CELL- AND TISSUE-SPECIFIC GLYCOSYLATION IN HUMAN STEM CELL MODELS

OF DEVELOPMENT

by

HARRISON ELLIOTT GRACE

(Under the Direction of Michael Tiemeyer)

ABSTRACT

Glycosylation is cell and tissue-type specific, regulated throughout cellular differentiation. Glycoproteins and glycolipids adorn the surface of cells where they mediate cell-cell and cell-environment interactions and have known roles in adhesion, migration, and signaling events that are essential for normal tissue development. Defects of glycosylation, as seen in Congenital Disorders of Glycosylation, often results in severe developmental and intellectual disabilities. The role of glycosylation in other neurodevelopmental disorders such as Autism spectrum disorders has not been well explored. Characterizing the dynamics of cell-specific glycosylation patterns during human development provides a framework for understanding the pathophysiologic relevance of altered glycosylation in human diseases. Human pluripotent stem cells are a powerful tool that can be used to generate different human cell types and tissues. Human embryonic stem cells differentiated into multiple cell were used to study the regulation of glycosylation in human cellular development. Combined analysis of transcriptomes and glycomes of these cells revealed shifts in biosynthetic pathways between pluripotent, multipotent, and differentiated cells leading to the generation of

glycan structural profiles unique to the different cell types. This data indicates a significant role for the regulation of glycan structures in development. In neural cultures generated from induced pluripotent stems cells of an Autism Spectrum Disorder patient with a mutation in synaptic adhesion gene NLGN4x, glycan profiles were altered and dysfunction in early neural development was evident. Investigation of these ASD patient cells reveals an undescribed early role for NLGN4 in neural development and provides a platform for studying the role of glycosylation in the poorly understood neurodevelopmental disorder of Autism.

INDEX WORDS: Glycosylation, N-Glycans, Glycosphingolipids, Stem cells, Development, Ectoderm, Mesoderm, Endoderm, Autism

CELL- AND TISSUE-SPECIFIC GLYCOSYLATION IN HUMAN STEM CELL MODELS

OF DEVELOPMENT

by

HARRISON ELLIOTT GRACE

B.S., The University of Georgia, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial

Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSPHY

ATHENS, GEORGIA

2018

CELL- AND TISSUE-SPECIFIC GLYCOSYLATION IN HUMAN STEM CELL MODELS

OF DEVELOPMENT

by

HARRISON ELLIOTT GRACE

Major Professor: Michael Tiemeyer
Committee: James Lauderdale
Lance Wells
Steven Dalton

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

ACKNOWLEDGEMENTS

The results documented in this work were the product of a collaborative effort. Cell culture and in-vitro stem cell differentiations were conducted by Michael Kulick and collaborators in the laboratory of Steven Dalton. RNA isolation and gene expression analysis experiments were performed by Alison Nairn and collaborators in the laboratory of Kelley Moreman.

I would also like to recognize my ever patient and positive advisor Michael Tiemeyer, and my many lab mates and colleagues, without whose support—intellectual, technical, moral, and psychological—this work would not be possible. Science is a team effort and I am honored to have been on your team.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Glycobiology, Glycans, and Complex Carbohydrates:

Glycobiology, glycosylation, and glycans all refer to the functions, biosynthesis, and structures of saccharides--sugar molecules. Saccharides can be of the basic, familiar sort: monosaccharide building blocks and metabolic fuel like glucose, disaccharides like lactose and sucrose, and polysaccharide polymers like starches or glycogen. However, the complex carbohydrates of glycobiology are not limited to these. There is a vast, structurally complex world of saccharide molecules, known as glycans, that are just as important for the biological structure and function of cells as proteins, nucleic acids, or any other type of biological molecule. Complex carbohydrates, glycan structures of multiple saccharide building blocks, are usually conjugated to proteins or lipids, forming glycoproteins and glycolipids respectively, where they modulate function with their unique biochemical properties. The hydrocarbon nature of carbohydrates and their many -OH hydroxyl groups allow them to form many different linkages and modifications. Thus, the biological functions of glycans are mediated by their diverse structures, which are made by enzymatic synthesis of oligosaccharide chains of various monosaccharaides with multiple possible linkages and branching structures. This structural and biosynthetic complexity that generates the biological versatility of glycan

modifications has also made their analysis difficult, limiting our full understanding of the human glycome.

The surface of mammalian cells, and indeed almost any cell, is covered in a coat of glycan sugar polymers that mediate cellular reception of soluble signaling compounds, connection to other cells and interaction with the extracellular matrix. Thus, glycoproteins and glycolipids play a role in many of the cellular processes of development such as cell fate decisions of self-renewal versus differentiation[3, 4], cell migration[5], and regulation of chemical signaling events through growth factor receptors[6] or neurotransmitter receptor[7]. These developmental processes have been implicated in Autism, yet the relevance of glycans in the pathology of autism and other neuropsychiatric disorders has been underappreciated. Still, accumulating evidence points to their essential role in human development and health. Human congenital disorders of glycosylation (CDGs) resulting in the loss of primary glycosylation enzymes are rarely survivable and often carry severe neurological pathology, especially when the mutation falls in a gene that acts in the early processes of glycosylation[8].

