic machinery and other factors may contribute to expression of the truncated O-glycans, Tn, STn, and T. Further, these structures have been shown to contribute to the tumor phenotype in various model systems.

i. Tn antigen

Background

Tn antigen is expressed on a majority of carcinomas arising from every tissue evaluated to date and not expressed on normal adult tissues. The Tn antigen was first identified by Dausset in 1959 on red blood cells from a patient with a rare autoimmune hemolytic anemia and polyagglutinability, later shown to result from anti-Tn antibodies in serum(142). Tn, or T nouvelle, was named in reference to T antigen, which was

discovered earlier through a similar RBC polyagglutinability. However, in contrast to Tn, T is present on all individuals after treatment with neuraminidase. Dahr described GalNAc-Ser/Thr as the Tn antigen and Springer and others identified the biosynthetic relationship between Tn and T(19, 143). Tn was first observed on tumor cells in 1969 but established as a pan-carcinoma antigen in the 1970s – 80s through the work of Springer(132). Many groups have since verified these early observations. Tn is highly expressed on carcinomas, but less common on blood cancers(8). Interestingly, many of the tissues that express Tn also express other truncated O-glycans, such as STn, and T.

Histology

Tn antigen is a pan-carcinoma antigen. It is expressed in ≥90% of carcinomas of breast, pancreas, and lung, ≥60% of carcinomas of colon, stomach, and bladder, 45–80% of cervical carcinomas, 20–90% of salivary carcinomas, depending on the tumor type, and <1% of primary melanomas and blood cancers(144-157). Tn is expressed in <20% of normal tissues, with lower percentages observed with more specific reagents(144, 146, 148, 149, 151, 152, 157). Furthermore, Tn is expressed highly on carcinoma cell surfaces and in luminal content, in addition to being upregulated in the cytoplasm of cancers. Expression of Tn on the cell surface and in luminal content is extremely specific for carcinomas as normal tissues do not express Tn in these locations(151). Tn is a normal biosynthetic precursor but not a normal terminal product. Therefore Tn is occasionally observed in the secretory apparatus but not found on cell surface or secreted proteins, except in cancer.

Tn antigen is expressed early in cancer development and its expression correlates with clinical progression. In the colon, Tn is found on 25–70% of transitional tissues and also in pre-malignant lesions and hyperplastic polyps (146-148, 158). In carcinogen induced rodent models of colorectal and breast cancers, Tn is observed in precursor lesions(159, 160). The ratio of Tn to T increases during cancer progression, largely due to increased Tn, and accordingly, Tn predicts cancer invasion more readily than T(161). In comparing Tn-positive and negative tumors, Tn expression correlates with poor survival in carcinomas of the cervix, lung, colon, and stomach(150, 153, 154, 162).

Mechanisms for expression

Tn is highly expressed in carcinomas, expressed early in tumorigenesis, and correlates with disease progression and survival. However, the mechanism for Tn expression is less clear. Approximately 10 mechanisms have been proposed for expression of Tn antigen in cancer and some of these have been observed in primary tumors. The proposed mechanisms include genetic/transcriptional alterations in proteins required for extension of the Tn antigen, alterations – either genetic or cell biologic – of the enzymes that synthesize the Tn antigen, changes in expression of nucleotide-sugars or transporters, and changes in vesicular transport, retention, and structure of the Golgi/ER as well as alterations in glycosyltransferase oligomerization(8).

Tn is normally extended by the T-synthase or C3GnT to form core 1 or core 3 structures. Loss of T-synthase and C3GnT activity has been observed in tumor cells and tissue(86, 111, 113). Defects in Cosmc, rather than T-synthase, appear to result in loss of T-synthase activity. Cosmc mutations, LOH, and epigenetic silencing have been observed in cancer cell lines and mutations have been observed in cervical cancer specimens(86, 163). Recently, methylation of the *Cosmc* promoter was observed in ~ 2/5 of a series of pancreatic adenocarcinoma samples, correlating with Tn expression and loss of T-synthase protein(164).

Various mechanisms that result in altered Golgi structure or altered glycosyltransferase localization within in the Golgi have also been suggested to result in Tn expression. These include growth factor induced relocalization of ppGalNAcTs to the ER, relocalization of ppGalNAcTs to the trans Golgi, alterations in proteins required for glycosyltransferase retention, and altered Golgi pH(165-167). However, unlike loss of T-synthase and C3GnT activities, the potential roles of altered Golgi structure and glycosyltransferase localization in Tn expression have not been evaluated in primary tumors so their contributions to Tn expression are unknown.

ii. Sialyl-Tn (STn)

Background

STn, or Neu5Acα2,6Tn, is a pan-carcinoma antigen that is expressed early in tumorigenesis and correlates with disease expression. STn is often co-expressed with Tn and, similar to Tn, not expressed on normal cells or tissues (**Figure 1.3**). STn was first identified along with Tn in Tn syndrome but discovered in tumors with generation of a monoclonal antibody, B72.3, generated from immunization with cell membrane fractions from metastatic breast cancer(168). B72.3 reacts with a sialidase sensitive epitope on STn-containing ovine submaxillary mucin (OSM) and binding to OSM is blocked by STn-Ser(169, 170). Other STn antibodies, such as MLS102 and TKH2, were also generated and have been used to evaluate STn in tumors(171).

Histology

Most carcinomas express STn, including >60% of carcinomas of the breast, colon, stomach, pancreas, lung, ovaries, and endometrium, 20–90% of salivary carcinomas, and, in contrast to Tn, few bladder carcinomas(128, 146-149, 151, 152, 156, 172-182). STn is also expressed early in tumorigenesis and correlates with poor survival. Aberrant crypt foci (the earliest sign of cellular atypia) and ulcerative colitis lesions that will progress to colorectal cancer express STn(64, 183, 184). Similar to Tn, STn is also expressed in precursor lesions in carcinogen induced rodent models of breast and colorectal cancer, and STn has been shown to correlate with histologic and clinical cancer progression, including poor survival in colorectal, gastric, and ovarian cancers(159, 160, 174-177, 183-185). In contrast to Tn, STn is not a normal biosynthetic precursor. Therefore, both intracellular and cell surface expression are necessarily pathologic.

Mechanisms for expression

STn is often co-expressed with Tn and therefore mechanisms that result in Tn expression likely apply to STn, including alteration in Cosmc/T-synthase, defects in C3GnT, and alterations in Golgi structure and glycosyltransferase dynamics, as discussed previously. Tn-independent mechanisms for STn expression have been proposed, including ST6GalNAc-I upregulation and de-esterification of acetyl STn to form STn. Engineered overexpression of ST6GalNAc-I in cell lines results in STn expression(186, 187). While some tumors exhibit elevated ST6GalNAc-I transcript, protein, or activity correlating with STn expression, other tumors, such as colorectal carcinoma, have a reduction in ST6GalNAc-I activity despite robust STn expression, and tissues expressing STn almost always co-express Tn, which could not be explained by ST6GalNAc-I(111, 173, 188). Another Tn-independent mechanism for STn expression derives from studies using alkaline, de-esterifying, conditions, which results in STn expression in normal colon (where it is usually absent) but not other normal tissues such as pancreas(171, 189). This suggests that in some tissues acetyl STn may normally conceal STn antigen and therefore STn may arise in tumors from increased esterase activity. However, this assumes that all STn antibodies evaluated in normal tissue to date are blocked by STn acetylation, which is unlikely but untested.

iii. T antigen

Background

T antigen is highly expressed on tumors, but reagents used to define T are more variable and less defined than those used for Tn and STn, resulting in more cross-reactivity with normal tissues and non-T structures. Parts of the CNS and germinal center B cells normally express T antigen(161, 190). Structurally, T consists of Galβ1-3GalNAcα1-Ser/Thr, which forms after transfer of Gal to Tn by the T-synthase (**Figure 1.3**). Historically, T was sometimes considered Galβ1-3GalNAc in α or β linkage to a glycoprotein or -lipid, in contrast to the current definition of α linkage to a glycoprotein. Normally, T is extended by addition of Neu5Ac and/or GlcNAc to form mono or disialyl T and core 2-based structures. The T antigen was first discovered by Hübener, Thomsen, and Friedenreich in the 1920s and 30s when studying blood agglutination(191). They found that a microbial contaminant, later attributed to neuraminidase, results in cold agglutination of RBCs when mixed with serum due to normal anti-T antibodies. Georg Springer first identified T as a pan-carcinoma antigen, along with his studies of Tn in the 1970s(143-145, 192).

Histology

T antigen is a pan-carcinoma antigen that is expressed in greater than 60% of tumors of breast, colon, pancreas, and lung, and ~20% of gastric tumors. Normal tissues express 0–40% of T, likely due to cross-reactivity of anti-T reagents, some of which react with α and β linked Galβ1-3GalNAc on glycoprotein or –lipids(144-146, 148, 149, 151, 193).

Nonetheless, chemical and enzymatic studies have confirmed that T is a genuine carcinoma antigen(194). In contrast to Tn and STn, T expression is less helpful to prognosticate tumors(161). Early studies from Springer using DTHRs suggested that cellular immune responses to T are extremely sensitive and specific for carcinoma(143, 161).

Mechanisms for expression

The expression of T antigen could occur through a variety of mechanisms. Colon tumors upregulate UDP-Gal transporters, resulting in elevated T antigen, as well as $SLe^{a/x}$, in cancer cells(195). Reduction of sulfotransferase activity in colon tumors has been proposed to result in conversion of endogenous sulfo-T in colon to T antigen(196). Alterations in Golgi pH increase T expression in cell lines, presumably due to changes in Golgi structure and glycosyltransferase localization(166). In breast cancer, decrease in C2GnT results in a shift from core 2 to core 1-based structures, which along with other mechanisms may enhance T expression(197). However, the importance of any of these mechanisms in T expression in primary tumors is currently unclear.

iv. Comparing Tn, STn, and T expression and function in tumor biology

Expression

The truncated O-glycans Tn, STn, and T are pan-carcinoma antigens and many tumors co-express these structures. Over half of colorectal cancers express all three antigens while most pancreatic cancers express both Tn and STn(146, 151). Interestingly, benign lesions sometimes express one of these antigens but rarely co-express multiple antigens(151). Some tumors, such as breast, exhibit a shift from core 2 to core 1-based structures including elevation of normal mono and di-sialylated T antigens(197). Hence, Tn, STn, T, and normal core 1-based structures can all be expressed on the same tumor. This probably reflects heterogeneity across some tumors, in which different cells express different antigens, as well as co-expression of these antigens on individual cells and individual proteins within a tumor. Whether one of these situations predominates is unclear but may highlight which mechanism(s) drive expression of truncated O-glycans.

Function

Tn, STn, and T are highly expressed in tumors and Tn and STn expression correlates with disease progression, suggesting that these antigens may contribute to the tumor phenotype. Deletion of *C3GnT* in a mouse results in expression of Tn and STn in the GI-tract and increased susceptibility to chemically induced colorectal cancer(27). Further, deletion of *Cosmc* in pancreatic cancer cells or an organotypic tissue model results in features of cellular transformation and tumorigenesis *in vitro* as well as increased tumor growth in xenotransplant studies(164). Alterations in cell adhesion and oncogenic signaling were also observed in these engineered cells, consistent with the behavioral alterations. Expression of Tn or STn may be immunomodulatory. STn facilitates

resistance to NK cell killing, and Tn interacts with MGL on dendritic cells and inhibits migration of immature APCs(198, 199). T antigen interacts with galectin-3, facilitating tumor cell interaction with endothelia and platelets(200). Many additional mechanisms likely contribute to the participation of truncated O-glycans in tumor biology, given that ~85% of proteins entering the secretory pathway are O-glycoproteins(14). Identifying these mechanisms may provide new therapeutic strategies.

c) Altered terminal and extended structures

Overview

Cancers exhibit alterations in terminal glycans, including Lewis antigens, blood group structures, as well as recently identified terminal α-GlcNAc on core 2 (**Table 1.2, 1.3**). Many tumors overexpress sialyl Lewis$^{a/x}$ and delete, overexpress, or ectopically express blood group structures, while gastric carcinomas delete terminal α-GlcNAc on core 2. Many of these changes correlate with survival; however, unlike truncated O-glycans and terminal α-GlcNAc, alterations in Lewis and blood group antigens are observed on N-glycoproteins and glycolipids in addition to O-glycoproteins. Determining the glycan carrier is not always possible, but a few examples highlight the importance of O-glycans as platforms for these alterations. O-glycans are major carriers of sialylated Lewis antigens and ABH structures in some tumors and of sialylated Lewis antigens in serum, and expression of these structures on O-glycans correlates with poor prognosis(201-204).

Further, expression of SLe^x on core 2 O-glycans facilitates interaction with E-selectin, which is important for tumor metastasis(205).

Table 1.2. Altered terminal O-glycans in cancer

Antigen	Tissue	Change	% Tumor positive	% Normal positive	Outcome	Notes	Citation
Colorectal							
SLe^a	Colorectal		12/21 (57%)			CA19-9, CA52a; RIA	129
SLe^a	Colorectal		40/68 (59%)	0/15 (0%)		CA19-9	215
SLe^a	Colorectal		1. Primary: 233/309 (75.4%) 2. Regional LN: 99/126 (78.5%)		1. Met to regional LN: Primary-SLe^a(+) = 48/86 (54.7%); Primary-SLe^a(-) 2. Recurrence: SLe^a(+) 36/147 (24.5%), SLe^a(-) 3/76 (3.9%), 5-yr survival: Primary-SLe^a(-) 93.0%, Primary-SLe^a(+), (++), (+++) = 74.8, 64.7, 71.0%	CA19-9; SLe^a(+) in primary predict met to regional LN	216
SLe^a	Colorectal		110/159 (69.2%)		Yes, 5-yr – DFS: + = 73%, - = 84.7%	CA19-9	217
Lewis	Colorectal		1. 67 (85.9%) 2. 53 (68%) 3. 57 (73.1%) 4. 51 (65.4%)	1. 1/42 (2.4%) 2. 39/42 (92.9%) 3. 32/42 (76.2%) 4. 24/42 (57.1%)	1. SLe^x 2. Le^a 3. SLe^a 4. Le^x	SLe^x predictor of non-polyploid growth type v. polyploid growth-type	218
SLe^x	Colorectal		76% (n = 17)	Weak staining		CSLEX1; IHC	221
SLe^x	Colorectal		1. 79% 2. 65%	1. 0% 2. 26%		1. FH6 (Sialyl-dimeric Le^x) 2. !B9	220
Sialyl-dimeric Le^x	Colorectal					FH6; increased in met	121
SLe^x	Colorectal		50/132 (37.9%)	Rare in normal or transitional mucosa	Yes; 5-yr survival: SLe^x(+) 58.3%, SLe^x(-) 93.0%	FH6; 6/8 liver mets had greater % cells FH6(+), 20/25 mets	222

						to LNs (80%) stronger FH6 staining; sLe^x correlate with depth of invasion, LN met, LN invasion, tumor stage; sLe^x(+) greater recurrence and recurrence to distant sights than sLe^x(-)	
SLe^x	Colorectal		58/159 (36.5%)		Yes, 5yr - DFS: SLe^x(+) = 55.6%, SLe^x(-) = 89%	FH6	217
Le^x	Colorectal		1. 82% 2. 65% 3. 88%	1. 74%, 63% 2. 11% 3. 5%		1. Short-chain mono fucosylated Le^x (SSEA-1, AH8-183) 2. Long chain Le^x (FH1) 3. Long chain, poly fucosylated Le^x, dimeric Le^x (FH4)	220
Incompat-ible BG-A or B	Colorectal		>50%				228
Deletion of BG structure	Colorectal						228
Precursor BG-H accumulation	Colorectal		80% (n = 25)				228
ABH	Colorectal		46/82 (56.1%) expressors	Proximal: 18/23	Yes, 5-yr survival: ABH “deletors”		229

				(78.3%) "deleted" Distal: 57/59 (96.6%) "deleted"	75.4%, ABH "expressors" 33.5%		
ABH	Colorectal		1. A: 28/78 (35.9%) 2. B: 6/78 (7.7%) 3. H: 29/78 (37.2%) No incompatible expression	1. A: 1/18 (5.6%) with A/AB 2. B: 0/23 with B/AB 3. H: 1/42 (2.4%)		A expression predictor of non-polyploid growth v. polyploid growth-type	218
Gastric							
SLe[a]	Gastric		4/5			CA19-9, CA52a; RIA	129
SLe[a]	Gastric		16/18 (89%)	7/19 (37%)		CA19-9	215
SLe[x]	Gastric		94% (n = 17)			CSLEX1	221
Pancreas							
SLe[a]	Pancreas		4/7 (57%)			CA19-9, CA52a; RIA	129
SLe[a]	Pancreas		19/22 (86%)	7/10 (70%)		CA19-9	215
SLe[x]	Pancreas		100% (n = 3)	Weak staining		CSLEX1	221
Esophagus							
SLe[a]	Esophagus		0/5			CA19-9, CA52a; RIA	129
SLe[x]	Esophagus		50% (n = 4)	Strong staining		CSLEX1	221
Liver							
SLe[a]	Liver		1/11 (9%)	7/11 (64%)		CA19-9	215
Gall bladder							
SLe[a]	Gall bladder		2/5	6/11 (54%)		CA19-9	215
Lung							
SLe[a]	Lung		28/66 (42%)			CA19-9	215
SLe[x]	Lung		63% (n = 16)			CSLEX1	221
BG-A	Lung (NSCLC)	Loss	43/71 with A or AB BG retain expression A in tumor; 28/71 (39%) lose expression A in tumor		Yes	Median survival – lose A: 15 months, retain A in tumor: 71 months; survival of B, O = same as retain A and loss of B or H not change survival	130
BG-A	Lung	loss	35/62 lose A (56%)		No		230

BG-A	Lung	loss			Yes, of A/AB BG patients, median survival: A loss (n = 36) 38 months v. A(+) (n = 54) 98 months		231
H/Le^{y}/Le^{b}	Lung		MIA15-5(+) (n = 91): 20.9% 5-yr survival v. Ag- (n = 58) 58.6% 5-yr survival		Yes	MIA15-5 Ab; when segregate by blood group A (i.e. A or AB) significant difference in survival but not for B or O	131
Breast							
SLe^{a}	Breast		1/18 (6%)			CA19-9	215
SLe^{x}	Breast		25% (n = 8)			CSLEX1	221
Mesothelioma							
SLe^{a}	Mesothelioma		1/12 (8%)			CA19-9	215
Ovary							
SLe^{x}	Ovary		50% (n = 6)			CSLEX1	221
Kidney							
SLe^{x}	Renal tubules			Strong staining		CSLEX1	221
Various epithelia							
Serum Ca19-9	Various (colon, gastric, breast, pancreas)					CEA more sensitive, except for pancreas	201
Myeloid							
ABH	Myeloid	loss	1. 16/25 (55%) of A, B, AB decreased A or B 2. 6/28 (21%) of O type lose H	1. No change (n = 127) 2. No change (n = 51)		Reason for reduced A, B: 8/29 (29%) primary loss of A or B, 5/29 (17%) indirect due to loss of H, 3/29 (10%) loss of A or B and H	232

Frequency of positive staining reported as number positive samples out of all samples unless indicated otherwise; % reported in parentheses for n>5; literature search performed for tumor antigen; abbreviations – BG: blood group, NSCLC: non-small cell lung cancer, LN: lymph node, yr: year, met: metastasis, DFS: disease free survival, RIA: radioimmunoassay

i. Terminal alpha-GlcNAc on Core 2

Gastric cancer is the fourth most common malignancy and the second leading cause of cancer deaths worldwide(206). *H. pylori* is a major risk factor for gastric cancer and interacts with surface but not gland mucins of the stomach(207-209). A unique glycan structure – α1-4GlcNAc terminating core 1 and core 2 branches of core 2 O-glycans – was identified on gland mucins, suggesting that this structure may prevent *H. pylori* colonization. Accordingly, terminal α1-4GlcNAc on a soluble protein inhibits *H. pylori* growth through disrupting the cell wall(210). Interestingly, deletion of *α1-4GlcNAcT* in a mouse results in spontaneous gastric adenocarcinoma, independent of *H. pylori*, along with pro-tumorigenic immune activation(211). Hence, α1-4GlcNAc on core 2 prevents gastric cancer through its antibiotic properties towards *H. pylori* and by preventing pro-tumorigenic inflammation in the stomach. In line with these mouse studies, loss of α1-4GlcNAc has been observed in patients with gastric carcinoma and Barrett's adenocarcinoma and found to correlate with worse survival in gastric cancer(212, 213).

ii. Lewis antigens

Background

Sialyl Lewis antigens consist of sialylated and fucosylated type 1, 2 chains (**Figure 1.2**), as previously described. Sialyl $Le^{a/x}$ ($SLe^{a/x}$) are upregulated in cancer; however, tumors express long chain and polyfucosylated forms, such as dimeric SLe^{x}, in contrast to short chain and monofucosylated forms observed in inflammation and normal tissues and cells.

Lewis antigens were first discovered in an RBC agglutination reaction; however, later observed as a synthetic product of internal epithelia and not RBCs, having been transferred to RBCs on glycolipids in serum(214). Generation of mAbs against tumor cells in the 1970s and 80s resulted in identification of tumor associated $SLe^{a/x}$ antigens.

SLe^a

SLe^a is a tumor associated carbohydrate antigen as first identified by Magnani and Koprowski using an antibody, CA19-9, generated from immunization with SW1116 cells(129, 134) (**Figure 1.2**). SLe^a is expressed by >50% of carcinomas of the colon, stomach, pancreas, ~40% of carcinomas of the lung, and ≤10% of carcinomas of esophagus, liver, breast and mesotheliomas. Normal colon does not express SLe^a in contrast to >30% of normal stomach, pancreas, and liver(129, 215-218). CA19-9 is mainly used as a serum test for pancreatic cancer. Although normal pancreas expresses CA19-9, pancreatic carcinomas secrete elevated levels of this antigen and expression correlates with poor survival(219). CA19-9 is not useful for all patients as 10–20% of the population is Lewis-negative and therefore cannot synthesize SLe^a.

SLe^x

SLe^x is a tumor marker and ligand for leukocyte adhesion. As described above, mono fucosylated short chain SLe^x is present in normal tissue and blood cells whereas long chain polyfucosylated SLe^x, such as sialyl-dimeric Le^x, is tumor specific. In colon for

example, antibodies against short chain SLe^x bind 60–80% of normal tissues whereas antibodies against long chain structures bind 5–10% of normal tissues(220). Expression of SLe^x in tumors largely resembles expression of SLe^a. SLe^x is expressed in >90% of tumors of stomach and pancreas, ≥40% of carcinomas of colon, esophagus, and ovaries and ~25% of breast carcinomas(121, 217, 218, 220-222). In colon, SLe^x expression correlates with poor survival(217, 222).

Functions of SLe^x and SLe^a in tumor biology

Sialyl Lewis antigens facilitate metastasis in model systems and patients. $SLe^{a/x}$ are selectin ligands, mediating tumor cell attachment to endothelia, platelets, and leukocytes(65, 223). Attachment to endothelia contributes to vessel invasion whereas attachment to platelets, and possibly leukocytes, contributes to survival in the vasculature. Cancer cells that express $SLe^{a/x}$ more readily metastasize in xenotransplants, and overexpression of E-selectin in the liver results in metastatic redirection of $SLe^{a/x}$(+) cells to the liver(66, 68, 224). In patients, expression of $SLe^{a/x}$ correlates with worse survival and using agents that block selectin expression or interactions with $SLe^{a/x}$ improves survival(216, 217, 219, 222, 225, 226).

Mechanisms for overexpression

Expression of $SLe^{a/x}$ arises from glycan truncation or overexpression(227). Some tissues contain 6-sulfated and 2-6 sialylated $SLe^{a/x}$. However, tumorigenesis results in reduced

sulfation and 2-6 sialylation of $SLe^{a/x}$ and increased expression of $SLe^{a/x}$(227). $SLe^{a/x}$ overexpression results from oncogene or environmental induction of biosynthetic enzymes, such as Cox-2 expression and hypoxia(227). Various glycosyltransferases can be affected, although upregulation of sialyltransferases appears to be a major mechanism for $SLe^{a/x}$ expression. Fut3, the major fucosyltransferase responsible for synthesizing SLe^{a}, is not usually upregulated in tumors(227). In addition to affecting glycosyltransferases, hypoxia induces UDP-Gal transporter overexpression, resulting in elevated levels of $SLe^{a/x}$ and T antigen(195).

iii. ABH structures

Background

Alterations in AB/H structures are observed in tumors with loss of blood groups in tissues normally containing structures (**Figure 1.2**), gain of blood groups in tissues normally deficient in structures, and occasionally mismatched expression of blood groups, e.g. expression of A in a BB/BO individuals. Some of these changes correlate with clinical progression. Landsteiner first described the ABO blood groups in the early 1900s, Watkins later characterized the structure and genetics of blood groups, and Masumune first observed alterations in ABO blood groups in tumors in the 1940s(127). AB/H structures are synthesized on type 1–4 chains, as previously discussed.

Histology

Colon and lung carcinomas exhibit different alterations in blood group structures due to different levels of normal expression. Colons (especially distal) are deficient in ABH while lungs express these structures. Accordingly, colon cancers exhibit ectopic and mismatched expression of blood groups structures, while lung carcinomas exhibit deletion of normal structures and expression of truncated structures, such as H in A individuals. A/B incompatibility is observed in 0 to >50% of colorectal carcinomas, depending on the study, and accumulation of blood group ABH structures is observed in 5–40% of tumors, depending on the structure, in contrast to 0–5% ABH expression in normal tissues(218, 228, 229). Lung carcinomas exhibit a loss of ABH structures, for example 40 – 55% of tumors exhibit a loss of A with a gain of H/Le^y/Le^b structures, due to precursor accumulation(130, 131, 230, 231) (**Figure 1.2**). In addition to carcinomas, myeloid cancers exhibit ~ 55% loss of A or B in A, B, A/B blood types and 21% loss of H in O blood types(232). Further, altered ABH structures correlate with disease progression. In particular, deletion of A in lung and ectopic expression of ABH in colon correlate with poor survival.

Mechanisms for altered expression

Loss of blood group structures appears to occur at the level of the A and B transferases. Loss of A/B transferase activities and epigenetic silencing at the A, B, H loci correlate with blood group deletion in carcinomas and myeloid cancers, respectively(233, 234). Mechanisms for incompatible blood group expression are less clear. A and B

glycosyltransferases differ by four amino acid residues, so mutation could result in conversion from A to B or vice versa. Alternatively, cross reactivity of anti-blood group reagents with other tumor associated antigens such as Forssman or Tn – both of which express terminal GalNAc – may result in apparent incompatible ABH expression. However, many anti-blood group reagents are specific so this is not likely.

Function of blood group structures in tumor biology

Correlation of survival with altered blood group structures suggests that these changes contribute to tumor biology. However, the mechanistic contributions of altered ABH to the tumor phenotype may differ depending on the tissue, since some tumors, such as colon, ectopically express ABH while others, such as lung, delete AB structures. In line with deleted AB structures in lung, A^-/H^+ cells display greater motility and proliferation *in vitro* as compared to A+/H- cells as well as changes in integrin function(235).

d) Genetic associations with glycogenes and cancer

Genomic studies including, GWAS, linkage analysis, transcriptomics, etc., have provided invaluable insights in biomedicine, including in cancer glycobiology, where it has been applied to investigate ppGalNAcTs in disease. ppGalNAcTs constitute a 20 enzyme family in mammals that expanded in recent evolutionary history. These enzymes exhibit partially overlapping specificities and expression, resulting in coordinated activity, which is poorly understood. Deletion of individual enzymes in higher organisms often results in

subtle defects that are lethal in simpler model systems, such as *Drosophila*, and some mammalian GalNAcTs are not able to compensate for their lower homologues bringing into question conservation of biological function across evolutionary history(77, 236-238). To address these limitations, researchers have performed genomic studies in humans to better understand roles of ppGalNAcTs and have observed a range of associations in cardiac development, obesity, dyslipidemia, calcium/phosphate imbalance, and cancer. SNPs in ppGalNAcT1 have been associated with ovarian cancer, mutations in ppGalNAcT12 have been observed in colorectal carcinoma, elevations in transcripts have been observed for ppGalNAcT6 and ppGalNAcT14 in breast and gastric cancers, and alterations in ppGalNAcT14 have been associated with apoptotic signaling in cancer cells, connecting enzyme level to cancer biology(239-246). However, it is currently unclear how or whether these observed alterations/associations in ppGalNAcTs contribute to altered glycan structure or tumor biology. More studies will be needed to confirm these initial observations and provide additional mechanistic insight.

e) Mucins

Mucins comprise a large gene family of ~20 molecules which can broadly be divided into secreted and membrane-associated mucins. Some mucins exhibit tissue-restricted expression while others are broadly expressed in a range of epithelia. Biochemically, mucins are large molecules characterized by tandem repeats rich in serine, threonine, and proline. These tandem repeats are highly O-glycosylated and mucins themselves exhibit >50% O-glycosylation by weight(20, 247). O-glycosylation is important in regulating

mucin biology as defects in O-glycosylation appear to result in decreased expression of mucins *in vivo* and enhanced bacterial degradation *in vitro*(27, 29, 248). Alterations in mucin structure, expression, and glycosylation have been observed in a range of carcinomas and these alterations have been hypothesized to be important in the tumor phenotype(20, 247). Changes in glycosylation include both underglycosylation and misglycosylation, for example expression of tumor glycans Tn, STn, and SLe^a. Mucins maintain barrier function, retain growth factors and signaling molecules, and engage in receptor-mediated signaling. Altered mucin structure is thought to result in alterations in all of these functions. Mucins play two opposing roles in cancer – as tumor suppressors and tumor promoters – probably depending on the tissue and mucin type. Deletion of *MUC2* in mice results in spontaneous tumorigenesis in intestine, colon, and rectum, while overexpression of *MUC1* in mouse mammary tissue results in spontaneous mammary carcinoma(249-251).

IV. Clinical applications

Overview

Discovery of tumor-associated O-glycans has led to development of clinical tools to target these antigens for detection and therapy. Application of tumor O-glycans for detection have been focused in three major areas: 1) tissue and serum biomarkers, 2) assessment of anti-tumor O-glycan immune responses, and 3) tumor O-glycan imaging. The second major application of tumor O-glycans is in cancer therapies. Tumor

associated O-glycans provide a unique epitope for targeted therapeutics. Cytotoxic antibodies and antibodies conjugated to radionuclides or toxins as well as therapeutic vaccines have been developed to target these TACAs. Further, efforts to interfere with the biologic contributions of tumor O-glycans to the cancer phenotype have been investigated, including interruption of sialyl Lewis – selectin interaction important in metastasis. Although many of the efforts focused on tumor detection and therapeutics are investigative, some have proven efficacious in patient treatment and/or management, including a few examples in which these technologies have become standard in patient management.

a) Cancer detection

Serum biomarkers

Serum biomarkers are important for diagnostics, prognostics, and following tumor response and burden. The first tests to identify tumor-associated O-glycans used immunohistochemistry to stain tumor samples. These include antibodies and lectins against truncated O-glycans Tn, STn, and T as well as altered terminal structures, such as SLe^x, SLe^a, and altered ABH antigens, as discussed above (**Figure 1.2**). Identification of tumor antigens in tissues led to the idea that these antigens may be present in serum, allowing non-invasive tumor detection.

Healthy epithelia are polarized with apical membranes often facing a luminal structure and basolateral membranes attached to basement membrane and extracellular matrix. This polarized architecture prevents secretion of large glycoproteins into serum. In contrast, transformation results in a loss of epithelial polarity and aberrant secretion of glycoproteins into serum. These include mucins and other aberrantly glycosylated glycoproteins which contain altered O-glycans.

Some of the earliest serum TACAs identified are CA15-3, CA125, CA19-9, and STn (CA72-4) (**Table 1.4**)(252). These are glycoproteins or glycans, and all but STn are widely used in clinic to follow tumor progression. CA15-3 and CA125 were cloned and discovered to recognize MUC1 and MUC16, which are highly O-glycosylated and aberrantly glycosylated in patient serum(135, 136, 253-255). CA19-9, in contrast, recognizes SLe^a(129) (**Figure 1.2**). CA15-3, CA125, and CA19-9 are most useful in the management of breast, ovarian, and pancreatic cancer, respectively. However, none of these antigens are sufficiently specific to be used for diagnostics. Further, expression of SLe^a requires a functional Lewis transferase. STn is thought to be an extremely specific marker; however, STn is expressed in most carcinomas (e.g. stomach, colon, pancreas, biliary tract, ovaries, and cervix), limiting its utility in identifying tissue of origin. Recent strategies have sought to increase specificity of mucin biomarkers as well as localization of STn by identifying glycoproteins in serum containing both of these antigens. Proof-of-principle experiments using sandwich ELISAs or arrays with STn and CA125 have established increased specificity of ovarian cancer detection as compared to CA125 alone(256, 257).

Table 1.4. Serum Biomarkers

Biomarker	Glycan/Glycoprotein Antigen	Tissue	Use
CA15-3	MUC1	Breast	Clinical, tumor burden
CA125	MUC16	Ovarian	Clinical, tumor burden
CA19-9	SLe[a]	Pancreas	Clinical, tumor burden
CA72-4	STn on Tag-72	Pan-carcinoma	Not in general use
STn/CA125	STn and CA125	Ovarian	In development

Imaging

Tumor imaging facilitates diagnosis, management, and planning as well as intra/post-operative assessment. Technologies such as MRI and X-ray evaluate features of the tissue, whereas scintigraphy, SPECT, and PET utilize radiopharmaceuticals, such as immunoradionuclides and metabolites, to assess tumor properties. Radionuclides have been optimized for various purposes and in general are transferable across probes, i.e antibodies or antibody fragments as well as metabolites such as FDG. Pharmacokinetic properties of antibodies have revealed that antibody fragments, such as scFvs, exhibit enhanced tissue penetration and reduced Fc-dependent clearance as compared to full-length antibodies. Nonetheless, radio and pharmacokinetic properties of an agent can be optimized once a selective probe is identified.

Anti-tumor O-glycan antibodies provide a unique opportunity for tumor imaging and as such have been developed for underglycosylated MUC1, Tn, STn, and STn/ST (**Table 1.5**). Reagents against all of these epitopes have been evaluated in rodents and humans, and a reagent targeting STn/ST has been shown to be useful in radioguided immunosurgery (RIGS)(258-291). Probes and radionuclides have been optimized for many of the anti-tumor O-glycans tested to date. Interestingly, one of the most researched antibodies for imaging is CC49, which reacts with ST > STn on TAG-72(269, 292, 293).

CC49 has been validated in RIGS and was initially developed as a second-generation antibody against TAG-72 (Tumor associated glycoprotein – 72)(290). CC49 binds TAG-72 and is inhibited by ST but not STn(292). Further, it binds ST > STn *in vitro*, with some reactivity against both antigens(293). Although ST is not generally considered to be a TACA, its presentation on TAG-72 appears to be tumor associated. In contrast to CC49, the antibody B72.3 reacts with STn on TAG-72(293). Based on the high expression of STn on tumors, this begs the question as to why there has been tremendous investment in CC49 as compared with B72.3 for imaging.

Table 1.5. Tumor imaging

Antigen	Antibody/ Agent	Probe/Source	Application	Stage of development	Tissue	Citation
MUC1						
MUC1	hPam4	111In/125/131I	SPECT, gamma planar imaging	Rodents, humans	PAC xenograft, in situ	266, 274 – 276, 279, 288

MUC1	TF10	125I	SPECT, gamma planar imaging	Rodents	PAC xenograft	261
MUC1	IMP-288 (TF10 – pre-target)	90Y	SPECT, gamma planar imaging	Rodents	PAC xenograft	264
MUC1	bsPAM4	125I	SPECT, gamma planar imaging	Rodents	PAC xenograft	258
MUC1	IMP-192 (bsPAM4-pre-target)	99mTc	SPECT, gamma planar imaging	Rodents	PAC xenograft	258
MUC1	IMP-166 (bsPAM4 – pre-target)	111In	SPECT, gamma planar imaging	Rodents	PAC xenograft	258
MUC1	PR81	99mTc	SPECT, gamma planar imaging	Rodents	Breast xenograft	262
Tn						
Tn	anti-Tn 2154F12A4 mAb	Qdot 800	Near-infrared fluorescence optical imaging	Rodents	Breast xenograft	263
Tn	MLS128	111In	SPECT, gamma planar imaging	Rodents	CRC xenograft	278
Tn	MLS128	125I/131I	SPECT, gamma planar imaging	Rodents	CRC xenograft	278
Tn	MLS128 (biotin-Ab, streptavidin-probe)	125I	SPECT, gamma planar imaging	Rodents	CRC xenograft	282
Tn	MLS128 (biotin-Ab, streptavidin, biotin-	111In	SPECT, gamma planar imaging	Rodents	CRC xenograft	281

	probe)					
STn						
	B72.3	131I	gamma planar imaging	Rodents, humans	CRC, CRC xenograft	267 – 269
ST > STn						
ST > STn	CC49 $(scFv)_2$	125I/131I	SPECT, gamma planar imaging	Rodents	CRC xenograft	283 – 285
ST > STn	CC49 Humanized ΔCH2	125I/131I	SPECT, gamma planar imaging	Rodents, humans		259, 272, 277, 280, 289, 290
ST > STn	CC49 Fab'	125I	SPECT, gamma planar imaging	Rodents		270, 271, 273
ST > STn	CC49 $sc(Fv)_2$	99mTc	SPECT, gamma planar imaging	Rodents	CRC xenograft	287
ST > STn	CC49 $sc(Fv)_2$	I125	SPECT, gamma planar imaging	Rodents	CRC xenograft	283
ST > STn	CC49 Humanized ΔCH2	111In	SPECT, gamma planar imaging	Rodents	CRC xenograft	260
ST > STn	CC49 $[sc(Fv)_2]_2$	125I/131I	SPECT, gamma planar imaging	Rodents	CRC xenograft	286
ST > STn	CC49 $(Fab')_2$	125I/131I	SPECT, gamma planar imaging	Rodents	CRC xenograft	270
ST > STn	CC49	124I	PET	Rodents	CRC xenograft	265
T						
T	JAA-F11	124I	PET	Rodents	Breast	291

					xenograf t	

Abbreviations – PAC: pancreatic adenocarcinoma, SPECT: single-photon emission computed tomography, PET: positron emission tomography, CRC: colorectal carcinoma

A major line of evidence that likely contributed to this shift in evaluation of CC49 highlights an important lesson in tumor glycobiology – that tumor cells are not always representative of tumor tissue glycosylation. Second generation CC series anti-TAG-72 antibodies were developed and tested in a colorectal cancer cell xenotransplant model(269). Many of the CC series, especially CC49, exhibited greater reactivity with the colorectal cancer cell *in vivo*. Importantly, the cell line evaluated was LS174T. LS174T comprise a mixed population of cells with a small fraction of cells exhibiting a loss of Cosmc and expression of Tn and STn antigens while the larger fraction exhibits intact O-glycosylation(86). This differs from real tumor biology in which STn, along with Tn, is highly expressed as a TACA in contrast to ST which is generally considered a normal O-glycan structure. Whether B72.3 or other STn/Tn immunoradionuclides would outperform CC49 in RIGS or other applications requires further investigation.

Assessing anti-O-glycan immune responses

The discovery that tumors elicit auto-antibodies has led to attempts to use these autoantibodies and other immunologic parameters for clinical evaluation and diagnosis(294, 295). Most of the early attempts to identify tumor auto-antibodies focused on protein recognitions, however, early evidence from Springer suggested that tumors

elicit glycan specific immune responses which result in changes in anti-glycan antibodies and induction of anti-glycan cellular immunity(143).

Recently, autoantibodies produced against tumor O-glycans have been evaluated in well-defined glycopeptide arrays and suggest that a portion of cancer patients with breast, ovarian, and prostate cancers generate IgG against MUC1-Tn, STn, T, and core 3 tumor glycopeptides(296). Antibodies appeared to be glycopeptide-specific, not reacting with peptide or glycan alone. Auto-antibodies against core 3 were surprising, as core 3 has not previously been shown to be expressed in many of the tissues evaluated and it is believed to be a normal O-glycan in the GI-tract. In contrast to IgG, anti-glycan and glycopeptide IgM antibodies were present in patients and controls, confirming Springer's early work that normal gut bacteria induce anti-glycan antibodies.

Auto-antibody arrays have also been used to predict clinical course. In a subset of breast cancer patients, elevated auto-antibodies against STn-MUC1 and core 3-MUC1 correlated with survival(297). This supports the intriguing possibility that normal immune responses against glycopeptides may contribute to cancer clearance.

b) Cancer therapeutics

Generation of mAbs against tumor antigens led to the idea that tumors express unique antigens that may be targeted by antibodies or therapeutic vaccines. These approached are currently being evaluated for a variety of tumor O-glycans and have resulted in a

variety of important lessons in the glycoimmunology and cancer glycoimmunology fields.

Passive immunotherapies

Passive cancer immunotherapies include unconjugated antibodies with intrinsic cytolytic activity (antibody dependent cell-mediated cytotoxicity, ADCC), antibodies conjugated to radionuclides (radioimmunotherapy), and antibodies conjugated toxins (immunotoxins). Reagents targeting tumor O-glycans have been developed for unconjugated antibodies and immunotoxins; however, the emerging success of radiolabeled antibodies for imaging highlight the potential for radioimmnotherapies.

Antibodies against Tn, STn, and mucins have been generated which exhibit ADCC-dependent and independent cytolytic activities, some of which have been evaluated *in vivo*(298-304). A major limitation of this approach, however, is that full-length antibodies with Fc are often required for cytolytic activities but exhibit poor tissue penetration. Further, not all highly specific antibodies exhibit cytolytic activities, although these can sometimes be engineered, for example by isotype swapping.

To address some of these challenges, researchers have developed immunotoxins in which antibody or antibody fragments have been conjugated to bacterial toxins. When these antibodies bind the cell surface, they are endocytosed and the toxins are activated within the cell, resulting in cell death. This is a powerful approach, but requires the antibodies to

recognize antigens that are selectively expressed on the tumor surface and not normal cells. Therefore, this approach would theoretically be promising for tumor specific O-glycans like Tn and STn but are less likely to be successful for tumor associated O-glycans, such as Lewis or blood group antigens. Nonetheless, an antibody specific for Le^y, which is elevated in some carcinomas, was generated and tested *in vivo*. Initial results were promising but dose-limiting toxicities, including gastritis, were observed and most likely resulted from reaction with low levels of antigen present in normal tissues(305-307).

Therapeutic vaccines

Although long controversial, landmark studies with melanoma and prostrate cancer established therapeutic vaccination as a viable approach for cancer treatment(308, 309). Further, early work from Springer suggested that vaccination with T/Tn containing substance (in his case RBC) could elicit anti-glycan immune responses and may improve patient survival(143). More recent work suggests that anti-cancer immune response may correlate with patient survival. Together, these lines of evidence suggest that therapeutic vaccination using tumor associated O-glycans may be able to generate anti-glycan and anti-tumor immunity.

In contrast to the traditional view that glycans elicit T cell-independent immunity, producing IgM and not IgG, and that glycoconjugates rely on peptide-MHC and not glycopeptide-MHC interactions for presentation, work over the last 10 – 20 years has

established that 1) certain classes of glycans, such as zwitterionic polysaccharides, elicit T-cell dependent help through ROS mediated endolysosomal processing and presentation by MHC-II to CD4 T cells, 2) glycopeptides are generated and processed in the endolysosome and presented as a glycopeptide in the MHC-II for presentation to glycopeptide specific TCRs of T helper cells, so called Tcarbs, 3) glycopeptides can be presented by MHC-I and recognized by glycopeptide specific cytolytic T cells, and 4) anti-glycan IgG is abundant in human sera(192, 310-313). Further, Tn and STn can elicit natural T-cell dependent anti-glycan or glycopeptide responses in cancer patients as well as in immunization while linked to carrier proteins or presented as glycopeptides(171, 296, 314-320). Although early attempts to generate anti-glycan and anti-O-glycan immune responses capable of impacting clinical course were empirical and exhibited mixed success, lessons learned over the last 20 years as well proof-of-principle glyco-vaccine engineering with other diseases and tumor models provide a basis for systematic development and evaluation of therapeutic vaccines targeting tumor associated O-glycans(310).

Selectin-Lewis interactions

SLe^a and SLe^x interact with selectins on endothelial cells, platelets, and leukocytes. Tumors overexpress sialyl Lewis antigens, facilitating metastasis via enhanced platelet interactions in blood and enhanced endothelial interaction during vascular extravasation. Platelets bind tumor in blood, forming a “cloak” that protects tumor cells from anti-tumor cellular immunity. Blocking $SLe^{a/x}$-selectin interactions in model systems through decoy

disaccharides or peptide mimics have provided proof-of-concept that blocking SLe-selectin interactions can block metastasis(66, 224). Further, work in mice and clinical trials suggested that interactions of P-selectin and $SLe^{a/x}$ could be blocked by heparin, resulting in improved course of disease in mice and patients(226, 321). These results appear independent of anti-coagulant effects as benefits were observed to a lesser degree in patients treated with coumarin anti-coagulant. Cimetidine, an H2 blocker, was also shown to increase survival among patients with colorectal cancer through downregulating E-selectin expressed on endothelia. This effect was most pronounced among patients with high level of $SLe^{a/x}$ expression in tumors and least pronounced among patients with low $SLe^{a/x}$ expression(225). Results from mouse models using specific SLe-selectin inhibitors as well as human studies using heparin and cimetidine provide strong evidence that SLe-selectin interactions may be a viable target for cancer treatment; however, no conclusions can be made until specific inhibitors are tested in patients in randomized controlled prospective studies.