Three major categories of glycan carrying compounds are N-linked glycoproteins, O-linked glycoproteins, and glycolipids (**Figure 1.1**). All three are known to play roles in systems related to neural function and disease. Here I will briefly describe these classes of glycoconjugates and examples of their roles.

N-linked glycans are co-translationally added to proteins via covalent linkage to the amide nitrogen(N) of asparagine residues as they are synthesized in the ER, where they help in protein folding and are an integral part of the system of protein quality control mediated through the Calnexin/Calreticulin and ER-associated degradation
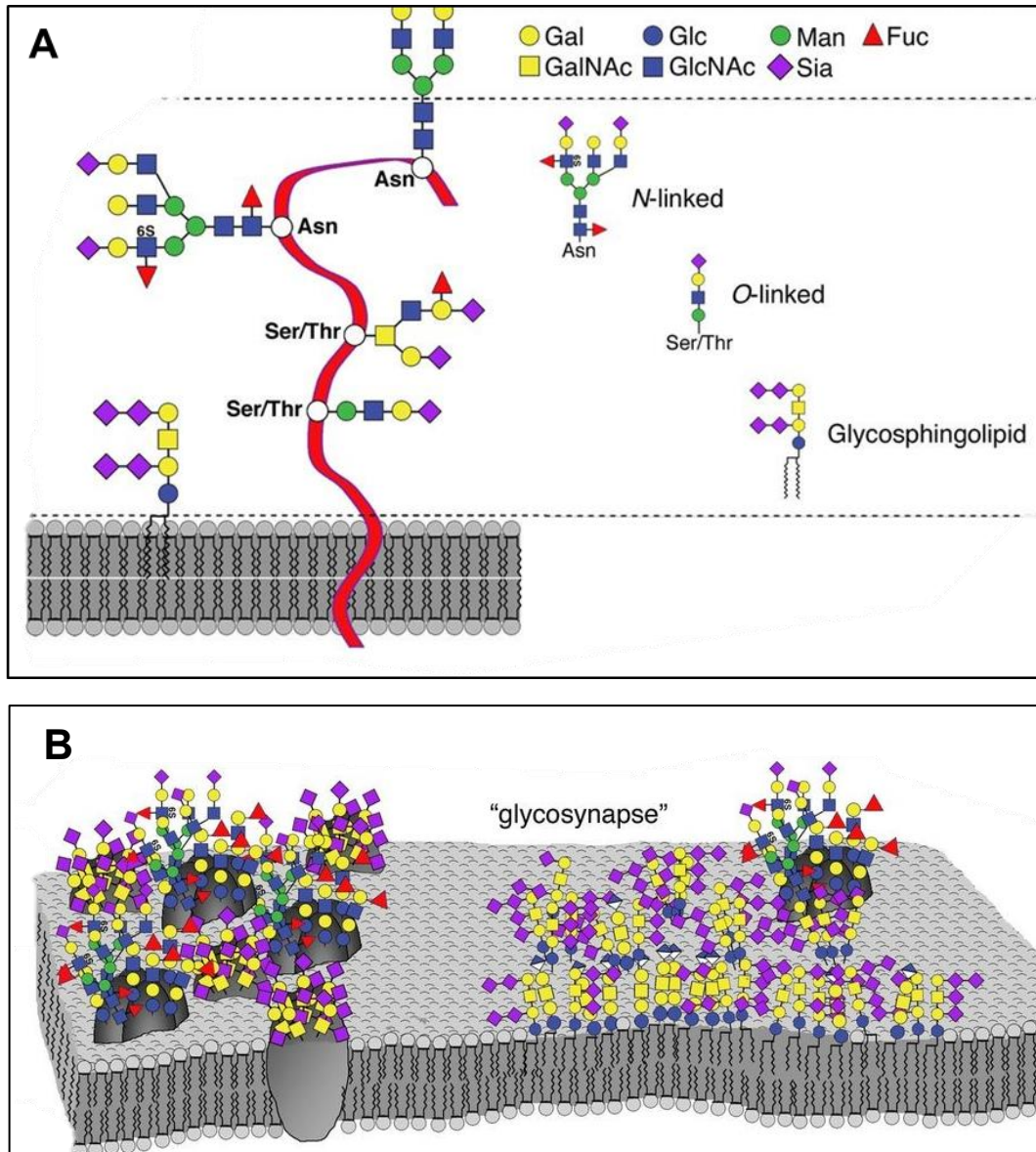
Figure 1.1 Glycoconjugate classes at the cell surface
A) Cartoon structures of glycan types. Glycoprotein glycans are
attached to a transmembrane protein (the red ribbon). N-linked glycans
at an asparagine (Asn) amino acid and O-linked glycans at a Serine
(Ser) or Threonine (Thr) amino acid. Glycosphingolipid (GSLs) are
tethered to the cell plasma membrane via they hydrophobic ceramide
lipid tail. **B)** Glycan cover a significant portion of cell surface. Here the
interaction of glycoproteins such as growth factor receptors is regulated
by their glycan interactions, and membrane bound GSLs may produces
specialized lipid rafts or "glycosynapses".
Adapted from (Schnaar, et al., 2014) [2] .