V. Conclusions

Since the discovery of altered glycosylation in cancer through lectin binding, monoclonal antibodies, and glycopeptide analysis, altered glycosylation has been observed in most every cancer to date establishing it as a hallmark of cancer. Altered O-glycosylation was arguably first observed with alterations in blood group structures in gastric cancer, but validated for truncated pan-carcinoma O-glycans T/Tn with the work of Springer and colleagues. Alterations include expression of truncated, tumor-specific O-glycans such as

Tn, STn, and T antigens, which are only found on O-glycans, loss of terminal α-GlcNAc on core 2 in gastric cancer, and altered terminal tumor-associated structures, such as ABH and $SLe^{a/x}$ structures, which are found on various classes of glycoconjugates.

Recent advances have highlighted possible mechanisms for altered expression of tumor O-glycans, including deletion, epigenetic silencing, and LOH of *Cosmc* resulting in Tn/STn expression, hypoxia-induced transcriptional changes resulting in increased expression of $SLe^{a/x}$ and T antigens, as well as epigenetic silencing of *A, B transferases*. The extent of these and other mechanistic alterations is currently unclear and will require further investigation.

New insights into whether and how altered O-glycans contribute to the tumor phenotype have also developed. Expression of Tn and STn results in increased susceptibility to tumorigenesis in core 3 knockout mice, and loss of α-GlcNAc contributes to gastric tumorigenesis in *H. pylori*-dependent and independent manners. T antigen expression has been implicated in tumorigenesis in xenotransplant studies as have deletion of ABH structures and overexpression of $SLe^{a/x}$. Tn, STn, α-GlcNAc, ABH, and $SLe^{a/x}$ correlate with patient survival across studies and blocking $SLe^{a/x}$ through heparin or cimetidine increases patient survival in clinical studies.

Antibodies against glycoproteins and glycans, including CA15-3, CA19-9, and CA125, are in clinical use but used for prognostics and not diagnostics due to poor specificities. Development of new antibodies against tumor O-glycans and use of multiple antibodies,

such as CA125 and anti-STn, will increase sensitivities and specificities, allowing broader application of these reagents. Imaging with radiolabeled anti-glycan antibodies provides a unique opportunity to localize tumors and has proven useful in surgical planning. Although no agent targeting tumor O-glycans is in standard clinical use, various approaches including radioimmunotherapies, immunotoxins, therapeutic vaccines, and glycan-GBP inhibitors are promising in pre-clinical and some clinical studies. The discovery of tumor O-glycans and associated applications for tumor detection, evaluation, and therapeutics offer great promise in the management of this terrible disease.

Chapter 2

An Introduction: Epithelial Glycosylation in Inflammatory Bowel Disease

Inflammatory bowel disease

Inflammatory bowel disease, including Crohn's and ulcerative colitis (UC), affects 3.6 million people and arises from inappropriate immune activation to gut microbiota in susceptible individuals(322). Disease onset occurs in the second decade of life and presents with abdominal pain, diarrhea, and weight loss. The main distinction between Crohn's and UC is based on the distribution of disease. UC is limited to the colorectum and only affects mucosal and submucosal tissue with continuous lesions. In contrast, Crohn's can affect both small and large intestine with transmural, skip lesions(323). Immunologic distinctions may in part explain the varied distributions. Crohn's exhibits elevated Th1 cytokines in contrast to elevated Th2 cytokines in UC(324).

Risk factors in IBD

Despite differences in gross pathology, Crohn's and UC share many environmental and genetic risk factors. Epidemiologic associations fall into three categories: those that impact the microbiota (e.g. antibiotics, diet), those that impact the gut barrier (e.g. food additives), and those that influence immunity (e.g. vitamin D, LPS, and parasite exposure). Twin studies have demonstrated that monozygotic twins of affected Crohn's patients have 50% risk of disease and those of UC patietns have 15% risk of disease in contrast to <10% risk for dizygotic twins(325). Genome wide association studies have

identified 163 genetic risk loci to date(326). 110 of these are shared between Crohn's and UC, while 30 are specific for Crohn's and 23 are specific for UC.

Multiple lines of evidence implicate an altered gut microbiota in IBD. IBD risk loci overlap with risk loci for infectious diseases, such as tuberculosis, and include genes involved in bacterial sensing such as NOD2 and ATG16L1(326). Anti-bacterial and yeast antibodies are elevated in IBD suggesting elevated immune activation against the microbiota, and administration of probiotics and fecal transplantion benefit some patients(327, 328). Further, early exposure to antibiotics leads to long-term changes in the gut microbiota and increased risk for IBD(329, 330).

The gut microflora in IBD

Despite being incompletely defined, a variety of changes have been observed in the IBD gut microbiota. Some changes occur consistently across multiple studies while others appear to be study- and thus population-specific. This may in part reflect the lack of a core microbiota community structure. For example, in a study of 154 healthy individuals, no single species was identified in all individuals (>0.5% abundance) despite consistent representation of microbiome gene content, arguing that functionality rather than species are conserved across individuals(331). In line with this observation, loss of species diversity is the most consistent change in IBD(332). Nonetheless, a variety of taxonomic and functional changes have been observed. These include reduced Bacteroides, Firmicutes, Clostridia, Ruminococcaceae, Bifidobacterium, Lactobacillus and F.

prausnitzii and increased Gammaproteobacteria, E. Coli, and Fusobacterium. Functionally, these taxonomic changes coincide with loss of SCFA production (a major nutrient for coloncytes) and increased auxotrophy (a feature of pathobionts), sulfate transport and oxidative stress (likely due to an inflammatory environment), and toxin secretion(332). Although poorly characterized, serotype-specific factors, such as LPS structure, are likely disrupted and contribute to disease biology. For example, different LPS molecules can differentially imprint an inflammatory versus tolerogenic tone in the gut(333).

Introduction to epithelial glycosylation in IBD

Host glycans have emerged as a potential mechanism and target in IBD. GWAS studies have implicated glycogenes, such as *FUT2* and *Cosmc*, and glycan structural characterization by antibodies or mass spectrometric analysis has identified loss of extended O-glycans with expression of truncated structures such as T and STn antigens. Disease-associated changes are observed in unaffected twins and worsen during disease flares arguing that in some patients altered glycomes precede and possibly contribute to disease initiation while in other patients these changes may result from epithelial disruption. In addition to a disruption in epithelial glycosylation, disrupted glycosylation has been observed in immune cell in IBD patients and ant-glycan antibodies are altered in disease, distinguishing Crohn's from UC patients.

The healthy gut glycome

The epithelial glycome is developmentally and regionally regulated by host and environmental factors. Mucin-type O-glycans, or O-glycans, are the major class of glycosylation in the gut, accounting for 80% of the mass of MUC2, the major intestinal mucin. In humans, the stomach and duodenum contain core 1 and 2-based structures, the remaining small intestine contains core 3-based structures, and the colon contains core 3 and 4-based structures with core 3 primarily in the sigmoid colon(334). These structures are further modified by Gal, GlcNAc, GalNAc, Fucose, sialic acid, and sulfate. Terminal epitopes follow a rostral-caudal gradient with increasing sialic acid, sulfate, and Cad antigen and decreasing fucose, blood group, and Lewis antigens(335). Because Lewis antigens and blood group antigens depend on secretor status (20% of people are secretor negative) as well as Lewis and ABO blood-types, the proximal gut glycome, including the cecum and small intestine, are highly variable in the population in contrast to relatively invariant distal gut glycome. The variability in the proximal gut mirrors other secretory materials, such as milk, saliva, lungs, and cervix, leading to the hypothesis that the distal gut glycome selects for commensals while the proximal glycome and that of other secretory organs defend against pathogens by providing decoy glycan receptors(21).

Different glycans are expressed in the fetal versus adult gut, indicating that epithelial glycosylation is developmentally regulated(336). In particular, the fetus mainly expresses neutral Core 2-based glycans in contrast to adults. Additionally, no Cad antigen is expressed in the fetus and there is an absence of a gradient of sulfate, sialic acid, and

fucose. Also, some oncofetal antigens are expressed in fetuses but not adults, including T and STn antigens.

Regulation of epithelial glycosylation in the gut

As described above, studies investigating fetal versus adult intestinal glycosylation indicate that the glycome is developmentally regulated. However, whether this regulation solely depends on host cues or also environmental signals has been a major question in the field. In contrast to humans, the mouse intestinal glycome primarily consists of core 2-based glycans and the regional distribution of some terminal epitopes varies compared to humans (337, 338). Nonetheless, both mice and humans express terminal Cad, sialic acid, sulfate, and fucose. Thus, mouse is a reasonable model to study regulation of epithelial glycosylation.

Epithelial fucosylation has been a model epitope to understand regulation of intestinal glycosylation. Fucose can be added to proteins in alpha 1,6, alpha 1,3, and alpha 1,2 linkages. UEA-I lectin recognizes alpha 1,2-linked fucose whose expression depends on two enzymes Fut1 and Fut2. Alpha 1,2-linked fucose is the major class of fucosylation in the intestinal epithelia. Importantly, Fut2 regulates fucosylation in most of the intestinal epithelia while Fut1-dependent fucosylation is restricted to paneth cells and M-cells in the small intestine(339, 340). Interestingly, in preliminary work from Kamioka and Kiyono, distinct populations of Paneth cells express Fut1, or both Fut1 and Fut2(ICMI, 2015).

One of the first observations that the microbiota can induce epithelial glycosylation arose from studies comparing UEA-I staining in conventionally housed versus germ-free mice(341). These studies demonstrated that conventionally housed mice express alpha 1,2 fucose in the terminal ileum but that this is not seen in germ-free mice. Further, colonization of germ-free mice with a model gut microbe, *Bacteroides thetaiotaomicron*, also induced alpha 1,2 fucose. Other microbes can also induce Fut2 expression, as has been shown for the commensal Segmented Filamentous Bacteria and pathogen Salmonella typhimurium(342). Bacteria are able to metabolize fucose and interestingly deleting a fucose sensor or fucose metabolism genes in the bacteria inhibits the ability of *Bacteroides thetaiotaomicron* to induce epithelial fucosylation(341, 343). Some bacteria, such as *Bacteroides fragilis,* can salvage free fucose and incorporate it into its own glyconjugates(344). This provides a competitive advantage in colonization experiments, perhaps by providing an immunologic mask on the bacterial surface in a process termed "molecular mimicry."

Although initial work proposed that bacteria directly interact with the epithelia to induce fucosylation, more recent work has proposed an indirect mechanism involving immune cells. Using a variety of KO mice, Goto and Kiyono showed that bacteria interact with ILC3s to induce IL-22 that presumably interacts with IL22R on epithelial cells to induce Fut2 expression in epithelia (342, 345). In addition to the microbe-inducible IL-22 signal, ILC3s also constitutively express lymphotoxin, which is also required for Fut2 expression. Indeed, IL-22 administration to ILC3 deficient mice was sufficient to induce

epithelial fucosylation throughout the small intestine, including ectopic expression in the duodenum and jejunum. Other immunologic signals may also regulate epithelial fucosyaltion. The same group showed that *Rag1* and T cell deficient mice, but not B cell deficient mice, had elevated epithelial fucosylation and that loss of IL-10 producing CD4 T cell subset was responsible for this induction(346). Thus components of the innate immune system induce fucosylation while components of the adaptive immune system repress it. This microbe-immune-IEC axis appears to be critical in the small intestine but not the colon. In contrast to absence of alpha 1,2 fucose in germ-free mice in the ileum, colonic alpha 1,2 fucose is similar between germ-free and conventionalized mice(341). Apparently, distinct mechanisms regulate epithelial glycosylation in different regions of the gut. Importantly, the microbiota-induced glycosylation is not limited to fucose. Conventionalization of germ-free mice results in expression of a more complex, extended O-glycome, dependent on microbiota induction of many, though not all, glycosyltransferases(338).

Altered epithelial glycosylation in IBD

Multiple lines of evidence implicate altered epithelial glycosylation in IBD. GWAS studies have identified genes that directly regulate epithelial glycosylation such as *FUT2* and *Cosmc* and those that indirectly regulate glycosylation via microbiota-immune-epithelia axis, for example IL-23 and Stat3 that help regulate Fut2 expression in epithelia(339, 347). In addition to genes associated with O-glycan synthesis, genes important for N-glycan synthesis, such as components of the oligosaccharyltransferase

complex, are also implicated(348). Immunohistochemistry indicates that IBD associates with a simplified and truncated O-glycome, such as T and STn antigens(184, 349, 350). Interestingly, STn expression seems to predict progression from UC to colitis-associated-cancer(351). In some cases these changes precede inflammation. Unaffected monozygotic twins of both Crohn's and UC patients exhibit elevated T antigen expression in the crypt surface with co-conmittant increased NFkB in the same cells, suggesting that a genetic or possibly environmental factor induces T antigen expression prior to disease onset and that T antigen may lead to increased inflammatory tone(352). Structural analysis similarly supports the idea of a shift to a simplified glycome, including expression of truncated glycans like STn and loss of extended glycans containing terminal epitopes such as Lewis and Cad antigens(353). In contrast to the interpretation of T antigen expression as an early step in inflammation, glycan structural alterations correlated with inflammatory status. Glycomes of UC patients in remission resembled those of controls while inflammation-associated changes were observed in patients undergoing a flare. Importantly, these changes were reversible within the same individual observed over time(353). Thus, it is likely that some alterations in epithelial glycosylation precede inflammation and others result from inflammation.

Mechanisms for altered epithelial glycosylation in IBD

Besides genomic alterations, for example polymorphisms in *FUT2*, the mechanisms for glycome alterations in IBD are currently unknown. Cytokines have been shown to reprogram the glycome in other tissues, such as airway bronchial epithelial cells;

however, cytokine induced changes often result in increased glycome complexity due to increased transcription of sialyltransferases and sulfotransferases that, for example, induce Sialyl Lewis x expression(354). This type of change contrasts from the simplification of the glycome observed in IBD and thus most likely represents a different mechanism. Rather, the IBD glycome may result from disruption in pathways involved in regulating normal epithelial glycosylation, such as microbiota-immune-epithelial crosstalk. Indeed, the glycome of a GF mouse is simplified, in some ways resembling the IBD glycome, compared to that of a conventionally housed mouse.

Consequences of altered glycosylation in IBD

MUC2 Synthesis and Stability

Epithelial glycosylation contributes to barrier formation, host-microbe symbiosis, and immunity. Thus it is no surprise that altered glycosylation plays a central role in IBD. Glycans are critical for synthesis of the MUC2 mucus layer, which forms a single loose layer in the small intestine and an outer loose layer and inner attached layer in the colon(355). The loose layers in the small intestine and colon serve as habitats for the microbiota and are penetrable to bacteria while the inner mucus layer of the colon is impenetrable to bacteria and prevents bacterial-epithelial interactions in the distal gut where bacterial loads reach 10^{12} bacteria per gram of stool.

MUC2 contains hundreds of O-glycans accounting for ~80% of its mass(356). In contrast, MUC2 only contains ~20 possible N-glycan attachment sites(357). Nonetheless, both N- and O-glycosylation are important for MUC2 biology. Muc2 is translated and dimerizes in the ER. N-glycosylation in ER is then critical for transfer to the Golgi(358). In the Golgi, MUC2 gets highly O-glycosylated and obtains many charged residues such as sialic acid and sulfate. Each MUC2 protein in the dimer then participates in additional interactions to form trimers. This along with a high pH and Ca^{2+} content in the Golgi facilitates dense compaction of the MUC2 network(359). MUC2 is stored in a special structure called the theca in goblet cells and is released homeostatically or from environmental triggers, for example by acetylcholine or TLR ligands(360, 361). Nearby CFTR channels release bicarbonate into the lumen that neutralizes the high pH and chelates Ca^{2+}(362), facilitating unfolding of the MUC2 network into stratified sheets(363). In the SI in a microbiota-dependent fashion, additional processing by the meprin beta protease releases the attached SI mucus layer from the epithelial to form an unattached loose layer(364). The mechanisms are currently unknown for how a loose layer is formed in the colon but likely involvse host protease cleavage of MUC2. Germ-free mice have an outer and inner mucus layer but in explant cultures, full mucus expansion does not occur in the presence of protease inhibitors(356).

O-glycans likely play two major roles in MUC2 synthesis. First O-glycans contain the charged residues that interact with Ca^{2+} in the Golgi and theca thereby facilitating tight packing and storage of MUC2(363). Although not experimentally tested, loss of charge residues in MUC2 likely results in a reduced amount of MUC2 that can be stored and

thus released. Second, MUC2 may prevent mucus degradation by bacterial proteases. This is certainly true for model proteases such as cysteine protease RgpB from Porphyromonas gingivalis as shown with glycopeptides or pronase treated mucin from T-synthase KO mice(248, 365); however, whether commensal proteases are able to access and degrade MUC2 deficient in O-glycans in in the inner mucus layer has not been demonstrated. Nonetheless, *T-synthase*-KO mice and *Cosmc*-KO mice (as described in Ch. 4), which are deficient in extended O-glycans beyond a singled GalNAc or its sialylated counterpart, both have loss of the mucus layer and increased bacterial epithelial-contact in the distal colon(29). Loss of this mucus layer results in increased gut permeability(29). However, whether increased permeability results from an indirect role of cytokine-induced remodeling of the tight junctions or a direct role of O-glycans on an adhesion molecule is not known(366). Thus defects in N and O-glycosylation could affect mucin synthesis and stability upon secretion, leading to compromised barrier function.

Host-microbe interactions

Epithelial O-glycans can directly regulate host-microbe interaction by providing ligands for bacterial adhesins and nutrients for bacterial metabolism. Indirectly, O-glycans have diverse roles, ranging from controlling epithelial gene and protein expression and regulating inflammatory tone via the mucosal barrier, as described above, as well as through antigen uptake and immune imprinting, as described below. Ultimately, these

diverse mechanisms converge to select commensals and facilitate or inhibit pathogen binding.

Historically, glycan binding proteins were first studied in pathogens, including bacteria and viruses. Many of these bind fucosylated residues that are only present in secretor-positive individuals. This includes BabA in *Helicobacter pylori* as well as adhesins in Norovirus and Rotavirus(367-370). As described in Ch. 1, *H. pylori* is an interesting example because glycans both facilitate and inhibit binding. The deep glands of the stomach express O-glycans containing terminal alpha 1,4 GlcNAc that block synthesis of a cholesterol-containing cell wall component by inhibiting a bacterial enzyme(210, 371). This normally prevents gastric invasion and inflammation; however, loss of this terminal alpha 1,4 GlcNAc facilitates invasion, inflammation and cancer.

More recently, a number of glycan-binding adhesins on commensal bacteria have also been identified. This has been seen for multiple Lactobacillus species: *L. plantarum* binds blood group A and B containing mucins through its cells surface expressed GAPDH, *L. mucosae* binds A and B blood groups through a newly identified Lam29 adhesin with homology to ABC transporters, and *LGG* binds mucin through SpaC in a glycan-dependent manner(372-374). In addition to Lactobacilli, a collection of FS1BPs from *Bifidobacterium longum* subsp. *Infantis* bind various host glycans(375). Although glycan metabolism will be described in more detail below, polysaccharide utilization loci require glycan-binding proteins homologous to SusC and SusD molecules from the starch

utilization locus. These molecules likely facilitate bacterial adhesion to host glycans in addition to aiding glycan metabolism(376).

The small intestine degrades and absorbs proteins and simple sugars but many of the complex plant polysaccharides and mucosal glycans are indigestible by the host, providing a food source for the gut microbiota. These bacteria contain polysaccharide utilization loci that bind, import, and degrade glycans. The Bacteroides contain a broad repertoire of PULs with *B. thetaoitamicron* containing 88 different loci(377). Studies of this model organism led to the idea that some bacteria can degrade both dietary and host glycans but when both are present dietary polysaccharides are preferentially utilized(378). Analysis of other organisms revealed that this is true for some but not all species of *Bacteroides* and that while some bacteria like *B. thetaiotomicron* are able to metabolize multiple classes of glycans other organisms are more limited(379). For example, *Bifidobacterium* and *A. muciniphilia* contains a more limited repertoire of PULs that restrict foraging to host glycans(380), corresponding to A. muciniphilia's enrichment in the mucosa.

Another emerging concept is that some microbes share resources. This was initially observed in pathogens in which pathogenic microbes contain monosaccharide transporters but not the glycosidases to release them and thus require commensals for colonization(381, 382). However, this idea of resource sharing was more recently observed among members of Bacteriodales arguing that this may be a more common feature among commensals than previously believed(383). Theoretical models of

commensal selection historically predicted that one bacteria will eventually dominate the population. This of course contrasts with the experimental evidence that the gut microbiota is incredibly diverse and that diversity is generally maintained over time. Resource sharing is one explanation that would allow mutualism and persistence of two species in addition to the idea that the gut contains many distinct ecological niches or biogeographies in both the mucosal-lumenal and rostral-caudal axis(384).

Multiple studies have shown that bacterial inoculation into animals induces PUL expression and that such expression is critical for colonization, transmission, and maintenance of a long-term reservoir following antibiotic or pathogen challenge (377, 385, 386). Collectively, these studies suggest that a simplified glycome observed in IBD may contribute to an altered microbiota and reduced diversity due to both disrupted bacterial binding and reduced resource availability.

Altered Immunity

Epithelial glycans and the mucins they decorate are involved in antigen uptake, immune imprinting, and inflammatory cell recruitment. Antigen is taken up in the small intestine by multiple routes. The first route described takes up antigen in the Peyer's patches utilizing a specialized cell called the M cell(387). M cells lack a mucin layer and have a relatively thin glycocalyx, facilitating interaction of these cells with the luminal contents. Antigen uptake by this route is critical for generating bacterial coated IgA as genetic M cell depletion results in a substantial reduction of IgA coated commensals(388).

In another route of antigen uptake, CD11c+ dendritic cells in the lamina propria extend dendrites into the lumen and sample antigen. During this process, dendritic cells maintain transepithelial resistance by expressing tight junction molecules that intercalate with a reorganized epithelial TJ(389). In contrast to M cells of Peyer's patches, this route occurs in villous epithelia and thus drastically expands the potential surface area for antigen sampling. In a similar way, in vivo imaging recently identified goblet associated passages (GAPs) that expand the surface area available for antigen sampling but mechanistically differ from DC sampling by transepithelial dendrites(390). While sampling by transepithelial dendrites likely plays an important role during infection, GAPs are involved in steady state sampling and result in tolerogenic responses(391).

Although these routes can sample uncoated antigen, FcRn is involved in uptake of immune complexes(392, 393). FcRn transports IgG in the basal-apical direction and subsequently retrotranslocates immune complexes in the apical-basal direction. The translocated antigen is subsequently able to prime CD4 T cell responses. FcRn mediates this uptake throughout absorptive cells of the epithelium. The different routes of antigen uptake differ in the type/size of antigens as well as in the quality of immune response – whether inflammatory or regulatory. Thus these modes differentially participate in uptake of pathogens, commensals, and food antigens.

Of these four routes, M cell uptake depends on epithelial glycosylation. The bacterial FimH lectin interacts with glycoprotein 2 of M cells in a carbohydrate-dependent

manner(394). Although not experimentally tested, paracellular dendritic cell sampling and GAP uptake also likely depend on glycosylation. Adhesion molecules are critical for maintaining TER during DC sampling and epithelial-epithelial interactions rely on glycans for stability(395, 396). Glycans are also critical for multiple aspects of MUC2 synthesis and stability in the goblet cell and mucus layer(29). Thus one would predict that disruption in epithelial glycosylation would perturb GAP-dependent uptake, although this will need to be tested.

Epithelial glycosylation also facilitates tolerogenic imprinting and PMN-epithelial interactions. Recent work found that dendritic cells in the lamina propria and Peyer's patches of the small intestine can take up MUC2 along with antigen, resulting in tolerogenic imprinting(397). Mechanistically, this involved a Galectin-3-Dectin-1-DC interaction. Galectin-3 interacts with glycosylated MUC2 but not deglycosylated peptide. Because galectin-3 is soluble, it is tethered to DCs in complex with dectin-1 and FcgammaRII. Binding of glycosylated MUC2 to this complex induces nuclear translocation of beta catenin, which is required to generate tolerogenic DCs.

PMN extravasation into the lumen can be important in fighting infections but also causes pathology in diseases like IBD. PMNs bind a glycosylatylated CD44 isoform (CD44v6) on epithelia and blockade of this interaction disrupts PMN-epithelial interactions(398). This epitope is only expressed in inflammatory conditions, providing a molecular basis for inflammation-induced PMN recruitment to the epithelium.

Conclusions

IBD results from dysregulation of immune cells, microbiota, and epithelia. Current evidence suggests that epithelial glycosylation plays a central role in all of these processes. Further, studies in IBD patients indicates that disease correlates with expression of truncated O-glycans although mechanisms for these alterations are currently unclear. Based on twin studies, altered glycosylation in some cases precedes disease initiation while for other glycan changes altered structures emerge during active disease. Glycosylation is critical for biosynthesis and stability of MUC2 the major glycoprotein in the intestinal mucus layer, and loss of MUC2 results in spontaneous inflammation and cancer(250, 399). The mucus layer is often compromised in IBD suggesting the possibility that altered glycosylation may be one explanation for this. Epithelial glycosylation along with dietary glycans likely synergize to establish microbial networks in the intestine and alterations in epithelial glycosylation may explain some aspects of dysbiosis in IBD. Additional work suggests that epithelial glycosylation may be a key component regulating the intestinal immune system. Glycans are involved in antigen uptake, tolerogenic imprinting, and immune cell interactions with the epithelia. Alterations in the normal glycome may lead to disruptions in these processes that tip the gut from a regulatory to an inflammatory environment. Many important questions remain: 1) How does IBD genetics contribute to the glycan structural changes observed in patient tissues? 2) How do host glycans interact with dietary glycans to regulate the microbiota? 3) What host factors are able to compensate for changes in glycosylation to maintain homeostasis? 4) How does compromised epithelial glycosylation affect the

intestinal immune system? Work presented in this thesis begins to address all of these questions and lays the groundwork for understanding and targeting altered glycosylation in IBD.

Chapter 3

Development of the first technology to amplify and sequence the cellular glycome:

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

(Adapted from Kudelka et al, Nature Methods, 2016.)

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 an approach termed Cellular O-Glycome Reporter/Amplification (CORA), a sensitive method to amplify and profile mucin-type O-glycans synthesized by living cells. Cells convert added peracetylated benzyl-α-N-acetylgalactosamine (GalNAc-α-Benzyl) 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 from *Homo sapiens* and *Mus musculus*. 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(7), but systems-level approaches for glycomics are lacking in most biological settings. A primary need are simple and sensitive technologies to analyze all glycans synthesized by cells (the cellular glycome)(42). 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 require relatively large amounts of biological samples for detailed structural analyses, limiting widespread application. One way to at least partly overcome these challenges would be to "amplify" the glycome, similar to PCR-hybridization-based technologies for the genome(400). 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 or Thr where R represents extended structures), which is present on >80% of proteins(14) that traverse the secretory apparatus and important in many normal and pathologic settings(8, 10, 92, 401). Nonetheless, little is known about either the repertoire of O-glycans or 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(402).

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). 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. Using CORA, we identified 76 O-glycans in a variety of cell types, including those observed by traditional alkaline β-elimination/MS.

Results

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

Mucin-type O-glycan biosynthesis begins with transfer of GalNAc to Ser or Thr residues in glycoproteins to generate GalNAcα1-O-Ser or Thr (Tn antigen), which T-synthase extends to the dominant Core 1 O-glycan Galβ3GalNAcα1-O-Ser or Thr.

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, Ac_3GalNAc-α-Bn), 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. 3.1**). Benzyl-α-GalNAc structurally mimics the precursor GalNAcα1-O-Ser or Thr (Tn antigen) in O-glycoproteins(403), was previously used *in vitro* as an acceptor for T-synthase and Core 3 GnT(113, 404), and its peracetylation promotes transport across the plasma membrane(66).

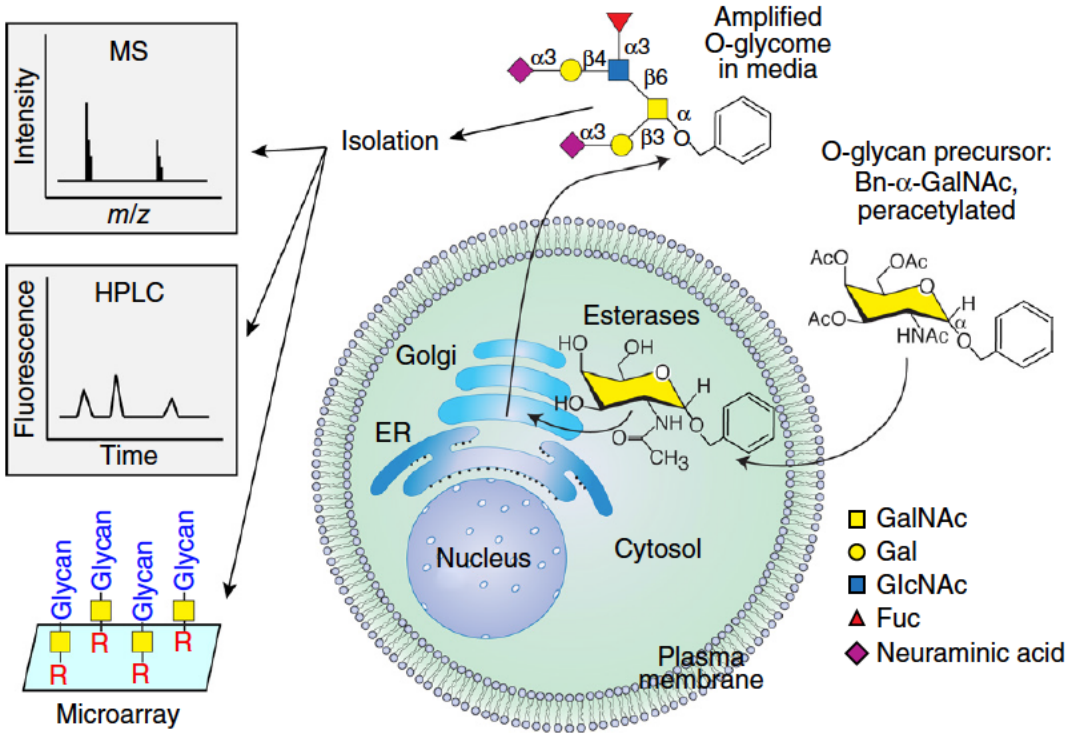


Figure 3.1. Overview of CORA. Cells are incubated with peracetylated chemical precursor (Ac3GalNAc-α-Bn). Cytosolic esterases generate Bn-α-GalNAc, which is taken up in the Golgi and modified by native glycosyltransferases during anterograde transport. Elaborated Bn-O-glycans are secreted into the media, purified and then

analyzed by MS (reported here) or HPLC or printed on glycan microarray for interrogation by glycan-binding proteins. ER, endoplasmic reticulum.

Prior studies have shown that monosaccharides linked to hydrophobic aglycones can prime glycan biosynthesis(405, 406), We hypothesized that Bn-α-GalNAc at low concentrations, could be used as a surrogate acceptor by the T-synthase to allow formation of free Bn-O-glycans representing the cellular O-glycome. We utilized Ac_3GalNAc-α-Bn, a more hydrophobic derivative of GalNAc-α-Bn (**Supplementary Fig. 3.1a**), to enhance cellular uptake as shown for other peracetylated carbohydrate compounds(66, 406) and predicted that upon entry into cells Ac_3GalNAc-α-Bn would become activated by cytosolic esterases to regenerate Bn-α-GalNAc. Bn-α-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 (**Fig. 3.1**).

We cultured adherent (HEK293) and suspension (Molt-4) cells in complete media containing 50 μM Ac_3GalNAc-α-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 (**Supplementary Fig. 3.2**) 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 0–250 μM of Ac_3GalNAc-Bn or Bn-α-GalNAc for 3 days. We observed Bn-O-glycans with predicted sialylated Core 1 structure at a concentration as low as 25 μM for Ac_3GalNAc-Bn, but only at the highest concentration of 250 μM for Bn-α-GalNAc (**Supplementary Fig. 3.3**). Further, peracetylation of Bn-α-GalNAc was stable in complete media for at least 3 days (**Supplementary Fig. 3.4**). Thus, peracetylation improved the sensitivity, and at such low concentrations should limit potential side effects, as explored below.

To optimize conditions, we cultured MDA-MB-231 cells with 0–250 μM of Ac_3GalNAc-Bn for 2–4 days. Bn-O-glycans in culture media were seen at all time points and concentrations down to 5 μM of Ac_3GalNAc-Bn for 2 days (**Supplementary Fig. 3.5**), and their profiles were stable over time. Remarkably, increasing Ac_3GalNAc-Bn concentration shifted abundance from disialylated to monosialylated Core 2 (**Supplementary Fig. 3.5d**), supporting observations that glycosyltransferase:substrate ratios drive glycan microheterogeneity(407). Thus, CORA should be performed at low concentrations of Ac_3GalNAc-Bn. For most subsequent studies, we incubated cells with 50 μM Ac_3GalNAc-α-Bn for 3 days. At these conditions, Ac_3GalNAc-α-Bn was not toxic to cells (**Supplementary Fig. 3.6**), did not alter cellular morphology or granularity (**Supplementary Fig. 3.7**), did not alter cell surface O- or N-glycosylation

(**Supplementary Fig. 3.8, 3.9**), and Bn-O-glycans were stable after secretion from cultured cells (**Supplementary Fig. 3.10**).

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

To confirm that Bn-α-GalNAc can only be utilized by glycosyltransferases involved in mucin-type O-glycan biosynthesis, but not other irrelevant or unknown pathway(s), we performed CORA on cells with mutant or wild-type *Cosmc*, a chaperone known to be essential for T-synthase activity(408). Only cells with functional *Cosmc* and active T-synthase secreted Bn-O-glycans when we administered Ac_3GalNAc-Bn (**Fig. 3.2**). Furthermore, cells incubated with the isomer Ac_3GlcNAc-β-Bn (**Supplementary Fig. 3.1b**) secreted no Bn-O-glycans or only the simple trisaccharide Neu5Ac-Gal-GlcNAc-Bn (**Supplementary Fig. 3.11**), indicating that Ac_3GalNAc-α-Bn is specific for mucin-type O-glycans.

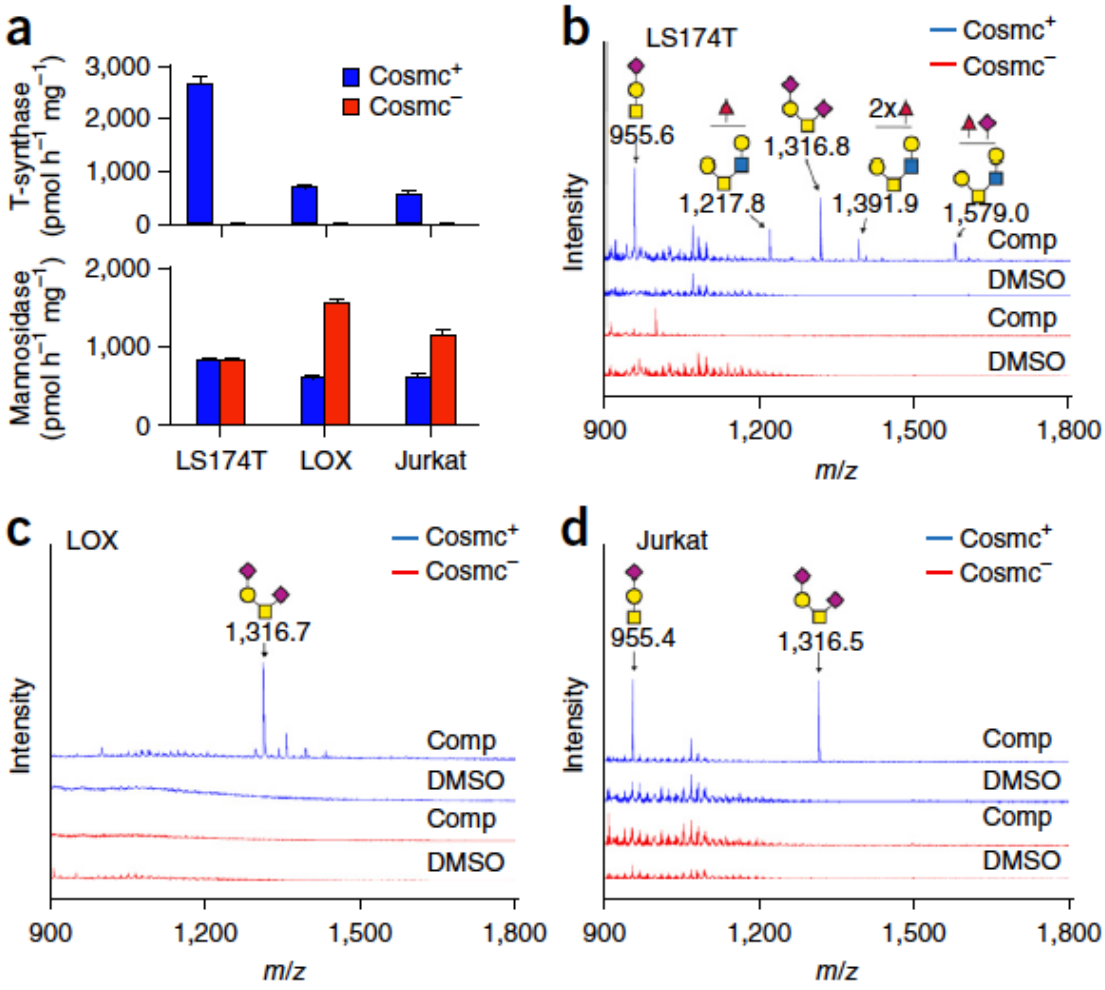


Figure 3.2. The chaperone Cosmc and active T-synthase are required for production of core 1– and core 2–based Bn-O-glycans. (**a**) Activity of T-synthase relative to mannosidase control enzyme for LS174T colorectal cells, LOX melanoma cells and Jurkat T cells with and without functional Cosmc. Enzymes were assayed twice in triplicate; shown are results of a representative experiment (mean and s.d. of triplicates). (**b–d**) Bn-O-glycans in media from the same LS174T (**b**), LOX (**c**) and Jurkat (**d**) cells after incubation with 50 μM Ac3GalNAc-α-Bn compound (Comp) or vehicle (DMSO) for 3 d. Only major glycans are annotated for LS174T for clarity; highly fucosylated minor species (**Supplementary Fig. 3.10b**) were also observed. Spectra for b–d are

offset but scaled to the same absolute intensity for each cell; representative profiles are shown (n = 2).

Accuracy of CORA

We compared O-glycome profiles from CORA to β-elimination, available through the CFG database (http://www.functionalglycomics.org). WEHI-3 and HL-60 cells were analyzed because they have complex O-glycomes with unique structures, such as Cad, GalNAcβ1,4-(α2,3Neu5Ac)Galβ1,3/4-R, and extended poly-N-acetyllactosamines, $[3Galβ4GlcNAcβ-]_n$, that are challenging to detect by β-elimination of lysates but observed on purified glycoproteins(409, 410).

MALDI-TOF-MS and MS/MS profiles (**Fig. 3.3, Supplementary Fig. 3.12-14**) show that WEHI-3 and HL-60 cells produced 40 and 11 glycan structures, respectively, including sialylated Core 1 and 2-based glycans for both cells and Cad antigen in WEHI-3 (**Fig. 3.3a**). CORA detected most of the compositions seen in β-eliminated samples (12 of 13 for WEHI-3, 4 of 7 for HL-60) and many additional compositions, 16 for WEHI-3 (**Fig. 3.3b&c**), and 6 for HL-60 (**Supplementary Fig. 3.12b&c)**, which were generally the most complex, including poly-N-acetyllactosamines with ~3 repeats and I antigen. The 4 masses observed by β-elimination but not CORA (**Fig. 3.3b&c, Supplementary Fig. 3.12b&c**) are Core 1 and 2-based glycans, lacking terminal sialylation, and therefore are likely biosynthetic intermediates, as might be observed for glycoproteins within an intermediate Golgi compartment prior to secretion.

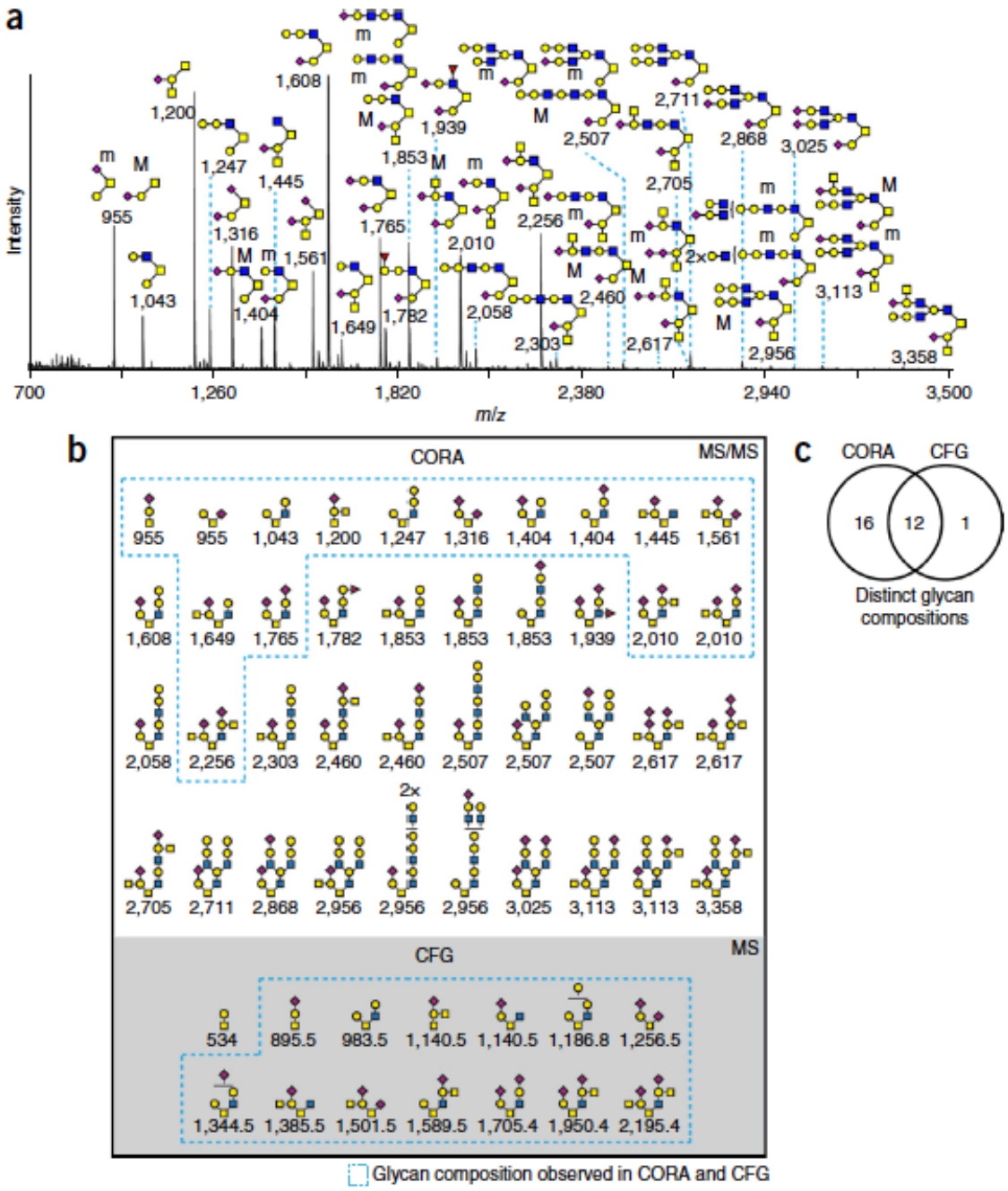


Figure 3.3. Accuracy of CORA for profiling the O-glycome. (**a**) MALDI-TOF-MS characterization of Bn-O-glycans from WEHI-3 cells incubated with 50 μM Ac3GalNAc-α-Bn. Here as in all of our experiments, we analyzed only the nonsulfated and/or nonphosphorylated glycans. (**b**) Bn-O-glycans from CORA (MS/MS structures) compared to profiles from alkaline β-elimination (MS compositions) (CFG data); blue

dashed outlines indicate glycan compositions observed in both CORA and β-elimination. (**c**) The number of glycans (by composition) observed with CORA, β-elimination (CFG data) and both. Putative structures are based on composition, MS/MS and biosynthetic knowledge.

To confirm our results, we repeated the experiments twice and obtained nearly identical O-glycome profiles (**Supplementary Fig. 3.15**) and performed ESI-MS and obtained similar results to MALDI-MS (**Supplementary Fig. 3.16**). We evaluated a range of cell types (**Supplementary Table 3.1**) with diverse glycosyltransferases (as exemplified by C2GnT1–3, **Supplementary Fig. 3.17**) and obtained O-glycome profiles from all the cells indicating that Bn-GalNAc can access most if not all of the O-glycan machinery. Thus, CORA reflects the cellular O-glycome, which is relatively stable under optimal culture conditions.

Sensitivity of CORA

Beta-elimination used for O-glycan release often requires $\geq 10^7$ cells and produces many unassignable peaks(411). To determine how many cells are needed to get clean, interpretable profiles with CORA, we profiled four cell lines each seeded at 5×10^5, 10^5, or 2×10^4 cells. We obtained O-glycomes from all lines seeded at $\geq 10^5$ cells and 3 of 4 lines seeded at 2×10^4 cells (**Fig. 3.4, Supplementary Fig. 3.18**). Notably, profiles did not change with different cell numbers. The detection of Bn-O-glycans from 2×10^4 cells

cultured for 3 days (to a total of ~8 × 10^4 cells assuming ~24 hour doubling times) represents a ~100–1000-fold increase in sensitivity compared to β-elimination.

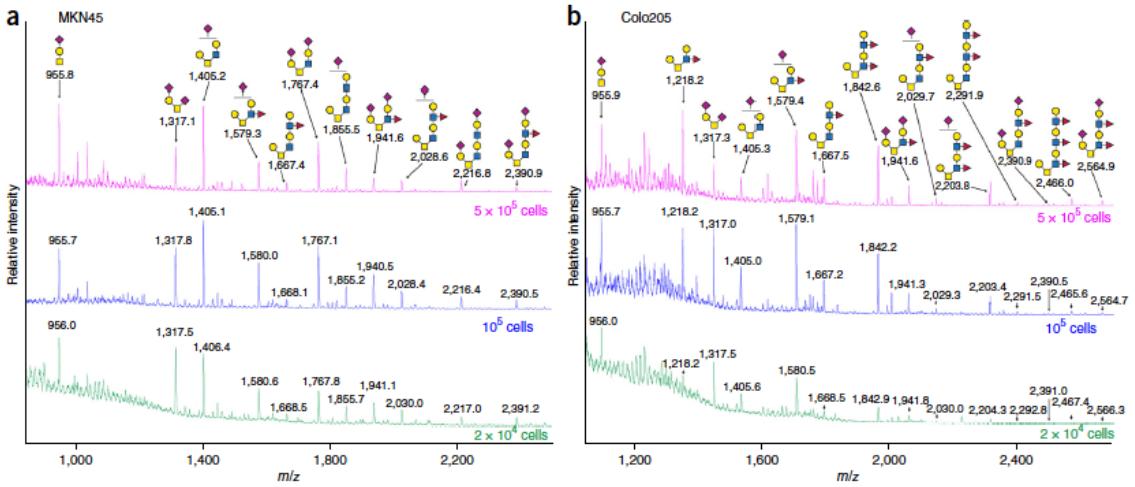


Figure 3.4. Sensitivity of CORA. (**a**,**b**) Bn-O-glycans purified from (**a**) MKN45 and (**b**) Colo205 cells seeded at the indicated cell densities and incubated with 50 μM Ac3GalNAc-α-Bn for 3 d. A fraction of total glycans were analyzed by MALDI-MS (composition). Spectra are offset for each seeding density and scaled relative to the maximum intensity. Representative profiles are shown (n = 2).

Profiling the O-glycome of mouse and human primary cells

Because primary cells differ metabolically from cancer cells(412), 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(32) to determine if Bn-O-glycan synthesis requires functional *Cosmc*. Only mPECs with a functional *Cosmc* secreted Bn-O-glycans (**Supplementary Fig. 3.19a,b**). Next we

isolated Tn(-) mPECs from mice (Tie2-*Cre*$^{-}$;*Cosmc*$^{F/y}$) and performed CORA (**Supplementary Fig. 3.19c,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 alter glycan biosynthesis, as in human tumors(413).

Next, we evaluated primary human dermal fibroblasts and umbilical vein endothelial cells (HUVECs) (**Fig. 3.5, Supplementary Fig. 3.20, 3.21**). HUVECs produced 43 O-glycan structures, including those containing poly-N-acetyllactosamine, Lewis structures, blood group antigens, and I antigen, as confirmed by MS/MS (**Fig. 3.5, Supplementary Fig. 3.20a, 3.21**). Fibroblasts also produced at least 18 glycans (unique masses), including poly-N-acetyllactosamine, Lewis structures, and blood group antigens (**Supplementary Fig. 3.22, Supplementary Fig. 3.20b**). The remarkable diversity of O-glycans in these cells indicates their potential importance.

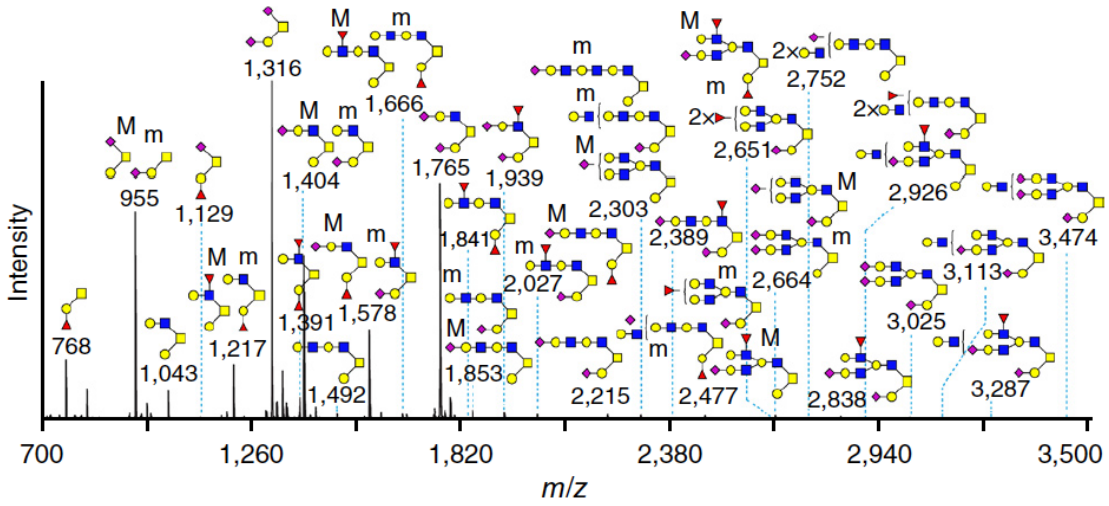


Figure 3.5. MALDI-TOF-MS/MS profiling of the O-glycome of primary cells. Bn-O-glycans from HUVECs seeded at a density of 12.5 . 10^4 cells, cultured for 2 d and then incubated with 50 μM Ac3GalNAc-α-Bn for 3 more days. Spectra are offset and scaled relative to maximum intensity. Putative structures are based on composition, MS/MS and biosynthetic knowledge.

CORA promotes the discovery of novel glycans

MS/MS sequencing is needed for definitive determination of glycan structure. However, this is often not possible for glycans derived by 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 glycans (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**. **3.3, 3.5, Supplementary Fig. 3.13, 3.14, 3.21**). 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(414-416).

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(417). 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, there are insufficient amounts of O-glycans released by β-elimination to identify larger O-glycans in the high mass range, and when larger O-glycans are identified, MS/MS usually is not feasible. CORA permits analyses of small numbers of cells since they continuously produce more Bn-O-glycans, thus amplifying their O-glycomes.

CORA can evaluate the complexity of the cellular O-glycome

The complete repertoire of O-glycans in human and animal cells 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 3.1, 3.2**). Detailed MS/MS structural characterization on 3 of these cells (HUVEC, HL-60, WEHI-3) revealed 76 unique structures, derived from 48 unique compositions (**Supplementary Table 3.3**). From our 18 cell lines, we evaluated the frequency of cell lines expressing a given glycan composition (**Supplementary Fig. 3.23a,b, Supplementary Table 2**) and determined that a few outstanding cell lines encompassed most of the complexity of all 18 glycomes (**Supplementary Fig. 3.23c**). We developed a computational model to determine the number of unique O-glycan compositions we would observe on average from a given number of cell lines and predicted that the estimated 200 cell types in the human body produce 235 unique compositions (**Supplementary Fig. 3.24**). 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 1.6 (76/48) × 235 = 376 non-sulfated/phosphorylated structurally different O-glycans in the human body. While

speculative to a degree, this is the first estimate of the size of the human cellular O-glycome based on comparative cellular O-glycomics and provides a roadmap to comprehensively sequence the human cellular O-glycome.

Discussion

CORA is an approach to both amplify and profile the mucin-type O-glycome in living cells. Its high sensitivity and amplification made it possible to identify novel complex O-glycans, which were not seen in the β-elimination released samples, indicating that β-elimination may not provide enough material for MS or MS/MS evaluation of these low abundance species. For example, MS analyses of β-eliminated O-glycans from WEHI-3 cells have not identified Sialyl-Le^X on core 2 O-glycans, a critical component on PSGL-1 recognized by P-selectin, which binds WEHI-3 cells. However, this minor structure was previously confirmed by an enrichment analysis(410), and now detected by CORA, directly demonstrating the improved sensitivity of CORA over conventional methods. Thus, the novel structures we identified are natural O-glycans, mostly present on glycoproteins at low abundance. Interestingly, CORA using Ac_3GalNAc-α-Bn is specific for O-glycans, since *Cosmc*-deficient cells do not produce Bn-O-glycans and culturing cells with the isomer Ac_3GlcNAc-β-Bn produced no glycans or a trisaccharide (Sia-Gal-GlcNAc-Bn) only. The latter result suggests that the enzymes responsible for poly-N-acetyllactosamine production, I antigen synthesis, and α3-fucosylation are relatively specific to extended O-glycans in the cell lines examined.

It is interesting that CORA is effective at low concentrations (<250 μM) 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(418, 419). 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. It is noteworthy that this approach defines the global ability of cells to make O-glycans, but may not always reflect their natural relative abundances, which may be influenced by the concentration of protein substrates or expression of O-glycosylated polypeptide cores.

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(420). 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(137). 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 largely drives O-glycan heterogeneity.

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(421-424), yet glycan release is often not sufficiently sensitive to analyze such precious specimens. CORA could address this challenge by amplifying the glycome of these cells in addition to analyzing tumor cells in a high-throughput manner. 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

Analyses

Sample size: At least two different cell lines were selected for most experiments in order to validate across cell types. More cell lines were used in some instances when we thought this could improve the generalizability of the result. To provide a global view of the O-glycome and validate CORA across cell lines and types, we analyzed a total of eighteen cell lines from distinct organs (skin, kidney, breast, colon, stomach) and tissues (epithelia and connective, including blood cells, fibroblasts, and endothelia) from multiple species (*H. sapiens*, *M. musculus*).

Randomization and blinding: Cell lines and primary cells were selected to capture a diversity of tissues and analyzed at different time points throughout the course of the experiments. Randomization and blinding were not performed for glycan analyses, as this would not have changed the interpretation of mass spectrometry data.

Exclusion criteria: No data were excluded. Data shown were randomly selected from replicate experiments.

Compounds

Benzyl-α-D-GalNAc was purchased from Sigma. Benzyl-β-D-GlcNAcAc$_3$ and 4MU-α-D-GalNAc were purchased from Carbosynth. Benzyl-α-D-GalNAcAc$_3$ 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 1 ml 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 (Oslo University Hospital). MKN45, Colo205, MDA-MB-231, and MCF7 were a kind gift from Dr. Henrik Clausen (University of Copenhagen). Tn(-) and Tn(+) LS174T were subcloned from a mixed population as previously described(86). 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 0.01 mg/ml insulin. WEHI-3 were cultured in Iscove's (Corning) supplemented with 10% FBS, 0.05 mM 2-ME, 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. We did not perform independent verification of cell lines or testing for mycoplasma.

Administration of compound

Ac_3GalNAc-α-Bn or Bn-α-GalNAc was dissolved to 50 or 100 mM in DMSO and further diluted to 5–250 μM 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 2465 × g and flow through was collected. Bn-O-glycans were subsequently purified from flow through by Sep-Pak 3 cc C18 cartridge (Waters) by gravity chromatography. The column was equilibrated with 2 × 2 ml acetonitrile, then 4 × 2 ml 0.1% TFA. Media was applied and then column washed with 4 × 2 ml 0.1% TFA. Bn-O-glycans were then eluted with 2 × 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(425). Two hundred µl NaOH/DMSO slurry was added to samples followed by 200 µl methyl iodide. Samples were shaken for 30 minutes and then spun down at 5000 × g for 5 minutes. Supernatant was collected and chloroform extraction performed to isolate permethylated glycans. Five hundred µl chloroform and 500 µl water were added to supernatant, mixed, and

centrifuged 5000 × g for 1 minute. Two more washes with 500 µl water were performed before evaporating chloroform by centrivap for 30 minutes. Bn-O-glycans were then resuspended in 25 or 50 µl 50% methanol. 0.5 µl matrix (10 mg/ml 2,5-dihydrobenzoic acid (Sigma), 50% acetonitrile, 0.1% TFA) and 0.5 µl 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 assays were performed as previously described(403). 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.

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(32) and isolated from lung as previously described, with minor modifications(426, 427). Mouse work was performed according to approved Emory IACUC protocols. 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, Clone MEC 13.3; BD), and anti-CD102 (clone 3C4; BD) using Dynabeads®. Tn(+)-ECs were isolated from 4-week old male or female mice 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 (BaGs6, a

kind gift from the late Georg Springer). 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, B-1305), *Ricinus communis agglutinin I* (RCA-I, B-1085), and *Maackia amurensis I* (MAL-I, B-1315) lectins were purchased from Vector and HRP-labeled *Peanut agglutinin* (PNA, L7759) lectin was purchased from Sigma. Cells were treated with 50 μM 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 16000 × g 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 1 μl 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 (Vector, SA5004) 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 25 mM 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, sc-47778) 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.

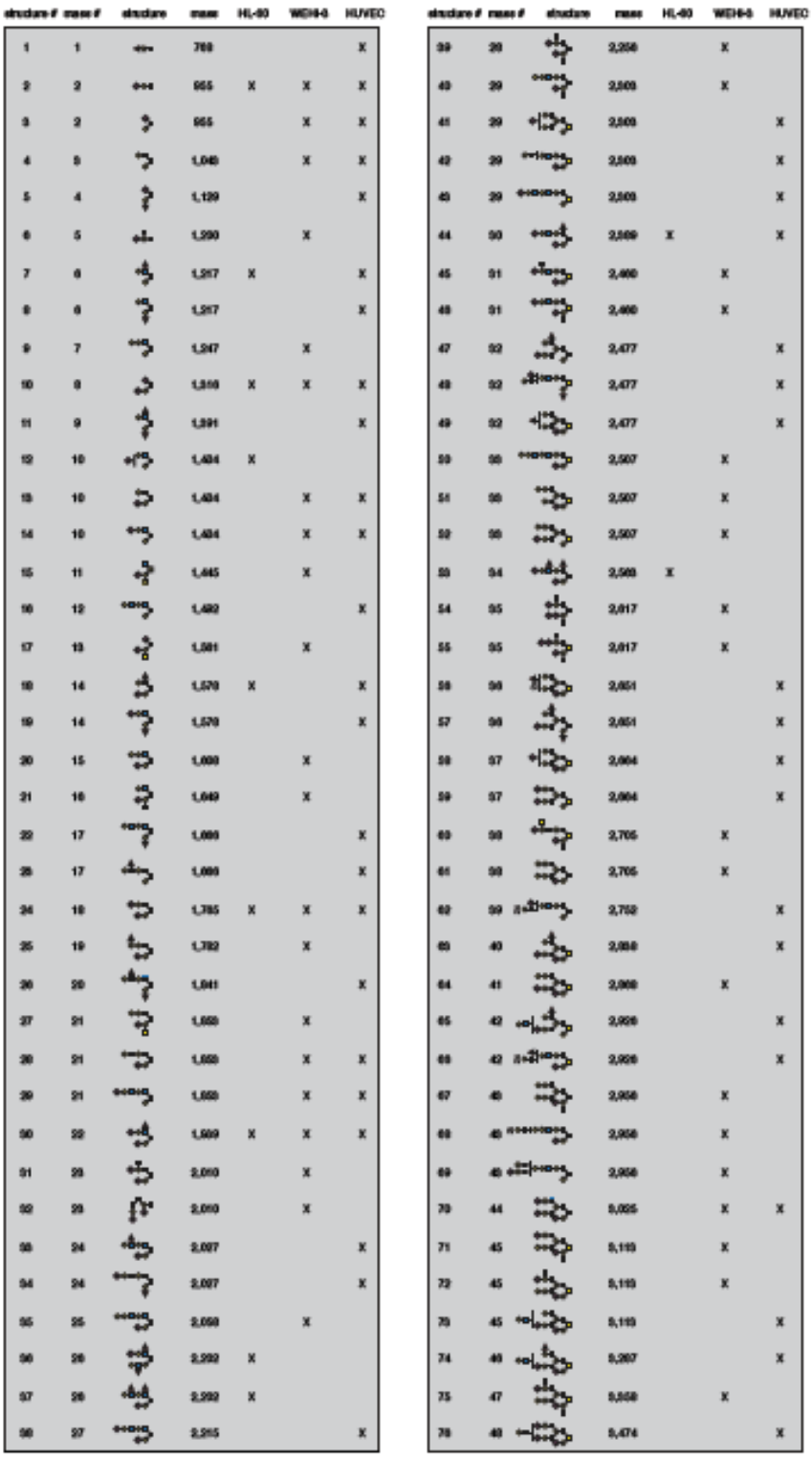
Supplemental Information

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, S10B
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	S12, S14, S15
10	7	WEHI-3	Murine	Myelomonocytic leukemia	I	MALDI-MS/MS, ESI-MS	28	40	4A, S13, S16C
11	8	MKN45	Human	Gastric cancer	I	MALDI-MS, ESI-MS	11	N/A	5A, S16B
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	S3, S5, S18A
14	11	MCF7	Human	Breast cancer	I	MALDI-MS	5	N/A	S18B
15	12	HUVEC	Human	Endothelial cell	P	MALDI-MS/MS, ESI-MS	29	43	5, S16A, S20A, S21
16	13	Human dermal fibroblasts	Human	Fibroblast	P	MALDI-MS	18	N/A	S20B, S22
17	14	Immortalized mPEC	Murine	Endothelial cell	I	MALDI-MS	5	N/A	S19B
18	15	Primary mPEC	Murine	Endothelial cell	P	MALDI-MS	6	N/A	S19D
Number of unique compositions or structures across all samples							57	76	

Supplementary Table 3.1. Information on cell lines and primary cells, the MS analyses, and data presented in the figures.

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	1,316.8	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	1,766.1	11
4	1,043.7	2	2	0	0	1,043.9									1,043					1,043	1,043.5			4	1,404.9	11
5	1,129.7	1	1	1	1															1,129				1	1,940.2	10
6	1,200.8	1	2	0	1										1,200						1,200.5			2	1,579.0	10
7	1,217.8	2	2	1	0	1,218.0		1,217.8						1,217			1,218.2			1,217	1,217.6			6	2,388.4	7
8	1,247.8	3	2	0	0										1,247									1	1,217.8	6
9	1,316.8	1	1	0	2	1,317.1	1,316.7	1,316.8			1,316.7		1,316.5	1,316	1,316	1,317.1	1,317.3	1,317.1	1,317.3	1,316	1,316.7	1,316.7	1,317.1	15	2,028.2	5
10	1,391.9	2	2	2	0			1,391.9												1,391	1,390.4			3	1,854.1	5
11	1,404.9	2	2	0	1	1,405.1	1,404.8	1,404.9						1,404	1,404	1,405.2	1,405.3	1,405.1	1,405.3	1,404	1,404.9			11	1,667.0	5
12	1,445.9	1	3	0	1										1,445									1	2,215.3	4
13	1,492.9	3	3	0	0															1,492	1,492.8			2	1,841.1	4
14	1,562.0	1	2	0	2										1,561									1	1,043.7	4
15	1,566.0	2	2	3	0			1,566.1																1	2,058.2	3
16	1,579.0	2	2	1	1	1,579.3	1,578.8	1,579						1,578		1,579.3	1,579.4	1,579.3	1,580.3	1,578	1,579.3			10	1,609.0	3
17	1,609.0	3	2	0	1										1,608							1,609.5	1,609.3	3	1,391.9	3
18	1,650.0	2	3	0	1										1,649									1	3,113.8	2
19	1,667.0	3	3	1	0			1,667								1,667.4	1,667.5			1,666	1,666.8			5	3,025.8	2
20	1,678.0	1	1	0	3																		1,678.4	1	2,563.5	2
21	1,753.1	2	2	2	1			1,753.2																1	2,303.4	2
22	1,766.1	2	2	0	2	1,766.4	1,765.9							1,765	1,765	1,767.4		1,766.4	1,767.5	1,765	1,766	1,766.6	1,766.3	11	2,202.3	2
23	1,783.1	3	2	1	1										1,782									1	1,901.2	2
24	1,841.1	3	3	2	0			1,841.2									1,842.6			1,841	1,840.1			4	1,462.9	2
25	1,854.1	3	3	0	1	1,854.4									1,853	1,855.5				1,853	1,854.3			5	1,200.8	2
26	1,901.2	5	3	0	0																	1,901.9	1,902.3	2	3,475.0	1
27	1,940.2	2	2	1	2	1,941.5	1,939.9	1,940.7						1,939	1,939	1,941.6	1,941.6	1,940.6		1,939	1,940.5			10	3,359.0	1
28	2,011.2	2	3	0	2										2,010									1	3,287.9	1
29	2,015.2	3	3	3	0			2,015.3																1	2,956.7	1
30	2,028.2	3	3	1	1	2,029.6										2,028.6	2,029.7			2,027	2,028.4			5	2,906.7	1
31	2,058.2	4	3	0	1										2,058							2,059.4	2,059.4	3	2,868.7	1
32	2,202.3	3	3	2	1									2,202			2,203.8							2	2,838.7	1
33	2,215.3	3	3	0	2	2,216.7										2,216.8				2,215	2,215.4			4	2,732.6	1
34	2,256.3	2	4	0	2										2,256									1	2,711.6	1
35	2,303.4	4	4	0	1										2,303					2,303				2	2,705.6	1
36	2,290.4	4	4	2	0												2,291.9							1	2,664.6	1
37	2,389.4	3	3	1	2	2,390.7	2,390.1							2,389		2,390.9	2,390.9			2,389	2,389.3			7	2,651.6	1
38	2,460.5	3	4	0	2										2,460									1	2,617.5	1
39	2,464.5	4	4	3	0												2,466.0							1	2,507.5	1
40	2,477.5	4	4	1	1															2,477				1	2,477.5	1
41	2,507.5	5	4	0	1										2,507									1	2,464.5	1
42	2,563.5	3	3	2	2									2,563			2,564.9							2	2,460.5	1
43	2,617.5	2	4	0	3										2,617									1	2,290.4	1
44	2,651.6	4	4	2	1															2,651				1	2,256.3	1
45	2,664.6	4	4	0	2															2,664				1	2,015.2	1
46	2,705.6	3	4	0	2										2,705									1	2,011.2	1
47	2,711.6	6	4	0	1										2,711									1	1,783.1	1
48	2,752.6	5	5	0	1															2,752				1	1,753.1	1
49	2,838.7	4	4	1	2															2,838				1	1,678.0	1
50	2,868.7	5	4	0	2										2,868									1	1,650.0	1
51	2,926.7	5	5	1	1															2,926				1	1,566.0	1
52	2,956.7	6	5	0	1										2,956									1	1,562.0	1
53	3,025.8	4	4	0	3										3,025					3,025				2	1,448.9	1
54	3,113.8	5	5	0	2										3,113					3,113				2	1,247.8	1
55	3,287.9	5	5	1	2															3,287				1	1,129.7	1
56	3,359.0	5	6	0	2										3,358									1	768.5	1
57	3,475.0	5	5	0	3															3,474				1	594.4	1

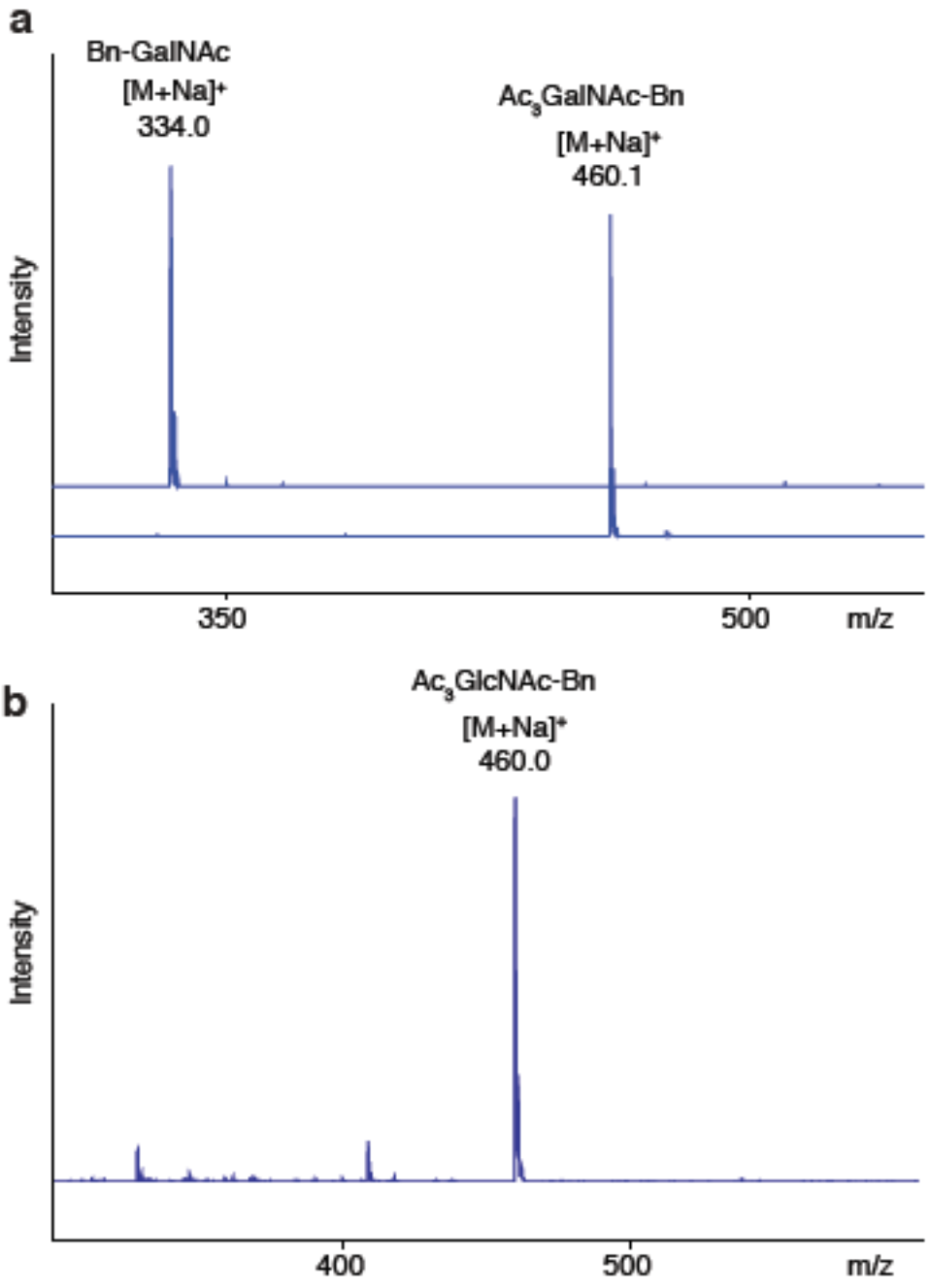
Supplementary Table 3.2. Summary of glycan masses and compositions observed in the cell lines.



structure #	mass #	structure	mass	HL-60	WEHI-3	HUVEC
1	1		708			X
2	2		955	X	X	X
3	2		955		X	X
4	3		1,040		X	X
5	4		1,129			X
6	5		1,290		X	
7	6		1,317	X		X
8	6		1,317			X
9	7		1,347		X	
10	8		1,318	X	X	X
11	9		1,391			X
12	10		1,434	X		
13	10		1,434		X	X
14	10		1,434		X	X
15	11		1,445		X	
16	12		1,482			X
17	13		1,581		X	
18	14		1,570	X		X
19	14		1,570			X
20	15		1,608		X	
21	16		1,649		X	
22	17		1,688			X
23	17		1,688			X
24	18		1,785	X	X	X
25	19		1,782		X	
26	20		1,841			X
27	21		1,858		X	
28	21		1,858		X	X
29	21		1,858		X	X
30	22		1,909	X	X	X
31	23		2,010		X	
32	23		2,010		X	
33	24		2,027			X
34	24		2,027			X
35	25		2,058		X	
36	26		2,222	X		
37	26		2,222	X		
38	27		2,215			X

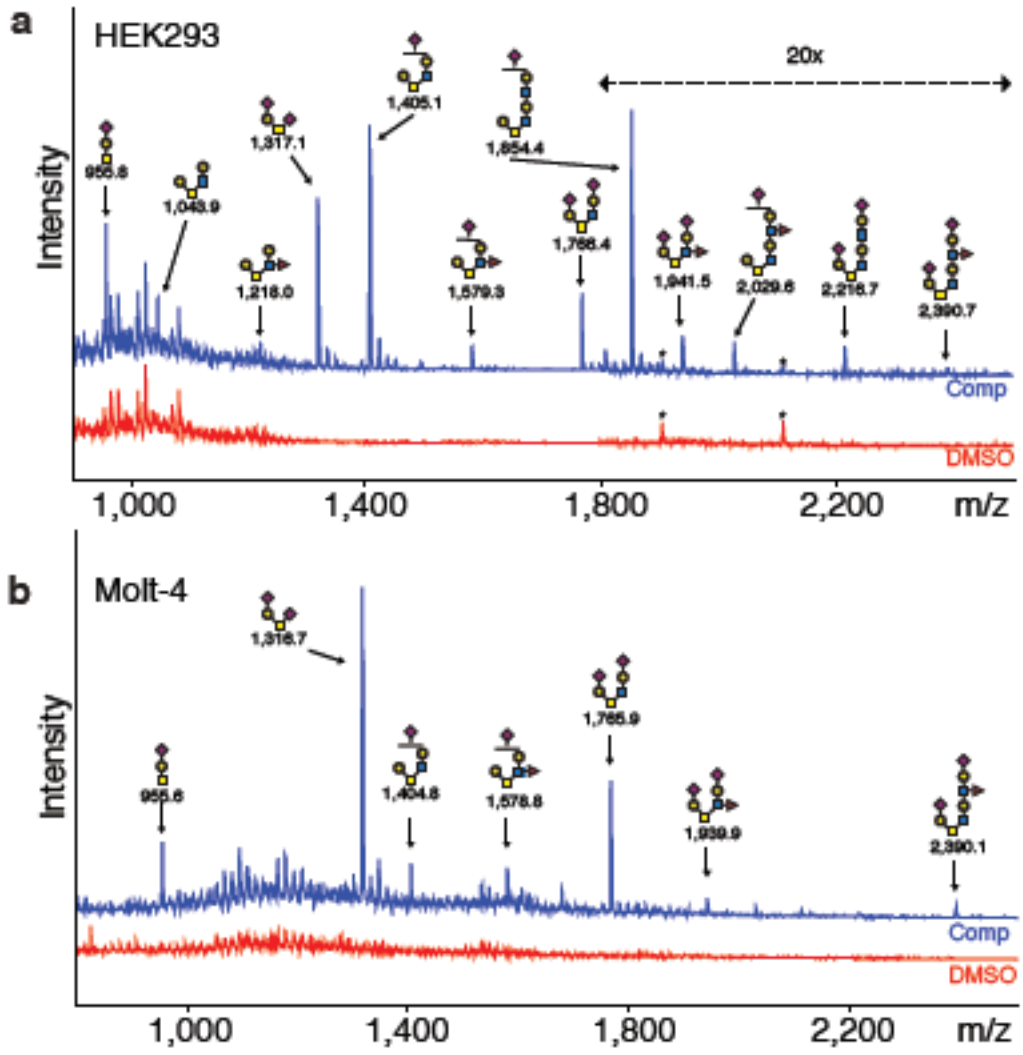
structure #	mass #	structure	mass	HL-60	WEHI-3	HUVEC
39	28		2,256		X	
40	29		2,503		X	
41	29		2,503			X
42	29		2,503			X
43	29		2,503			X
44	30		2,569	X		X
45	31		2,460		X	
46	31		2,460		X	
47	32		2,477			X
48	32		2,477			X
49	32		2,477			X
50	33		2,507		X	
51	33		2,507		X	
52	33		2,507		X	
53	34		2,568	X		
54	35		2,617		X	
55	35		2,617		X	
56	36		2,651			X
57	36		2,651			X
58	37		2,664			X
59	37		2,664			X
60	38		2,705		X	
61	38		2,705		X	
62	39		2,752			X
63	40		2,854			X
64	41		2,868		X	
65	42		2,956			X
66	42		2,956			X
67	43		2,956		X	
68	43		2,956		X	
69	43		2,956		X	
70	44		3,055		X	X
71	45		3,110		X	
72	45		3,110		X	
73	45		3,110			X
74	46		3,257			X
75	47		3,658		X	
76	48		3,474			X

Supplementary Table 3.3. Glycan structures from three cell lines determined by MS/MS.

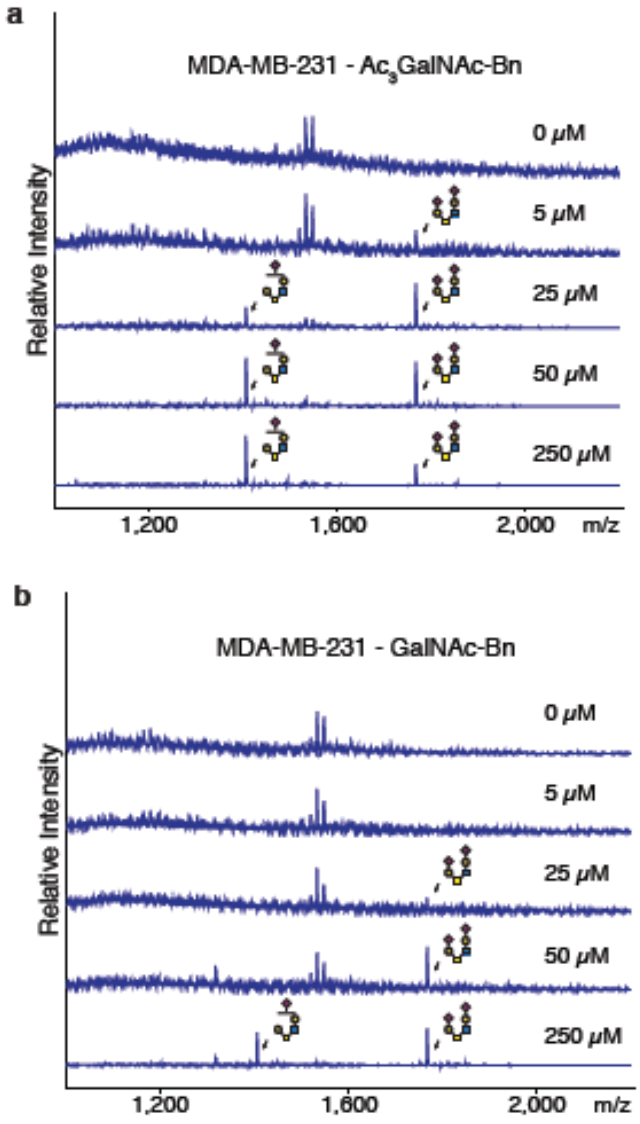


Supplementary Fig. 3.1. MALDI profiles of compounds used in this study.

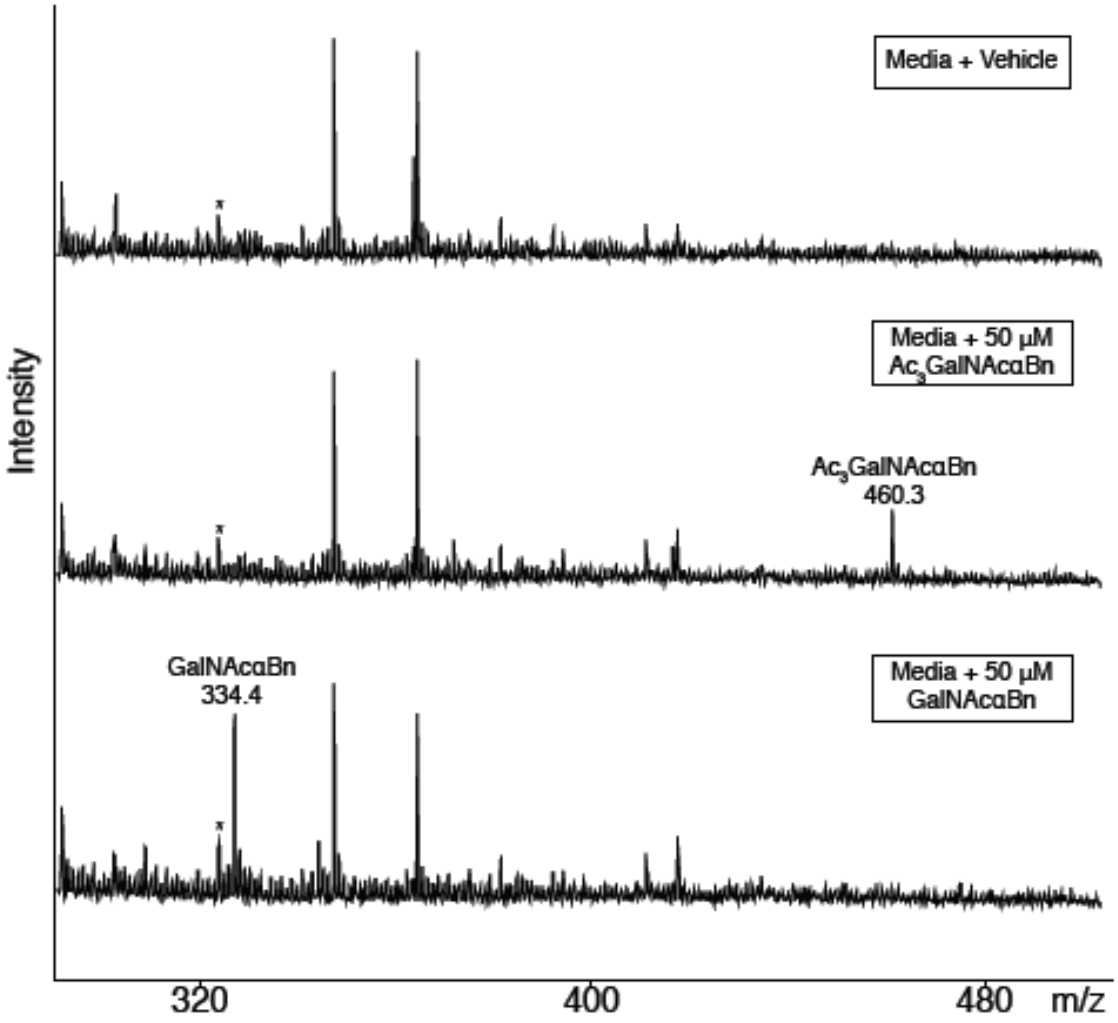
(a) Bn-GalNAc, Ac3GalNAc-Bn, and (b) Ac3GlcNAc-Bn.



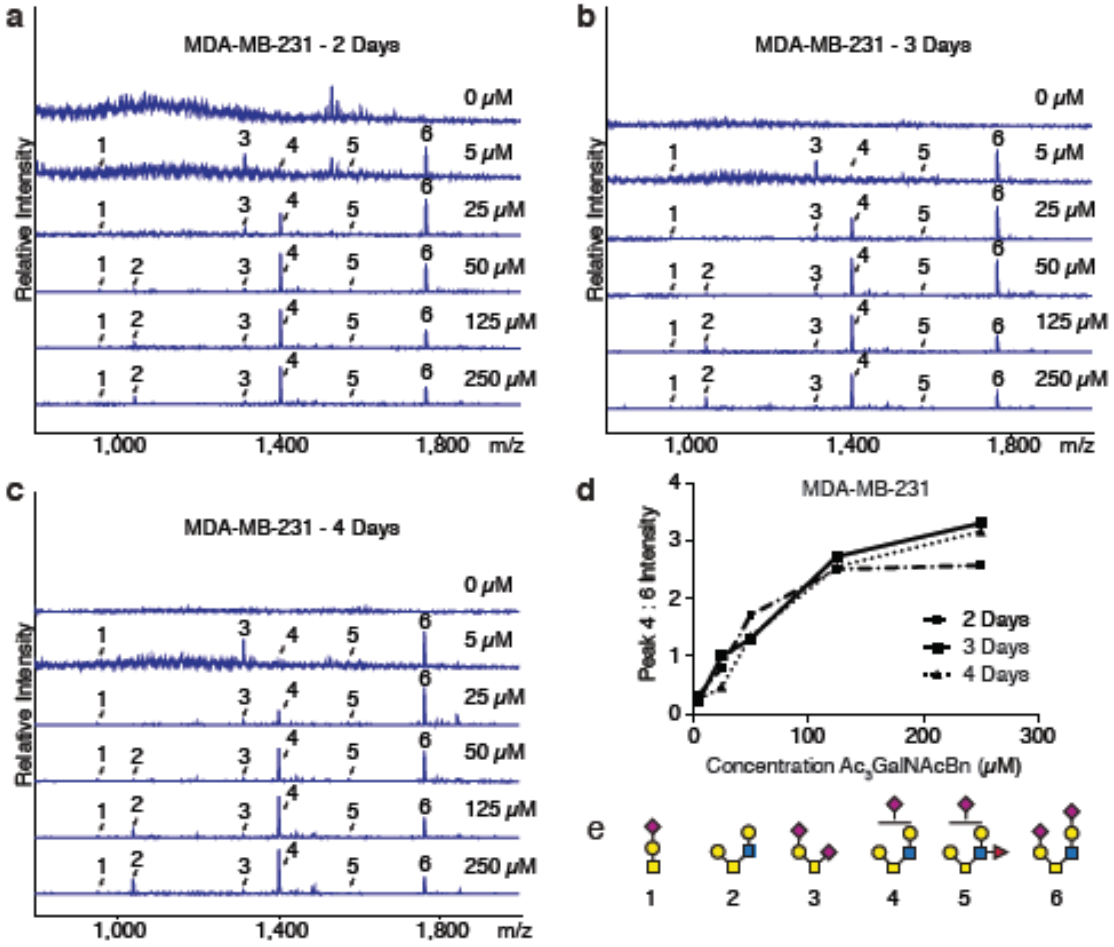
Supplementary Fig. 3.2. Bn-O-glycans produced by HEK293 and Molt-4 cells. (**a,b**) 50 μM Ac3GalNAc-Bn or DMSO was administered to HEK293 (**a**) or Molt-4 (**b**) cells and incubated for 3 days. Bn-O-glycans were purified from media, and one-third of eluent was permethylated and analyzed by MALDI-MS (composition). Spectra for Ac3GalNAc-Bn and DMSO are off-set, but scaled to same absolute intensity for each cell. Representative profiles are shown (n = 2).



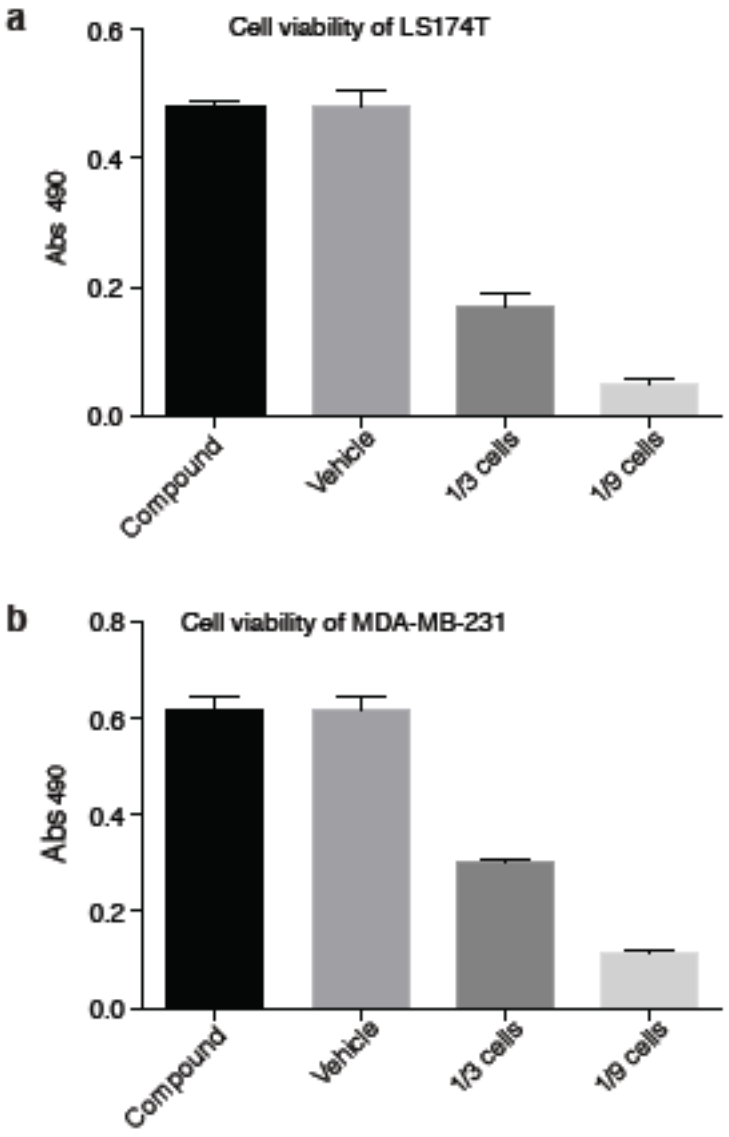
Supplementary Fig. 3.3. Bn-GalNAc peracetylation increases CORA sensitivity. (**a,b**) 0–250 μM of Ac3GalNAc-Bn (**a**) or Bn-GalNAc (**b**) was incubated with MDA-MB-231 cells for 3 days before collecting, purifying, permethylating, and analyzing Bn-O-glycans by MALDI to assess sensitivity of the compounds for reporting the O-glycome. Spectra are off-set for each concentration and scaled relative to maximum intensity. Representative profiles are shown (n = 2).