(ERAD) pathways[9]. As these nascent glycoproteins travel through the Golgi to their cell surface or extracellular destination, their glycan structures are sequentially modified by glycan processing enzymes that add or subtract saccharide units such as glucose, mannose, fucose, and N-acetylgalactosamine, resulting in a multitude of diverse branching patterns and sugar compositions. Specific N-glycan structures have been shown to have distinct functional roles, such as the regulation of cellular adhesion/migration by the reciprocally inhibiting enzymes that make branching (MGAT5) and bisecting(MGAT3) N-glycans[10]. Because N-glycosylation mediates crucial roles such as protein folding and trafficking, altered N-glycosylation is seen in many disease associated protein mutations, including autism linked synaptic proteins[11].

O-linked glycans are also attached to proteins, though the hydroxyl oxygen(O) of a serine or threonine. O-glycans may be added to proteins or modified by enzymes inside and outside of the ER, and thus serve many diverse roles ranging from the major structural unit of mucin-type proteins to a dynamic post-translational modification that regulates protein function and links to cellular metabolism[12]. An example of O-linked glycans and glycoproteins that are highly relevant to cellular adhesion and neural development are α-dystroglycan[13] and the LARGE matriglycan[14] that mediate binding to laminin and the extracellular matrix. Proper attachment via O-mannosyl linkage and elongation of this α-dystroglycan is necessary for proper neural function as evidenced by the neurodevelopmental defects seen in congenital muscular dystrophies[15]. Demonstrating cell type specificity, other O-mannose linked glycans have been found to be differentially expressed on neurons and glia[16].

Glycosphingolipids (GSL) are an important component of the cell membrane and are made by the addition of a glycan structure onto ceramide, a lipid molecule composed of a sphingosine base with an amide-linked fatty acid. Like many glycoproteins, they are made in the ER, modified as they pass through the Golgi, and are trafficked to the plasma membrane surface[17]. Relevant roles for glycolipids include the stabilization of growth factor receptors and cell surface plasma membrane functional domains by the formation of "lipid rafts"[18, 19] (See **Figure 1.1b**). Lipid rafts and specialized lipid domains are shared features of two important signaling domains: neuronal synapses and the primary cilia[20].

Previous[21] and ongoing studies from our lab, and others at the CCRC have demonstrated that glycan structures and the Golgi enzymes that produce them are developmentally regulated and cell and tissue type specific. In this work, I will focus on N-linked glycoprotein glycans (N-Glycans) and glycosphingolipids (GSLs), examining their diversity and biosynthetic regulation in normal human cells and their dysregulation in ASD patient cells.

N-Linked Glycoprotein Glycans:

N-Linked Glycan Precursor Biosynthesis:

The initial fourteen-saccharide $Glc_3Man_9GlcNAc_2$ structure of asparagine(N)-linked glycans is co-translationally transferred to developing glycoproteins at an Asn-X-Ser/Thr amino-acid sequence as the protein is synthesized into the Endoplasmic Reticulum(ER). The biosynthesis of this proto N-glycan begins at the cytoplasmic face of the ER, where the lipid-linked oligosaccharide (LLO) is built on a dolichol-phosphate lipid structure to a $Man_5GlcNAc_2$ structure, translocated into the ER lumen, further

elaborated into the $Glc_3Man_9GlcNAc_2$ structure. This pre-made glycan structure is co-translationally transferred to the glycoprotein as it is being made at the ER ribosome through the actions of a protein complex known as the Oligosaccharide Transferase Complex (OST) [9] (**Figure 1.2**).