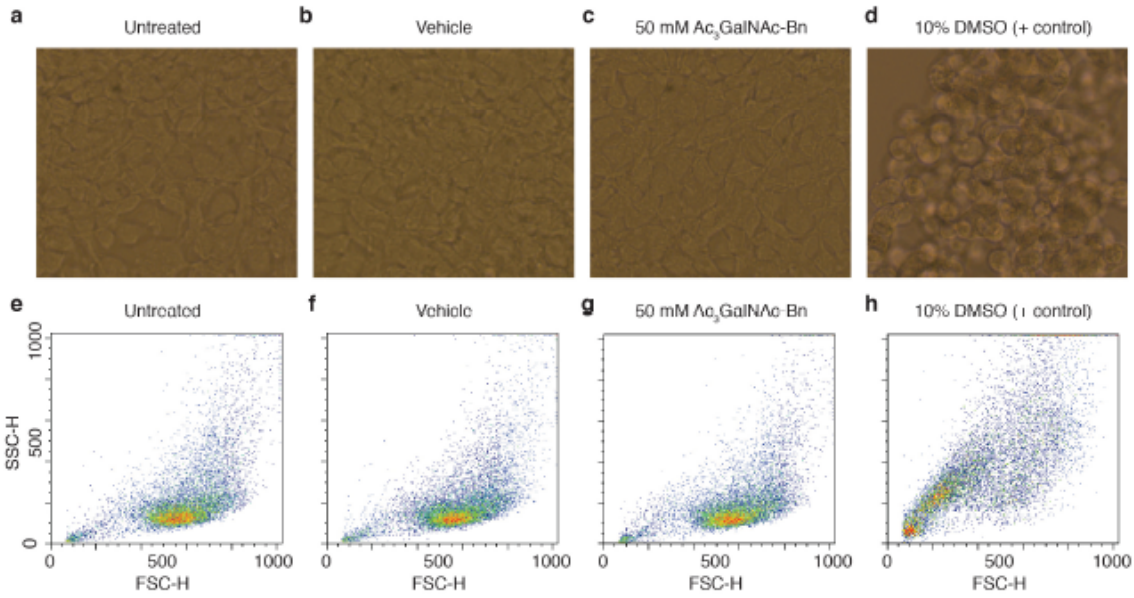
Supplementary Fig 3.4. Stability of acetylation in complete media. 50 µM Bn-GalNAc, Ac3GalNAc-Bn, or vehicle was added to DMEM with 5% FBS and incubated at 37°C, 5% CO2 for 3 days. Media was directly analyzed by MALDI (n = 2).



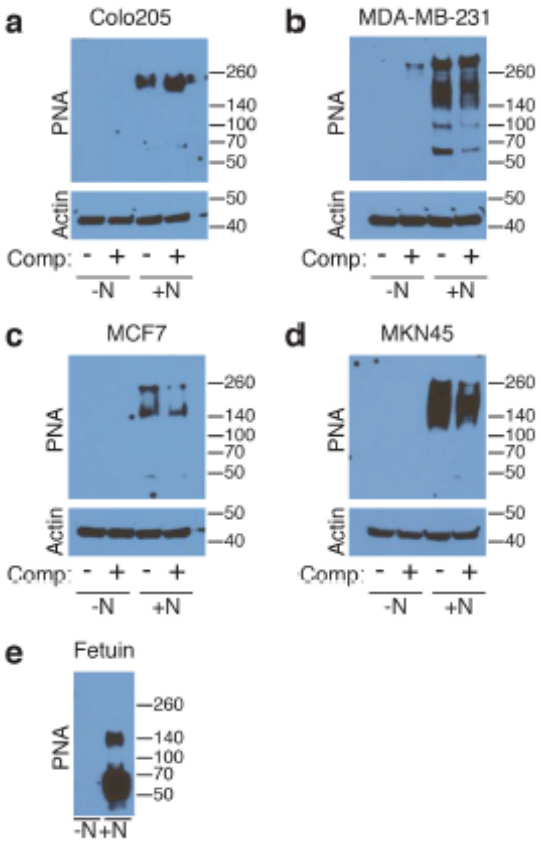
Supplementary Fig. 3.5. Optimizing concentration of Ac3GalNAc-Bn and incubation time. (**a–c**) MDA-MB-231 breast cancer cells were incubated with 0 – 250 μM Ac3GalNAc-Bn for 2 (**a**), 3 (**b**), or 4 days (**c**). Profiles did not change across time. (**d**) Increasing concentration resulted in a shift in relative intensities of disialylated Core 2 (glycan #6) to monosialylated Core 2 (glycan #4). Glycans are labeled 1 – 6 (**a–d**) 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).



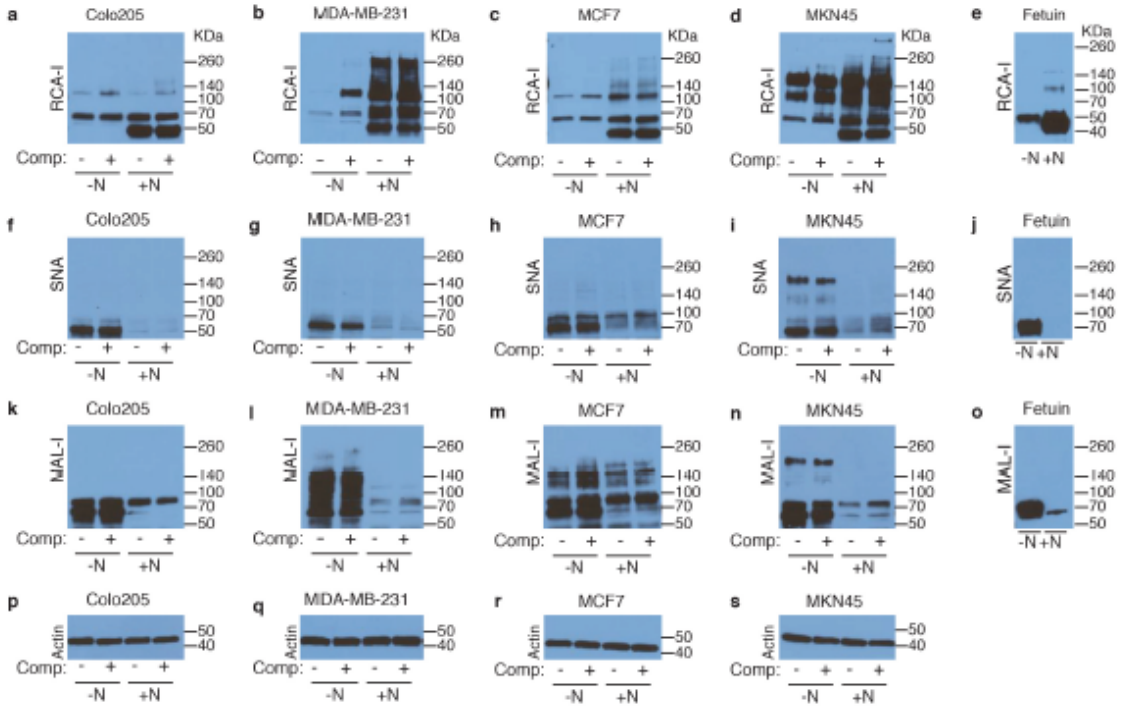
Supplementary Fig. 3.6. CORA does not alter cell viability. (**a**) LS174T-Tn(-) and (**b**) MDA-MB-231 cells were seeded in 96 well plates in triplicate. 50 μM Ac3GalNAc-Bn or DMSO vehicle was added to cells for 3 days and XTT assay was performed. One-third and 1/9th dilutions of cells were plated as a positive control for reduced cell numbers. No difference was observed between Ac3GalNAc-Bn and DMSO for either cell line. Representative experiments are shown (n = 2), mean ± SD of triplicates.



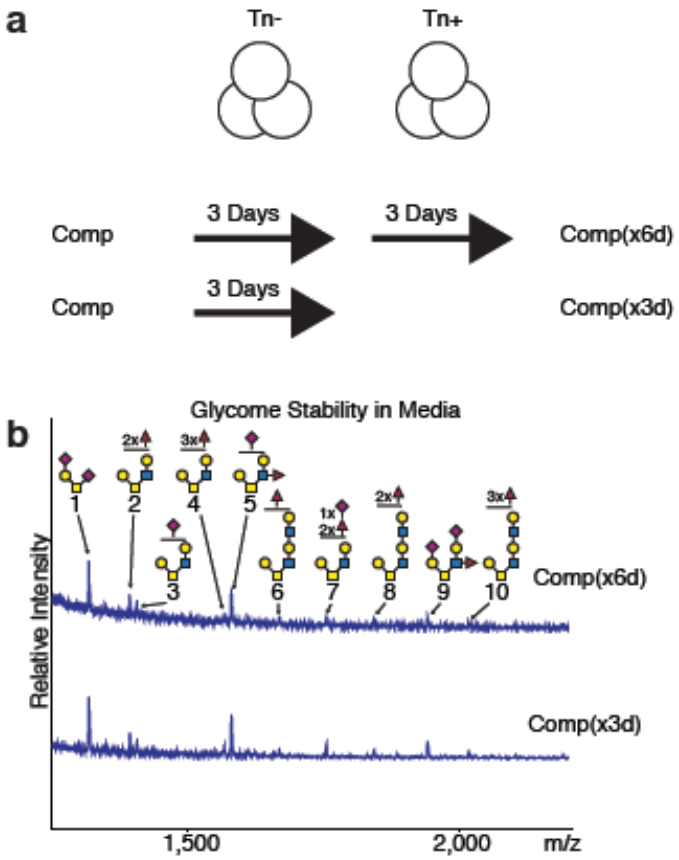
Supplementary Fig. 3.7. CORA does not alter cell morphology or granularity. (**a–h**) HEK293 cells were incubated for 3 days with no treatment (**a, e**), vehicle (**b, f**), 50 μM Ac3GalNAc-Bn (**c, g**), or 10% DMSO (**d, h**) as a positive control. Bright field microscopy was used to assess changes in cell morphology (**a–d**) and side scatter was determined by FACS to assess cell granularity (**e–h**). No changes were observed upon treatment with 50μM Ac3GalNAc-Bn (n = 2).



Supplementary Fig. 3.8. CORA does not alter cell surface O-glycosylation. (**a**) Colo205, (**b**) MDA-MB-231, (**c**) MCF7, and (**d**) MKN45 cells were incubated with 50 μM Ac3GalNAc-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. Actin was used as loading control (**a**–**d**) and fetuin was used as a control for neuraminidase treatment and lectin staining (**e**). Without neuraminidase treatment, Ac3GalNAc-Bn did not reveal significant PNA binding, indicating that 50 μM Ac3GalNAc-Bn did not inhibit addition of sialic acid to Core 1 structure. With neuraminidase treatment, Ac3GalNAc-Bn did not reduce PNA binding, indicating that 50 μM Ac3GalNAc-Bn did not block Core 1 O-glycan synthesis. Representative blots are shown (n = 2).

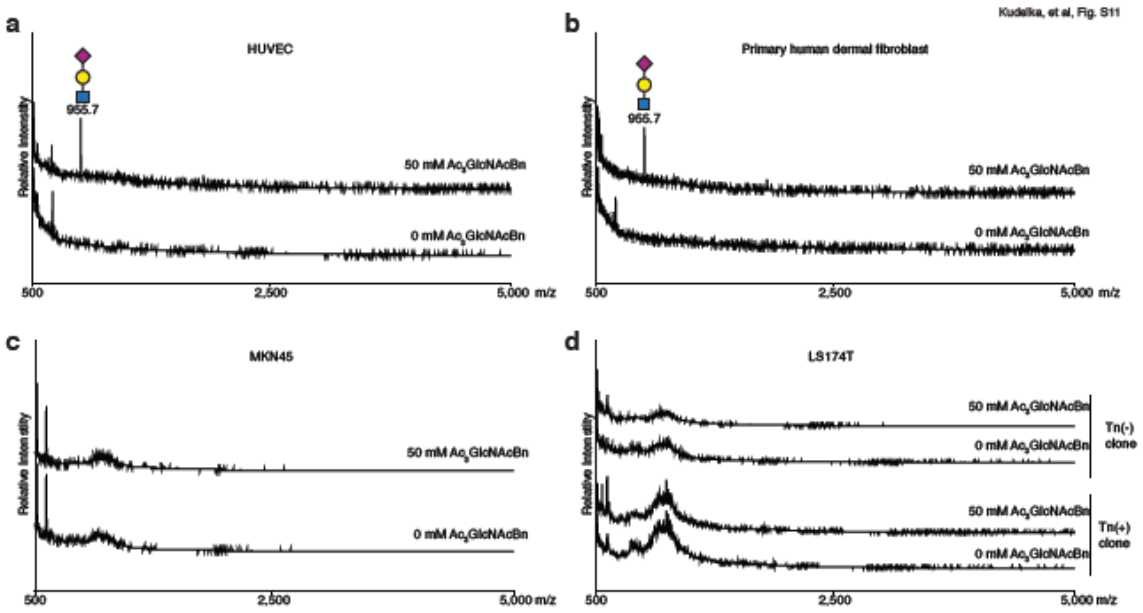


Supplementary Fig. 3.9. CORA does not alter cell surface N-glycosylation. (**a–s**) Colo205 (**a, f, k, p**), MDA-MB-231 (**b, g, l, q**), MCF7 (**c, h, m, r**), and MKN45 (**d, i, n, s**) cells were incubated with 50 μM Ac3GalNAc-Bn or DMSO for 3 days. Cells were then collected, lysed, and one fraction was treated with neuraminidase, before running a denaturing, reducing gel. Lysates were then blotted for RCA-I (**a–e**), which recognizes Galβ4GlcNAc of N-glycans, SNA (**f–j**), which recognizes α6-linked sialic acids, MAL-I (**k–o**), which recognizes α3-linked sialic acids, and actin as loading control (**p–s**). (**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 Ac3GalNAc-Bn or DMSO incubated cells, indicating that 50 μM Ac3GalNAc Bn did not disrupt addition of sialic acid to N-glycan termini or biosynthesis of N glycans. Representative blots are shown (n = 2).

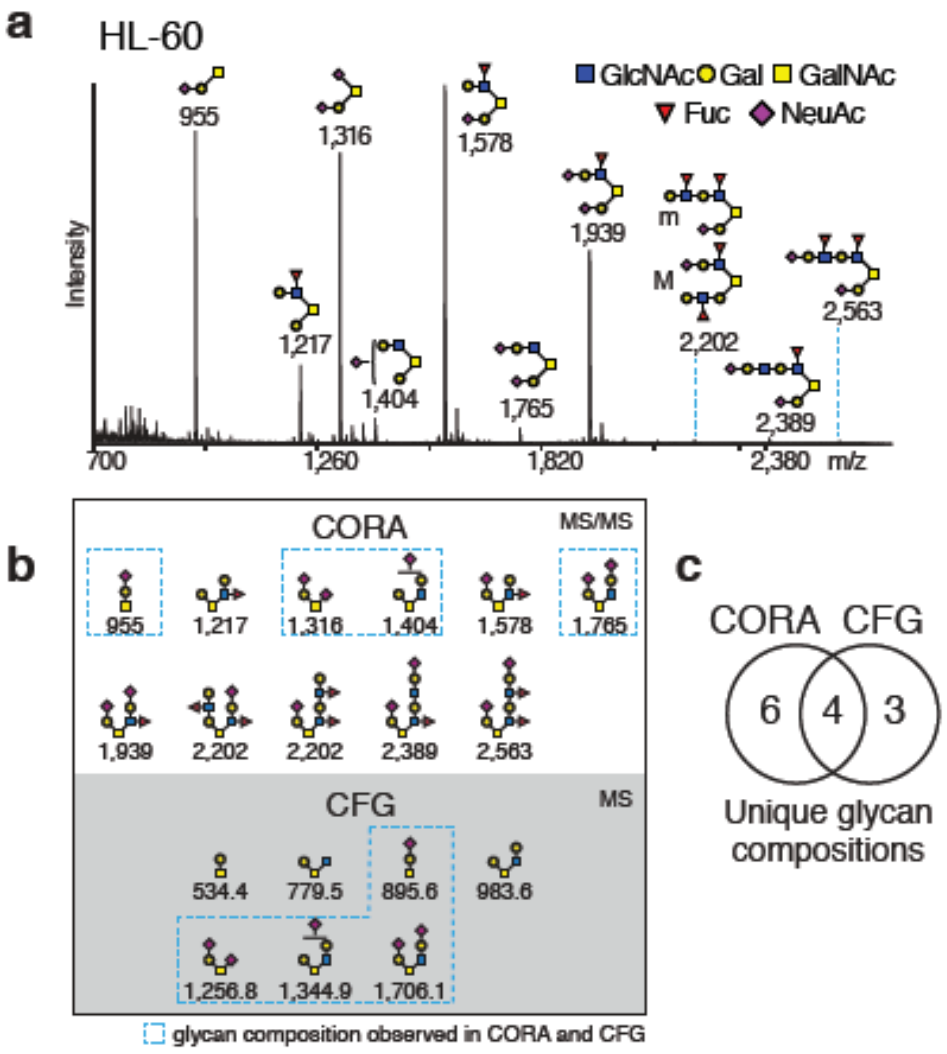


m/z	Glycan # 1	2	3	4	5	6	7	8	9	10
Comp(x6)	1,316.7	1,391.8	1,404.8	1,565.8	1,578.9	1,666.9	1,753.8	1,841.8	1,940.8	2,015.9
Comp(x3)	1,316.8	1,392.0	1,404.9	1,566.1	1,579.1	1,667.0	1,753.2	1,841.2	1,940.7	2,015.3

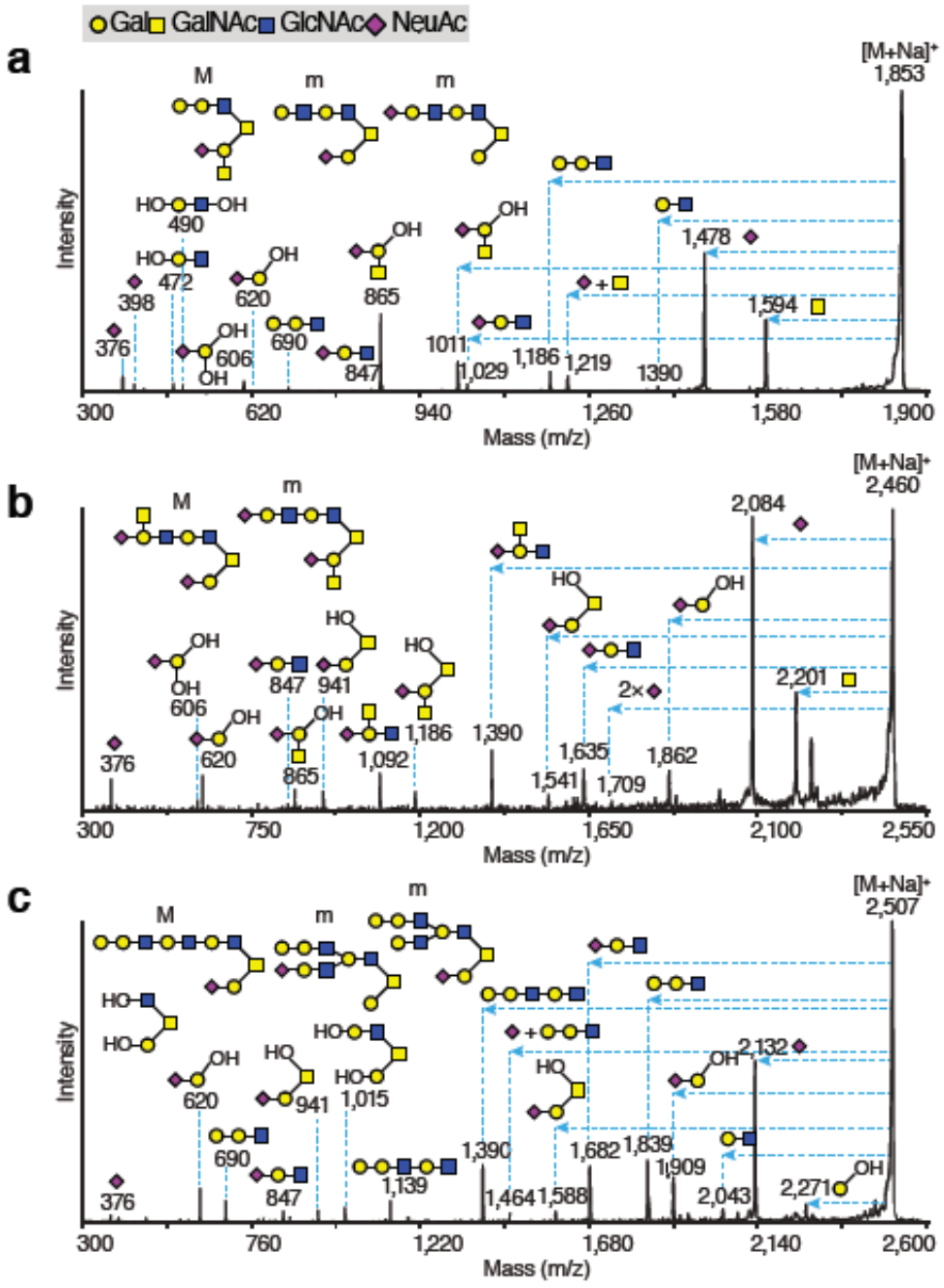
Supplementary Fig. 3.10. Stability of O-glycans in media with cells. (**a**) LS174T-Tn(-) cells were seeded overnight. Ac3GalNAc-Bn was added. After 3 days, media was collected and stored for further analysis or mixed 1:1 with fresh media and incubated for 3 days with LS174T-Tn(+). (**b**) Bn-O-glycans 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).

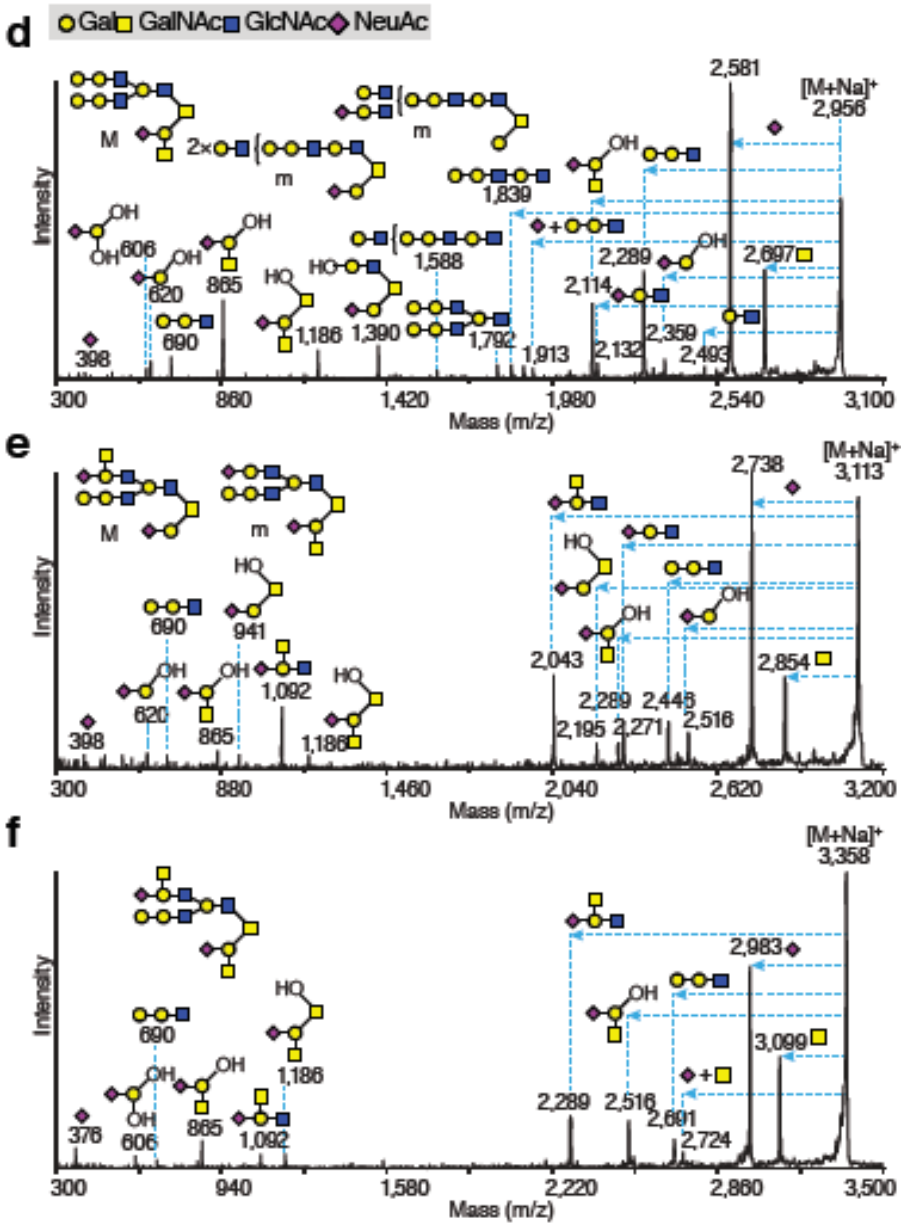


Supplementary Fig. 3.11. Evaluation of potential Bn-glycans produced from cells incubated with Ac3GlcNAc-Bn. 50 μM Ac3GlcNAc-Bn or vehicle was incubated with (**a**) HUVEC, (**b**) primary human dermal fibroblasts, (**c**) MKN45, and (**d**) LS174T-Tn(-) and Tn(+) cells. Potential Bn-glycans were purified, permethylated, and analyzed by MALDI.

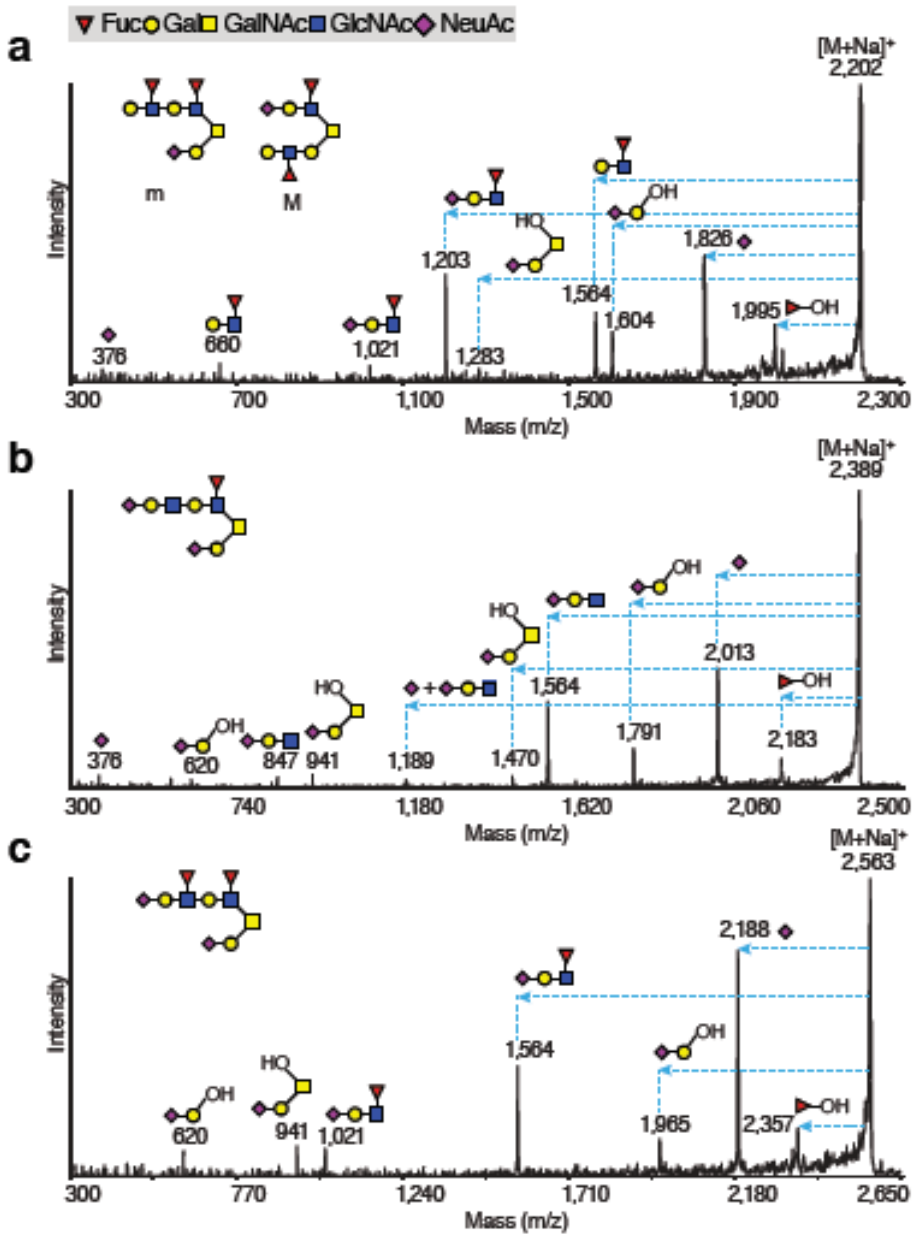


Supplementary Fig. 3.12. Accuracy of CORA for profiling the O-glycome. (**a**) 50 μM Ac3GalNAc-Bn was incubated with HL-60 for 3 days and Bn-O-glycans were purified, permethylated, and analyzed by MALDI-TOF-MS. (**b**) Bn-O-glycans from CORA (MS/MS structures) were compared to profiles from alkaline β-elimination (MS compositions) (CFG); blue boxes indicate glycan compositions observed in both CORA and β-elimination. (**c**) The number of glycans (by composition) observed in CORA, β elimination (CFG), or both are indicated. Putative structures are based on composition, tandem MS, and biosynthetic knowledge.

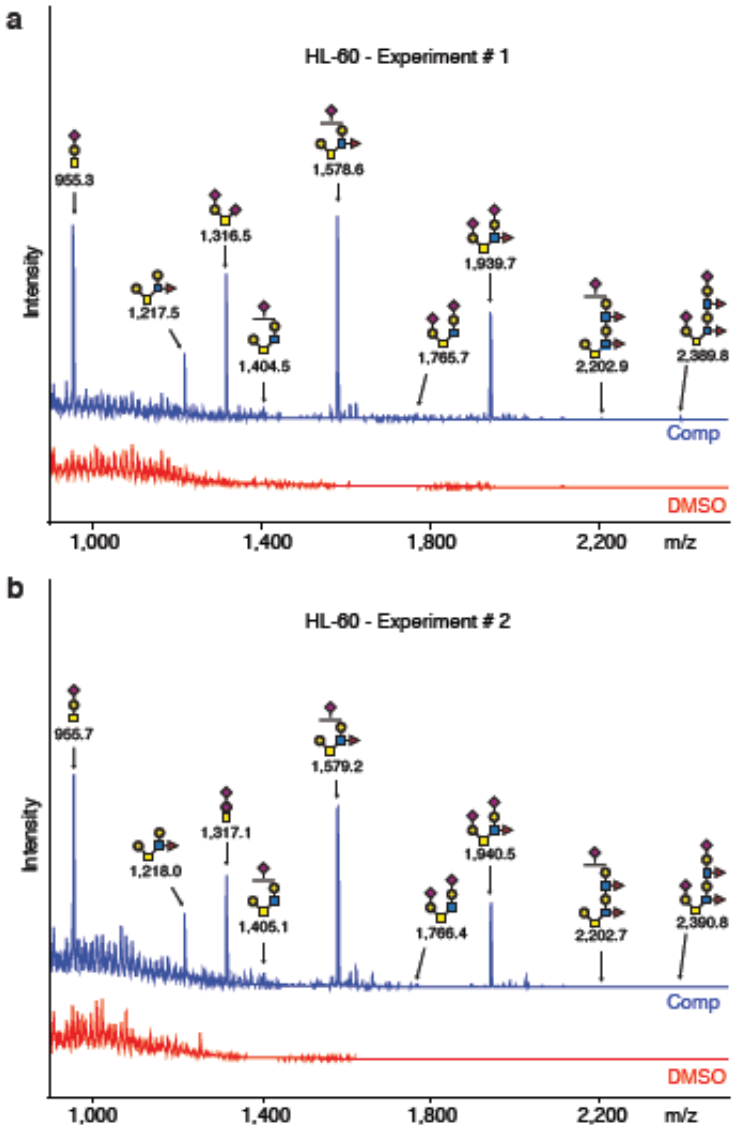

Gal GalNAc GlcNAc NeuAc
a
M
m
m
[M+Na]+
1,853
Intensity
490
472
398
376
620
606
690
865
847
1011
1,029
1,186
1,219
1390
1,478
1,594
300
620
940
1,260
1,580
1,900
Mass (m/z)
b
M
m
[M+Na]+
2,460
2,084
2,201
606
620
847
941
865
376
1,092
1,186
1,390
1,541
1,635
1,709
1,862
300
750
1,200
1,650
2,100
2,550
Mass (m/z)
c
M
m
m
[M+Na]+
2,507
620
941
1,015
690
847
376
1,139
1,390
1,464
1,588
1,682
1,839
1,909
2,043
2,132
2,271
300
760
1,220
1,680
2,140
2,600
Mass (m/z)



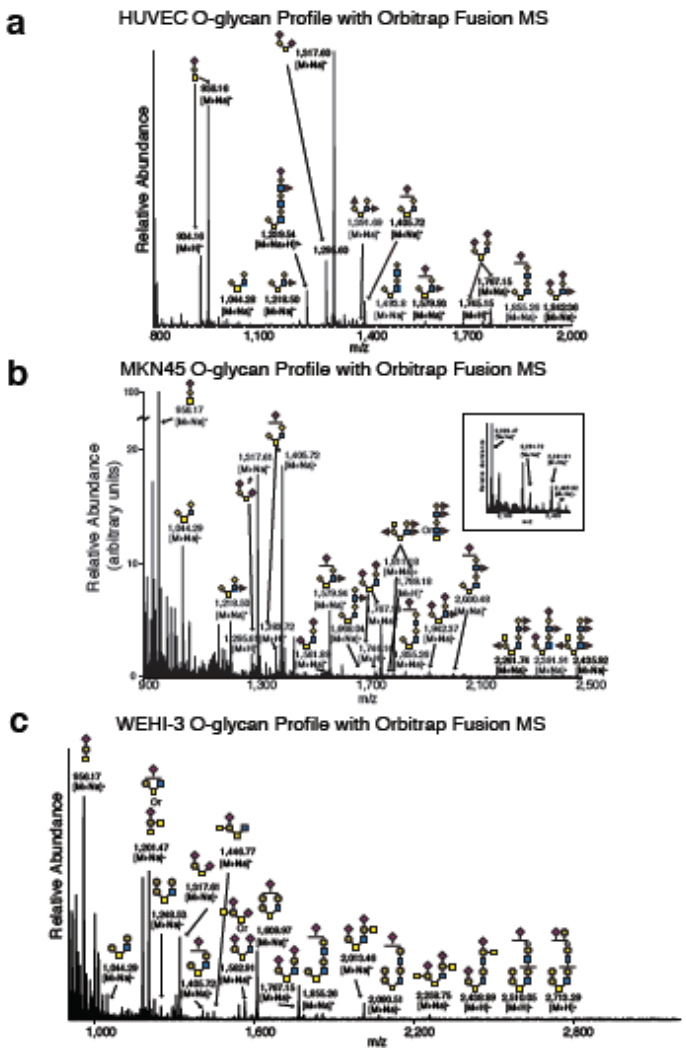
Supplementary Fig. 3.13. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from WEHI-3 cells from **Fig. 3.4a**: (**a**) m/z 1853, (**b**) m/z 2460, (**c**) m/z 2507, (**d**) m/z 2956, (**e**) m/z 3113, (**f**) m/z 3358. All molecular ions are [M+Na]+. Horizontal blue arrows indicate the 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.



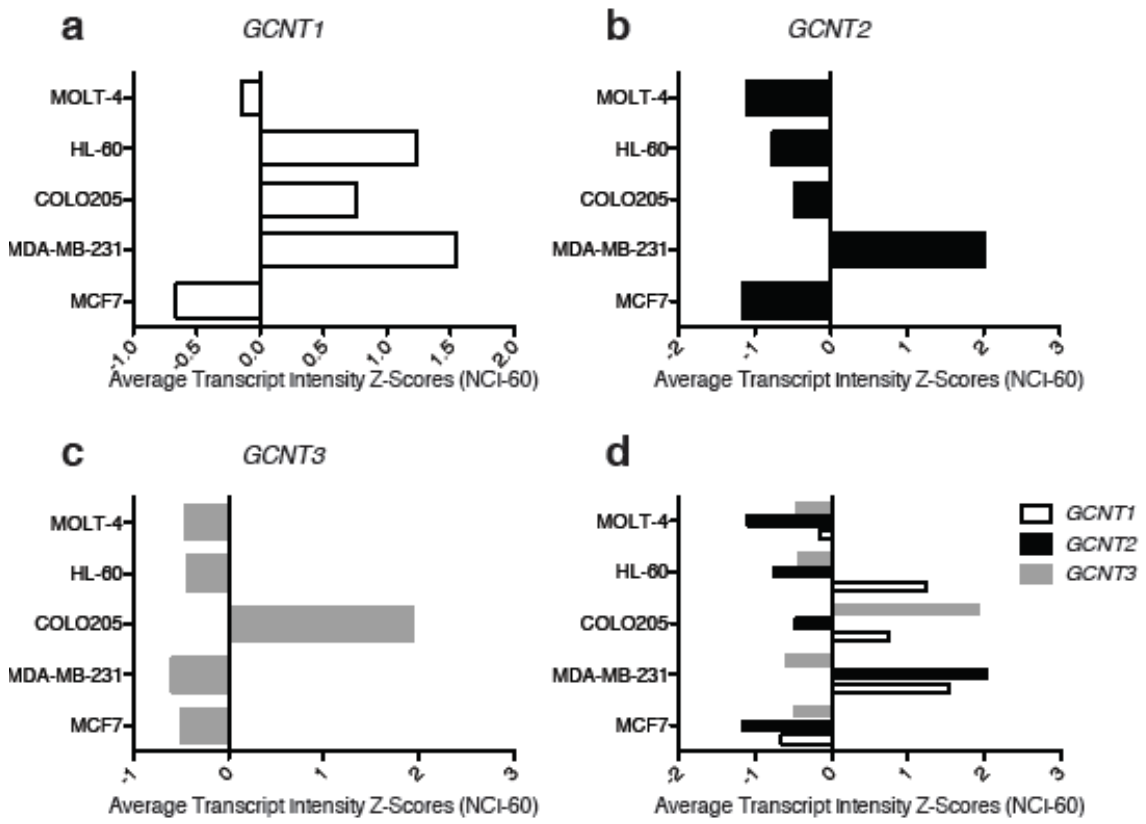
Supplementary Fig. 3.14. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from HL-60 cells. (**a**) m/z 2202, (**b**) m/z 2389, (**c**) m/z 2563. All molecular ions are [M+Na]+. Horizontal blue arrows indicate the losses indicated from the molecular ion. Vertical blue lines indicate the corresponding peak ion. Letters "m" and "M" in bold characters suggest minor and major abundances respectively.



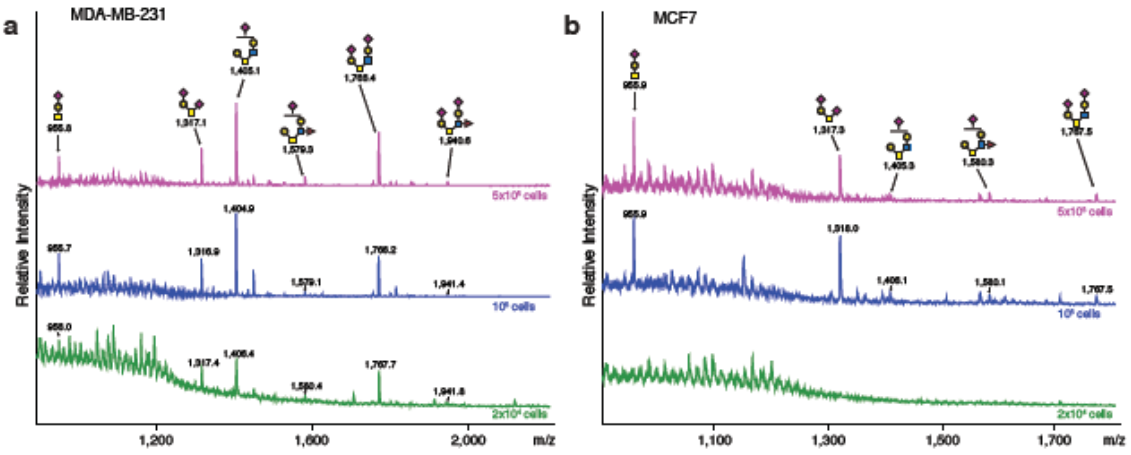
Supplementary Fig. 3.15. Reproducibility of CORA. (**a, b**) Ac3GalNAc-Bn was added to HL-60 cells and Bn-O-glycans were purified, permethylated and analyzed. MALDI-MS profiles from two independent experiments are shown and nearly identical. MALDI MS/MS experiments described in **Supplementary Fig. 3.12** were independently performed but used same material as those used for (**a**) here. Ac3GalNAc-Bn and DMSO are off-set, but scaled to same absolute intensity for each experiment.



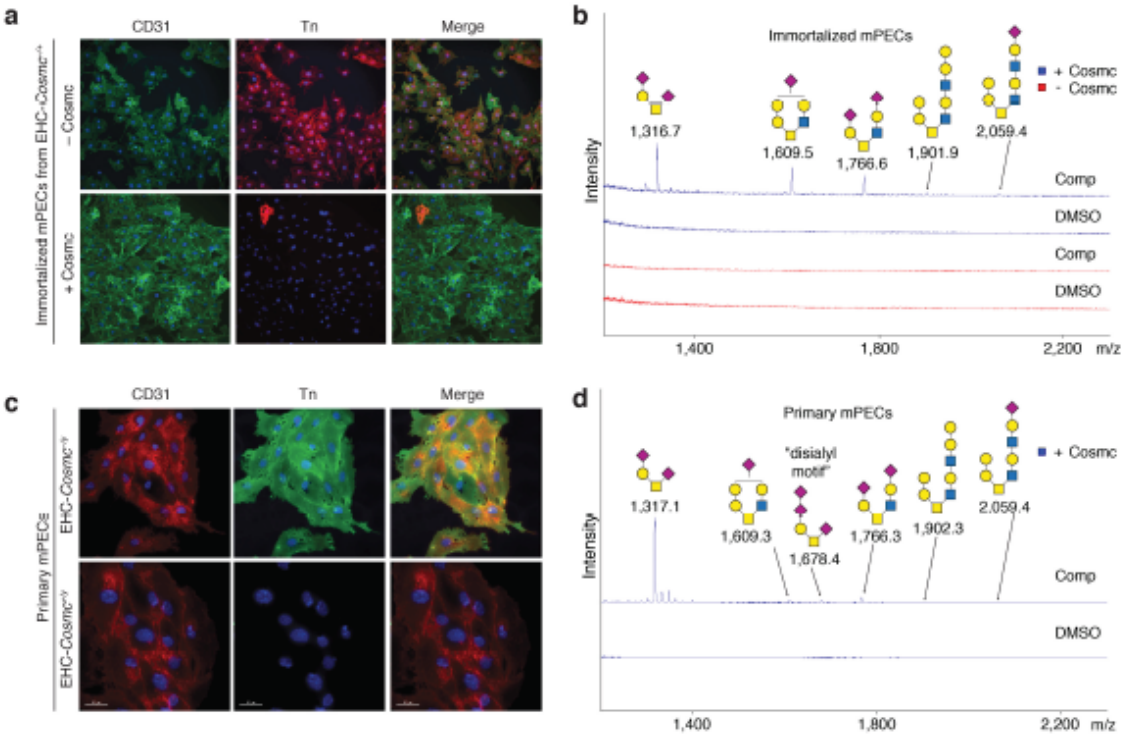
Supplementary Fig 3.16. Analysis of HUVEC, MKN45, and WEHI-3 O-glycans by ESI MS. 50 µM Ac3GalNAcBn was incubated with (**a**) HUVEC, (**b**) MKN45, and (**c**) WEHI 3 cells for 3 days and Bn-O-glycans were purified and permethylated. To validate our MALDI-MS data from **Fig. 3.3a**, **3.4a**, and **3.5** on a different mass spectrometry platform, we analyzed permethylated samples by ESI-MS on an Orbitrap Fusion mass spectrometer in positive ion mode with full MS1 scan. Spectrum is enlarged for 2000-2500 m/z in inset for (**b**). Profiles were similar from ESI-MS versus MALDI-MS. Glycans indicated were not detected in vehicle only control (not shown).