Early N-Glycan Trimming:

Early in the Endoplasmic Reticulum(ER), the attached $Glc_3Man_9GlcNAc_2$ N-glycan structure facilitates the folding and oligomerization of the nascent poly-peptide chain, adding bulky hydrophilic groups to modulate protein conformation and keeping the developing glycoprotein in solution. Trimming and extension of the α-linked glucose residues also play a key role in ER protein folding quality control and ERAD, as the mono-glucosylated form is recognized by ER QC lectins Calnexin and Calreticulin. As folding progresses, ER α-glucosidase-I (MOGS) removes the distal most α-linked glucose, and ER α-glucosidase-II (GANAB and PRKCSH) trims the second glucose, allowing for binding of CNX and CRT, and eventually the final proximal glucose. If properly folded, the $Man_9GlcNAc_2$ bearing glycoprotein exits the ER for further processing in the Golgi. However, if unfolded, UGGT glucotransferase re-adds an α-glucose residue and the cycle of chaperone binding begins again [22]. Further trimming of the High Mannose Structure form $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ is accomplished by ER and early-Golgi mannosidases. ER MANEA and Golgi MAN1B1 trim from Man9 to Man8, and Golgi mannosidases MAN1A1, MAN1A2, and MAN1C1 cleave the remaining mannose residues to the $Man_5GlcNAc_2$ form that becomes the substrate for further elaboration into Hybrid or Complex Structures.
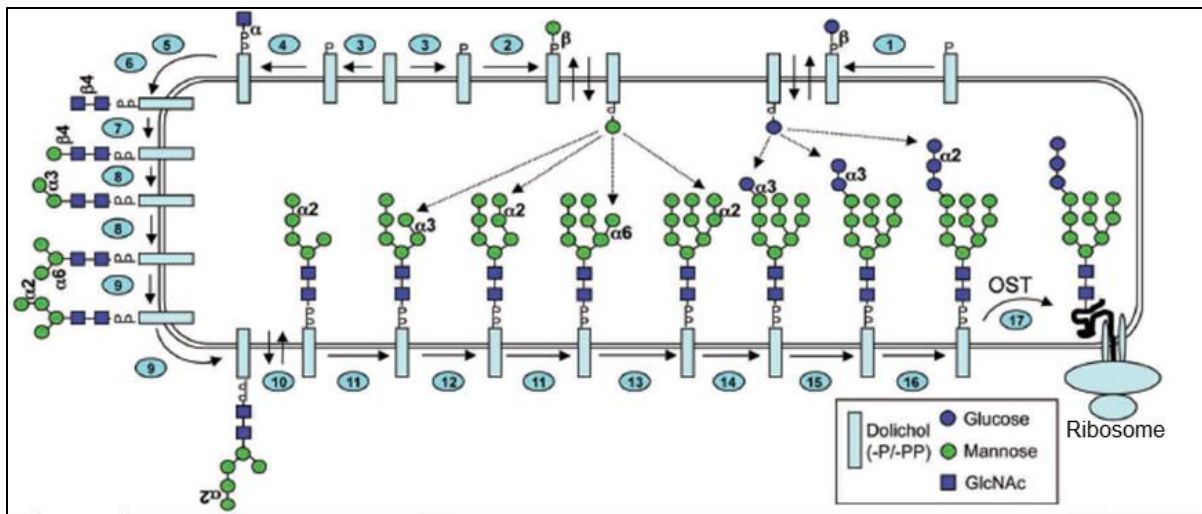
Figure 1.2 N-Glycan LLO Precursor Synthesis
 A diagram of the synthesis of the N-Glycan lipid linked oligosaccharide
(LLO) on the membrane of the ER. LLO synthesis begins on the
cytoplasmic side of the ER with the addition of a GlcNac to the membrane
bound dolicol-phosphate (step 4). Stepwise construction of the N-glycan
proceeds to step 9, before being translocated to the inside of the ER lumen
for further processing to the Glc3Man9GlcNAc2 structure which is attached
pre-assembled onto the translating protein via the OST complex.

N-Glycan Branching and Modification:

Continuing processing as it passes through the Golgi system, the $Man_5GlcNAc_2$ structure may be extended with a GlcNAc on the Man(α1-3) arm by MGAT1, producing Hybrid-type structures. Further trimming of the hybrid arm mannoses by MAN2A1 and MAN2A2 and the addition of another GlcNAc to the Man(α1-6) arm by MGAT2 continue processing to produce a biantennary complex-type structure.

Complex-type biantennary structures may be further branched to tri- and tetra-antennary structures with further additions of GlcNAc to the branching mannose arms by MGAT4 (β4-linked on the 3-arm) and MGAT5(β6-linked on the 6-arm). The MGAT4b isozyme is the more ubiquitously expressed form compared to MGAT4a. MGAT4a, is found to be more specifically expressed in tissues of endoderm descent[23]. Substrate specificity studies have shown that MGAT5 is the main N-linked glycan form, while MGAT5b is more specific for branching on O-linked mannose structures[24].

Both Hybrid and Complex type structures may also be modified with a "Bisecting" GlcNAc added by MGAT3 to the β-linked core mannose. Unique from the other GlcNAc branches the MGAT3 bisecting branch is not further extended and its presence blocks the action of many other glycosyltransferases including MGAT4, MGAT5, FUT8, and a2,3-Sialyl transferases [23, 25]. Variations in branching has been shown to mediate cellular adhesion and migration, as has been studied extensively in cancer and metastatic progression [23].