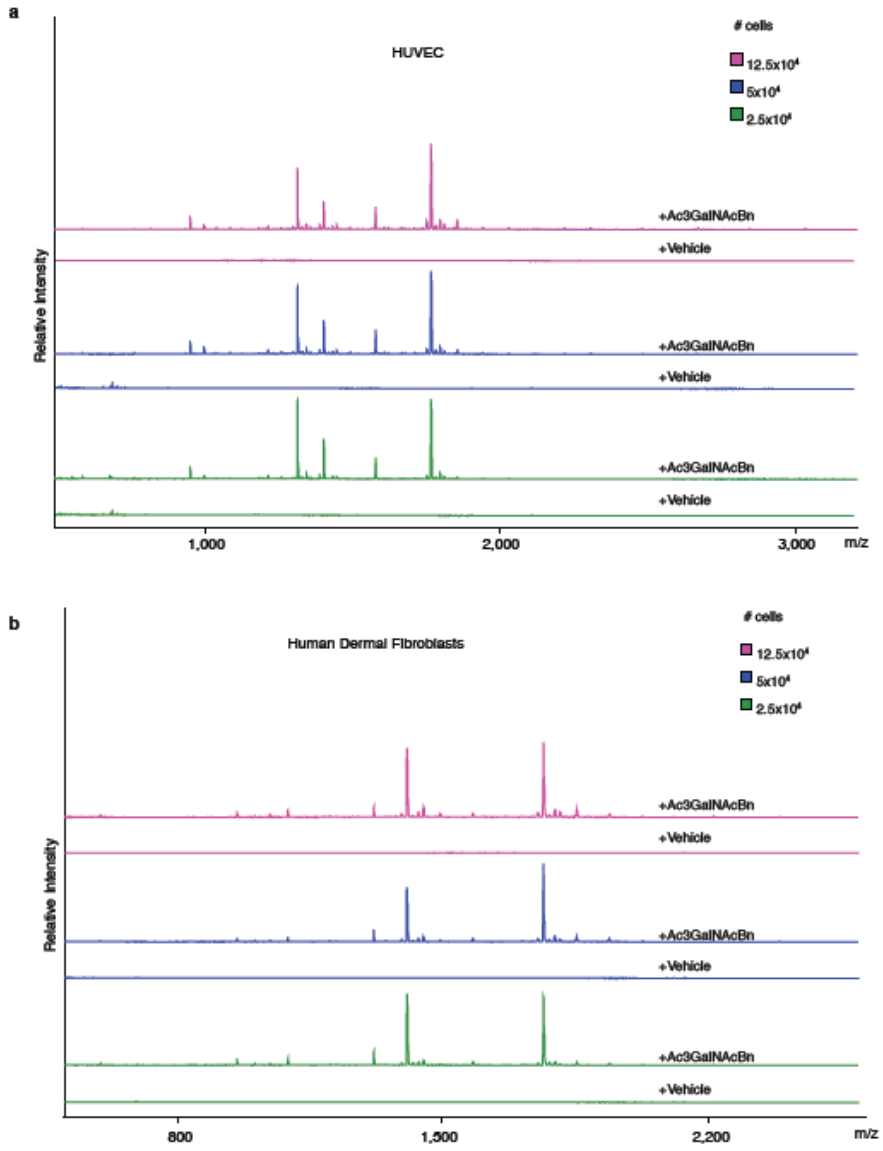
Supplementary Fig. 3.17. Transcript levels of Core 2 GnT1–3 in cell lines. (**a–d**) Average Transcript Intensity Z-Scores based on NCI-60 cells for MOLT-4, HL-60, COLO205, MDA-MB-231, and MCF7 for GCNT1 (C2GnT1) (**a**), GCNT2 (C2GnT2) (**b**), GCNT3 (C2GnT3) (**c**), and GCNT1–3 (**d**). Analysis was performed on NCI-60 public data through CellMiner(428, 429).



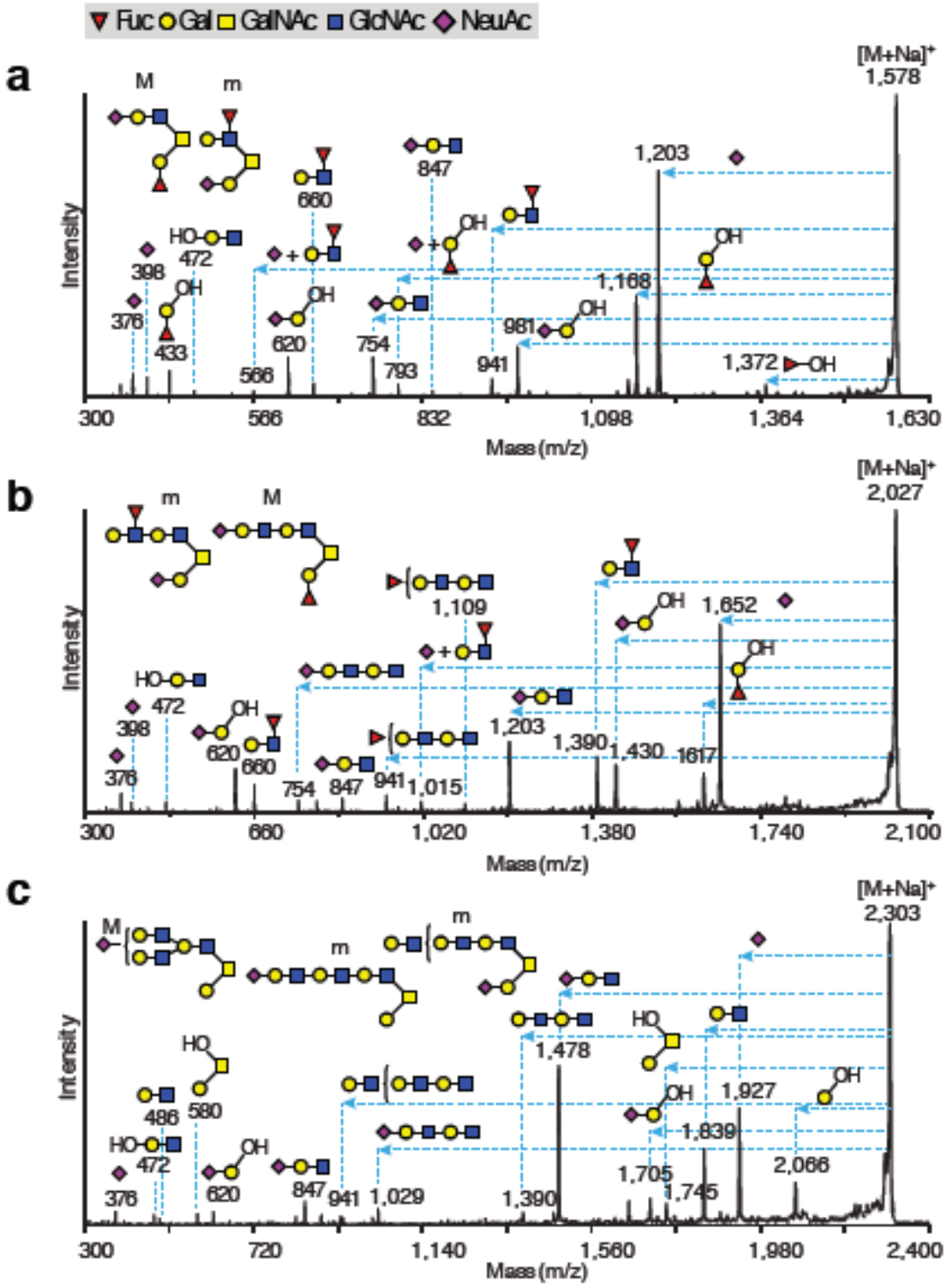
Supplementary Fig. 3.18. Sensitivity of CORA. (**a**) MDA-MB-231 and (**b**) MCF7 cells were seeded at the indicated cell densities, 50 μM Ac3GalNAc-Bn was added, and Bn-O glycans were purified and permethylated after 3 days. A fraction of total glycans were analyzed by MALDI-MS (composition). Spectra are off-set for each seeding density and scaled relative to maximum intensity. Representative profiles are shown (n = 2).

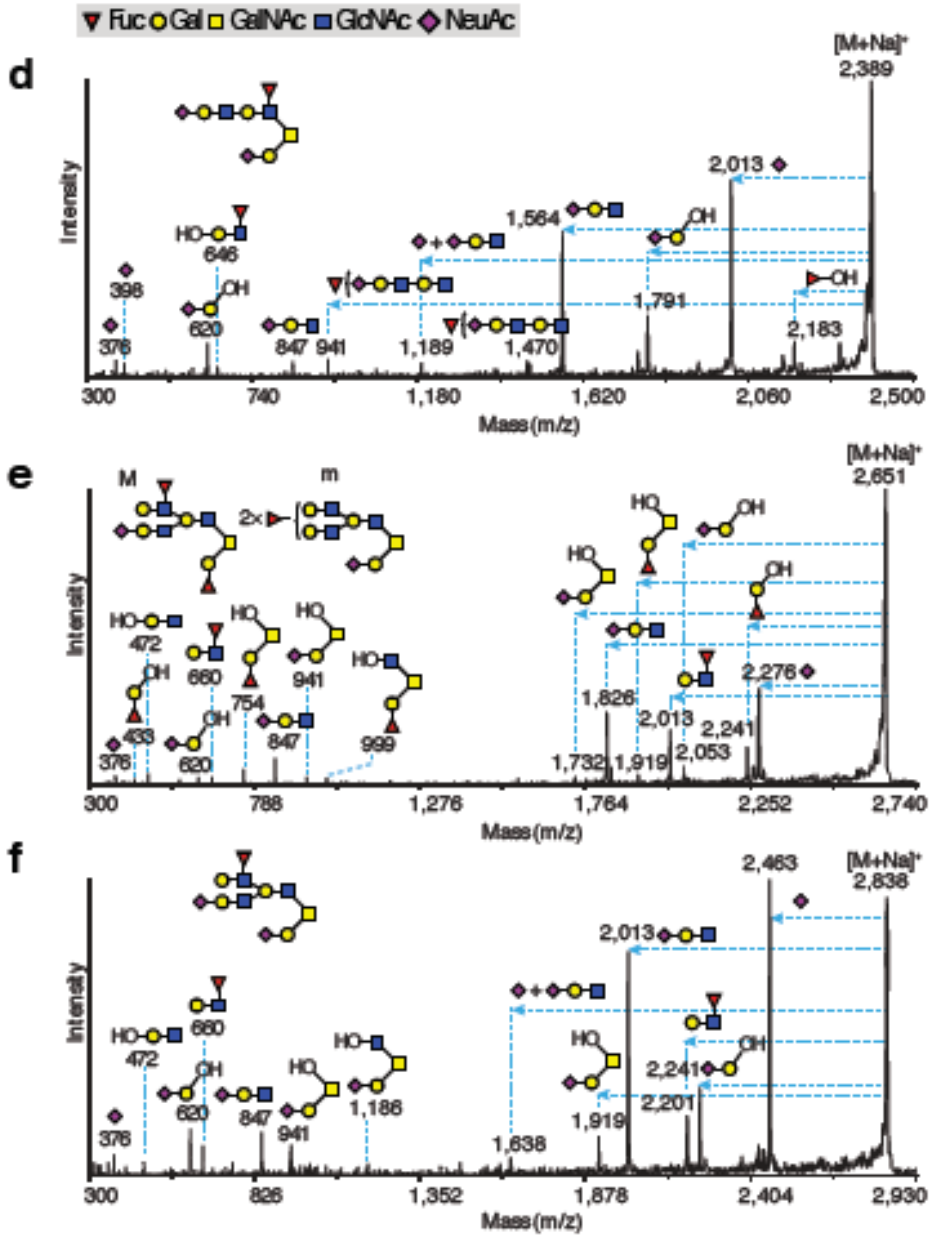


Supplementary Fig. 3.19. O-glycome of murine pulmonary endothelial cells. (**a–d**) Tn(-) and Tn(+) murine pulmonary endothelial cells (mPECs), with or without Cosmc, respectively, were obtained from mosaic EHC-Cosmc+/- mice (immortalized) or EHC Cosmc+/y or EHC-Cosmc-/y mice (primary). Tn expression and CD31 (an endothelial marker) was analyzed by immunofluorescence for immortalized (**a**) and primary cells (**c**). Ac3GalNAc or DMSO was added to immortalized (**b**) or primary cells (**d**) 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).

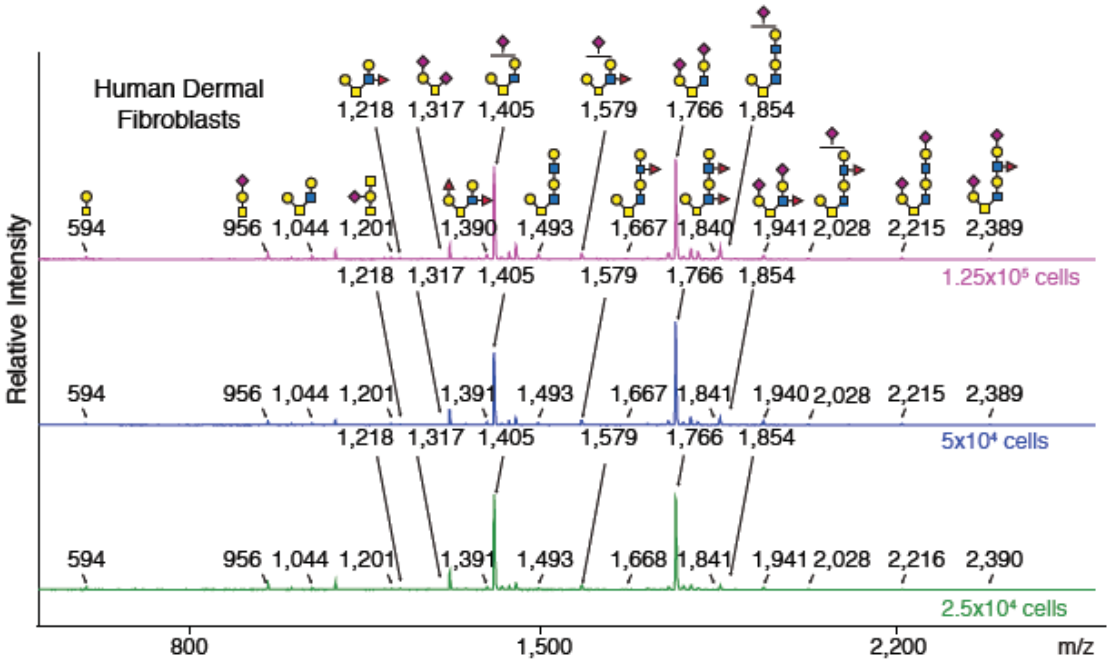


Supplementary Fig. 3.20. Profiling the O-glycome of Primary Cells. Ac3GalNAc-Bn or DMSO was added to (**a**) HUVECs or (**b**) primary human dermal fibroblasts seeded at 12.5 Å~ 104, 5 Å~ 104, or 2.5 Å~ 104 cells/well. Glycan structures and masses are omitted for clarity, but similar to those indicated in **Fig. 3.5** and **Supplementary Fig. 3.22**. 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).

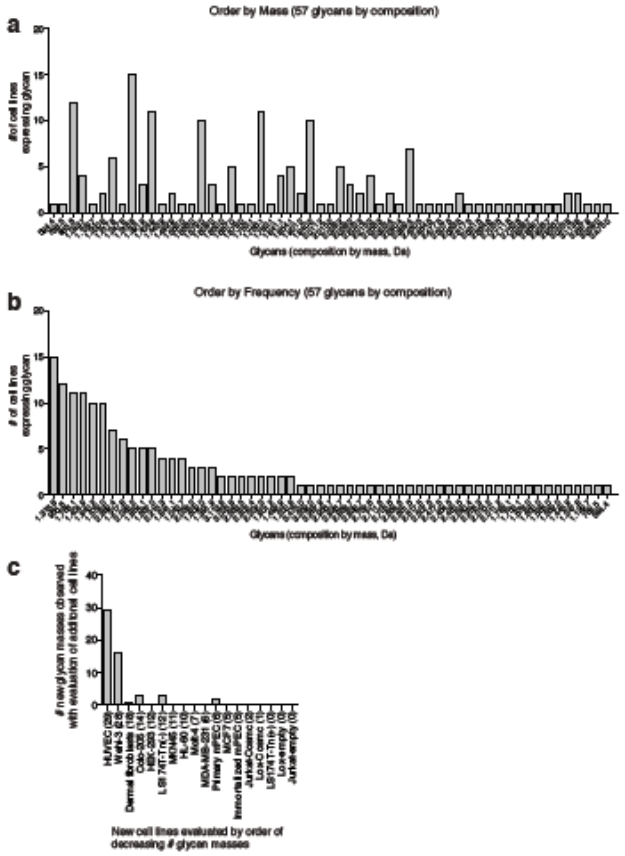

Fuc Gal GalNAc GlcNAc NeuAc
a
M
m
[M+Na]+
1,578
660
847
1,203
HO
472
OH
398
376
433
620
566
754
793
941
981
1,168
1,372
Intensity
300
566
832
1,098
1,364
1,630
Mass (m/z)
b
m
M
[M+Na]+
2,027
1,109
1,652
1,203
1,390
1,430
1617
1,015
660
1,020
1,380
1,740
2,100
c
M
m
[M+Na]+
2,303
1,478
580
486
1,927
1,839
2,066
1,705
1,745
1,029
720
1,140
1,560
1,980
2,400



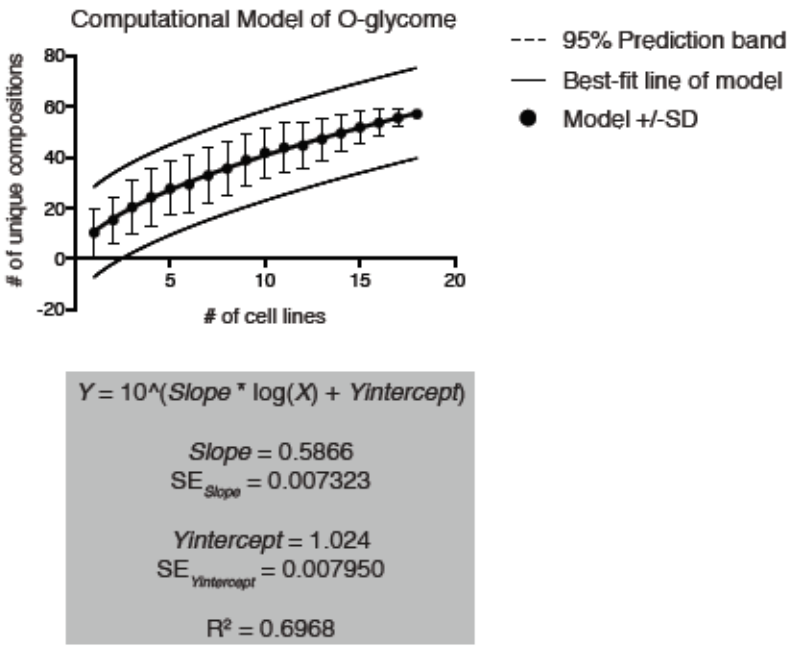
Supplementary Fig. 3.21. MALDI-TOF/TOF-MS/MS analysis of permethylated O-glycans derived from HUVEC cells from **Fig. 3.5**: (**a**) m/z 1578, (**b**) m/z 2027, (**c**) m/z 2303, (**d**) m/z 2389, (**e**) m/z 2651, (**f**) m/z 2838. All molecular ions are [M+Na]+. Horizontal blue arrows indicate the 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. 3.22. MALDI-TOF-MS profiling of the O-glycome of primary human dermal fibroblasts. Different numbers of cells, 12.5 Å~ 104, 5 Å~ 104, and 2.5 Å~ 104 were seeded in T25, 6 well, or 12 well flasks, respectively. 50 μM Ac_3GalNAc-Bn was added after 2 days, and Bn-O-glycans were purified, permethylated, and analyzed by MALDI after 3 more days. MS (composition) analysis was performed. Spectra are off-set for each seeding density and scaled relative to maximum intensity. Putative structures are based on composition, and biosynthetic knowledge.



Supplementary Fig. 3.23. Glycan frequency across cell lines evaluated in CORA. (**a,b**) Histograms were generated based on data from Supplementary Table 2 indicating number of cells lines expressing a given glycan (mass listed on x-axis) (**a**); glycans were reordered from glycans expressed in greatest number of cells to glycans expressed in least number of cells (**b**). Eight unique glycans were each observed in >5 cell lines, 17 glycans in 2 – 5 cell lines, and 32 glycans in at most one 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. 3.24. Computational model to estimate the size of the 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. 256 iterations were plotted for each xi = 1 to 18 and the average of all yi's were 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 Y intercept as well as R2 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 nonsulfated and non-phosphorylated O-glycans.

Chapter 4

Discovery of a new risk gene for IBD involved in epithelial glycosylation:

Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk.

(Adapted from Kudelka et al, PNAS, 2016.)

Abstract

Inflammatory bowel disease (IBD) results from aberrant immune stimulation against a dysbiotic mucosal but relatively preserved luminal microbiota and preferentially affects males in early onset disease. However, factors contributing to sex-specific risk and the pattern of dysbiosis are largely unexplored. Cosmc, which encodes an X-linked chaperone important for glycocalyx formation, was recently identified as an IBD risk factor by GWAS. We deleted Cosmc in mouse intestinal epithelial cells (IECs) and found marked reduction of microbiota diversity in progression from the proximal to the distal gut mucosa, but not in the overlying lumen, as seen in IBD. This coincided with local emergence of a pro-inflammatory pathobiont and distal gut restricted pathology. Mechanistically, we found that Cosmc regulates host genes, bacterial ligands, and nutrient availability to control microbiota biogeography. Loss of one Cosmc allele in males (IEC-*Cosmc*$^{-/y}$) resulted in a compromised mucus layer, spontaneous microbe-dependent inflammation, and enhanced experimental colitis; however, females with loss of one allele and mosaic deletion of Cosmc in 50% of crypts (IEC-*Cosmc*$^{+/-}$) were protected from spontaneous inflammation and partially protected from experimental colitis, likely due to lateral migration of normal mucin glycocalyx from WT cells over KO crypts. These studies functionally validate Cosmc as an IBD risk factor and implicate it in regulating the spatial pattern of dysbiosis and sex bias in IBD.

Significance Statement

Inflammatory bowel disease (IBD) is a devastating illness that affects 1.6 million people in the United States and disproportionately affects males in early onset disease. However, IBD genes that contribute to sex-specific risk are unexplored. The gut microbiota interfaces the host with its environment and exhibits alterations in spatial organization in IBD with dysbiosis in the mucosa but a relatively unaffected lumen. Cosmc was recently identified as an IBD risk gene on the X-chromosome by GWAS. We functionally evaluated Cosmc in IBD and discovered that loss of Cosmc leads to gut inflammation in males but not females and a spatial pattern of dysbiosis resembling IBD. Thus, Cosmc contributes to sex bias in IBD and spatially regulates the gut microbiota.

Introduction

Inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), is a devastating disease that affects 1.6 million people in the United States. Symptoms include abdominal pain, diarrhea, rectal bleeding, and growth failure. Severe cases require bowel resection, with current treatments aimed at management rather than cure. Mechanistically, immune hyperactivation to gut bacteria damages the intestine. Early onset or pediatric IBD (<15 years old) occurs in 10 - 20% of patients and is a model to understand genetic contributions to disease(430). CD and possibly UC in patients <15 or <8 years old, respectively, preferentially affects males (M:F is ~1.5:1 for CD), indicating that genetic factors, such as risk genes on the X-chromosome, may explain this sex bias and provide new insights into disease mechanisms and targets(430).

The gut microbiota can contribute to IBD with changes in community structure and functional capacity(431). Normally, the bacterial distribution and composition vary greatly from ~$10^{11\text{-}12}$ cells/gram within the ascending colon, to ~$10^{7\text{-}8}$ in the distal ileum, and ~$10^{2\text{-}3}$ in proximal ileum and jejunum(431). Specific members of the microbiota may occupy a planktonic niche, free-living in the luminal fecal stream, or adherent to the mucosa(384). The distinct geographical and ecological niches in the intestine (biogeographies) expand microbial metabolic potential and help maintain diversity in a competitive ecosystem(384). A recent study of pediatric IBD identified a dysbiotic microbiota in the intestinal mucosa but not in the lumen, suggesting that specific factors regulate the gut microbe biogeography in the healthy intestine and that these factors are

altered in IBD(432). Nevertheless, no IBD risk genes are currently known to regulate gut microbe biogeography or the pattern of dysbiosis in patients.

Cosmc on the X-chromosome encodes a chaperone for the T-synthase glycosyltransferase that extends O-glycans on >80% of secretory pathway proteins to form the glycocalyx and mature mucins (**Fig. 4.1A**) (14, 33, 80). Cosmc was recently implicated in sex-specific risk in Crohn's and UC by GWAS(433); however, its functional role in the intestine and IBD is unexplored. To address this, we deleted *Cosmc* in mouse intestinal epithelial cells (IECs). Loss of one allele in male mice (IEC-*Cosmc*$^{-/y}$) led to spontaneous inflammation, enhanced experimental colitis, and altered microbe biogeography, as seen in IBD. Significantly, female mice with loss of one allele (IEC-*Cosmc*$^{-/+}$) were protected from spontaneous inflammation and partially protected from experimental colitis, despite complete KO of *Cosmc* in ~50% of crypts (due to random X inactivation). Thus, *Cosmc* is an IBD risk gene that regulates microbe biogeography and disproportionately affects males.

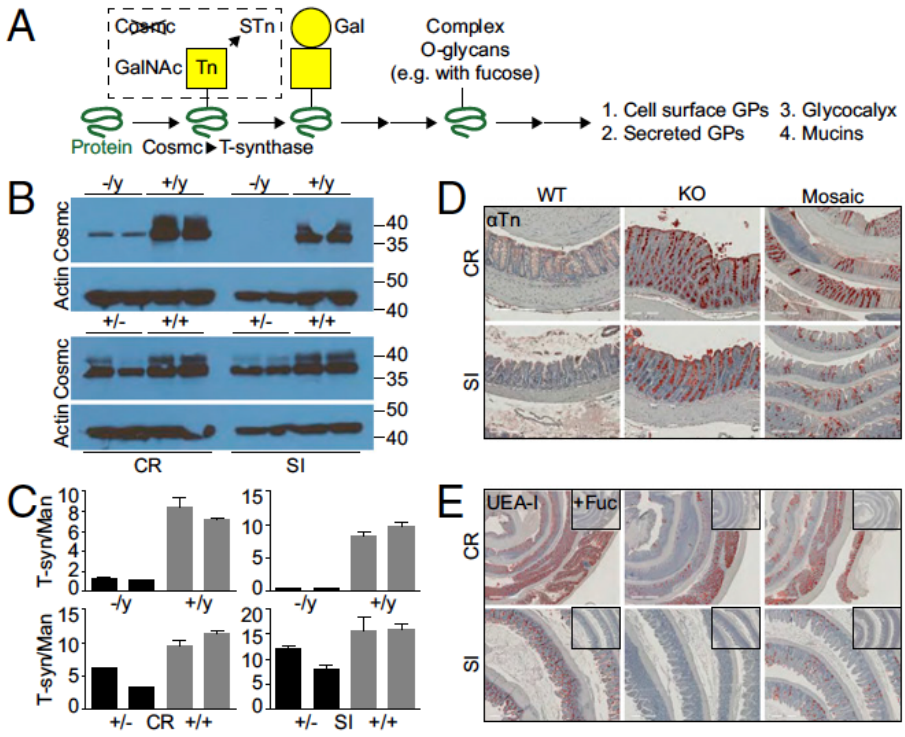


Fig. 4.1. Characterization of IEC-*Cosmc* mice. (**A**) Schematic of O-glycan biosynthesis: Cosmc is the chaperone for the T-synthase and required to extend O-glycans beyond the Tn antigen. GPs, glycoproteins. Purified IECs from CR and SI from KO, mosaic, and gender-matched WT mice were probed with anti-Cosmc and anti-Actin antibodies by Western blot (**B**) and assayed for T-synthase and mannosidase activities (2 mice pooled per group; 2 groups evaluated/tissue per genotype; age 2–3 mo) (**C**). (**D**) Tn expression was analyzed by IHC on formalin fixed colorectum (CR) and ileum (SI) from KO, mosaic, and WT mice (n = 3–5mice per group, 3 mo old). (Scale bars: Left and Center, 200 μM; Right, 500 μM.) (**E**) UEA-I lectin was used to analyze α1,2-fucose expression in the colorectum and ileum from KO, mosaic, and WT mice; Insets show UEA-I preincubated with 100 mM free fucose denoted as "+Fuc" (n = 3–5 mice per group, 3 mo old). Representative images shown. (Scale bars: Top, 1 mm; Bottom, 500 μM.)

Results

Spontaneous inflammation in male KO mice

To delete *Cosmc* in the intestinal epithelia, we crossed *Vil-Cre*$^{+}$ male mice with *Cosmc*$^{f/+}$ female mice and obtained *Vil-Cre*$^{+}$*; Cosmc*$^{f/y}$ (KO), *Vil-Cre*$^{+}$*; Cosmc*$^{f/+}$ (mosaic), and *Vil-Cre*$^{+}$ (WT) controls (**Fig. S4.1A**) at Mendelian ratios (n = 500, Pχ2 = 0.17). We confirmed near complete loss of Cosmc protein (**Fig. 4.1B**) and T-synthase activity (Fig. 1C) in purified intestinal epithelial cells (IECs) from male KOs and ~50% loss in female mosaics. This led to >90% expression of the Tn (**Fig. 4.1D**) and STn (**Fig. S4.1B**) truncated O-glycans by IHC (**Fig. 4.1A**) in colorectum and small intestine in male KOs and ~50% expression in female mosaics with patches of Tn/STn(+) crypts and Tn/STn(-) crypts, as predicted from random X-inactivation of *Cosmc*(33).

To examine whether glycan truncation results in loss of terminal epitopes, including known bacterial nutrients and ligands(434), we assessed terminal α1,2-linked fucose and sulfates/sialic acids with UEA-I and alcian blue (AB), respectively. WT colon and ileum IECs strongly bound UEA-I (**Fig. 4.1E**) and AB (**Fig. S4.1C**), and UEA-I was specific as it was blocked by free fucose (**Fig. 4.1E, inset**). In contrast, male KO crypts lost binding to UEA-I and AB and female mosaic crypts partially lost binding in patches. Unexpectedly, both male KOs and female mosaics retained binding of UEA-I and AB in the proximal colon (**Fig. 4.1E** and **Fig. S4.1C**). Thus, Cosmc differentially regulates

expression of terminal epitopes of the glycocalyx in different regions of the gut that are lost in the distal colon of male KOs.

We next assessed the consequence of *Cosmc* deletion and glycocalyx loss in the intestine, which results in profound pathology when deleted in other tissues(32, 33). Beginning at 3 mo of age, male KOs exhibited multiple signs of inflammation, including weight loss (**Fig. 4.2A** and **Fig. S4.2A**), elevated stool index (stool softness, blood in the stool) (**Fig. 4.2B** and **Fig. S4.2B**), rectal prolapse (RP) in ~40% of mice (**Fig. 4.2C** and **Fig. S4.2C**), inflammatory infiltrates limited to the distal colon (**Fig. S4.2D**), and elevated fecal lipocalin-2, an inflammatory biomarker(435) (**Fig. 4.2D**). Notably, RP alone did not drive inflammation, as male KOs with or without RP both had weight loss and an elevated stool index (**Fig. S4.3**).

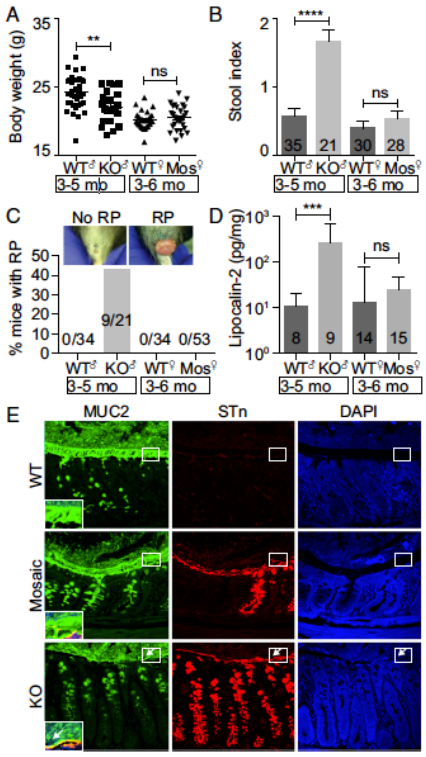


Fig. 4.2. Spontaneous inflammation in IEC-*Cosmc*-KO but not mosaic mice. Body weight (n = 35 WT males, 21 KOs; 30 WT females, 28 mosaics) (**A**), stool index (average of softness and blood content) (**B**), and rectal prolapse for KO, mosaics, and gender matched WTs (**C**). (**D**) Fecal lipocalin-2 in KO, mosaics, and gender-matched WTs. (**E**) Distal colon from WT, mosaic, and KO were fixed with Carnoy's reagent and stained with antibodies against MUC2 (green), STn (red), or with DAPI (blue) (3 mice per group, representative images shown); white boxes indicate the inner mucus layer, enlarged and merged in lower left corner; white arrows show where the inner mucus layer was lost. (Magnification: 40Å~.) **P ≤ 0.01; ***P ≤ 0.001; ****P ≤ 0.0001; ns, not significant from two-tailed Student's t test (**A**) or Mann–Whitney (**B** and **D**); mean ± SE (**A** and **B**) or median ± interquartile range (**D**); number of mice are indicated on the graph (**B**–**D**).

To further evaluate inflammation, we analyzed cell proliferation, which can increase by an uncontrolled injury-repair cycle or through immune drivers(436). Prior to the development of RP, male KOs showed histologic evidence of mucosal regeneration in the distal colon (**Fig. S4.4**) characterized by crypt elongation (154.7 ± 5.5 μ m in WT versus 384.3 ± 28.1 μ m in KO, mean ± SE) (**Fig. S4.4A**), epithelial hyperplasia (**Fig. S4.4B**), and increased mitoses and Ki67 proliferation indices (**Fig. S4.4B**). Other regions of the small and large bowel were unaffected, showing that loss of Cosmc in male KOs leads to spontaneous inflammation that is unexpectedly localized to the distal colon.

O-glycans, dependent on Cosmc, account for 80% of MUC2's mass and thus may be important for its function(437). To explore a possible mechanism for spontaneous inflammation(399), we preserved the mucus layer with Carnoy's fixative and analyzed mucus integrity in the distal colon with immunofluorescence (IF) (**Fig. 4.2E**), AB staining (**Fig. S4.1C and D**), and a 16S universal fluorescence in situ hybridization (FISH) probe against bacteria (Fig. S5). WT mice contained a thick, stratified AB(+) (**Fig. S4.1C and D**) and MUC2(+) inner mucus layer that separated the DAPI(+) luminal contents(438) (**Fig. 4.2E, top panels, box**) and FISH(+) bacteria (**Fig. S4.5A and B**) from the epithelia. In contrast, male KOs showed thinning, fragmentation, and/or absence of this layer (**Fig. 4.2E, bottom panels; Fig. S4.1C; S4.5A and B**) with 85% loss by AB staining (**Fig. S4.1D**), resulting in direct bacterial-epithelial contact, including helical-shaped bacteria nearing or contacting superficial colonocytes (**Fig. S4.5A, bottom two arrows**). Thus, Cosmc is required to form an intact mucus layer in the distal colon.

No spontaneous inflammation in female mosaic mice

KO males with loss of one allele have near complete loss of Cosmc throughout the intestinal tract. To understand how loss of one allele affects mosaic females, we assessed inflammation. We expected to see an intermediate level of inflammation, since ~50% of crypts lack Cosmc and the mucin glycocalyx. Surprisingly, we found no difference between female mosaic and control mice in multiple metrics of inflammation (body weight (**Fig. 4.2A**), stool index (**Fig. 4.2B**), RP (**Fig. 4.2C**), fecal lipocalin-2 levels (**Fig. 4.2D**), and IEC proliferation (**Fig. S4.4B**)), indicating that female mosaic mice were protected from inflammation.

To determine the mechanism, we fixed distal colons from female mosaic mice with Carnoy's and performed IF (**Fig. 4.2E**), AB staining (**Fig. S4.1C and D**), and FISH (**Fig. S4.5B and C**) to identify STn(+) KO cells, AB(+) WT cells, and their products in a female mosaic colon. In contrast to male KOs, female mosaic mice had a continuous, intact MUC2(+) mucus layer that covered both KO and WT cells, clearly separating luminal contents (**Fig. 4.2E, middle panels, box**) and bacteria (**Fig. S4.5B and C**) from the epithelium. Mechanistically, STn(+) KO MUC2 (**Fig. 4.2E and Fig. S4.5**) and AB(+) WT MUC2 (**Fig. S4.1C and D**) formed a continuous layer over both WT and KO crypts, indicating that WT MUC2 intermixed with KO MUC2 to form an intact mucus layer. Thus, local disruptions in Cosmc do not result in spontaneous inflammation because adjacent crypts are able to produce WT MUC2 that diffuses laterally over the crypt surface to protect KO cells.

To further investigate the extent of WT Cosmc compensation in female mosaic mice, we induced colitis with a four day course of 2.5% DSS in drinking water and assessed body weight and disease activity index (average of body weight change, blood in stool, and stool softness) in male KO, female mosaic, and control mice (**Fig. 4.3**). Male WTs were largely unaffected by DSS treatment, but male KOs exhibited profound weight loss (**Fig. 4.3A**) and rise in DAI (**Fig. 4.3B**) throughout the acquisition and recovery phase, necessitating euthanasia in all of the male KOs but none of the controls (**Fig. 4.3C**). Histologically, ~75% of the tissue was ulcerated or eroded in male KOs compared to less than 5% in male WTs (**Fig. 4.3D**). Female mosaics had less severe colitis than male KOs that did not differ from female WTs in the acquisition phase but had enhanced colitis in the recovery phase (**Fig. 4.3E and F**). Notably, fecal lipocalin-2 from female mosaics remained elevated for at least 18 days in contrast to female WTs that recovered by day 11 (**Fig. 4.3G**). Although we could only generate a few KO females ($VilCre^{+};Cosmc^{F/F}$) precluding statistical analysis, we saw RP and histologic inflammation in one of these mice, and WT females had more severe DSS colitis than WT males, arguing that WT females are not inherently protected from gut inflammation(439). These data suggest that the apparently normal mucus layer in female mosaic mice is able to protect from spontaneous inflammation and enhanced acquisition of colitis seen in male KOs, but once DSS disrupts the mucus barrier that developed in the absence of bacteria, WT MUC2 is unable to protect KO cells that have now contacted bacteria of an adult gut for the first time, resulting in prolonged inflammation.

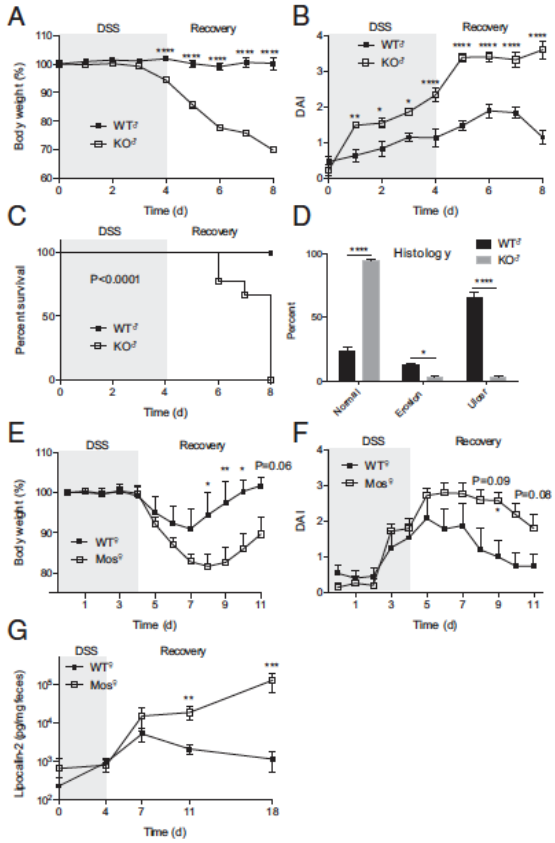


Fig. 4.3. DSS colitis in KO and mosaic mice. KO (**A–D**), mosaic (**E–G**), and WT mice (**A–G**) were treated with 2.5% DSS for 4 d and then switched to drinking water. Body weight (**A and E**), disease activity index (% weight change, stool softness, blood in stool) (**B and F**), survival (Mantel–Cox test) (**C**), histology (**D**), and fecal lipocalin-2 (**G**) were measured. n = 9 mice per group, age 3 mo (**A–C**); n = 6 KO, 9 WT, killed on day 8 and the percent normal, eroded, and ulcerated epithelia on H&E-stained CR was quantified (**D**); n = 5–6 mice per group, age 3 mo (**E–G**): repeated two independent times, one representative experiment shown (**E and F**) or pooled data (**G**); *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, ****P ≤ 0.0001 from ANOVA with Sidak's post hoc test for multiple comparisons (**A, B, and D–F**) or Mann–Whitney (**G**); 0.05 < P < 0.1 listed; mean ± SE (**A, B, and D–G**).

Microbes drive inflammation

To assess whether microbes drive inflammation, we treated male KO and WT mice with NMVA antibiotics for 4 weeks. Antibiotics depleted bacteria to undetectable levels (**Fig. S4.6**), resulting in normalization of fecal lipocalin-2 to WT levels (**Fig. 4.4A**) and partial normalization of IEC proliferation (**Fig. 4.4B and C**). We then examined whether inflammation was transmissible by fecal-oral route(440). Clinical signs of inflammation (stool index, body weight) were similar between WTs housed with other WTs or WTs housed with KO males (**Fig. S4.3**). Interaction of the microbiota with small intestine epithelia induces fecal IgA(441). However, fecal IgA levels were equivalent in 3-5 mo old male KO and WT mice (**Fig. S4.7**). Collectively, these data show that whereas microbes drive inflammation in male KOs, the inflammation is not transmissible by coprophagy nor associated with a change in fecal IgA.

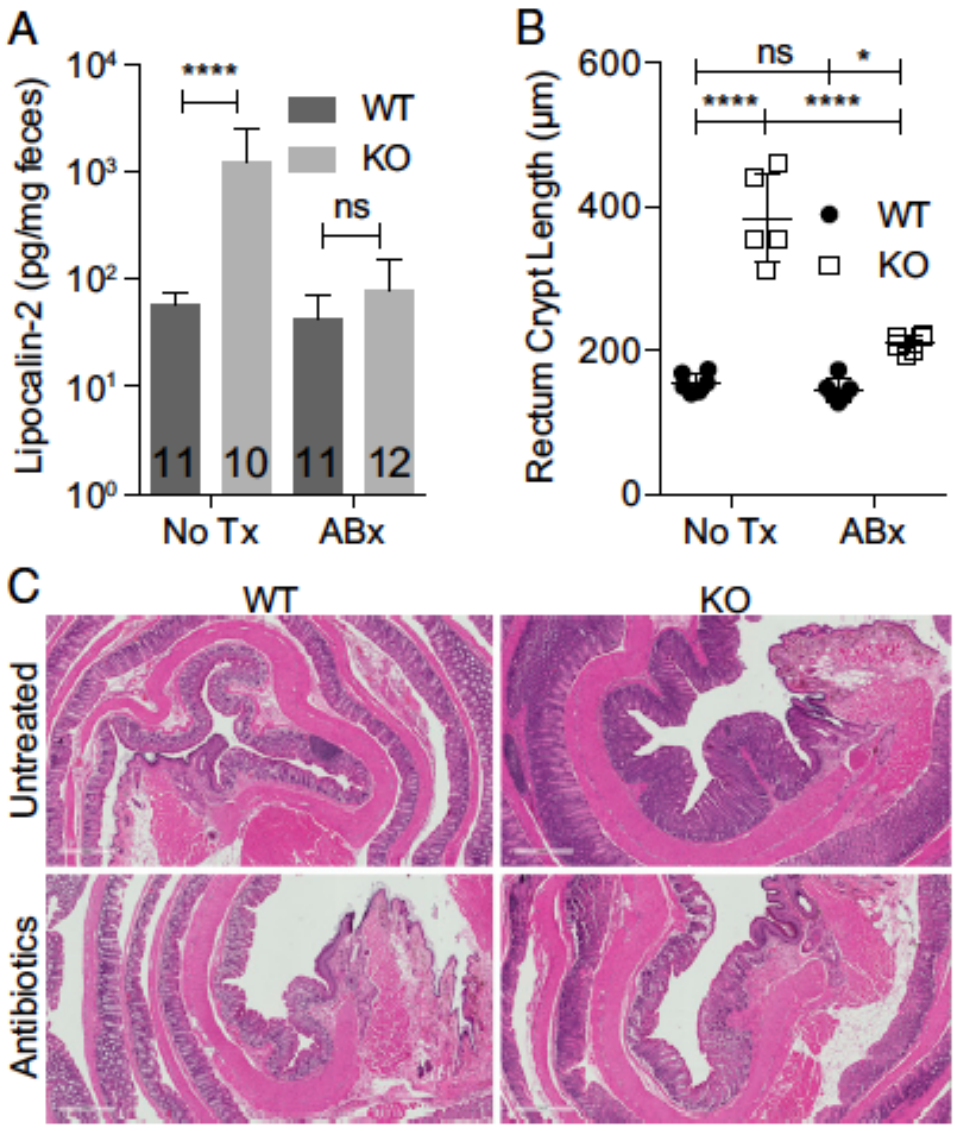


Fig. 4.4. Microbes drive inflammation in KO mice. Fecal lipocalin-2 (number of mice indicated on graph, data pooled from two independent experiments) (**A**), summary data (**B**), and representative images of rectum crypt length (**C**) in WT and KO mice treated with NMVA antibiotics or untreated for 4 wk beginning at 2 mo of age (n = 4–6 mice per group, 20 individual measurements/region per mouse for each replicate value); *P ≤ 0.05, ****P ≤ 0.0001 from two-way ANOVA with Sidak's post hoc test for multiple comparisons (**A and B**). ns, not significant. (Scale bars: 500 µM.)

Regional changes in the microbiota

Distal gut localized pathology led us to ask whether we could identify pro-inflammatory microbes or regional changes in microbiota community structure. We sequenced the V4 region of 16S rRNA gene from multiple luminal and mucosal regions from ~2 mo old male KO and WT mice (**Fig. 4.5**) to identify changes that precede clinical signs of inflammation. Composition of the microbiota from control mice compared well with previous reports, including greater abundance of Bacteroidiales in the lumen versus mucosa of the colon(384).

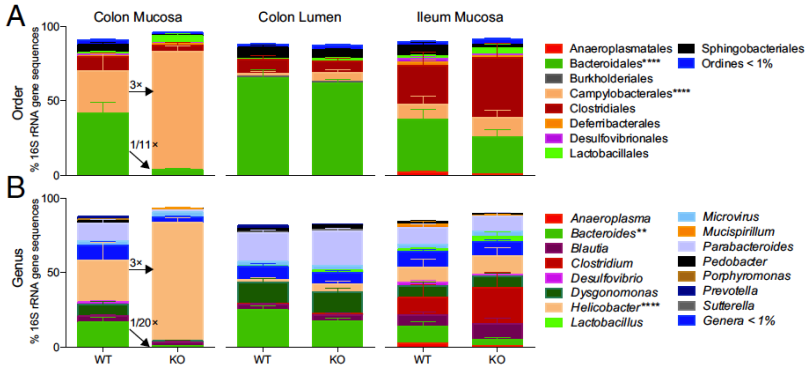


Fig. 4.5. Cosmc regionally regulates the gut microbiota community structure. (**A and B**) The microbiota was harvested from the gut lumen (distal colon) and mucosa (distal colon, ileum) from KO and WT mice for amplification of the V4 region and 16S rRNA gene sequencing (n = 7–8 mice per group, 2 mo old). Microbiota composition at the order (**A**) and genus (**B**) levels were plotted for all regions. Only colon mucosa contained significantly altered taxa (indicated with asterisks); ****$P \leq 0.0001$ from two-way ANOVA with Sidak post hoc test for multiple comparisons.

Male KOs had a dramatic loss of diversity in the colonic mucosa, as indicated by a loss of comparative taxonomic representation at both the order (**Fig. 4.5A**) and genus (**Fig. 4.5B**) levels. Notably, male KOs had an 11-fold loss of the Bacteroidiales (42.0 ± 7.1 % in WT to 3.8 ± 1.2 % in KO, mean ± SE), including a 20-fold loss of *Bacteroides* (16.7 ± 9.5 % in WT to 0.8 ± 0.2 % in KO, mean ± SE), coinciding with markedly increased representation of *Helicobacter* (28.1 ± 11.6 % in WT to 79.2 ± 3.3 % in KO, mean ± SE), a proinflammatory pathobiont.

We next examined whether similar changes were observed along the mucosal-luminal and cephalo-caudal axis. Unexpectedly, no changes were found in the luminal contents directly overlying the colonic mucosa or in the mucosa of the ileum. Our functional data support the sequencing results: 1) Distal gut localized pathology corresponds to the site of dysbiosis; 2) Inflammation was not transferrable by coprophagy as predicted by an unaffected lumen; and 3) IgA levels, regulated by the SI microbiota, were similar in male KO and WT mice along with a normal SI microbiota. Collectively, these results demonstrate that Cosmc differentially regulates the microbiota in different regions of the gut.

To identify host pathways that may contribute to regional changes in the microbiota, we isolated colorectal (CR) and small intestine (SI) epithelia from ~2 mo old male WT and KO mice, prior to clinical inflammation and at the same age as 16S sequencing, and analyzed transcripts with Illumina Mouse-Ref v2.0 gene chip array (n = 4

mice/group/region) (**Fig. S4.8**). Male KOs had a decrease in 79 genes in the CR, 22 genes in the SI, and only 5 genes in both (including Cosmc) (**Fig. S4.8A and B**; FDR 5%, fold-change ≥1.5). Further, different gene sets (small molecule metabolism, redox, stress response, digestion) (**Fig. S4.8C**, Metacore GSEA), pathways (**Fig. S4.8D**, Ingenuity Pathway Analysis), and transcripts (*RELMB*, *MEP1A*, *CHKA*, *IL18*, *ANG4*, *INTL-1*, *INTL-2*) (442-449) (**Fig. S4.8E**) important for regulating the microbiota were altered in CR versus SI. These data suggest that Cosmc differentially regulates gene expression in different regions of the gut to help control microbial biogeography.

Discussion

Here, in the first report of IEC-specific deletion of *Cosmc* in mice, we have provided insight into whether the IBD risk gene *Cosmc* functionally contributes to IBD and sex bias in pathology. Deleting one allele of *Cosmc* in males resulted in loss of T-synthase activity and of O-glycan extension, compromising the mucus layer, epithelial glycoproteins, and downstream host pathways important for maintaining homeostasis of the gut and its microbiota. Ultimately these direct and indirect effects led to dysbiosis and bacterial-epithelial contact, resulting in spontaneous inflammation and enhanced experimental colitis.

Unexpectedly, loss of one allele of *Cosmc* in females did not induce spontaneous inflammation, despite *Cosmc* KO and glycocalyx loss in 50% of crypts. Normally, the colonic mucus has an apparently attached sterile inner layer and an outer loose layer

occupied by bacteria(355). Although prior work hypothesized that the inner layer covalently links to goblet cells(355, 356, 364, 450), we found that MUC2 from WT cells migrates over multiple crypts to cover and protect KO cells while maintaining bacterial-epithelial separation. Thus, cohesive/adhesive forces may be more important than mucus-goblet cell linkages in attaching the inner mucus layer to the epithelia and covalent linkage may not be required to maintain a sterile inner mucus layer in female mosaic mice.

We also made the surprising discovery that Cosmc spatially regulates the gut microbiota, but in a very regio-specific manner. Deletion of *Cosmc* throughout the intestine led to a loss of diversity in the colonic mucosa but not in the overlying lumen or small intestine mucosa. Similarly, loss of *Bacteroides*, reduced mucosal diversity, and a dysbiotic mucosal but relatively preserved luminal microbiota are seen in human IBD(432, 451).

Mechanistically, our results indicate that nutrient selection, host genes, and bacterial adhesion function to establish the microbial gradient via Cosmc. In the mucosal-luminal axis, we found that the host glycocalyx, dependent on Cosmc, selects commensals in the mucosa despite a diet rich in polysaccharides, but not in the lumen(452). This contrasts with the conventional view that bacteria only rely on host glycans following dietary polysaccharide depletion and suggests that dietary polysaccharides may prevent migration of a dysbiotic mucosal microbiota to the lumen and subsequent fecal-oral transmission(385, 452). In the cephalo-caudal axis, we found that Cosmc differentially controls host genes in the small versus large bowel by indirect mechanisms downstream

of altered glycoproteins to spatially regulate the microbiota and facilitates expression of host ligands for bacteria by directly regulating glycocalyx synthesis. These ligands select commensals in the slow transit times of the colon but not in the small intestine(453). Our data functionally validate *Cosmc* as an IBD risk gene that spatially regulates the gut microbiota and contributes to sex bias in pathology and sets the stage to evaluate the interaction of Cosmc, the glycocalyx, and the gut microbiota in human IBD.

Methods

Mice: *Cosmc*$^{f/+}$ females(33) were crossed with B6.SJL-Tg(Vil-cre)997Gum/J transgenic males from Jackson Lab. All mice were backcrossed to C57BL/6J for at least 10 generations. All studies adhered to approved IACUC protocols (Emory University). Mouse genotypes were determined by PCR with primer for *Vil-cre* (F: 5'GTGTGGGAC AGAGAACAAACC, R: 5'ACATCTTCAGGTTCTGCGGG) and *Cosmc*^{flox} (F: 5'GCAACA CAAAGAAACCCTGGG, R: 5'TCGTCTTTGTTAGGGGCTTGC).

Immunoblot: Epithelia were purified(454) and immunoblots performed(455) with antibodies against Cosmc (C4, Santa Cruz) and actin (H-10, Santa Cruz).

Enzyme Assays: Enzyme activity assays for T-synthase and mannosidase were performed as described(455).

Tissue Staining: Intestine was dissected and prepared by swiss-roll (FFPE, distal portion of segment in center) for staining with monoclonal antibodies against Tn (CA3638, a kind gift from the late Georg Springer) and STn (TAG72, B72.3, Santa Cruz) antigens by immunohistochemistry(33). For mucus analysis, Carnoy's was used to preserve the mucus layer and tissues were stained with PAS/Alcian Blue; MUC2 (sc-15334, H-300), STn (TAG72, B72.3, Santa Cruz), UEA-I (B-1065, Vector Laboratories), and DAPI; or with a universal fluorescent in situ hybridization probe against bacteria(456). Crypt length in colon or crypt-villus length in small intestine was measured on H&E sections as described in text.

ELISA: ELISA for lipocalin-2 was performed as instructed (R&D, DuoSet). Fecal IgA was performed as described(457).

DSS Colitis: 2.5% DSS (Affymetrix) was administered in drinking water to mice for 4 days before switching to autoclaved drinking water. Body weight, stool softness, and blood in the stool (Hemoccult SENSA, Beckman Coulter) was assessed daily. DAI was determined by averaging score for %body weight change (no loss (0), 1-5% (1), 5-10% (2), 10-20% (3), >20% (4)), stool consistency (hard (0), soft (2), diarrhea (4)), and blood (negative (0), positive (2), macroscopic (4)). Histology was assessed on H&E stained swiss-roll of colorectum (CR) from FFPE sections. Aperio digital microscope was used to measure CR length and length of ulcerated (complete loss of crypt architecture with inflammatory infiltrate) and eroded (partial loss of crypt architecture with inflammatory

infiltrate) epithelia. Mortality was determined in mice whose body weight dropped >25%, necessitating euthanasia per Emory IACUC guidelines.

Microbiota: Intestines were removed and dissected longitudinally. Fecal pellet was collected from distal colon for analysis. Otherwise, feces were gently removed and tissue was collected from distal colon and ileum for analysis of mucosal-associated bacteria. V4 region of rRNA gene was amplified by PCR and sequenced by MiSeq (Illumina). Taxonomic divisions were determined by MacQIIME(458). Antibiotics (NMVA; neomycin, metronidazole, vancomycin, ampicillin) were administered as described(459). Feces were cultured on LB in aerobic conditions to confirm elimination of microbiota.

Gene expression: Intestinal epithelia were purified(454) and total RNA isolated by RNeasy (Qiagen) with RNA quality and purity assessed by Agilent BioAnalyzer and nanodrop, respectively. UCLA genomics core performed RNA amplification and hybridization on Illumina mouse ref 8 v2.0 expression chips. Individual genes with fold-change >1.5 and FDR <0.05 were determined by SAM. GSEA for disease pathways and GO processes (Metacore, only relevant findings included) and gene networks (Ingenuity Pathway Analysis) were determined. Data was deposited in Gene Expression Omnibus (GEO) under accession number GSE84416.

Statistics: Statistics were calculated with Prism GraphPad software as described in the figure legends.

Supplemental Information

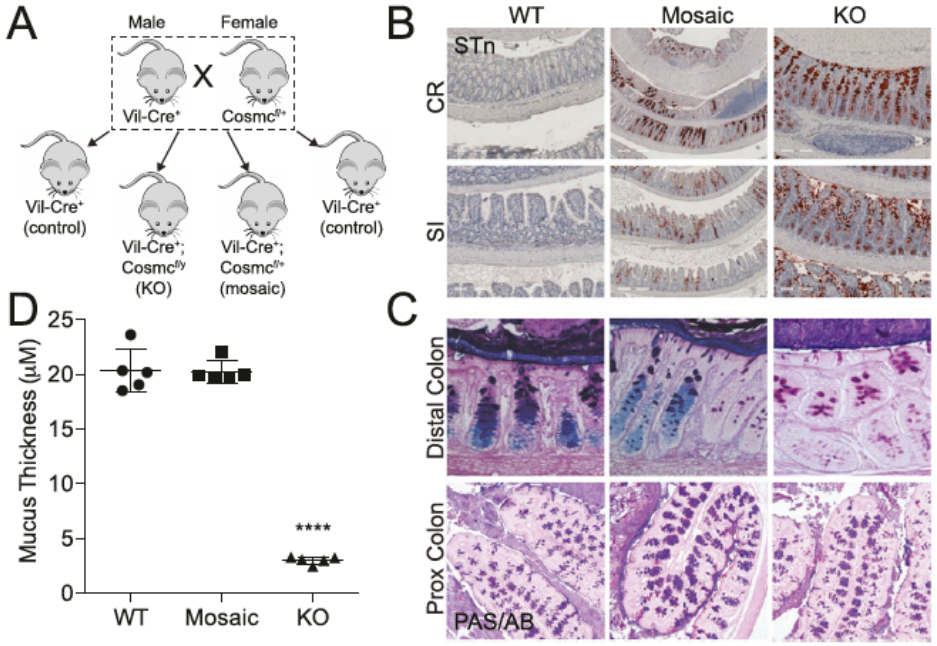


Fig. S4.1. Breeding strategy and characterization of IEC-Cosmc mice. (**A**) Breeding strategy: *Vil-Cre*$^{+}$ males were bred to *Cosmc*$^{f/+}$ females to generate intestinal epithelia specific KO (*Vil-Cre*$^{+}$*;Cosmc*$^{f/y}$), mosaic (*Vil-Cre*$^{+}$*;Cosmc*$^{f/+}$), and WT control (*Vil Cre*$^{+}$*;Cosmc*$^{+/y}$ OR $^{+/+}$) mice. (**B**) Sialyl-Tn expression was analyzed by monoclonal antibody on FFPE, swiss-rolled CR, and SI from KO, mosaic, and WT mice (distal segment in center of swiss-roll). (Scale bars: Left and Center, 200 µM; Right, 500 µM.) PAS/AB staining was performed on Carnoy's fixed distal (magnification: 40×) and proximal (magnification: 10×) colon from WT, mosaic, and KO mice (**C**), and the inner mucus layer thickness was quantified in the distal colorectum (5 measurements per mouse per replicate, n = 5 mice per group) (**D**). n = 3–5 mice per group, age 3 mo (**B and C**). ****P ≤ 0.0001 from one-way ANOVA with Tukey's post hoc test for multiple comparisons; mean ± SD.

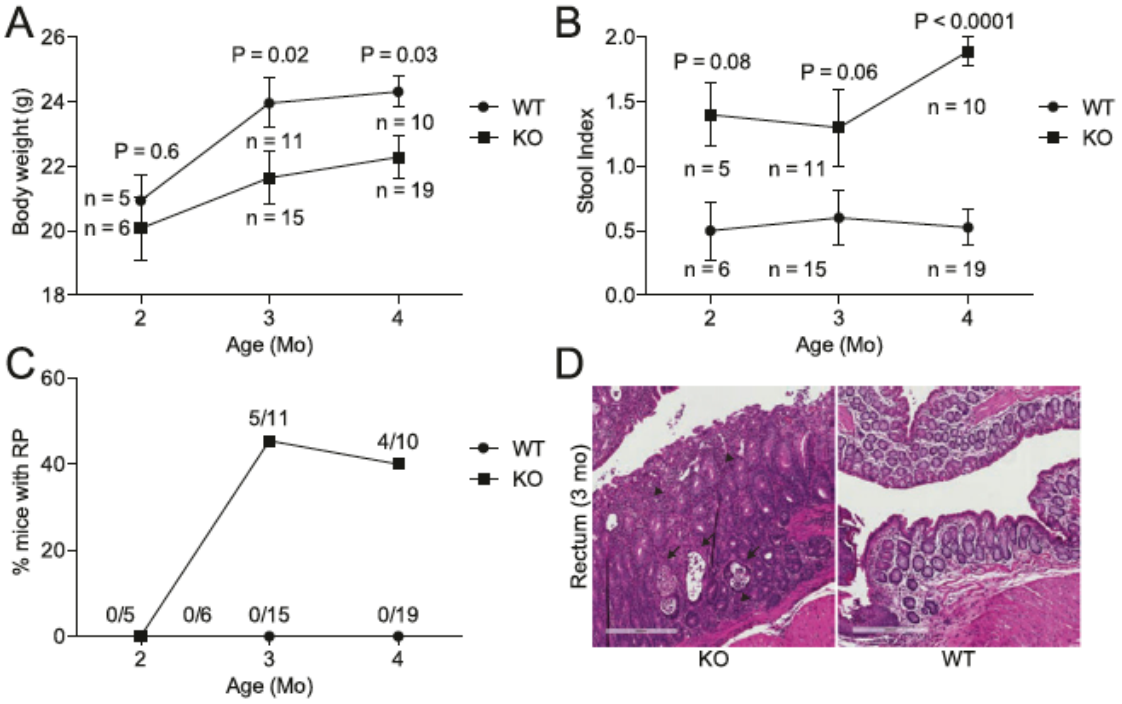


Fig. S4.2. Time course of clinical inflammation. Body weight (**A**), stool index (**B**), and percent of mice with RP (**C**) were measured in KO and WT mice at 2, 3, and 4 mo of age. Measurements are the same as in **Fig. 4.2**. (**D**) Histologic evidence for active inflammation in distal colon of KO but not WT: crypt abscesses (arrows) and inflammatory infiltrate consisting of neutrophils, lymphocytes, and plasma cells (arrowheads) (n = 4 mice, < 1 y, 3 mo old mice shown). (Scale bars: 300 μM.) Number of mice indicated on graphs (**A-C**); P values from two-tailed Student's t test (**A**) or Mann–Whitney (**B**); mean ± SE (**A and B**).

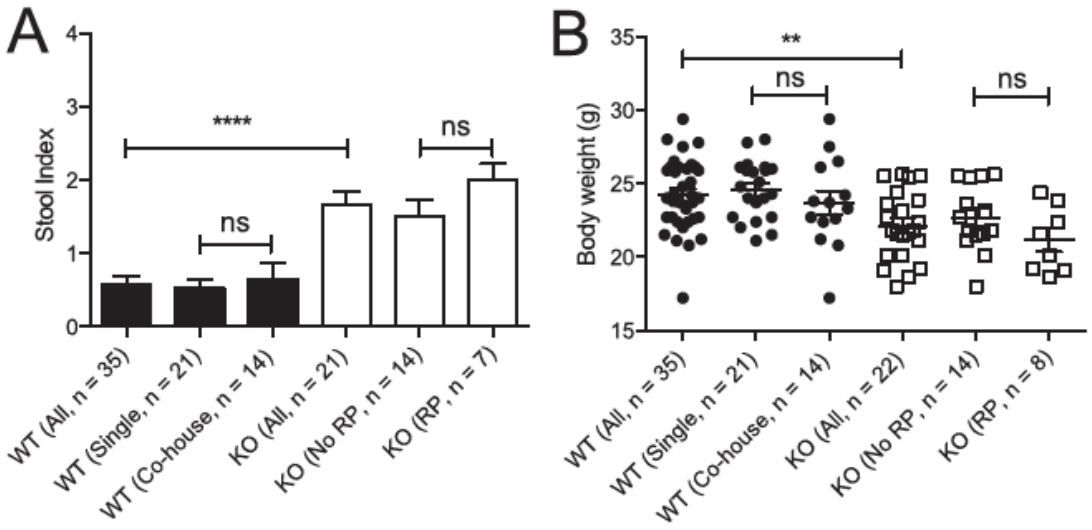


Fig. S4.3. Effects of cohousing and rectal prolapse on inflammation in KO mice. Stool index (average of softness and blood content) (**A**) and body weight (**B**) in 3- to 5-mo-old KO and WT mice. WT mice segregated by single housing (WT with WT) or cohousing (WT with KO); KO mice segregated by presence or absence of rectal prolapse. **P ≤ 0.01, ****P ≤ 0.0001 from Kruskal Wallis test with Dunn's post hoc test for multiple comparisons (**A**) or one-way ANOVA with Sidak's post hoc test for multiple comparisons (**B**); mean ± SE (**A and B**); number of mice indicated on graph (**A and B**). ns, not significant.

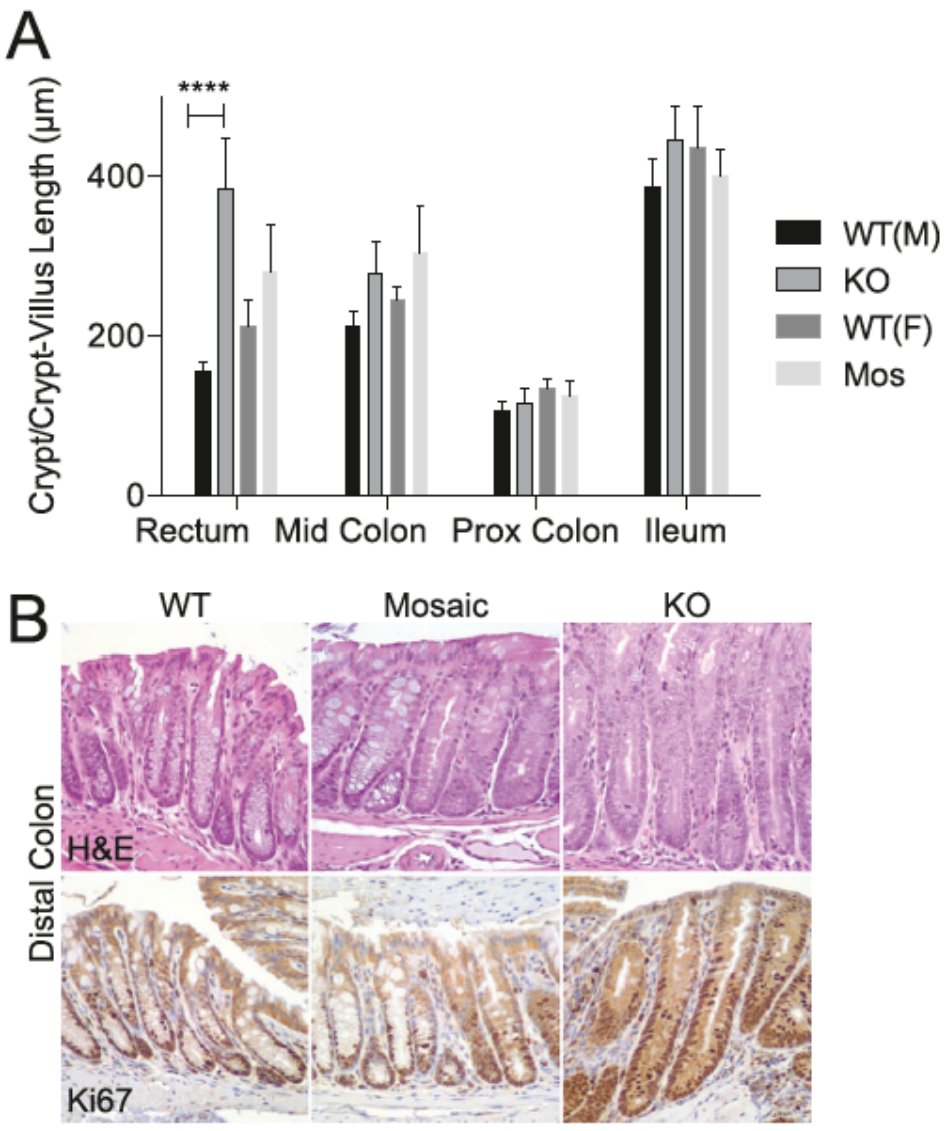


Fig. S4.4. Hyperplasia in KO mice but not mosaic mice. Crypt length (**A**) and Ki67 (**B**) were analyzed in WT, mosaic, and KO mice. Summary data (n = 4–6 mice per group, 20 individual measurements/region per mouse for each replicate value) (**A**) and representative images (n = 3 mice per group) (**B**) are shown (magnification: 40×); ****P ≤ 0.0001 from two-way ANOVA with Sidak's post hoc test for multiple comparisons (**A and B**).

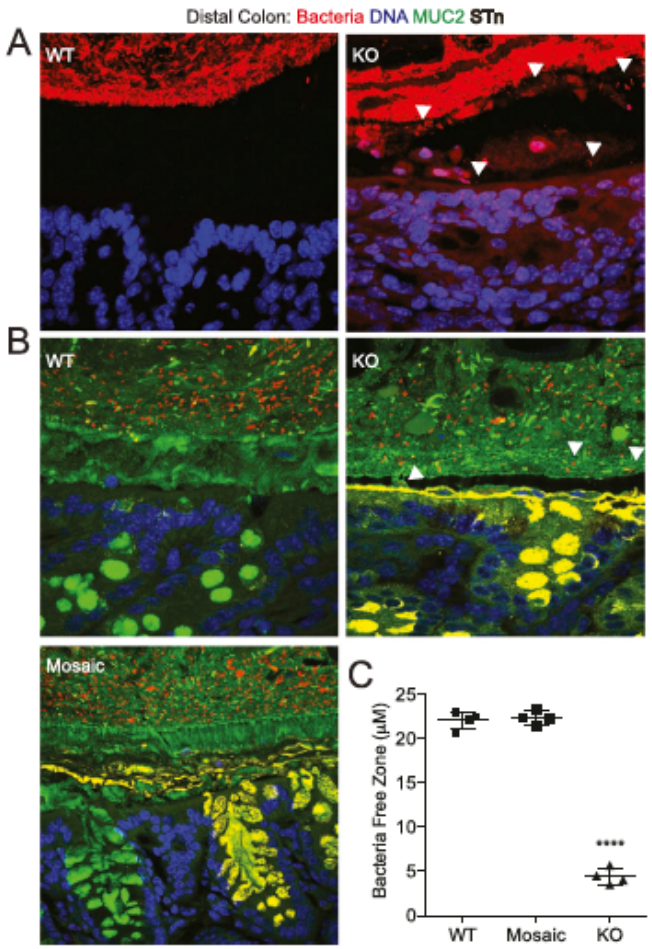


Fig. S4.5. Bacteria–epithelial interactions in the distal colon. Bacteria were visualized with EUB-16S rRNA Universal Bacteria FISH probe (red) and epithelia by DAPI (blue) (n = 2–3 mice per group, representative images shown) in **A** or additionally with antibodies against MUC2 (green) and STn (yellow) in Lower (n = 3 mice per group, representative images shown) (**B**) and the bacterial–epithelial distance was quantified (5 measurements per mouse per replicate from B, n = 4 mice per group) (**C**). White arrows indicate invading bacteria. ****$P \leq 0.0001$ from one-way ANOVA with Tukey's post hoc test for multiple comparisons; mean ± SD. (Magnification: A, 100×; B, 40×.)

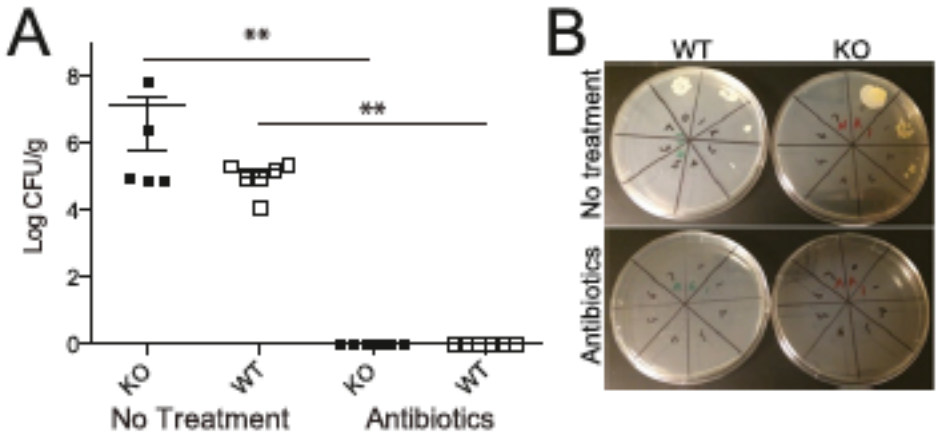


Fig. S4.6. Bacterial counts of mice treated with antibiotics. (**A**) Two-mo-old KO and WT mice were treated with NMVA antibiotics for 4 wk, and CFUs were assessed from feces (n = 6 per group) with untreated KO (n = 5) and WT (n = 6) mice for comparison (two independent experiments, one representative shown; all antibiotic-treated groups had 0 CFUs/g, which were set at log 0 for display on the graph). (**B**) Representative plates are shown. **P ≤ 0.01 from Mann–Whitney test; mean ± SE.

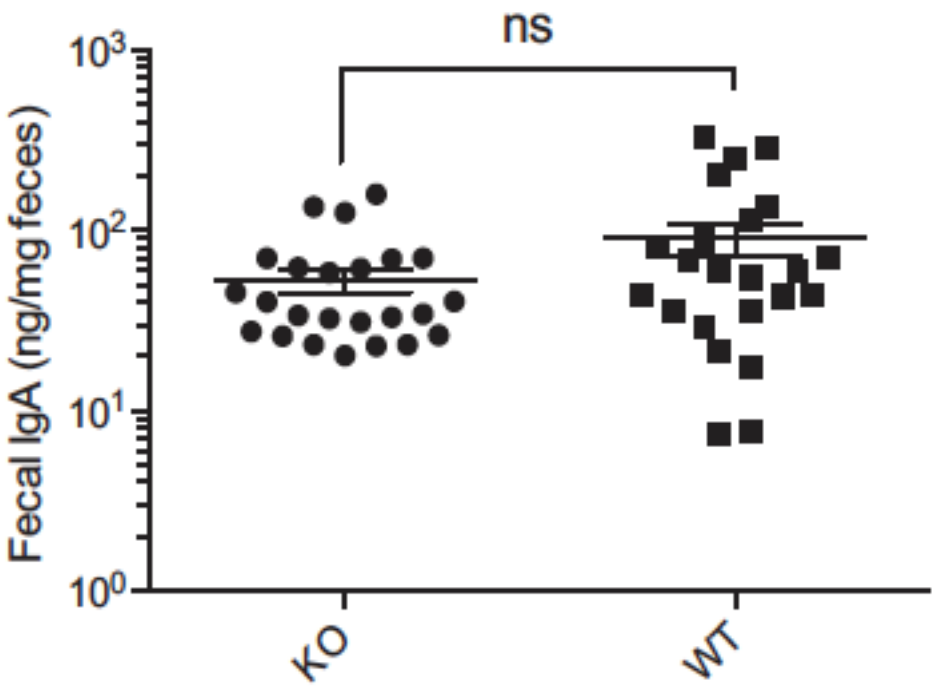


Fig. S4.7. Analysis of fecal IgA. Fecal IgA was analyzed in KO (n = 24) and WT (n = 23) mice by ELISA (age-matched, 3–5 mo). ns, not significant from Mann–Whitney; mean ± SE.

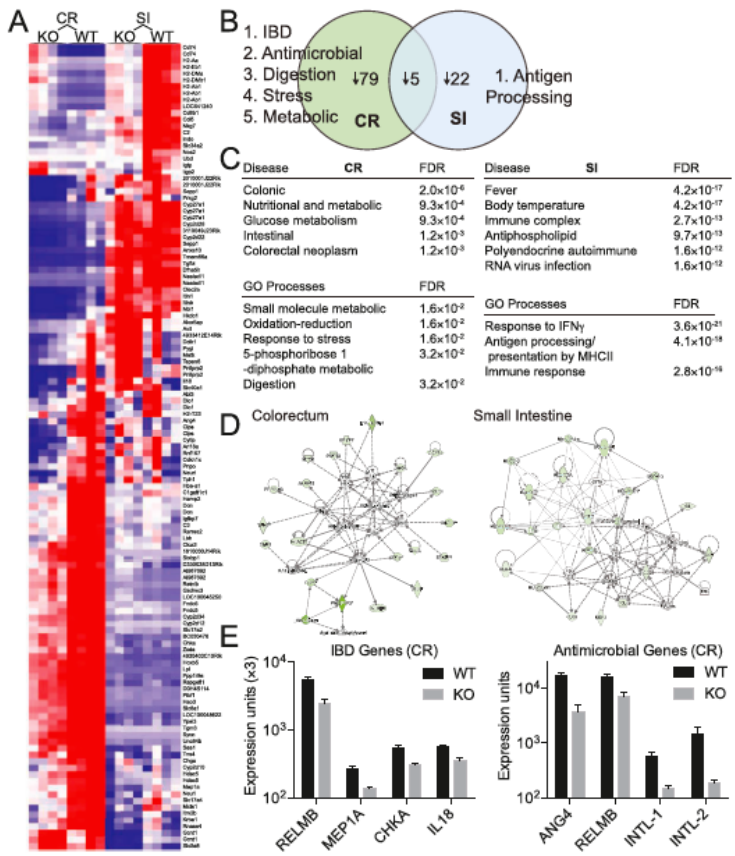


Fig. S4.8. Cosmc controls diverse host pathways that regulate the microbiota and inflammation in the distal colon. Epithelial cells were purified from the CR or SI from KO and WT mice, and gene expression was performed on Illumina mouse ref 8 v2.0 expression chip. (**A**) Heat map indicating genes significantly altered by *Cosmc* KO (n = 4 mice per group, age 2 mo; fold change > 1.5, FDR < 0.05). (**B**) The number of genes altered in only CR, SI, or both are indicated with the most relevant gene sets from C or individual genes from E bulleted. No change in *ST6GALNAC1* was seen. (**C**) Metacore GSEA was performed on CR and SI for gene sets enriched for diseases or GO processes. (**D**) Ingenuity pathway analysis used for gene expression from KO and WT epithelia isolated from colorectum. (**E**) Expression of individual IBD or antimicrobial genes from the CR as determined from the expression array.

Chapter 5

Discussion

Discussion

Cells are coated with a dense glycocalyx rich in carbohydrates that facilitate diverse biological processes. Work in cell lines, animal models, and *in vitro* systems indicate that glycans confer direct and indirect roles, facilitating interactions with glycan binding proteins or indirectly modulating protein oligomerization, stability, confirmation, cell surface resident time, and turnover(42). Nonetheless, sufficiently sensitive, high-throughput technologies to analyze glycans in similar scale to genomics have been lacking, limiting our understanding of cell surface glycans. Similarly, only a few animal models have fully integrated human disease genetics and molecular analyses to obtain an ecological view of disease.

We were motivated by three critical questions: 1) What glycans are normally present in cells? 2) How do they change in disease? 3) What are the consequences of those changes? To address the first two questions, we developed a new technology allowing us for the first time to amplify and sequence the cellular glycome(455). We believe that such an amplification strategy could have far reaching impact similar to how DNA amplification revolutionized genomics. We used this technology to begin to define the cellular glycome in normal and diseased states and estimate the size of the human glycome. In the second study, we integrated disease genetics with defined biochemical alterations to develop a mouse model of IBD and discovered multiple roles of altered glycosylation in human disease, ranging from regulation of host biology to altered ecology of the host microbiota(460). Thus, this thesis represents a major advance in

understanding what glycans are present in cells, how glycans change in disease, and the consequences of those changes.

Defining the cellular glycome

Cells surface glyconjugates include a diverse array of molecules, including O-glycans, N-glycans, GAGs, and glycolipids. The function of glycans can either be restricted to glycan class (e.g. N v. O-glycan for galectins), glycoprotein acceptor (e.g. Core 2 sialyl-Lewis x on PSGL1), or have terminal epitopes with broad activities across multiple classes of glycosylation (e.g. sialyl-Lewis x on the zonu pallucida for sperm binding)(99, 461, 462). Therefore, studying one class of glyconjugates provides broad insights.

Historically, glycomics has required a biochemical tour de force. Glyconjugates are purified and processed (e.g. into glycopeptides) and glycans are chemically or enzymatically released, purified, labeled, and analyzed. This approach fails at small scales of less than 1 – 10 million cells because too few glycans are present in a sample and also at large scales greater than a gram because the chemistries are inefficient and enzymes become prohibitively expensive. Our work outside the scope of this thesis recently addressed the latter problem by devising efficient chemistries to release glycans from diverse glyconjugates in preparative quantities(463). We became interested in addressing the first challenge, glycomics at small scale, because many primary cells and clinical specimens were impossible to analyze with the existing technologies thus limiting our understanding of the role of glycans in health and disease.

O-glycans are an important class of glycosylation, but their cellular repertoires are largely unknown due to limitations in glycomic analysis. O-glycans are present on >80% of proteins that traverse the secretory apparatus and are critical in mammalian development(14). Deletion of O-glycans in a mouse results in embryonic lethality in part due to defects in angiogenesesis, and disruption in a tissue specific manner results in gross abnormalities in all tissues studied to date(33, 401). Further, these structures are frequently altered in disease. Cancer coincides with expression of both large and small tumor-associated and tumor-specific glycans, such as CA19-9 (sialyl Lewis A), T, STn, and Tn antigens, and colitis coincides with expression of a simplified O-glycome, including T and STn antigens and loss of Cad and Lewis antigen expression; however a few larger glycans such as sialyl Lewis A on a CD44 isoform are also expressed(398, 401).

No general O-glycanase has been found, in contrast to PNGaseF for N-glycans, restricting O-glycan release to chemical approaches. Alkaline β-elimination is the most common strategy to release O-glycans but is inefficient (releases at most 50% of O-glycans present in a sample) and can result in glycan degradation through the peeling reaction. Thus, there is a balance between increasing efficiency of release with harsher condition and avoiding glycan degradation. Because of these limitations, at least 10 million cells are generally needed for O-glycomic analysis.

To address this challenge, we took advantage of the O-glycan biosynthetic machinery and synthesized a chemical O-glycan precursor (peracetylated benzyl-α-GalNAc) that could be taken up into the cell, de-esterified, taken up into the Golgi, and modified by endogenous glycosyltransferases utilizing endogenous sugar-nucleotide donors. Extended Bn-O-glycans representing the normal repertoire of cellular O-glycans are then secreted from cells for easy purification and analysis by mass spectrometry and other strategies. We term this technology Cellular O-glycome Reporter/Amplification or CORA(455).

CORA immediately provides a few advantages over existing technologies. First and perhaps most significantly, CORA increases sensitivity of O-glycan analysis by at least 100 – 1000 fold. This was clear from being able to analyze multiple cells lines in a 48 well dish down to $2x10^4$ cells compared to 10^7 cells in a T75 flask for traditional methods. Further, in addition to being able to identify most of the O-glycans observed by β-elimination, CORA also identified many higher molecular weight structures that were known to exist in these cell lines but that are notoriously difficult to detect. Being able to perform O-glycomics in a 48-well dish is a landmark advance. Although we did not attempt smaller cell numbers or dishes such as 96 or 384-well plates, it is very likely that this would be possible. Advances in robotics and primary cell culture and engineering would make it possible to evaluate the cellular O-glycome across thousands of cell lines, primary cells, and patient-derived organoids. This opens a new frontier for next-generation glycomics.

The fact that we were able to generate the cellular repertoire of O-glycans off of a simple chemical precursor provides fundamental insights into the mechanisms regulating O-glycan biosynthesis. Historically, two processes were thought to determine the repertoire of O-glycans present in cells – protein backbone and biosynthetic machinery. The finding that alpha-dystroglycan contains a specific class of glycosylation initiated by the attachment of O-mannose argues that some glycans are restricted to specific proteins(464). Although this has not been rigorously demonstrated for mucin-type O-glycans, it is certainly the case that most though not all secretory path proteins contain O-glycans and that this can be predicted based on machine learning algorithms, although an exact sequon is not clear(14). Thus some feature of the protein presumably determines whether it receives an O-glycan. Further, certain glycan binding proteins interact with glycopeptides rather than glycans alone (e.g. P-selectin with glycosylated PSGL1) and the degree of glycan heterogenity differs in a protein specific manner such that some proteins contain different glycoforms than others from the same cell(462, 465, 466). Collectively, these data suggest that glycoprotein structure may in part determine cell-type specific glycosylation. On the other hand, analyses of cellular glycomes has shown that different glycomes correspond to differential expression of glycosyltransferases(467). This suggests that the enzymes that encode the glycans rather than the protein acceptors determines glycan repertoire.

In our studies, we found that the cellular O-glycome could be accurately recapitulated onto a simple chemical glycan acceptor incubated with live cells. The diversity of glycans we saw across cells mirrored the diversity that is seen when analyzing the glycans

released from glycoproteins from those cells. This finding argues that the enzymatic machinery rather than the glycoprotein acceptor is the major determinant of the cellular glycome. Even so, this does not exclude the possibility that for a handful of proteins within a cell, that the protein structure drives glycan structure, keeping in mind that the cellular glycome represents the glycans attached to all proteins in the cell. Interestingly, although cell number did not affect the relative abundance of Bn-O-glycans obtained from cells, increasing the concentration of chemical precursor resulted in a shift of relative abundances of some glycan structures, but not all. Identifying what structures changed may provide insight into rate limiting enzymes in glycan biosynthesis and competition for enzymes across different glycan substrates. Related findings were observed comparing AFP glycoforms produced from the yolk sac versus the fetal liver, showing that protein expression level shifts the relative abundance of glycans on a glycoprotein(468). Although protein structure and thus repertoire of proteins may not generally determine the cellular glycome, the amount of protein is important. This observation may inform our understanding of human disease. A common feature of cancer is upregulation of the ribosome resulting in increased protein levels(469). This may in part explain altered glycosylation in cancer although this has not been tested.

The high-throughput nature of CORA allowed us to examine a large set of primary and cancer cells and combine that with computational approaches to predict the size of the human glycome. Our preliminary model based on experimental data predicted 376 unique O-glycan structures, corresponding well to theoretical predictions based on biosynthetic possibilities(137). Our initial report analyzed 18 cell types representing a diversity of

species (mouse, human) and tissue types (epithelia: breast, intestine; endothelial, fibroblasts; and hematopoietic cells), giving us a broad view of the cellular glycome across tissues and species. However, we had insufficient representation from each tissue type to assess cell-type specific glycosylation. Clues from our initial report nonetheless motivated us to address this in further studies not described in this thesis: In preliminary work evaluating murine macrophage subsets and human colorectal cells our data support the idea of cell-type specific glycosylation both by looking at glycan structures and diversity. For example, whereas we identified over 100 different O-glycan structures in colorectal cells, we only identified two O-glycans from primary macrophages including those coming from distinct developmental lineages and representing diverse activation states. This suggests that colorectal cells use their glycomes to interact with host and microbial lectins, while macrophages maintain a simple glycome to avoid self-recognition with their own galectins and Siglecs.

In sum, the development of CORA has been an enabling technology that has helped us begin to assess the repertoire of cellular glycans in health and disease.

Investigating the role of altered glycosylation in disease

Cosmc encodes a chaperone for the T-synthase on the X-chromosome that is critical to extend O-glycans beyond the simple Tn and STn antigens. GWAS studies have implicated *Cosmc* in sex-specific risk for IBD(433); however the functional consequences of Cosmc disruption were unknown. We crossed *Villin-Cre* mice with

Cosmc$^{+/flox}$ to generate mice with an intestinal epithelial-specific deletion of *Cosmc*(460). We generated KO male mice (*Vil-Cre*$^{+}$*;Cosmc*$^{flox/y}$), female mosaic mice (*Vil-Cre*$^{+}$*;Cosmc*$^{flox/+}$) and *Vil*$^{+}$ WT control mice. KO mice exhibited >90% deletion of *Cosmc* in intestinal epithelia while mosaic mice had complete loss of Cosmc in ~50% of crypts in a mosaic pattern with ~50% WT crypts intermixed with KO crypts. We found that KO mice had dysbiosis, loss of MUC2, enhanced bacterial-epithelial contact, spontaneous colitis, and enhanced sensitivity to experimental colitis. In contrast, mosaic mice were completely protected from spontaneous colitis and partially protected from experimental colitis due to the presence of an intact MUC2 layer. WT MUC2 from WT crypts migrated over adjacent KO crypts to form an intact mucus layer.

Deletion of *Cosmc* in the intestinal epithelia and engineered expression of a simplified IBD-like glycome led to dysbiosis, mirroring many of the changes observed in patients: when examining the colonic mucosa, we observed dramatic loss of bacterial diversity, especially of the glycan consuming Bacteriodales. This disruption led to monocolonization with a pathobiont, *Helicobacter*, which others have implicated in intestinal inflammation and the rectal prolapse observed in our mice(470). These studies experimentally demonstrate that disruption in *Cosmc* and expression of a simplified O-glycome can drive some aspects of dysbiosis in IBD.

The alterations we observed were regionally localized. Different microbes are present in different regions of the gut in the rostral-caudal and mucosal-lumenal axis, expanding the functional capacity of the gut microbiota. Although a few mechanisms have been

proposed for this so called gut microbe biogeography, only few have been rigorously tested(384). Further, alterations in this gut microbe biogeography are observed in IBD with more prominent dysbiosis in the intestinal mucosa versus the overlying lumen or fecal content(432). Intestinal glycosylation varies along the length of the intestine with increase in Cad and sialic acid and reduced fucose and fucose-containing blood and Lewis antigen in progression to the distal colon. This results in an invariant glycome in the distal gut that does not change from person-to-person while the proximal gut glycome expresses the full repertoire of blood group and Lewis antigens and thus varies across the population. This regional pattern of glycosylation was hypothesized to help establish the spatial distribution of microbe in the intestine(21); however, this had not been tested.

To address this, we performed 16S rRNA gene sequencing in the colonic mucosa, overlying lumen, and ileal mucosa of KO and WT mice and discovered that alterations in the microbiota were restricted to the colonic mucosa but that the community membership in the overlying lumen and small intestine mucosa were unaffected, despite equivalent deletion of *Cosmc* throughout the intestine. Further, the changes we saw were not simply driven by inflammation as we examined KO mice at two months of age prior to the clinical onset of colitis. As hypothesized(21), we showed that O-glycans establish commensals in the distal gut but were dispensable in the small intestine, suggesting that SI glycosylation may instead be important in pathogen recognition. These results provide the first direct evidence that epithelial glycosylation spatially regulates the gut microbiota. Moreover, dysbiosis restricted to the mucosa resembles the spatial disruption

observed in IBD(432), suggesting that altered epithelial glycosylation may drive spatial dysbiosis in patients.

Prior work evaluating model members of the gut microbiota led to the hypothesis that intestinal bacteria prioritize dietary glycans over host glycans(380). The Bacteriodales contain a diverse repertoire PULs capable of degrading both host and dietary polysaccharides(471). We found that genetic deletion of host glycans led to a dramatic loss in diversity, especially of the Bacteriodales, in the mucosa but not in the lumen, indicating that bacteria in different spatial geographies differentially depend on host versus dietary glycans, despite genetic metabolic potential, rather than globally prioritizing dietary glycans. Based on ecological selection in the colon, dietary glycans are prioritized in the lumen while host glycans are prioritized in the mucosa. While this may seem intuitive, this had not been examined in an ecologically complex systems taking into account the full repertoire of gut biogeographies. We evaluated community membership by 16S rDNA sequencing, but it would also be interesting to perform metagenomic sequencing to examine total gene content in our model and evaluate the impact of dietary polysaccharide restriction on gene content and community membership.