Glycosphingolipids:

GSL Classes:

GSL classes are defined by the glycan structures attached to Ceramide moiety. The initial monosaccharide added may be either a β1-linked galactose resulting in GalCer or a β1-linked glucose GlcCer. GalCer, also known as Cerebroside or oligodendrocyte marker O4, is major component of the myelin in the nervous system and a precursor to other myelin glycolipids such as sialyl-GalCer (GM4) and sulfo-GalCer (Sulfatide). However, these galacto-lipids are seldom further extended. It is the GlcCer structure that is capable of being extended into the major glycosphingolipids classes that generate the diversity of structures and make up the bulk of the structures detected in this study [17]. Generation of further structures begins with the synthesis of Lactosylceramide (LacCer) by addition of β4-linked Gal onto GlcCer via Lactosylceramide synthase B4GALT5. The extension of LacCer defines the core structures defining the major human GSL classes: gangliosides, globosides, and lacotosylceramides (**Figure 1.3a**). Addition of α4-Gal by A4GALT gives rise to globo-series structures, β3-GlcNAc by B3GNT5 leads to lacto-series structures, and the addition of β4-GalNAc and/or sialic acid generates ganglio-series structures. Further elongation generates different ganglio- and lacto-series subdivisions, as detailed in **Figure 1.3a and 1.3b,** and elaboration of a variety of GSL structures.

GSL Nomenclature:

Throughout this document I will be using a widely adopted ganglioside nomenclature system [26] with a "Class--SialicAcid#--Size" code. For example, the GSL

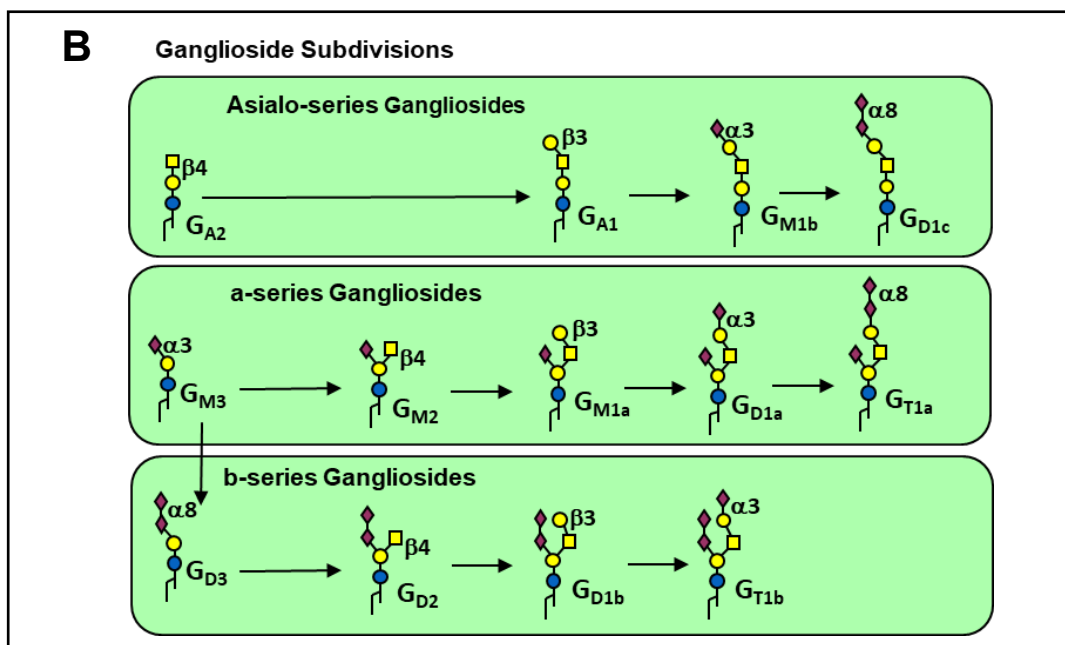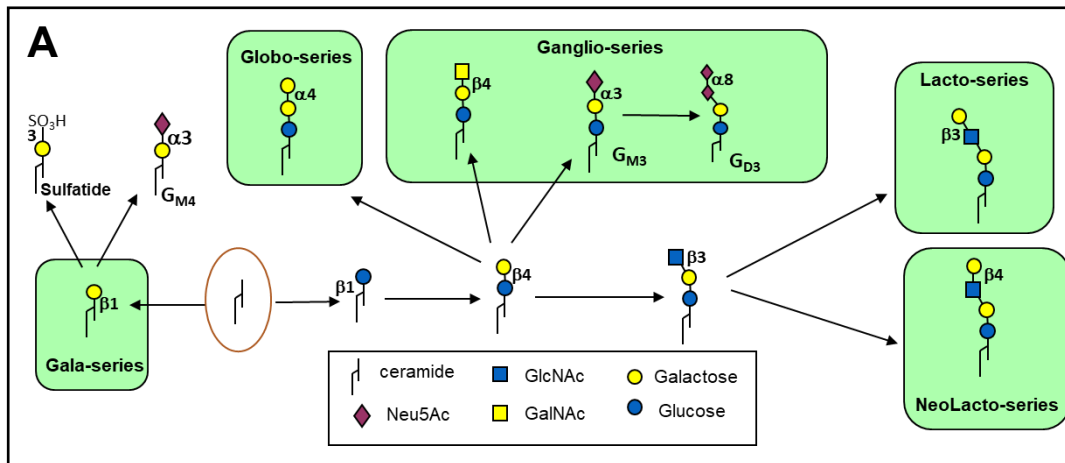**Figure 1.3 Glycosphingolipid Synthesis and Major Classes**
 A) Capping of a ceramide with a Galactose generates Gala-series glycosphingolipids (GSLs), while addition of Glucose (GlcCer) followed by a Galactose generates the LacCer backbone upon which the Globo-, Ganglio-, Lacto- and neoLacto-series GSL are synthesized.
B) Ganglioside GSLs are further subdivided into asialo-, a-, and b-series.