Another interesting discovery we made was that compared to KO males, females mosaics are completely protected from spontaneous inflammation and partially protected experimental colitis. Prior work hypothesized that the inner mucus layer of the colon is covalently attached to the epithelia. This hypothesis arose from studies evaluating whether mucus could be removed from epithelia by gentle aspiration or whether it

required scraping(471). Mucus in the small intestine and the outer layer of the colon could be removed by gentle aspiration, indicating that it is loose and unattached. In contrast, the inner mucus layer could only be removed by scraping, suggesting that it is attached. When we mechanistically investigated the basis for protection in female mosaics we found that WT mucus was able to migrate or diffuse over multiple crypts to cover and protect KO cells. Moreover mucus overlaying KO and WT crypts in mosaics was functionally intact, separating the gut bacteria from the epithelia. This finding suggests that the inner mucus layer is not covalently attached to the epithelia and that such attachments are not necessary for maintaining an intact mucus layer. It is more likely that the physical properties of the inner mucus layer provide non-covalent interactions with the epithelia that prevent aspiration. Thus a more appropriate description of the inner mucus layer would be that it is tightly associated but not attached.

Cosmc was identified as a sex-specific risk gene in colitis by GWAS(433). Pediatric IBD is thought to represent a genetically pure form of IBD, making it an ideal subpopulation to gain genetic insights. Although the frequency of adult IBD is similar between males and females, pediatric IBD is slightly more common in males(430), suggesting that there may be sex-specific risk genes. Indeed, we found that male mice with loss of one allele of *Cosmc* are more susceptible than female mice with loss of one allele to inflammation. Mechanistically, as we described above, WT mucus migration protect KO cells. Thus we have validated *Cosmc* as a sex-specific risk gene in IBD.

Loss of Cosmc results in expression of a truncated O-glycome and expression of the Tn and STn antigens. While these have been observed in IBD, they are also widely expressed in carcinomas including ~70% of colorectal cancer(401). In work not presented in this thesis, we asked whether engineered expression of Tn/STn antigens in the gut could initiate colorectal cancer. We found that intestinal-epithelial-specific deletion of *Cosmc* resulting in Tn/STn expression led to spontaneous colorectal adenocarcinoma by 6 – 9 months of age. This coincided with elevated TGF-β signaling, increased ROS, and ultrastructural loss of the glycocalyx but did not lead to overt β-catenin nuclear translocation. Interestingly, TGF-β was only activated in KO crypts but not WT crypts in mosaic mice. Although the mechanisms for tumorigenesis downstream of Tn are not completely clear, we hypothesize that enhanced bacterial-epithelial contact leads to elevated ROS, resulting in epithelial proliferation and ROS-dependent mutation in the highly proliferative stem cells of the intestinal tract. These changes at a genetic level combine with altered signaling, such as TGF-β, to promote tumor formation and progression.

Tn antigen is a promising marker to detect carcinomas and colitis. Colorectal cancer is the second most deadly cancer but preventable if detected early. Although colonoscopy is effective for early detection, it has poor adherence; thus, non-invasive screening methods are needed. In work not presented in this thesis, we generated a monoclonal antibody against Tn and developed a microarray platform with our Tn antibody to detect Tn antigen in feces and serum. Using our mouse model, we were able to detect Tn in feces of

KO and mosaic mice at ng levels prior to the onset of colitis/cancer suggesting that this may be a promising approach for early detection of cancer or IBD.

In sum, we discovered that Cosmc is a sex-specific risk gene in IBD that regulates the microbiota biogeography and that mucins of the mucus layer have an unappreciated ability to migrate along the crypt surface. Further, in work not presented, we discovered that Tn can initiate colorectal cancer and that an antibody against Tn can be used for early detection of cancer in a mouse model.

Concluding remarks

The studies in this thesis were motivated by three central questions: What glycans are present in cells? How do they change in disease? How do these changes contribute to disease biology? We developed a technology, termed CORA, that enabled us to predict the human cellular glycome and begin to investigate its regulation in disease. In separate studies, we deleted a glycogene implicated in IBD and discovered a molecular and ecological basis for how altered epithelial glycosylation contributes to IBD. This thesis makes major strides in answering our three central questions, providing a foundation to understand the role of glycosylation in health and disease.

References

1. Warren L, Buck CA, & Tuszynski GP (1978) Glycopeptide changes and malignant transformation. A possible role for carbohydrate in malignant behavior. *Biochimica et biophysica acta* 516(1):97-127.

2. Meezan E, Wu HC, Black PH, & Robbins PW (1969) Comparative studies on the carbohydrate-containing membrane components of normal and virus-transformed mouse fibroblasts. II. Separation of glycoproteins and glycopeptides by sephadex chromatography. *Biochemistry* 8(6):2518-2524.

3. Buck CA, Glick MC, & Warren L (1971) Glycopeptides from the surface of control and virus-transformed cells. *Science* 172(3979):169-171.

4. Aub JC, Tieslau C, & Lankester A (1963) Reactions of Normal and Tumor Cell Surfaces to Enzymes. I. Wheat-Germ Lipase and Associated Mucopolysaccharides. *Proceedings of the National Academy of Sciences of the United States of America* 50:613-619.

5. Ozanne B & Sambrook J (1971) Binding of radioactively labelled concanavalin A and wheat germ agglutinin to normal and virus-transformed cells. *Nature: New biology* 232(31):156-160.

6. Feizi T (1985) Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. *Nature* 314(6006):53-57.

7. Ohtsubo K & Marth JD (2006) Glycosylation in cellular mechanisms of health and disease. *Cell* 126(5):855-867.

8. Ju T, Aryal RP, Kudelka MR, Wang Y, & Cummings RD (2014) The Cosmc connection to the Tn antigen in cancer. *Cancer biomarkers : section A of Disease markers* 14(1):63-81.

9. Ju T, Otto VI, & Cummings RD (2011) The Tn antigen-structural simplicity and biological complexity. *Angewandte Chemie* 50(8):1770-1791.

10. Ju T*, et al.* (2013) Tn and sialyl-Tn antigens, aberrant O-glycomics as human disease markers. *Proteomics. Clinical applications*.

11. Schjoldager KT & Clausen H (2012) Site-specific protein O-glycosylation modulates proprotein processing - deciphering specific functions of the large polypeptide GalNAc-transferase gene family. *Biochimica et biophysica acta* 1820(12):2079-2094.

12. Hansen JE*, et al.* (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate journal* 15(2):115-130.

13. Julenius K, Molgaard A, Gupta R, & Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15(2):153-164.

14. Steentoft C*, et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* 32(10):1478-1488.

15. Gottschalk A & Murphy WH (1961) Studies on mucoproteins. IV. The linkage of the prosthetic group to the protein core in ovine submaxillary gland mucoprotein. *Biochimica et biophysica acta* 46:81-90.

16. Tanaka K, Bertolini M, & Pigman W (1964) Serine and threonine glycosidic linkages in bovine submaxillary mucin. *Biochemical and biophysical research communications* 16(5):404-409.

17. Carubelli R, Bhavanandan P, & Gottschalk A (1965) Studies on Glycoproteins. Xi. The O-Glycosidic Linkage of N-Acetylgalactosamine to Seryl and Threonyl Residues in Ovine Submaxillary Gland Glycoprotein. *Biochimica et biophysica acta* 101:67-82.

18. Schauer H & Gottschalk A (1968) Studies on glycoproteins. XVII. Purification of O-seryl-N-acetylgalactosaminide glycohydrolase. Its complete separation from N-acetyl-beta-D-hexosaminidase and proteases. *Biochimica et biophysica acta* 156(2):304-310.

19. Dahr W, Uhlenbruck G, & Bird GW (1974) Cryptic A-like receptor sites in human erythrocyte glycoproteins: proposed nature of Tn-antigen. *Vox Sang* 27(1):29-42.

20. Hollingsworth MA & Swanson BJ (2004) Mucins in cancer: protection and control of the cell surface. *Nature reviews. Cancer* 4(1):45-60.

21. Larsson JM, Karlsson H, Sjovall H, & Hansson GC (2009) A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 19(7):756-766.

22. Cummings RD*, et al.* (1983) Biosynthesis of N- and O-linked oligosaccharides of the low density lipoprotein receptor. *The Journal of biological chemistry* 258(24):15261-15273.

23. Sasaki H, Bothner B, Dell A, & Fukuda M (1987) Carbohydrate structure of erythropoietin expressed in Chinese hamster ovary cells by a human erythropoietin cDNA. *The Journal of biological chemistry* 262(25):12059-12076.

24. Do SI & Cummings RD (1992) Presence of O-linked oligosaccharide on a threonine residue in the human transferrin receptor. *Glycobiology* 2(4):345-353.

25. Do SI, Enns C, & Cummings RD (1990) Human transferrin receptor contains O-linked oligosaccharides. *The Journal of biological chemistry* 265(1):114-125.

26. Alexander WS, *et al.* (2006) Thrombocytopenia and kidney disease in mice with a mutation in the C1galt1 gene. *Proceedings of the National Academy of Sciences of the United States of America* 103(44):16442-16447.

27. An G, *et al.* (2007) Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *The Journal of experimental medicine* 204(6):1417-1429.

28. Ellies LG, *et al.* (1998) Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation. *Immunity* 9(6):881-890.

29. Fu J, *et al.* (2011) Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. *J Clin Invest* 121(4):1657-1666.

30. Priatel JJ, *et al.* (2000) The ST3Gal-I sialyltransferase controls CD8+ T lymphocyte homeostasis by modulating O-glycan biosynthesis. *Immunity* 12(3):273-283.

31. Tenno M, *et al.* (2007) Initiation of protein O glycosylation by the polypeptide GalNAcT-1 in vascular biology and humoral immunity. *Molecular and cellular biology* 27(24):8783-8796.

32. Wang Y, *et al.* (2012) Platelet biogenesis and functions require correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* 109(40):16143-16148.

33. Wang Y, *et al.* (2010) Cosmc is an essential chaperone for correct protein O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America* 107(20):9228-9233.

34. Xia L, *et al.* (2004) Defective angiogenesis and fatal embryonic hemorrhage in mice lacking core 1-derived O-glycans. *The Journal of cell biology* 164(3):451-459.

35. Yeh JC, *et al.* (2001) Novel sulfated lymphocyte homing receptors and their control by a Core1 extension beta 1,3-N-acetylglucosaminyltransferase. *Cell* 105(7):957-969.

36. Fakhro KA, *et al.* (2011) Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proceedings of the National Academy of Sciences of the United States of America* 108(7):2915-2920.

37. Higgins EA, Siminovitch KA, Zhuang DL, Brockhausen I, & Dennis JW (1991) Aberrant O-linked oligosaccharide biosynthesis in lymphocytes and platelets from patients with the Wiskott-Aldrich syndrome. *The Journal of biological chemistry* 266(10):6280-6290.

38. Ju T & Cummings RD (2005) Protein glycosylation: chaperone mutation in Tn syndrome. *Nature* 437(7063):1252.

39. Teslovich TM*, et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466(7307):707-713.

40. Topaz O*, et al.* (2004) Mutations in GALNT3, encoding a protein involved in O-linked glycosylation, cause familial tumoral calcinosis. *Nature genetics* 36(6):579-581.

41. Schjoldager KT*, et al.* (2012) Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proceedings of the National Academy of Sciences of the United States of America* 109(25):9893-9898.

42. Cummings RD & Pierce JM (2014) The challenge and promise of glycomics. *Chem Biol* 21(1):1-15.

43. Varki A & Angata T (2006) Siglecs--the major subfamily of I-type lectins. *Glycobiology* 16(1):1R-27R.

44. Kingsley DM, Kozarsky KF, Hobbie L, & Krieger M (1986) Reversible defects in O-linked glycosylation and LDL receptor expression in a UDP-Gal/UDP-GalNAc 4-epimerase deficient mutant. *Cell* 44(5):749-759.

45. Kingsley DM & Krieger M (1984) Receptor-mediated endocytosis of low density lipoprotein: somatic cell mutants define multiple genes required for expression of surface-receptor activity. *Proceedings of the National Academy of Sciences of the United States of America* 81(17):5454-5458.

46. Kozarsky K, Kingsley D, & Krieger M (1988) Use of a mutant cell line to study the kinetics and function of O-linked glycosylation of low density lipoprotein receptors. *Proceedings of the National Academy of Sciences of the United States of America* 85(12):4335-4339.

47. Bremer EG, Schlessinger J, & Hakomori S (1986) Ganglioside-mediated modulation of cell growth. Specific effects of GM3 on tyrosine phosphorylation of the epidermal growth factor receptor. *The Journal of biological chemistry* 261(5):2434-2440.

48. Dall'Olio F & Chiricolo M (2001) Sialyltransferases in cancer. *Glycoconjugate journal* 18(11-12):841-850.

49. Dennis J, Waller C, Timpl R, & Schirrmacher V (1982) Surface sialic acid reduces attachment of metastatic tumour cells to collagen type IV and fibronectin. *Nature* 300(5889):274-276.

50. Dennis JW & Laferte S (1989) Oncodevelopmental expression of--GlcNAc beta 1-6Man alpha 1-6Man beta 1--branched asparagine-linked oligosaccharides in murine tissues and human breast carcinomas. *Cancer research* 49(4):945-950.

51. Dennis JW, Laferte S, Waghorne C, Breitman ML, & Kerbel RS (1987) Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis. *Science* 236(4801):582-585.

52. Fernandes B, Sagman U, Auger M, Demetrio M, & Dennis JW (1991) Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia. *Cancer research* 51(2):718-723.

53. Fuster MM & Esko JD (2005) The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature reviews. Cancer* 5(7):526-542.

54. Ganzinger U & Deutsch E (1980) Serum sialyltransferase levels as a parameter in the diagnosis and follow-up of gastrointestinal tumors. *Cancer research* 40(4):1300-1304.

55. Granovsky M*, et al.* (2000) Suppression of tumor growth and metastasis in Mgat5-deficient mice. *Nature medicine* 6(3):306-312.

56. Guo HB, Lee I, Kamar M, Akiyama SK, & Pierce M (2002) Aberrant N-glycosylation of beta1 integrin causes reduced alpha5beta1 integrin clustering and stimulates cell migration. *Cancer research* 62(23):6837-6845.

57. Hakomori S (1996) Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer research* 56(23):5309-5318.

58. Nagy P*, et al.* (2002) Lipid rafts and the local density of ErbB proteins influence the biological role of homo- and heteroassociations of ErbB2. *Journal of cell science* 115(Pt 22):4251-4262.

59. Partridge EA*, et al.* (2004) Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis. *Science* 306(5693):120-124.

60. Santer UV, Gilbert F, & Glick MC (1984) Change in glycosylation of membrane glycoproteins after transfection of NIH 3T3 with human tumor DNA. *Cancer research* 44(9):3730-3735.

61. Tai T, Paulson JC, Cahan LD, & Irie RF (1983) Ganglioside GM2 as a human tumor antigen (OFA-I-1). *Proceedings of the National Academy of Sciences of the United States of America* 80(17):5392-5396.

62. van Beek WP, Smets LA, & Emmelot P (1973) Increased sialic acid density in surface glycoprotein of transformed and malignant cells--a general phenomenon? *Cancer research* 33(11):2913-2922.

63. Yamashita K, Tachibana Y, Ohkura T, & Kobata A (1985) Enzymatic basis for the structural changes of asparagine-linked sugar chains of membrane glycoproteins of baby hamster kidney cells induced by polyoma transformation. *The Journal of biological chemistry* 260(7):3963-3969.

64. Wargovich MJ*, et al.* (2004) Expression of cellular adhesion proteins and abnormal glycoproteins in human aberrant crypt foci. *Applied immunohistochemistry & molecular morphology : AIMM / official publication of the Society for Applied Immunohistochemistry* 12(4):350-355.

65. Takada A*, et al.* (1993) Contribution of carbohydrate antigens sialyl Lewis A and sialyl Lewis X to adhesion of human cancer cells to vascular endothelium. *Cancer research* 53(2):354-361.

66. Fuster MM, Brown JR, Wang L, & Esko JD (2003) A disaccharide precursor of sialyl Lewis X inhibits metastatic potential of tumor cells. *Cancer research* 63(11):2775-2781.

67. Kim YJ, Borsig L, Varki NM, & Varki A (1998) P-selectin deficiency attenuates tumor growth and metastasis. *Proceedings of the National Academy of Sciences of the United States of America* 95(16):9325-9330.

68. Biancone L, Araki M, Araki K, Vassalli P, & Stamenkovic I (1996) Redirection of tumor metastasis by expression of E-selectin in vivo. *The Journal of experimental medicine* 183(2):581-587.

69. Freeze HH & Ng BG (2011) Golgi glycosylation and human inherited diseases. *Cold Spring Harbor perspectives in biology* 3(9):a005371.

70. Bennett EP*, et al.* (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22(6):736-756.

71. Rottger S*, et al.* (1998) Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *Journal of cell science* 111 (Pt 1):45-60.

72. Roth J, Wang Y, Eckhardt AE, & Hill RL (1994) Subcellular localization of the UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-mediated O-glycosylation reaction in the submaxillary gland. *Proceedings of the National Academy of Sciences of the United States of America* 91(19):8935-8939.

73. Tabak LA (2010) The role of mucin-type O-glycans in eukaryotic development. *Seminars in cell & developmental biology* 21(6):616-621.

74. Young WW, Jr., Holcomb DR, Ten Hagen KG, & Tabak LA (2003) Expression of UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase isoforms in murine tissues determined by real-time PCR: a new view of a large family. *Glycobiology* 13(7):549-557.

75. Gerken TA, Raman J, Fritz TA, & Jamison O (2006) Identification of common and unique peptide substrate preferences for the UDP-GalNAc:polypeptide alpha-N-acetylgalactosaminyltransferases T1 and T2 derived from oriented random peptide substrates. *The Journal of biological chemistry* 281(43):32403-32416.

76. Wandall HH*, et al.* (1997) Substrate specificities of three members of the human UDP-N-acetyl-alpha-D-galactosamine:Polypeptide N-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *The Journal of biological chemistry* 272(38):23503-23514.

77. Orr SL*, et al.* (2013) A phenotype survey of 36 mutant mouse strains with gene-targeted defects in glycosyltransferases or glycan-binding proteins. *Glycobiology* 23(3):363-380.

78. Ju T, Aryal RP, Stowell CJ, & Cummings RD (2008) Regulation of protein O-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *The Journal of cell biology* 182(3):531-542.

79. Ju T, Brewer K, D'Souza A, Cummings RD, & Canfield WM (2002) Cloning and expression of human core 1 beta1,3-galactosyltransferase. *The Journal of biological chemistry* 277(1):178-186.

80. Ju T & Cummings RD (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* 99(26):16613-16618.

81. Aryal RP, Ju T, & Cummings RD (2010) The endoplasmic reticulum chaperone Cosmc directly promotes in vitro folding of T-synthase. *The Journal of biological chemistry* 285(4):2456-2462.

82. Aryal RP, Ju T, & Cummings RD (2012) Tight complex formation between Cosmc chaperone and its specific client non-native T-synthase leads to enzyme activity and client-driven dissociation. *The Journal of biological chemistry* 287(19):15317-15329.

83. Ju T, Zheng Q, & Cummings RD (2006) Identification of core 1 O-glycan T-synthase from Caenorhabditis elegans. *Glycobiology* 16(10):947-958.

84. Aryal RP, Ju T, & Cummings RD (2014) Identification of a novel protein binding motif within the T-synthase for the molecular chaperone Cosmc. *The Journal of biological chemistry* 289(17):11630-11641.

85. Narimatsu Y*, et al.* (2011) Co-translational function of Cosmc, core 1 synthase specific molecular chaperone, revealed by a cell-free translation system. *FEBS letters* 585(9):1276-1280.

86. Ju T*, et al.* (2008) Human tumor antigens Tn and sialyl Tn arise from mutations in Cosmc. *Cancer research* 68(6):1636-1646.

87. Bierhuizen MF & Fukuda M (1992) Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen. *Proceedings of the National Academy of Sciences of the United States of America* 89(19):9326-9330.

88. Schwientek T*, et al.* (1999) Control of O-glycan branch formation. Molecular cloning of human cDNA encoding a novel beta1,6-N-acetylglucosaminyltransferase forming core 2 and core 4. *The Journal of biological chemistry* 274(8):4504-4512.

89. Schwientek T*, et al.* (2000) Control of O-glycan branch formation. Molecular cloning and characterization of a novel thymus-associated core 2 beta1, 6-n-acetylglucosaminyltransferase. *The Journal of biological chemistry* 275(15):11106-11113.

90. Stone EL*, et al.* (2009) Glycosyltransferase function in core 2-type protein O glycosylation. *Molecular and cellular biology* 29(13):3770-3782.

91. Yeh JC, Ong E, & Fukuda M (1999) Molecular cloning and expression of a novel beta-1, 6-N-acetylglucosaminyltransferase that forms core 2, core 4, and I branches. *The Journal of biological chemistry* 274(5):3215-3221.

92. Tian E & Ten Hagen KG (2009) Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconjugate journal* 26(3):325-334.

93. Fukuda M (2006) Roles of mucin-type O-glycans synthesized by core2beta1,6-N-acetylglucosaminyltransferase. *Methods in enzymology* 416:332-346.

94. Falkenberg VR, Alvarez K, Roman C, & Fregien N (2003) Multiple transcription initiation and alternative splicing in the 5' untranslated region of the core 2 beta1-6 N-acetylglucosaminyltransferase I gene. *Glycobiology* 13(6):411-418.

95. Sekine M, Nara K, & Suzuki A (1997) Tissue-specific regulation of mouse core 2 beta-1,6-N-acetylglucosaminyltransferase. *The Journal of biological chemistry* 272(43):27246-27252.

96. Brockhausen I (2006) Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO reports* 7(6):599-604.

97. Lefebvre JC*, et al.* (1994) Altered glycosylation of leukosialin, CD43, in HIV-1-infected cells of the CEM line. *The Journal of experimental medicine* 180(5):1609-1617.

98. Iwai T*, et al.* (2002) Molecular cloning and characterization of a novel UDP-GlcNAc:GalNAc-peptide beta1,3-N-acetylglucosaminyltransferase (beta 3Gn-T6), an enzyme synthesizing the core 3 structure of O-glycans. *The Journal of biological chemistry* 277(15):12802-12809.

99. Pang PC*, et al.* (2011) Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. *Science* 333(6050):1761-1764.

100. Leppanen A*, et al.* (1999) A novel glycosulfopeptide binds to P-selectin and inhibits leukocyte adhesion to P-selectin. *The Journal of biological chemistry* 274(35):24838-24848.

101. Somers WS, Tang J, Shaw GD, & Camphausen RT (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* 103(3):467-479.

102. Kumar R, Camphausen RT, Sullivan FX, & Cumming DA (1996) Core2 beta-1,6-N-acetylglucosaminyltransferase enzyme activity is critical for P-selectin glycoprotein ligand-1 binding to P-selectin. *Blood* 88(10):3872-3879.

103. Bruehl RE, Bertozzi CR, & Rosen SD (2000) Minimal sulfated carbohydrates for recognition by L-selectin and the MECA-79 antibody. *The Journal of biological chemistry* 275(42):32642-32648.

104. Hemmerich S, Butcher EC, & Rosen SD (1994) Sulfation-dependent recognition of high endothelial venules (HEV)-ligands by L-selectin and MECA 79, and adhesion-blocking monoclonal antibody. *The Journal of experimental medicine* 180(6):2219-2226.

105. Yang RY, Rabinovich GA, & Liu FT (2008) Galectins: structure, function and therapeutic potential. *Expert reviews in molecular medicine* 10:e17.

106. Watanabe K, Hakomori SI, Childs RA, & Feizi T (1979) Characterization of a blood group I-active ganglioside. Structural requirements for I and i specificities. *The Journal of biological chemistry* 254(9):3221-3228.

107. Fukuda M*, et al.* (1985) Structures of sialylated fucosyl polylactosaminoglycans isolated from chronic myelogenous leukemia cells. *The Journal of biological chemistry* 260(24):12957-12967.

108. Fukuda M, Carlsson SR, Klock JC, & Dell A (1986) Structures of O-linked oligosaccharides isolated from normal granulocytes, chronic myelogenous leukemia cells, and acute myelogenous leukemia cells. *The Journal of biological chemistry* 261(27):12796-12806.

109. Clausen H & Hakomori S (1989) ABH and related histo-blood group antigens; immunochemical differences in carrier isotypes and their distribution. *Vox Sang* 56(1):1-20.

110. Inaba N*, et al.* (2003) A novel I-branching beta-1,6-N-acetylglucosaminyltransferase involved in human blood group I antigen expression. *Blood* 101(7):2870-2876.

111. Yang JM*, et al.* (1994) Alterations of O-glycan biosynthesis in human colon cancer tissues. *Glycobiology* 4(6):873-884.

112. Thomsson KA*, et al.* (2012) Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1- and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* 22(8):1128-1139.

113. Vavasseur F, Yang JM, Dole K, Paulsen H, & Brockhausen I (1995) 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(3):351-357.

114. Podolsky DK (1985) Oligosaccharide structures of isolated human colonic mucin species. *The Journal of biological chemistry* 260(29):15510-15515.

115. Capon C, Maes E, Michalski JC, Leffler H, & Kim YS (2001) Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. *The Biochemical journal* 358(Pt 3):657-664.

116. Marsh WL (1961) Anti-i: a cold antibody defining the Ii relationship in human red cells. *British journal of haematology* 7:200-209.

117. Henry S, Oriol R, & Samuelsson B (1995) Lewis histo-blood group system and associated secretory phenotypes. *Vox Sang* 69(3):166-182.

118. Stowell SR*, et al.* (2010) Innate immune lectins kill bacteria expressing blood group antigen. *Nature medicine* 16(3):295-301.

119. Kukowska-Latallo JF, Larsen RD, Nair RP, & Lowe JB (1990) A cloned human cDNA determines expression of a mouse stage-specific embryonic antigen and the Lewis blood group alpha(1,3/1,4)fucosyltransferase. *Genes & development* 4(8):1288-1303.

120. Stanley P & Cummings RD (2009) Structures Common to Different Glycans. *Essentials of Glycobiology*, eds Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, Hart GW, & Etzler MECold Spring Harbor (NY)), 2nd Ed.

121. Matsushita Y, Cleary KR, Ota DM, Hoff SD, & Irimura T (1990) Sialyl-dimeric Lewis-X antigen expressed on mucin-like glycoproteins in colorectal cancer metastases. *Lab Invest* 63(6):780-791.

122. Tangvoranuntakul P*, et al.* (2003) Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid. *Proceedings of the National Academy of Sciences of the United States of America* 100(21):12045-12050.

123. Hedlund M, Padler-Karavani V, Varki NM, & Varki A (2008) Evidence for a human-specific mechanism for diet and antibody-mediated inflammation in carcinoma progression. *Proceedings of the National Academy of Sciences of the United States of America* 105(48):18936-18941.

124. Harduin-Lepers A*, et al.* (2001) The human sialyltransferase family. *Biochimie* 83(8):727-737.

125. Muthana SM, Campbell CT, & Gildersleeve JC (2012) Modifications of glycans: biological significance and therapeutic opportunities. *ACS chemical biology* 7(1):31-43.

126. Yu H & Chen X (2007) Carbohydrate post-glycosylational modifications. *Organic & biomolecular chemistry* 5(6):865-872.

127. 大内清太 (1949) Polysaccharides and a glycidamin in the tissue of gastric cancer. *The Tohoku Journal of Experimental Medicine* 51(3-4):297-304.

128. Nuti M*, et al.* (1982) A monoclonal antibody (B72.3) defines patterns of distribution of a novel tumor-associated antigen in human mammary carcinoma cell populations. *Int J Cancer* 29(5):539-545.

129. Magnani JL*, et al.* (1982) A monoclonal antibody-defined antigen associated with gastrointestinal cancer is a ganglioside containing sialylated lacto-N-fucopentaose II. *The Journal of biological chemistry* 257(23):14365-14369.

130. Lee JS*, et al.* (1991) Expression of blood-group antigen A--a favorable prognostic factor in non-small-cell lung cancer. *N Engl J Med* 324(16):1084-1090.

131. Miyake M, Taki T, Hitomi S, & Hakomori S (1992) Correlation of expression of H/Le(y)/Le(b) antigens with survival in patients with carcinoma of the lung. *N Engl J Med* 327(1):14-18.

132. Prokop O & Uhlenbruck G (1969) [N-acetyl-D-galactosamine in tumor cell membranes: demonstration by means of Helix agglutinins]. *Die Medizinische Welt* 46:2515-2519.

133. Takahashi HK, Metoki R, & Hakomori S (1988) Immunoglobulin G3 monoclonal antibody directed to Tn antigen (tumor-associated alpha-N-acetylgalactosaminyl

epitope) that does not cross-react with blood group A antigen. *Cancer research* 48(15):4361-4367.

134. Magnani JL, Steplewski Z, Koprowski H, & Ginsburg V (1983) Identification of the gastrointestinal and pancreatic cancer-associated antigen detected by monoclonal antibody 19-9 in the sera of patients as a mucin. *Cancer research* 43(11):5489-5492.

135. Gendler SJ*, et al.* (1990) Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *The Journal of biological chemistry* 265(25):15286-15293.

136. Yin BW & Lloyd KO (2001) Molecular cloning of the CA125 ovarian cancer antigen: identification as a new mucin, MUC16. *The Journal of biological chemistry* 276(29):27371-27375.

137. Cummings RD (2009) The repertoire of glycan determinants in the human glycome. *Mol Biosyst* 5(10):1087-1104.

138. Heimburg-Molinaro J*, et al.* (2013) Microarray analysis of the human antibody response to synthetic Cryptosporidium glycopeptides. *International journal for parasitology* 43(11):901-907.

139. Han BW, Herrin BR, Cooper MD, & Wilson IA (2008) Antigen recognition by variable lymphocyte receptors. *Science* 321(5897):1834-1837.

140. Hong X*, et al.* (2013) Sugar-binding proteins from fish: selection of high affinity "lambodies" that recognize biomedically relevant glycans. *ACS chemical biology* 8(1):152-160.

141. Hirohashi S, Clausen H, Yamada T, Shimosato Y, & Hakomori S (1985) Blood group A cross-reacting epitope defined by monoclonal antibodies NCC-LU-35 and -81 expressed in cancer of blood group O or B individuals: its identification as Tn antigen. *Proceedings of the National Academy of Sciences of the United States of America* 82(20):7039-7043.

142. Dausset J, Moullec J, & Bernard J (1959) Acquired hemolytic anemia with polyagglutinability of red blood cells due to a new factor present in normal human serum (Anti-Tn). *Blood* 14:1079-1093.

143. Springer GF (1984) T and Tn, general carcinoma autoantigens. *Science* 224(4654):1198-1206.

144. Springer GF, Desai PR, & Banatwala I (1975) Blood group MN antigens and precursors in normal and malignant human breast glandular tissue. *J Natl Cancer Inst* 54(2):335-339.

145. Springer GF, Murthy MS, Desai PR, & Scanlon EF (1980) Breast cancer patient's cell-mediated immune response to Thomsen-Friedenreich (T) antigen. *Cancer* 45(12):2949-2954.

146. Itzkowitz SH*, et al.* (1989) Expression of Tn, sialosyl-Tn, and T antigens in human colon cancer. *Cancer research* 49(1):197-204.

147. Itzkowitz SH, Bloom EJ, Lau TS, & Kim YS (1992) Mucin associated Tn and sialosyl-Tn antigen expression in colorectal polyps. *Gut* 33(4):518-523.

148. Cao Y*, et al.* (1995) Expression of Thomsen-Friedenreich-related antigens in primary and metastatic colorectal carcinomas. A reevaluation. *Cancer* 76(10):1700-1708.

149. David L, Nesland JM, Clausen H, Carneiro F, & Sobrinho-Simoes M (1992) Simple mucin-type carbohydrate antigens (Tn, sialosyl-Tn and T) in gastric mucosa, carcinomas and metastases. *APMIS Suppl* 27:162-172.

150. Kakeji Y, Tsujitani S, Mori M, Maehara Y, & Sugimachi K (1991) Helix pomatia agglutinin binding activity is a predictor of survival time for patients with gastric carcinoma. *Cancer* 68(11):2438-2442.

151. Osako M*, et al.* (1993) Immunohistochemical study of mucin carbohydrates and core proteins in human pancreatic tumors. *Cancer* 71(7):2191-2199.

152. Langkilde NC, Wolf H, Clausen H, Kjeldsen T, & Orntoft TF (1992) Nuclear volume and expression of T-antigen, sialosyl-Tn-antigen, and Tn-antigen in carcinoma of the human bladder. Relation to tumor recurrence and progression. *Cancer* 69(1):219-227.

153. Laack E*, et al.* (2002) Lectin histochemistry of resected adenocarcinoma of the lung: helix pomatia agglutinin binding is an independent prognostic factor. *The American journal of pathology* 160(3):1001-1008.

154. Hirao T, Sakamoto Y, Kamada M, Hamada S, & Aono T (1993) Tn antigen, a marker of potential for metastasis of uterine cervix cancer cells. *Cancer* 72(1):154-159.

155. Hamada S, Furumoto H, Kamada M, Hirao T, & Aono T (1993) High expression rate of Tn antigen in metastatic lesions of uterine cervical cancers. *Cancer letters* 74(3):167-173.

156. Therkildsen MH, Mandel U, Christensen M, & Dabelsteen E (1993) Simple mucin-type Tn and sialosyl-Tn carbohydrate antigens in salivary gland carcinomas. *Cancer* 72(4):1147-1154.

157. Roxby DJ, Pfeiffer MB, Morley AA, & Kirkland MA (1992) Expression of the Tn antigen in myelodysplasia, lymphoma, and leukemia. *Transfusion* 32(9):834-838.

158. Yuan M (1989) [The expression of Tn and S-Tn antigens in cancer and pre-malignant lesion of colorectal tissues by enzyme immunohistochemical method]. *Zhonghua bing li xue za zhi Chinese journal of pathology* 18(3):211-213.

159. Babino A*, et al.* (2000) Tn antigen is a pre-cancerous biomarker in breast tissue and serum in n-nitrosomethylurea-induced rat mammary carcinogenesis. *Int J Cancer* 86(6):753-759.

160. Berriel E*, et al.* (2005) Simple mucin-type cancer associated antigens are pre-cancerous biomarkers during 1,2-dimethylhydrazine-induced rat colon carcinogenesis. *Oncology reports* 14(1):219-227.

161. Desai PR (2000) Immunoreactive T and Tn antigens in malignancy: role in carcinoma diagnosis, prognosis, and immunotherapy. *Transfusion medicine reviews* 14(4):312-325.

162. Konno A, Hoshino Y, Terashima S, Motoki R, & Kawaguchi T (2002) Carbohydrate expression profile of colorectal cancer cells is relevant to metastatic pattern and prognosis. *Clinical & experimental metastasis* 19(1):61-70.

163. Mi R*, et al.* (2012) Epigenetic silencing of the chaperone Cosmc in human leukocytes expressing tn antigen. *The Journal of biological chemistry* 287(49):41523-41533.

164. Radhakrishnan P*, et al.* (2014) Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences of the United States of America*.

165. Gill DJ, Chia J, Senewiratne J, & Bard F (2010) Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *The Journal of cell biology* 189(5):843-858.

166. Kellokumpu S, Sormunen R, & Kellokumpu I (2002) Abnormal glycosylation and altered Golgi structure in colorectal cancer: dependence on intra-Golgi pH. *FEBS letters* 516(1-3):217-224.

167. Petrosyan A, Ali MF, & Cheng PW (2012) Glycosyltransferase-specific Golgi-targeting mechanisms. *The Journal of biological chemistry* 287(45):37621-37627.

168. Colcher D, Hand PH, Nuti M, & Schlom J (1981) A spectrum of monoclonal antibodies reactive with human mammary tumor cells. *Proceedings of the National Academy of Sciences of the United States of America* 78(5):3199-3203.

169. Johnson VG*, et al.* (1986) Analysis of a human tumor-associated glycoprotein (TAG-72) identified by monoclonal antibody B72.3. *Cancer research* 46(2):850-857.

170. Kjeldsen T*, et al.* (1988) Preparation and characterization of monoclonal antibodies directed to the tumor-associated O-linked sialosyl-2----6 alpha-N-acetylgalactosaminyl (sialosyl-Tn) epitope. *Cancer research* 48(8):2214-2220.

171. Julien S, Videira PA, & Delannoy P (2012) Sialyl-tn in cancer: (how) did we miss the target? *Biomolecules* 2(4):435-466.

172. Thor A, Ohuchi N, Szpak CA, Johnston WW, & Schlom J (1986) Distribution of oncofetal antigen tumor-associated glycoprotein-72 defined by monoclonal antibody B72.3. *Cancer research* 46(6):3118-3124.

173. Vazquez-Martin C, Cuevas E, Gil-Martin E, & Fernandez-Briera A (2004) Correlation analysis between tumor-associated antigen sialyl-Tn expression and ST6GalNAc I activity in human colon adenocarcinoma. *Oncology* 67(2):159-165.

174. Itzkowitz SH*, et al.* (1990) Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients. *Cancer* 66(9):1960-1966.

175. Ma XC*, et al.* (1993) Expression of sialyl-Tn antigen is correlated with survival time of patients with gastric carcinomas. *Eur J Cancer* 29A(13):1820-1823.

176. Werther JL, Rivera-MacMurray S, Bruckner H, Tatematsu M, & Itzkowitz SH (1994) Mucin-associated sialosyl-Tn antigen expression in gastric cancer correlates with an adverse outcome. *Br J Cancer* 69(3):613-616.

177. Werther JL*, et al.* (1996) Sialosyl-Tn antigen as a marker of gastric cancer progression: an international study. *Int J Cancer* 69(3):193-199.

178. Kim GE*, et al.* (2002) Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas. *Gastroenterology* 123(4):1052-1060.

179. Inoue M*, et al.* (1991) Immunodetection of sialyl-Tn antigen in normal, hyperplastic and cancerous tissues of the uterine endometrium. *Virchows Arch A Pathol Anat Histopathol* 418(2):157-162.

180. Inoue M, Ton SM, Ogawa H, & Tanizawa O (1991) Expression of Tn and sialyl-Tn antigens in tumor tissues of the ovary. *Am J Clin Pathol* 96(6):711-716.

181. Motoo Y*, et al.* (1991) Serum sialyl-Tn antigen levels in patients with digestive cancers. *Oncology* 48(4):321-326.

182. Nanashima A*, et al.* (1999) High serum concentrations of sialyl Tn antigen in carcinomas of the biliary tract and pancreas. *J Hepatobiliary Pancreat Surg* 6(4):391-395.

183. Itzkowitz SH*, et al.* (1995) Sialosyl-Tn antigen: initial report of a new marker of malignant progression in long-standing ulcerative colitis. *Gastroenterology* 109(2):490-497.

184. Itzkowitz SH*, et al.* (1996) Sialosyl-Tn antigen is prevalent and precedes dysplasia in ulcerative colitis: a retrospective case-control study. *Gastroenterology* 110(3):694-704.

185. Kobayashi H, Terao T, & Kawashima Y (1992) Serum sialyl Tn as an independent predictor of poor prognosis in patients with epithelial ovarian cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 10(1):95-101.

186. Julien S*, et al.* (2001) Expression of sialyl-Tn antigen in breast cancer cells transfected with the human CMP-Neu5Ac: GalNAc alpha2,6-sialyltransferase (ST6GalNac I) cDNA. *Glycoconjugate journal* 18(11-12):883-893.

187. Marcos NT*, et al.* (2004) Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Cancer research* 64(19):7050-7057.

188. Marcos NT*, et al.* (2011) ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci (Elite Ed)* 3:1443-1455.

189. Jass JR, Allison LM, & Edgar S (1994) Monoclonal antibody TKH2 to the cancer-associated epitope sialosyl Tn shows cross-reactivity with variants of normal colorectal goblet cell mucin. *Pathology* 26(4):418-422.

190. Butcher EC*, et al.* (1982) Surface phenotype of Peyer's patch germinal center cells: implications for the role of germinal centers in B cell differentiation. *Journal of immunology* 129(6):2698-2707.

191. Reepmaker J (1952) The relation between polyagglutinability of erythrocytes in vivo and the Hubener-Thomsen-Friedenreich phenomenon. *Journal of clinical pathology* 5(3):266-270.

192. Springer GF, Desai PR, Tegtmeyer H, Spencer BD, & Scanlon EF (1993) Pancarcinoma T/Tn antigen detects human carcinoma long before biopsy does and its vaccine prevents breast carcinoma recurrence. *Ann N Y Acad Sci* 690:355-357.

193. Klein PJ*, et al.* (1979) The presence and significance of the Thomsen-Friedenreich antigen in mammary gland. II. Its topochemistry in normal, hyperplastic and carcinoma tissue of the breast. *J Cancer Res Clin Oncol* 93(2):205-214.

194. Campbell BJ, Finnie IA, Hounsell EF, & Rhodes JM (1995) Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic

adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *The Journal of clinical investigation* 95(2):571-576.

195. Kumamoto K*, et al.* (2001) Increased expression of UDP-galactose transporter messenger RNA in human colon cancer tissues and its implication in synthesis of Thomsen-Friedenreich antigen and sialyl Lewis A/X determinants. *Cancer research* 61(11):4620-4627.

196. Yu LG (2007) The oncofetal Thomsen-Friedenreich carbohydrate antigen in cancer progression. *Glycoconjugate journal* 24(8):411-420.

197. Brockhausen I, Yang JM, Burchell J, Whitehouse C, & Taylor-Papadimitriou J (1995) Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. *European journal of biochemistry / FEBS* 233(2):607-617.

198. Ogata S, Maimonis PJ, & Itzkowitz SH (1992) Mucins bearing the cancer-associated sialosyl-Tn antigen mediate inhibition of natural killer cell cytotoxicity. *Cancer research* 52(17):4741-4746.

199. van Vliet SJ, Paessens LC, Broks-van den Berg VC, Geijtenbeek TB, & van Kooyk Y (2008) The C-type lectin macrophage galactose-type lectin impedes migration of immature APCs. *Journal of immunology* 181(5):3148-3155.

200. Glinsky VV*, et al.* (2001) The role of Thomsen-Friedenreich antigen in adhesion of human breast and prostate cancer cells to the endothelium. *Cancer research* 61(12):4851-4857.

201. Gupta MK*, et al.* (1985) Measurement of a monoclonal-antibody-defined antigen (CA19-9) in the sera of patients with malignant and nonmalignant diseases. Comparison with carcinoembryonic antigen. *Cancer* 56(2):277-283.

202. Izumi Y*, et al.* (1995) Characterization of human colon carcinoma variant cells selected for sialyl Lex carbohydrate antigen: liver colonization and adhesion to vascular endothelial cells. *Experimental cell research* 216(1):215-221.

203. Shimodaira K*, et al.* (1997) Carcinoma-associated expression of core 2 beta-1,6-N-acetylglucosaminyltransferase gene in human colorectal cancer: role of O-glycans in tumor progression. *Cancer research* 57(23):5201-5206.

204. St Hill CA*, et al.* (2009) The high affinity selectin glycan ligand C2-O-sLex and mRNA transcripts of the core 2 beta-1,6-N-acetylglucosaminyltransferase (C2GnT1) gene are highly expressed in human colorectal adenocarcinomas. *BMC cancer* 9:79.

205. St Hill CA, Baharo-Hassan D, & Farooqui M (2011) C2-O-sLeX glycoproteins are E-selectin ligands that regulate invasion of human colon and hepatic carcinoma cells. *PloS one* 6(1):e16281.

206. Thun MJ, DeLancey JO, Center MM, Jemal A, & Ward EM (2010) The global burden of cancer: priorities for prevention. *Carcinogenesis* 31(1):100-110.

207. Anonymous (1993) An international association between Helicobacter pylori infection and gastric cancer. The EUROGAST Study Group. *Lancet* 341(8857):1359-1362.

208. Parsonnet J*, et al.* (1991) Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med* 325(16):1127-1131.

209. Nakayama J (2014) Dual Roles of Gastric Gland Mucin-specific O-glycans in Prevention of Gastric Cancer. *Acta Histochem Cytochem* 47(1):1-9.

210. Kawakubo M*, et al.* (2004) Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection. *Science* 305(5686):1003-1006.

211. Karasawa F*, et al.* (2012) Essential role of gastric gland mucin in preventing gastric cancer in mice. *The Journal of clinical investigation* 122(3):923-934.

212. Iwaya Y*, et al.* (2014) Reduced expression of alphaGlcNAc in Barrett's oesophagus adjacent to Barrett's adenocarcinoma--a possible biomarker to predict the malignant potential of Barrett's oesophagus. *Histopathology* 64(4):536-546.

213. Shiratsu K, Higuchi K, & Nakayama J (2014) Loss of gastric gland mucin-specific O-glycan is associated with progression of differentiated-type adenocarcinoma of the stomach. *Cancer Sci* 105(1):126-133.

214. Marcus DM & Cass LE (1969) Glycosphingolipids with Lewis blood group activity: uptake by human erythrocytes. *Science* 164(3879):553-555.

215. Atkinson BF*, et al.* (1982) Gastrointestinal cancer-associated antigen in immunoperoxidase assay. *Cancer research* 42(11):4820-4823.

216. Nakayama T, Watanabe M, Katsumata T, Teramoto T, & Kitajima M (1995) Expression of sialyl Lewis(a) as a new prognostic factor for patients with advanced colorectal carcinoma. *Cancer* 75(8):2051-2056.