Neu5Acα2-3Galβ1-4Glcβ-Cer is GM3. "G" for ganglioside, "M" for the number of sialic acids being one (mono-, di-, tri-), and a number "3" which designates the size of the underlying neutral core where larger number equals smaller size. While seemingly backwards, this is a reference to the ganglioside's migration pattern on thin layer chromatography (TLC), a technique that sorts by molecular size such that the order 1, 2, 3, etc. goes from largest to smallest. In practice this refers to the extension of the ganglioside with neutral Gal or GalNAc residues. Thus, the addition of a GalNAc to the Gal residue of GM3 produces the ganglioside GM**2**. Alternatively, the addition of another sialic acid to GM3 produces G**D**3. Asialo-series gangliosides are annotated e.g. G**A**# for their initial lack of sialic acid, Globo-series are Gb#, and Lacto- and NeoLActo-series are Lc#, and nLc# respectively.

Glycosyltransferases are the Glycan Biosynthetic Enzymes

Once the core structures of N-linked and GSL glycans are generated, they may be further elaborated as they pass through the Golgi secretory pathway. These glycan modifications impart biochemical properties, structural modifications, and information content to their underlying protein or lipid conjugates and are crucial to proper cellular function. A number of enzymes responsible for the synthesis of common glycan saccharide modifications and structural motifs, along with some examples of their biological relevance will be discussed below.

Lactosamine (LacNAc) Branching and Extension, Type-I versus Type-II:

Lactosamine structures, Galactose(Gal) linked to N-Acetyl-Glucose (GlcNAc) make up one of the most common structural motifs found in N-glycans as well as GSLs. Type-1 lactosamine is generated by the addition of a β1,3-linked Gal to a terminal GlcNAc by the action of the B3GALT family of enzymes.

The B4GALT galactosyltransferase family is responsible for the generation of type-2 lactosamine (LacNAc) structures on many different glycoconjugates, including N- and O-linked glycoproteins and glycolipids. This family catalyzes the addition of β1-4 linked galactose for the formation of Gal-(β1,4)-GlcNAc branches.

Extension of Type-2 lactosamine structures with Gal-(β1,4)-GlcNAc-(β1,3) repeats occurs through the action of β3-GlcNAc transferases B3GNT2, 4 and 8. B3GNT2 is the main poly-lactosamine extension enzyme for N-glycans and has been shown to form heteromeric complexes with B4GALT1, stabilizing their localization in the trans-Golgi and facilitating disaccharide repeat formation [27]. N-glycan poly-lactosamine extension in general favors extending the Gal-GlcNAc residues of the α6Man-arm, and this B3GNT2/B4GALT1 complex is equally efficient on either the MGAT2 extended β2GlcNAc -Man6, or MGAT5 extended β6GlcNAc -Man6 branch [28]. B3GNT8 has been shown to complex with B3GNT2 and to preferentially extend the β6 arm[29]. MGAT5, B3GNT8, and the extension of poly lactosamine on the β6-GlcNAc branch of N-glycans has been implicated in cell adhesion and migration colon cancer [30, 31].

Lactosamine structures on N-glycans and Lacto-/NeoLacto series GSLs can be further modified by the addition of Sialic acid (Neu5Ac) and/or fucose. These additions

may have enzymatic preferences for certain underlying structures, such as type-1 vs type-2 lactosamine [32, 33], glycan classes such as GSL vs N-linked vs O-linked [34], and even specific branches on N-glycans [35] or lengths of poly-lactosamine repeats [32].