217. Nakamori S*, et al.* (1997) Involvement of carbohydrate antigen sialyl Lewis(x) in colorectal cancer metastasis. *Dis Colon Rectum* 40(4):420-431.

218. Nakagoe T*, et al.* (2001) Comparison of the expression of ABH/Lewis-related antigens in polypoid and non-polypoid growth types of colorectal carcinoma. *J Gastroenterol Hepatol* 16(2):176-183.

219. Berger AC*, et al.* (2008) Postresection CA 19-9 predicts overall survival in patients with pancreatic cancer treated with adjuvant chemoradiation: a prospective validation by RTOG 9704. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 26(36):5918-5922.

220. Itzkowitz SH, *et al.* (1986) Lewisx- and sialylated Lewisx-related antigen expression in human malignant and nonmalignant colonic tissues. *Cancer research* 46(5):2627-2632.

221. Fukushima K, *et al.* (1984) Characterization of sialosylated Lewisx as a new tumor-associated antigen. *Cancer research* 44(11):5279-5285.

222. Nakamori S, *et al.* (1993) Increased expression of sialyl Lewisx antigen correlates with poor survival in patients with colorectal carcinoma: clinicopathological and immunohistochemical study. *Cancer research* 53(15):3632-3637.

223. Takada A, *et al.* (1991) Adhesion of human cancer cells to vascular endothelium mediated by a carbohydrate antigen, sialyl Lewis A. *Biochemical and biophysical research communications* 179(2):713-719.

224. Fukuda MN, *et al.* (2000) A peptide mimic of E-selectin ligand inhibits sialyl Lewis X-dependent lung colonization of tumor cells. *Cancer research* 60(2):450-456.

225. Matsumoto S, *et al.* (2002) Cimetidine increases survival of colorectal cancer patients with high levels of sialyl Lewis-X and sialyl Lewis-A epitope expression on tumour cells. *Br J Cancer* 86(2):161-167.

226. Lee AY, *et al.* (2005) Randomized comparison of low molecular weight heparin and coumarin derivatives on the survival of patients with cancer and venous thromboembolism. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 23(10):2123-2129.

227. Kannagi R (2004) Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression-The Warburg effect revisited. *Glycoconjugate journal* 20(5):353-364.

228. Yuan M, *et al.* (1985) Distribution of blood group antigens A, B, H, Lewisa, and Lewisb in human normal, fetal, and malignant colonic tissue. *Cancer research* 45(9):4499-4511.

229. Nakagoe T, *et al.* (2000) Expression of blood group antigens A, B and H in carcinoma tissue correlates with a poor prognosis for colorectal cancer patients. *J Cancer Res Clin Oncol* 126(7):375-382.

230. Gwin JL, *et al.* (1994) Loss of blood group antigen A in non-small cell lung cancer. *Ann Surg Oncol* 1(5):423-427.

231. Graziano SL, Tatum AH, Gonchoroff NJ, Newman NB, & Kohman LJ (1997) Blood group antigen A and flow cytometric analysis in resected early-stage non-small cell lung cancer. *Clin Cancer Res* 3(1):87-93.

232. Bianco T, Farmer BJ, Sage RE, & Dobrovic A (2001) Loss of red cell A, B, and H antigens is frequent in myeloid malignancies. *Blood* 97(11):3633-3639.

233. Bianco-Miotto T, Hussey DJ, Day TK, O'Keefe DS, & Dobrovic A (2009) DNA methylation of the ABO promoter underlies loss of ABO allelic expression in a significant proportion of leukemic patients. *PloS one* 4(3):e4788.

234. Orntoft TF, Wolf H, & Watkins WM (1988) Activity of the human blood group ABO, Se, H, Le, and X gene-encoded glycosyltransferases in normal and malignant bladder urothelium. *Cancer research* 48(15):4427-4433.

235. Ichikawa D, Handa K, & Hakomori S (1998) Histo-blood group A/B antigen deletion/reduction vs. continuous expression in human tumor cells as correlated with their malignancy. *Int J Cancer* 76(2):284-289.

236. Bennett EP*, et al.* (2010) Rescue of Drosophila Melanogaster l(2)35Aa lethality is only mediated by polypeptide GalNAc-transferase pgant35A, but not by the evolutionary conserved human ortholog GalNAc-transferase-T11. *Glycoconjugate journal* 27(4):435-444.

237. Ten Hagen KG & Tran DT (2002) A UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase is essential for viability in Drosophila melanogaster. *The Journal of biological chemistry* 277(25):22616-22622.

238. Tran DT*, et al.* (2012) Multiple members of the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase family are essential for viability in Drosophila. *The Journal of biological chemistry* 287(8):5243-5252.

239. Sellers TA*, et al.* (2008) Association of single nucleotide polymorphisms in glycosylation genes with risk of epithelial ovarian cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 17(2):397-404.

240. Wagner KW*, et al.* (2007) Death-receptor O-glycosylation controls tumor-cell sensitivity to the proapoptotic ligand Apo2L/TRAIL. *Nature medicine* 13(9):1070-1077.

241. Gomes J*, et al.* (2009) Expression of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 in gastric mucosa, intestinal metaplasia, and gastric carcinoma. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 57(1):79-86.

242. Wu C*, et al.* (2010) N-Acetylgalactosaminyltransferase-14 as a potential biomarker for breast cancer by immunohistochemistry. *BMC cancer* 10:123.

243. Phelan CM*, et al.* (2010) Polymorphism in the GALNT1 gene and epithelial ovarian cancer in non-Hispanic white women: the Ovarian Cancer Association

Consortium. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 19(2):600-604.

244. Patani N, Jiang W, & Mokbel K (2008) Prognostic utility of glycosyltransferase expression in breast cancer. *Cancer genomics & proteomics* 5(6):333-340.

245. Berois N*, et al.* (2006) UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 54(3):317-328.

246. Freire T*, et al.* (2006) UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (ppGalNAc-T6) mRNA as a potential new marker for detection of bone marrow-disseminated breast cancer cells. *Int J Cancer* 119(6):1383-1388.

247. Kufe DW (2009) Mucins in cancer: function, prognosis and therapy. *Nature reviews. Cancer* 9(12):874-885.

248. van der Post S*, et al.* (2013) Site-specific O-glycosylation on the MUC2 mucin protein inhibits cleavage by the Porphyromonas gingivalis secreted cysteine protease (RgpB). *The Journal of biological chemistry* 288(20):14636-14646.

249. Spicer AP, Rowse GJ, Lidner TK, & Gendler SJ (1995) Delayed mammary tumor progression in Muc-1 null mice. *The Journal of biological chemistry* 270(50):30093-30101.

250. Velcich A*, et al.* (2002) Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* 295(5560):1726-1729.

251. Schroeder JA*, et al.* (2004) MUC1 overexpression results in mammary gland tumorigenesis and prolonged alveolar differentiation. *Oncogene* 23(34):5739-5747.

252. Reis CA, Osorio H, Silva L, Gomes C, & David L (2010) Alterations in glycosylation as biomarkers for cancer detection. *Journal of clinical pathology* 63(4):322-329.

253. Kufe D*, et al.* (1984) Differential reactivity of a novel monoclonal antibody (DF3) with human malignant versus benign breast tumors. *Hybridoma* 3(3):223-232.

254. Swallow DM*, et al.* (1987) The human tumour-associated epithelial mucins are coded by an expressed hypervariable gene locus PUM. *Nature* 328(6125):82-84.

255. Siddiqui J*, et al.* (1988) Isolation and sequencing of a cDNA coding for the human DF3 breast carcinoma-associated antigen. *Proceedings of the National Academy of Sciences of the United States of America* 85(7):2320-2323.

256. Akita K*, et al.* (2012) Different levels of sialyl-Tn antigen expressed on MUC16 in patients with endometriosis and ovarian cancer. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society* 22(4):531-538.

257. Chen K*, et al.* (2013) Microarray Glycoprofiling of CA125 improves differential diagnosis of ovarian cancer. *Journal of proteome research* 12(3):1408-1418.

258. Cardillo TM*, et al.* (2004) Improved targeting of pancreatic cancer: experimental studies of a new bispecific antibody, pretargeting enhancement system for immunoscintigraphy. *Clin Cancer Res* 10(10):3552-3561.

259. Xiao J*, et al.* (2005) Pharmacokinetics and clinical evaluation of 125I-radiolabeled humanized CC49 monoclonal antibody (HuCC49deltaC(H)2) in recurrent and metastatic colorectal cancer patients. *Cancer biotherapy & radiopharmaceuticals* 20(1):16-26.

260. Chinn PC*, et al.* (2006) Pharmacokinetics and tumor localization of (111)in-labeled HuCC49DeltaC(H)2 in BALB/c mice and athymic murine colon carcinoma xenograft. *Cancer biotherapy & radiopharmaceuticals* 21(2):106-116.

261. Gold DV*, et al.* (2008) A novel bispecific, trivalent antibody construct for targeting pancreatic carcinoma. *Cancer research* 68(12):4819-4826.

262. Salouti M, Rajabi H, Babaei MH, & Rasaee MJ (2008) Breast tumor targeting with (99m)Tc-HYNIC-PR81 complex as a new biologic radiopharmaceutical. *Nuclear medicine and biology* 35(7):763-768.

263. Danussi C*, et al.* (2009) A newly generated functional antibody identifies Tn antigen as a novel determinant in the cancer cell-lymphatic endothelium interaction. *Glycobiology* 19(10):1056-1067.

264. Karacay H*, et al.* (2009) Pretargeted radioimmunotherapy of pancreatic cancer xenografts: TF10-90Y-IMP-288 alone and combined with gemcitabine. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 50(12):2008-2016.

265. Zou P*, et al.* (2010) 124I-HuCC49deltaCH2 for TAG-72 antigen-directed positron emission tomography (PET) imaging of LS174T colon adenocarcinoma tumor implants in xenograft mice: preliminary results. *World journal of surgical oncology* 8:65.

266. Gulec SA*, et al.* (2011) Treatment of advanced pancreatic carcinoma with 90Y-Clivatuzumab Tetraxetan: a phase I single-dose escalation trial. *Clin Cancer Res* 17(12):4091-4100.

267. Colcher D*, et al.* (1987) Quantitative analyses of selective radiolabeled monoclonal antibody localization in metastatic lesions of colorectal cancer patients. *Cancer research* 47(4):1185-1189.

268. Esteban JM*, et al.* (1987) Quantitative and qualitative aspects of radiolocalization in colon cancer patients of intravenously administered MAb B72.3. *Int J Cancer* 39(1):50-59.

269. Colcher D*, et al.* (1988) Radioimmunolocalization of human carcinoma xenografts with B72.3 second generation monoclonal antibodies. *Cancer research* 48(16):4597-4603.

270. Milenic DE*, et al.* (1991) Construction, binding properties, metabolism, and tumor targeting of a single-chain Fv derived from the pancarcinoma monoclonal antibody CC49. *Cancer research* 51(23 Pt 1):6363-6371.

271. Yokota T, Milenic DE, Whitlow M, & Schlom J (1992) Rapid tumor penetration of a single-chain Fv and comparison with other immunoglobulin forms. *Cancer research* 52(12):3402-3408.

272. Slavin-Chiorini DC*, et al.* (1993) Biologic properties of a CH2 domain-deleted recombinant immunoglobulin. *Int J Cancer* 53(1):97-103.

273. Yokota T*, et al.* (1993) Microautoradiographic analysis of the normal organ distribution of radioiodinated single-chain Fv and other immunoglobulin forms. *Cancer research* 53(16):3776-3783.

274. Alisauskus R, Wong GY, & Gold DV (1995) Initial studies of monoclonal antibody PAM4 targeting to xenografted orthotopic pancreatic cancer. *Cancer research* 55(23 Suppl):5743s-5748s.

275. Gold DV, Alisauskas R, & Sharkey RM (1995) Targeting of xenografted pancreatic cancer with a new monoclonal antibody, PAM4. *Cancer research* 55(5):1105-1110.

276. Mariani G*, et al.* (1995) Initial tumor targeting, biodistribution, and pharmacokinetic evaluation of the monoclonal antibody PAM4 in patients with pancreatic cancer. *Cancer research* 55(23 Suppl):5911s-5915s.

277. Slavin-Chiorini DC*, et al.* (1995) Biological properties of chimeric domain-deleted anticarcinoma immunoglobulins. *Cancer research* 55(23 Suppl):5957s-5967s.

278. Yao Z*, et al.* (1995) Radioimmunoimaging of colon cancer xenografts with anti-Tn monoclonal antibody. *Nuclear medicine and biology* 22(2):199-203.

279. Gold DV, Cardillo T, Vardi Y, & Blumenthal R (1997) Radioimmunotherapy of experimental pancreatic cancer with 131I-labeled monoclonal antibody PAM4. *Int J Cancer* 71(4):660-667.

280. Slavin-Chiorini DC*, et al.* (1997) A CDR-grafted (humanized) domain-deleted antitumor antibody. *Cancer biotherapy & radiopharmaceuticals* 12(5):305-316.

281. Nakamoto Y*, et al.* (1998) Three-step tumor imaging with biotinylated monoclonal antibody, streptavidin and 111In-DTPA-biotin. *Nuclear medicine and biology* 25(2):95-99.

282. Zhang M*, et al.* (1998) Effect of administration route and dose of streptavidin or biotin on the tumor uptake of radioactivity in intraperitoneal tumor with multistep targeting. *Nuclear medicine and biology* 25(2):101-105.

283. Beresford GW, Pavlinkova G, Booth BJ, Batra SK, & Colcher D (1999) Binding characteristics and tumor targeting of a covalently linked divalent CC49 single-chain antibody. *Int J Cancer* 81(6):911-917.

284. Pavlinkova G, Beresford GW, Booth BJ, Batra SK, & Colcher D (1999) Pharmacokinetics and biodistribution of engineered single-chain antibody constructs of MAb CC49 in colon carcinoma xenografts. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 40(9):1536-1546.

285. Pavlinkova G, Booth BJ, Batra SK, & Colcher D (1999) Radioimmunotherapy of human colon cancer xenografts using a dimeric single-chain Fv antibody construct. *Clin Cancer Res* 5(9):2613-2619.

286. Goel A*, et al.* (2000) Genetically engineered tetravalent single-chain Fv of the pancarcinoma monoclonal antibody CC49: improved biodistribution and potential for therapeutic application. *Cancer research* 60(24):6964-6971.

287. Goel A*, et al.* (2001) 99mTc-labeled divalent and tetravalent CC49 single-chain Fv's: novel imaging agents for rapid in vivo localization of human colon carcinoma. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* 42(10):1519-1527.

288. Gold DV, Cardillo T, Goldenberg DM, & Sharkey RM (2001) Localization of pancreatic cancer with radiolabeled monoclonal antibody PAM4. *Critical reviews in oncology/hematology* 39(1-2):147-154.

289. Forero A*, et al.* (2003) A novel monoclonal antibody design for radioimmunotherapy. *Cancer biotherapy & radiopharmaceuticals* 18(5):751-759.

290. Agnese DM*, et al.* (2004) Pilot study using a humanized CC49 monoclonal antibody (HuCC49DeltaCH2) to localize recurrent colorectal carcinoma. *Ann Surg Oncol* 11(2):197-202.

291. Chaturvedi R*, et al.* (2008) Tumor immunolocalization using 124 I-iodine-labeled JAA-F11 antibody to Thomsen-Friedenreich alpha-linked antigen. *Applied radiation and isotopes : including data, instrumentation and methods for use in agriculture, industry and medicine* 66(3):278-287.

292. Hanisch FG, Uhlenbruck G, Egge H, & Peter-Katalinic J (1989) A B72.3 second-generation-monoclonal antibody (CC49) defines the mucin-carried carbohydrate epitope Gal beta(1-3) [NeuAc alpha(2-6)]GalNAc. *Biological chemistry Hoppe-Seyler* 370(1):21-26.

293. O'Boyle KP*, et al.* (1996) Specificity analysis of murine monoclonal antibodies reactive with Tn, sialylated Tn, T, and monosialylated (2-->6) T antigens. *Hybridoma* 15(6):401-408.

294. Sahin U*, et al.* (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proceedings of the National Academy of Sciences of the United States of America* 92(25):11810-11813.

295. Wang X*, et al.* (2005) Autoantibody signatures in prostate cancer. *N Engl J Med* 353(12):1224-1235.

296. Wandall HH*, et al.* (2010) Cancer biomarkers defined by autoantibody signatures to aberrant O-glycopeptide epitopes. *Cancer research* 70(4):1306-1313.

297. Blixt O*, et al.* (2011) Autoantibodies to aberrantly glycosylated MUC1 in early stage breast cancer are associated with a better prognosis. *Breast cancer research : BCR* 13(2):R25.

298. Ando H*, et al.* (2008) Mouse-human chimeric anti-Tn IgG1 induced anti-tumor activity against Jurkat cells in vitro and in vivo. *Biological & pharmaceutical bulletin* 31(9):1739-1744.

299. Hubert P*, et al.* (2011) Antibody-dependent cell cytotoxicity synapses form in mice during tumor-specific antibody immunotherapy. *Cancer research* 71(15):5134-5143.

300. Kubota T, Matsushita T, Niwa R, Kumagai I, & Nakamura K (2010) Novel anti-Tn single-chain Fv-Fc fusion proteins derived from immunized phage library and antibody Fc domain. *Anticancer research* 30(9):3397-3405.

301. Morita N, Yajima Y, Asanuma H, Nakada H, & Fujita-Yamaguchi Y (2009) Inhibition of cancer cell growth by anti-Tn monoclonal antibody MLS128. *Bioscience trends* 3(1):32-37.

302. Welinder C, Baldetorp B, Borrebaeck C, Fredlund BM, & Jansson B (2011) A new murine IgG1 anti-Tn monoclonal antibody with in vivo anti-tumor activity. *Glycobiology* 21(8):1097-1107.

303. Pegram MD*, et al.* (2009) Phase I dose escalation pharmacokinetic assessment of intravenous humanized anti-MUC1 antibody AS1402 in patients with advanced breast cancer. *Breast cancer research : BCR* 11(5):R73.

304. Ibrahim NK*, et al.* (2011) Randomized phase II trial of letrozole plus anti-MUC1 antibody AS1402 in hormone receptor-positive locally advanced or metastatic breast cancer. *Clin Cancer Res* 17(21):6822-6830.

305. Pastan I*, et al.* (1991) Characterization of monoclonal antibodies B1 and B3 that react with mucinous adenocarcinomas. *Cancer research* 51(14):3781-3787.

306. Pai LH, Wittes R, Setser A, Willingham MC, & Pastan I (1996) Treatment of advanced solid tumors with immunotoxin LMB-1: an antibody linked to Pseudomonas exotoxin. *Nature medicine* 2(3):350-353.

307. Pastan I, Hassan R, Fitzgerald DJ, & Kreitman RJ (2006) Immunotoxin therapy of cancer. *Nature reviews. Cancer* 6(7):559-565.

308. Kantoff PW*, et al.* (2010) Sipuleucel-T immunotherapy for castration-resistant prostate cancer. *N Engl J Med* 363(5):411-422.

309. Schwartzentruber DJ*, et al.* (2011) gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N Engl J Med* 364(22):2119-2127.

310. Avci FY, Li X, Tsuji M, & Kasper DL (2011) A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. *Nature medicine* 17(12):1602-1609.

311. Stowell SR*, et al.* (2014) Microbial glycan microarrays define key features of host-microbial interactions. *Nature chemical biology* 10(6):470-476.

312. Cobb BA, Wang Q, Tzianabos AO, & Kasper DL (2004) Polysaccharide processing and presentation by the MHCII pathway. *Cell* 117(5):677-687.

313. Haurum JS*, et al.* (1994) Recognition of carbohydrate by major histocompatibility complex class I-restricted, glycopeptide-specific cytotoxic T lymphocytes. *The Journal of experimental medicine* 180(2):739-744.

314. Gilewski TA*, et al.* (2007) Immunization of high-risk breast cancer patients with clustered sTn-KLH conjugate plus the immunologic adjuvant QS-21. *Clin Cancer Res* 13(10):2977-2985.

315. Ingale S, Wolfert MA, Gaekwad J, Buskas T, & Boons GJ (2007) Robust immune responses elicited by a fully synthetic three-component vaccine. *Nature chemical biology* 3(10):663-667.

316. Napoletano C*, et al.* (2007) Tumor-associated Tn-MUC1 glycoform is internalized through the macrophage galactose-type C-type lectin and delivered to

the HLA class I and II compartments in dendritic cells. *Cancer research* 67(17):8358-8367.

317. Pedersen JW*, et al.* (2011) Seromic profiling of colorectal cancer patients with novel glycopeptide microarray. *Int J Cancer* 128(8):1860-1871.

318. Sabbatini PJ*, et al.* (2007) Pilot study of a heptavalent vaccine-keyhole limpet hemocyanin conjugate plus QS21 in patients with epithelial ovarian, fallopian tube, or peritoneal cancer. *Clin Cancer Res* 13(14):4170-4177.

319. Sorensen AL*, et al.* (2006) Chemoenzymatically synthesized multimeric Tn/STn MUC1 glycopeptides elicit cancer-specific anti-MUC1 antibody responses and override tolerance. *Glycobiology* 16(2):96-107.

320. von Mensdorff-Pouilly S*, et al.* (2000) Reactivity of natural and induced human antibodies to MUC1 mucin with MUC1 peptides and n-acetylgalactosamine (GalNAc) peptides. *Int J Cancer* 86(5):702-712.

321. Borsig L*, et al.* (2001) Heparin and cancer revisited: mechanistic connections involving platelets, P-selectin, carcinoma mucins, and tumor metastasis. *Proceedings of the National Academy of Sciences of the United States of America* 98(6):3352-3357.

322. Loftus EV, Jr. (2004) Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 126(6):1504-1517.

323. Kumar V, Abbas AK, & Aster JC (2015) *Robbins and Cotran pathologic basis of disease* (Elsevier/Saunders, Philadelphia, PA) Ninth edition. Ed pp xvi, 1391 pages.

324. Strober W & Fuss IJ (2011) Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases. *Gastroenterology* 140(6):1756-1767.

325. Halfvarson J, Bodin L, Tysk C, Lindberg E, & Jarnerot G (2003) Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 124(7):1767-1773.

326. Jostins L*, et al.* (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491(7422):119-124.

327. Reese GE*, et al.* (2006) Diagnostic precision of anti-Saccharomyces cerevisiae antibodies and perinuclear antineutrophil cytoplasmic antibodies in inflammatory bowel disease. *Am J Gastroenterol* 101(10):2410-2422.

328. Lodes MJ*, et al.* (2004) Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* 113(9):1296-1306.

329. Hviid A, Svanstrom H, & Frisch M (2011) Antibiotic use and inflammatory bowel diseases in childhood. *Gut* 60(1):49-54.

330. Jernberg C, Lofmark S, Edlund C, & Jansson JK (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* 156(Pt 11):3216-3223.

331. Turnbaugh PJ*, et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* 457(7228):480-484.

332. Kostic AD, Xavier RJ, & Gevers D (2014) The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146(6):1489-1499.

333. Vatanen T*, et al.* (2016) Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* 165(6):1551.

334. Tailford LE, Crost EH, Kavanaugh D, & Juge N (2015) Mucin glycan foraging in the human gut microbiome. *Front Genet* 6:81.

335. Robbe C*, et al.* (2003) Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *The Journal of biological chemistry* 278(47):46337-46348.

336. Robbe-Masselot C, Maes E, Rousset M, Michalski JC, & Capon C (2009) Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. *Glycoconjugate journal* 26(4):397-413.

337. Holmen Larsson JM, Thomsson KA, Rodriguez-Pineiro AM, Karlsson H, & Hansson GC (2013) Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *Am J Physiol Gastrointest Liver Physiol* 305(5):G357-363.

338. Arike L, Holmen-Larsson J, & Hansson GC (2017) Intestinal Muc2 mucin O-glycosylation is affected by microbiota and regulated by differential expression of glycosyltranferases. *Glycobiology*.

339. Pickard JM & Chervonsky AV (2015) Intestinal fucose as a mediator of host-microbe symbiosis. *Journal of immunology* 194(12):5588-5593.

340. Terahara K*, et al.* (2011) Distinct fucosylation of M cells and epithelial cells by Fut1 and Fut2, respectively, in response to intestinal environmental stress. *Biochemical and biophysical research communications* 404(3):822-828.

341. Bry L, Falk PG, Midtvedt T, & Gordon JI (1996) A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273(5280):1380-1383.

342. Goto Y*, et al.* (2014) Innate lymphoid cells regulate intestinal epithelial cell glycosylation. *Science* 345(6202):1254009.

343. Hooper LV, Xu J, Falk PG, Midtvedt T, & Gordon JI (1999) A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proceedings of the National Academy of Sciences of the United States of America* 96(17):9833-9838.

344. Coyne MJ, Reinap B, Lee MM, & Comstock LE (2005) Human symbionts use a host-like pathway for surface fucosylation. *Science* 307(5716):1778-1781.

345. Pham TA*, et al.* (2014) Epithelial IL-22RA1-mediated fucosylation promotes intestinal colonization resistance to an opportunistic pathogen. *Cell Host Microbe* 16(4):504-516.

346. Goto Y*, et al.* (2015) IL-10-producing CD4(+) T cells negatively regulate fucosylation of epithelial cells in the gut. *Sci Rep* 5:15918.

347. McGovern DP*, et al.* (2010) Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum Mol Genet* 19(17):3468-3476.

348. Graham DB*, et al.* (2016) TMEM258 Is a Component of the Oligosaccharyltransferase Complex Controlling ER Stress and Intestinal Inflammation. *Cell Rep* 17(11):2955-2965.

349. Rhodes JM, Black RR, & Savage A (1988) Altered lectin binding by colonic epithelial glycoconjugates in ulcerative colitis and Crohn's disease. *Dig Dis Sci* 33(11):1359-1363.

350. Campbell BJ, Finnie IA, Hounsell EF, & Rhodes JM (1995) Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative colitis mucin and its concealment in normal mucin. *J Clin Invest* 95(2):571-576.

351. Karlen P*, et al.* (1998) Sialyl-Tn antigen as a marker of colon cancer risk in ulcerative colitis: relation to dysplasia and DNA aneuploidy. *Gastroenterology* 115(6):1395-1404.

352. Bodger K*, et al.* (2006) Altered colonic glycoprotein expression in unaffected monozygotic twins of inflammatory bowel disease patients. *Gut* 55(7):973-977.

353. Larsson JM*, et al.* (2011) Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis* 17(11):2299-2307.

354. Dewald JH, Colomb F, Bobowski-Gerard M, Groux-Degroote S, & Delannoy P (2016) Role of Cytokine-Induced Glycosylation Changes in Regulating Cell Interactions and Cell Signaling in Inflammatory Diseases and Cancer. *Cells* 5(4).

355. Johansson ME*, et al.* (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 105(39):15064-15069.

356. Johansson ME, Larsson JM, & Hansson GC (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 1:4659-4665.

357. van der Post S, Thomsson KA, & Hansson GC (2014) Multiple enzyme approach for the characterization of glycan modifications on the C-terminus of the intestinal MUC2mucin. *Journal of proteome research* 13(12):6013-6023.

358. Asker N, Axelsson MA, Olofsson SO, & Hansson GC (1998) Dimerization of the human MUC2 mucin in the endoplasmic reticulum is followed by a N-glycosylation-dependent transfer of the mono- and dimers to the Golgi apparatus. *The Journal of biological chemistry* 273(30):18857-18863.

359. Godl K*, et al.* (2002) The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *The Journal of biological chemistry* 277(49):47248-47256.

360. Specian RD & Neutra MR (1980) Mechanism of rapid mucus secretion in goblet cells stimulated by acetylcholine. *The Journal of cell biology* 85(3):626-640.

361. Birchenough GM, Nystrom EE, Johansson ME, & Hansson GC (2016) A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion. *Science* 352(6293):1535-1542.

362. Gustafsson JK*, et al.* (2012) Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *The Journal of experimental medicine* 209(7):1263-1272.

363. Ambort D*, et al.* (2012) Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proceedings of the National Academy of Sciences of the United States of America* 109(15):5645-5650.

364. Schutte A*, et al.* (2014) Microbial-induced meprin beta cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. *Proceedings of the National Academy of Sciences of the United States of America* 111(34):12396-12401.

365. Bergstrom K*, et al.* (2017) Core 1- and 3-derived O-glycans collectively maintain the colonic mucus barrier and protect against spontaneous colitis in mice. *Mucosal Immunol* 10(1):91-103.

366. Capaldo CT, *et al.* (2014) Proinflammatory cytokine-induced tight junction remodeling through dynamic self-assembly of claudins. *Mol Biol Cell* 25(18):2710-2719.

367. Aspholm-Hurtig M, *et al.* (2004) Functional adaptation of BabA, the H. pylori ABO blood group antigen binding adhesin. *Science* 305(5683):519-522.

368. Yolken RH, *et al.* (1992) Human milk mucin inhibits rotavirus replication and prevents experimental gastroenteritis. *J Clin Invest* 90(5):1984-1991.

369. Hu L, *et al.* (2012) Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. *Nature* 485(7397):256-259.

370. Cao S, *et al.* (2007) Structural basis for the recognition of blood group trisaccharides by norovirus. *J Virol* 81(11):5949-5957.

371. Lee H, *et al.* (2008) Alpha1,4GlcNAc-capped mucin-type O-glycan inhibits cholesterol alpha-glucosyltransferase from Helicobacter pylori and suppresses H. pylori growth. *Glycobiology* 18(7):549-558.

372. Kinoshita H, *et al.* (2008) Cell surface glyceraldehyde-3-phosphate dehydrogenase (GAPDH) of Lactobacillus plantarum LA 318 recognizes human A and B blood group antigens. *Res Microbiol* 159(9-10):685-691.

373. Nishiyama K, Ueno S, Sugiyama M, Yamamoto Y, & Mukai T (2016) Lactobacillus rhamnosus GG SpaC pilin subunit binds to the carbohydrate moieties of intestinal glycoconjugates. *Anim Sci J* 87(6):809-815.

374. Watanabe M, *et al.* (2010) Identification of a new adhesin-like protein from Lactobacillus mucosae ME-340 with specific affinity to the human blood group A and B antigens. *J Appl Microbiol* 109(3):927-935.

375. Garrido D, Kim JH, German JB, Raybould HE, & Mills DA (2011) Oligosaccharide binding proteins from Bifidobacterium longum subsp. infantis reveal a preference for host glycans. *PloS one* 6(3):e17315.

376. Sonnenburg JL, Angenent LT, & Gordon JI (2004) Getting a grip on things: how do communities of bacterial symbionts become established in our intestine? *Nat Immunol* 5(6):569-573.

377. Martens EC, Chiang HC, & Gordon JI (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4(5):447-457.

378. Lynch JB & Sonnenburg JL (2012) Prioritization of a plant polysaccharide over a mucus carbohydrate is enforced by a Bacteroides hybrid two-component system. *Mol Microbiol* 85(3):478-491.

379. Pudlo NA, *et al.* (2015) Symbiotic Human Gut Bacteria with Variable Metabolic Priorities for Host Mucosal Glycans. *MBio* 6(6):e01282-01215.

380. Marcobal A, Southwick AM, Earle KA, & Sonnenburg JL (2013) A refined palate: bacterial consumption of host glycans in the gut. *Glycobiology* 23(9):1038-1046.

381. Ng KM, *et al.* (2013) Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502(7469):96-99.

382. Kamada N, *et al.* (2012) Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science* 336(6086):1325-1329.

383. Rakoff-Nahoum S, Foster KR, & Comstock LE (2016) The evolution of cooperation within the gut microbiota. *Nature* 533(7602):255-259.

384. Donaldson GP, Lee SM, & Mazmanian SK (2016) Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 14(1):20-32.

385. Sonnenburg JL, *et al.* (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307(5717):1955-1959.

386. Lee SM, *et al.* (2013) Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* 501(7467):426-429.

387. Owen RL (1977) Sequential uptake of horseradish peroxidase by lymphoid follicle epithelium of Peyer's patches in the normal unobstructed mouse intestine: an ultrastructural study. *Gastroenterology* 72(3):440-451.

388. Rios D, *et al.* (2016) Antigen sampling by intestinal M cells is the principal pathway initiating mucosal IgA production to commensal enteric bacteria. *Mucosal Immunol* 9(4):907-916.

389. Rescigno M, *et al.* (2001) Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol* 2(4):361-367.

390. McDole JR, *et al.* (2012) Goblet cells deliver luminal antigen to CD103+ dendritic cells in the small intestine. *Nature* 483(7389):345-349.

391. Knoop KA, Miller MJ, & Newberry RD (2013) Transepithelial antigen delivery in the small intestine: different paths, different outcomes. *Curr Opin Gastroenterol* 29(2):112-118.

392. Yoshida M, *et al.* (2004) Human neonatal Fc receptor mediates transport of IgG into luminal secretions for delivery of antigens to mucosal dendritic cells. *Immunity* 20(6):769-783.

393. Yoshida M, *et al.* (2006) Neonatal Fc receptor for IgG regulates mucosal immune responses to luminal bacteria. *J Clin Invest* 116(8):2142-2151.

394. Hase K, *et al.* (2009) Uptake through glycoprotein 2 of FimH(+) bacteria by M cells initiates mucosal immune response. *Nature* 462(7270):226-230.

395. Scott DW, *et al.* (2015) N-glycosylation controls the function of junctional adhesion molecule-A. *Mol Biol Cell* 26(18):3205-3214.

396. Jiang K, *et al.* (2014) Galectin-3 regulates desmoglein-2 and intestinal epithelial intercellular adhesion. *The Journal of biological chemistry* 289(15):10510-10517.

397. Shan M, *et al.* (2013) Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals. *Science* 342(6157):447-453.

398. Brazil JC, *et al.* (2013) alpha3/4 Fucosyltransferase 3-dependent synthesis of Sialyl Lewis A on CD44 variant containing exon 6 mediates polymorphonuclear leukocyte detachment from intestinal epithelium during transepithelial migration. *Journal of immunology* 191(9):4804-4817.

399. Van der Sluis M, *et al.* (2006) Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* 131(1):117-129.

400. Shriver Z, Raguram S, & Sasisekharan R (2004) Glycomics: a pathway to a class of new and improved therapeutics. *Nature reviews. Drug discovery* 3(10):863-873.

401. Kudelka MR, Ju T, Heimburg-Molinaro J, & Cummings RD (2015) Simple sugars to complex disease--mucin-type O-glycans in cancer. *Adv Cancer Res* 126:53-135.

402. Furukawa J, Fujitani N, & Shinohara Y (2013) Recent advances in cellular glycomic analyses. *Biomolecules* 3(1):198-225.

403. Ju T, *et al.* (2011) A novel fluorescent assay for T-synthase activity. *Glycobiology* 21(3):352-362.

404. Brockhausen I, *et al.* (1992) 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(2):99-108.

405. Okayama M, Kimata K, & Suzuki S (1973) The influence of p-nitrophenyl beta-d-xyloside on the synthesis of proteochondroitin sulfate by slices of embryonic chick cartilage. *J Biochem* 74(5):1069-1073.

406. Sarkar AK, Fritz TA, Taylor WH, & Esko JD (1995) Disaccharide uptake and priming in animal cells: inhibition of sialyl Lewis X by acetylated Gal beta 1-->4GlcNAc beta-O-naphthalenemethanol. *Proceedings of the National Academy of Sciences of the United States of America* 92(8):3323-3327.

407. Wilson JR & Zimmerman EF (1976) Yolk sac: site of developmental microheterogeneity of mouse alpha-fetoprotein. *Developmental biology* 54(2):187-199.

408. Ju T & Cummings RD (2002) A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Natl Acad Sci U S A* 99(26):16613-16618.

409. Wilkins PP, McEver RP, & Cummings RD (1996) Structures of the O-glycans on P-selectin glycoprotein ligand-1 from HL-60 cells. *The Journal of biological chemistry* 271(31):18732-18742.

410. Kawar ZS, Johnson TK, Natunen S, Lowe JB, & Cummings RD (2008) 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(6):441-446.

411. Morelle W & Michalski JC (2007) Analysis of protein glycosylation by mass spectrometry. *Nature protocols* 2(7):1585-1602.

412. Hsu PP & Sabatini DM (2008) Cancer cell metabolism: Warburg and beyond. *Cell* 134(5):703-707.

413. Dube DH & Bertozzi CR (2005) Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nature reviews. Drug discovery* 4(6):477-488.

414. Babu P*, et al.* (2009) Structural characterisation of neutrophil glycans by ultra sensitive mass spectrometric glycomics methodology. *Glycoconjugate journal* 26(8):975-986.

415. Macher BA, Buehler J, Scudder P, Knapp W, & Feizi T (1988) A novel carbohydrate, differentiation antigen on fucogangliosides of human myeloid cells recognized by monoclonal antibody VIM-2. *The Journal of biological chemistry* 263(21):10186-10191.

416. Sato C*, et al.* (2000) 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. *The Journal of biological chemistry* 275(20):15422-15431.

417. Karlsson NG & Thomsson KA (2009) Salivary MUC7 is a major carrier of blood group I type O-linked oligosaccharides serving as the scaffold for sialyl Lewis x. *Glycobiology* 19(3):288-300.

418. Delannoy P, *et al.* (1996) Benzyl-N-acetyl-alpha-D-galactosaminide inhibits the sialylation and the secretion of mucins by a mucin secreting HT-29 cell subpopulation. *Glycoconjugate journal* 13(5):717-726.

419. Zanetta JP, *et al.* (2000) Massive in vitro synthesis of tagged oligosaccharides in 1-benzyl-2-acetamido-2-deoxy-alpha-D-galactopyranoside treated HT-29 cells. *Glycobiology* 10(6):565-575.

420. Liu J, Jin C, Cherian RM, Karlsson NG, & Holgersson J (2015) 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.

421. Sato T, *et al.* (2009) Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 459(7244):262-265.

422. Gao D, *et al.* (2014) Organoid cultures derived from patients with advanced prostate cancer. *Cell* 159(1):176-187.

423. Park IH, *et al.* (2008) Disease-specific induced pluripotent stem cells. *Cell* 134(5):877-886.

424. Dimos JT, *et al.* (2008) Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* 321(5893):1218-1221.

425. Jang-Lee J, *et al.* (2006) Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies. *Methods in enzymology* 415:59-86.

426. Dong QG, *et al.* (1997) 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. *Arterioscler Thromb Vasc Biol* 17(8):1599-1604.

427. Hartwell DW, *et al.* (1998) Role of P-selectin cytoplasmic domain in granular targeting in vivo and in early inflammatory responses. *The Journal of cell biology* 143(4):1129-1141.

428. Shankavaram UT, *et al.* (2009) CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics* 10:277.

429. Reinhold WC, *et al.* (2012) CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer research* 72(14):3499-3511.

430. Sauer CG & Kugathasan S (2010) Pediatric inflammatory bowel disease: highlighting pediatric differences in IBD. *Med Clin North Am* 94(1):35-52.

431. Sekirov I, Russell SL, Antunes LC, & Finlay BB (2010) Gut microbiota in health and disease. *Physiol Rev* 90(3):859-904.

432. Gevers D*, et al.* (2014) The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15(3):382-392.

433. Chang D*, et al.* (2014) Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. *PloS one* 9(12):e113684.

434. Kashyap PC*, et al.* (2013) Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proceedings of the National Academy of Sciences of the United States of America* 110(42):17059-17064.

435. Chassaing B*, et al.* (2012) Fecal lipocalin 2, a sensitive and broadly dynamic non-invasive biomarker for intestinal inflammation. *PloS one* 7(9):e44328.

436. Grivennikov SI, Greten FR, & Karin M (2010) Immunity, inflammation, and cancer. *Cell* 140(6):883-899.

437. Axelsson MA, Asker N, & Hansson GC (1998) O-glycosylated MUC2 monomer and dimer from LS 174T cells are water-soluble, whereas larger MUC2 species formed early during biosynthesis are insoluble and contain nonreducible intermolecular bonds. *The Journal of biological chemistry* 273(30):18864-18870.

438. Earle KA*, et al.* (2015) Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* 18(4):478-488.

439. Khalili H (2016) Risk of Inflammatory Bowel Disease with Oral Contraceptives and Menopausal Hormone Therapy: Current Evidence and Future Directions. *Drug Saf* 39(3):193-197.

440. Elinav E*, et al.* (2011) NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 145(5):745-757.

441. Lecuyer E*, et al.* (2014) Segmented filamentous bacterium uses secondary and tertiary lymphoid tissues to induce gut IgA and specific T helper 17 cell responses. *Immunity* 40(4):608-620.

442. Siegmund B (2010) Interleukin-18 in intestinal inflammation: friend and foe? *Immunity* 32(3):300-302.

443. Pemberton AD*, et al.* (2004) Innate BALB/c enteric epithelial responses to Trichinella spiralis: inducible expression of a novel goblet cell lectin, intelectin-2, and its natural deletion in C57BL/10 mice. *Journal of immunology* 173(3):1894-1901.

444. Karner M*, et al.* (2014) First multicenter study of modified release phosphatidylcholine "LT-02" in ulcerative colitis: a randomized, placebo-controlled trial in mesalazine-refractory courses. *Am J Gastroenterol* 109(7):1041-1051.

445. Hooper LV, Stappenbeck TS, Hong CV, & Gordon JI (2003) Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* 4(3):269-273.

446. Hogan SP*, et al.* (2006) Resistin-like molecule beta regulates innate colonic function: barrier integrity and inflammation susceptibility. *J Allergy Clin Immunol* 118(1):257-268.

447. Bergstrom KS*, et al.* (2015) Goblet Cell Derived RELM-beta Recruits CD4+ T Cells during Infectious Colitis to Promote Protective Intestinal Epithelial Cell Proliferation. *PLoS Pathog* 11(8):e1005108.

448. Banerjee S*, et al.* (2009) MEP1A allele for meprin A metalloprotease is a susceptibility gene for inflammatory bowel disease. *Mucosal Immunol* 2(3):220-231.

449. Wesener DA*, et al.* (2015) Recognition of microbial glycans by human intelectin-1. *Nat Struct Mol Biol* 22(8):603-610.

450. Pelaseyed T*, et al.* (2014) The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev* 260(1):8-20.

451. Ott SJ*, et al.* (2004) Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53(5):685-693.

452. David LA*, et al.* (2014) Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505(7484):559-563.

453. McLoughlin K, Schluter J, Rakoff-Nahoum S, Smith AL, & Foster KR (2016) Host Selection of Microbiota via Differential Adhesion. *Cell Host Microbe* 19(4):550-559.

454. Neufert C*, et al.* (2013) Tumor fibroblast-derived epiregulin promotes growth of colitis-associated neoplasms through ERK. *J Clin Invest* 123(4):1428-1443.

455. Kudelka MR*, et al.* (2016) Cellular O-Glycome Reporter/Amplification to explore O-glycans of living cells. *Nat Methods* 13(1):81-86.

456. Johansson ME & Hansson GC (2012) Preservation of mucus in histological sections, immunostaining of mucins in fixed tissue, and localization of bacteria with FISH. *Methods Mol Biol* 842:229-235.

457. Knoop KA*, et al.* (2009) RANKL is necessary and sufficient to initiate development of antigen-sampling M cells in the intestinal epithelium. *Journal of immunology* 183(9):5738-5747.

458. Caporaso JG*, et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7(5):335-336.

459. Rakoff-Nahoum S, Paglino J, Eslami-Varzaneh F, Edberg S, & Medzhitov R (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* 118(2):229-241.

460. Kudelka MR*, et al.* (2016) Cosmc is an X-linked inflammatory bowel disease risk gene that spatially regulates gut microbiota and contributes to sex-specific risk. *Proceedings of the National Academy of Sciences of the United States of America* 113(51):14787-14792.

461. Croci DO*, et al.* (2012) Disrupting galectin-1 interactions with N-glycans suppresses hypoxia-driven angiogenesis and tumorigenesis in Kaposi's sarcoma. *The Journal of experimental medicine* 209(11):1985-2000.

462. Leppanen A, Yago T, Otto VI, McEver RP, & Cummings RD (2003) Model glycosulfopeptides from P-selectin glycoprotein ligand-1 require tyrosine sulfation and a core 2-branched O-glycan to bind to L-selectin. *The Journal of biological chemistry* 278(29):26391-26400.

463. Song X*, et al.* (2016) Oxidative release of natural glycans for functional glycomics. *Nat Methods* 13(6):528-534.

464. Yoshida-Moriguchi T*, et al.* (2010) O-mannosyl phosphorylation of alpha-dystroglycan is required for laminin binding. *Science* 327(5961):88-92.

465. Gray JS, Yang BY, & Montgomery R (1998) Heterogeneity of glycans at each N-glycosylation site of horseradish peroxidase. *Carbohydr Res* 311(1-2):61-69.

466. Manzella SM, Hooper LV, & Baenziger JU (1996) Oligosaccharides containing beta 1,4-linked N-acetylgalactosamine, a paradigm for protein-specific glycosylation. *The Journal of biological chemistry* 271(21):12117-12120.

467. Nairn AV*, et al.* (2008) Regulation of glycan structures in animal tissues: transcript profiling of glycan-related genes. *The Journal of biological chemistry* 283(25):17298-17313.

468. Janzen RG & Tamaoki T (1983) Secretion and glycosylation of alpha-foetoprotein by the mouse yolk sac. *The Biochemical journal* 212(2):313-320.

469. Silvera D, Formenti SC, & Schneider RJ (2010) Translational control in cancer. *Nature reviews. Cancer* 10(4):254-266.

470. Mukhopadhya I, Hansen R, El-Omar EM, & Hold GL (2012) IBD-what role do Proteobacteria play? *Nat Rev Gastroenterol Hepatol* 9(4):219-230.

471. Coyne MJ & Comstock LE (2008) Niche-specific features of the intestinal bacteroidales. *J Bacteriol* 190(2):736-742.