Neu5Ac Sialic Acid

Sialic acid can be α3-linked to Gal by ST3Gal sialyltransferases, α6-linked to Gal by ST6Gal sialyltransferases, α6-linked to GalNAc by ST6GalNAc sialyltransferase, or α8-linked to an underlying sialic acid by St8sia polysialic acid (PSA) transferases. Sialylation is a unique glycan modification in that it adds, in humans, a negatively charged Neuraminic acid (Neu5Ac). Brain and central nervous system tissues are found to be especially enriched in sialic acid, where a significant portion is found in GSL brain gangliosides as well as glycoproteins such as neural cell adhesion molecule (NCAM1)[2]. By virtue of its negative charge that can interact with other extracellular matrix structures and its ability to form long poly-sialic acid chains or dense clusters of ganglioside lipid rafts, sialic acids mediate many cell adhesion vs cell migration events. In the brain sialic acid plays roles in development with neural progenitor migration dependent on variable polysialylation of NCAM1, and in mature neuron function of neurite spreading and axon interactions with the electrically insulating myelin sheath [2]. Differential Sialylation and different linkages of sialic acid have also been shown to be important in cancer cell migration [36, 37].

Core and External Fucosylation:

Fucosyltransferases, grouped as α6- (FUT8), α2- (FUT1, FUT2) or α3/4-linked types (FUT3-7,9-11), add fucose to the Gal, GlcNAc, and sometimes GalNAc residues

of N-linked, O-linked, and GSL glycans. FUT8 is unique in humans for its ability to "core fucosylate" the inner most reducing end GlcNAc of the N-glycan chitobiose core. Here it can affect major functional modulations of the underlying glycoprotein function. Loss of FUT or core-fucosylation leads to B-cell receptor/antibody disfunction[38], dysregulation of growth factor and immune system signaling factors such as TGF-beta1[6], and altered cell-cell adhesion in cancer cells through modulation of cadherin binding[39]. The external branches of N-glycan or extended O-glycans and GSLs by α2- or α3/4-fucosyltransferases also mediate important functions at the cell surface. These fucosyltransferases generate biologically relevant glycan epitopes such as the ABO blood group antigens, and cell adhesion selectin ligands to name a few of their functions. External fucosylation can combine with sialylation to form sialyl-Lewis type structures, or it can compete with sialylation as a mutually exclusive modification [40, 41]. External fucose modifications such as the Lewis-X (LeX) structure with fucose α3-linked to the GlcNAc of a terminal lactosamine repeat, are developmentally regulated and mediate cell migration and maintenance of stem cell populations, especially in the developing brain. In the forebrain, α3- fucosyltransferase FUT9 is regulated by the neurodevelopmental "Master control" transcription factor Pax6 [42]. Loss of FUT9 function, decreases LeX glycan abundance and impairs neurite outgrowth in developing neurons[43]. Another LeX generator, Fut10 is required for the maintenance and proper migration of neural stem cell populations [44].

Given these and other examples of the functional roles and developmental regulation of glycan structures, an understanding of the specifics of their biosynthesis and its regulation is crucial. One impediment is the complexity and incompletely

understood nature of glycan biosynthesis. A variety of Golgi glycosyltransferases and enzyme families have been identified, with varying monosaccharide donor acceptors, linkage formations, and substrate specificities. Many of the glycan structural features generated by these classes of glycosyltransferases can modify GSLs as well as N-linked and O-linked glycoprotein glycans. Some of these enzymes may have overlapping specificities, a feature that would complicate the correlation of enzyme gene expression or function to glycan synthesis. In order to understand how these enzymes regulate glycosylation it is thus vital to first understand their functions and specificities towards lipid-, N-, or O-linked glycan substrates. Various studies have attempted to determine the substrate specificities of these enzymes, such as the various members of the B4GalT family, with in-vitro and in-vivo studies sometimes giving conflicting results[35, 45-47]. In **Table 1.1** below I will discuss the glycosyltransferase families and enzymes involved in N-glycan and GSL synthesis and a brief examination of their known properties. In compiling the specificity assignments in the table below, priority was given to in-vivo experiments or biochemical assays using full glycoprotein/glycolipid substrates.

Table 1.1: Glycosyltransferase Functions and Specificities:

Glycosyl-transferase Family	Gene Name	Function/ Product/ Substrate specificity	Glycan Class	Citations
B4GALT Family				
	B4GALT1	•Lactose Synthase. •Main N-glycan Type2 lactosamine, prefers β2-linked GlcNAc of α6-Man arm of N-linked glycans. •Main poly-lactosamine extension enzyme.	N	[48] (Ujita,2000) [47] (Qasba, 2008) [45] (Sato, 2001)
	B4GALT2	Major Type2 poly-lactosamine enzyme in the brain, complex-type N-glycans. O-glycans on α-dystroglycan, Notch.	N, O	[45] (Sato, 2001) [49] (Stanley, 2016) [50] (Furukawa, 2014)
	B4GALT3	NeoLacto GSL (Lc3 extension)	GSL	[51] (Schwientek, 1998)
	B4GALT4	NeoLacto GSL extension (nLc5> Lc3)	GSL	[51] (Schwientek, 1998)
	B4GALT5	Main LacCer Synthase. Galactosylation of β6-linked branch (MGAT5) of N-glycans.	GSL, N, O	[52](Yamaji, 2014) [35](Sato, 2000)
	B4GALT6	Accessory LacCer Synthase	GSL	[50] (Furukawa, 2014)
	B4GALT7	Involved in glycoaminoglycan synthesis	GAG	[50] (Furukawa,2014)
B3GALT Family				
	B3GALT1	GSL Type-1 Lc3 -> Lc4	GSL	[53] (Amado 1998)
	B3GALT2	GSL Type-1 Lc3 -> Lc4. N-links	GSL N	[54](Vester-Christensen, 2014)
	B3GALT3 (B3GALNT1)	Globoside (gb4) Synthase	GSL	[55]Okajima, 2000
	B3GALT4	Ganglioside core extension GM1, GD1b, GA1	GSL	[53] (Amado 1998) [56](Furukawa, 2014)
	B3GALT5	Type-1 GSL Lc3 -> Lc4 Gb4->Gb5 (SSEA3) Core-3 O-glycan extension for Sialyl-LeA synthesis.	GSL O	[57](Zhou, 2000) [58](Zhou, 1999) [59](Isshiki, 1999)
B3GNT-Poly Lac				

	B3GNT2	Poly-lactosamine, type II (major N-glycan type) GSL nLc4 and nLc6	N, GSL	[49](Stanley, 2016) [60] (Shiraishi, 2001)
	B3GNT3	Poly-lactosamine Core 1 O-links. Type1 GSL Lc4 > nLc4	O, GSL	[61] (Yeh, 2001) [60](Shiraishi, 2001)
	B3GNT4	Poly-lactosamine extension, highest affinity for extended LacNAc repeats.	N	[62] (Togayachi, 2014)
	B3GNT5	LC3 synthase	GSL	[63](Biellmann, 2008)
	B3GNT8	Poly-lactosamine Preference for tetraantennary N-glycans and 2,6 triantennary glycans	N	[31](Ishida, 2005), [29](Seko, 2005) [30](Seko, 2008)
B4GALNT Family				
	B4GALNT1	Gangliosides: GM3→GM2, GD3→GD2, LacCer→GA2	GSL	[64](Takamiya, 1996) [65](Furukawa, 2014)
	B4GALNT2	Sd(a) antigen from Sialylparagloboside (Sialyl-nLc4)	GSL	[66](Montiel, 2003)
	B4GALNT3	LacDiNAc: GalNAc-b1,4-GlcNAc.	N, maybe O, peptide specific	[67](Baenziger, 2014)
	B4GALNT4	LacDiNAc: GalNAc-b1,4-GlcNAc.	N, maybe O, peptide specific	[67](Baenziger, 2014)
FUT Family				
	FUT1 (H)	α2-fucose on Gal, H-type 1 and 2 (ectodermal and mesodermal lineages)	N, O, GSL	[68] (Liang, 2011)
	FUT2 (Se)	α2-fucose on Gal, H-type 1 and 2 (endodermal lineages)	N, O, GSL	[68] (Liang, 2011) [69](Satomaa,2009)
	FUT3	(a1-3/4)-fucose on GlcNac of Type1: LeA and LeB. Can also make type2 LeX, LeY, but is only one that can make 1	N, O, GSL	[70](Dupuy, 2004) [71](Kudo, 2014)

	FUT4	(α1-3)Fucose on GlcNAc. LeX/CD15/ SSEA-1 Proximal LacNAc of poly-lactosamine. (Myeloid lineage)	N, O, GSL	[72](Kudo,2014)
	FUT6	(α1-3)Fucose, Sialyl-Lex Type 2 only, Liver specific.	N, O, GSL	[73](Lauc, 2010)
	FUT7	(α1-3)Fucose on GlcNAc of distal LacNAc of sialylated poly-lactosamine. Type 2 Sialyl-LeX only.	N, O, GSL	[72](Kudo,2014) [74](Kudo,2014)
	FUT8	(α1-6) Core fucose-the only core fucose enzyme	N, O, GSL	[75](Ihara, 2014) [25](Calderon, 2016)
	FUT9	(α1-3)Fucose on GlcNAc. LeX Specific Distal LacNAc of poly-lactosamine. Most efficient.	N, O, GSL	[76](Nishihara, 1999)
	FUT10	Probable LeX (α1-3)Fucose on LacNAc arms of bisected, complex N-glycans.	N, O, GSL	[44](Kumar,2013)
	FUT11	Function unknown.	N, O, GSL	
	ABO	GalNAc/Gal transferase that generates A or B blood group antigen on H-type Fuc-(α1,2)-Gal.	N, O, GSL	[77](Hakomori)
ST3GAL Family				
	ST3GAL1	(Gal-β1,3-GalNAc) on O-links.	O	
	ST3GAL2	(Gal-β1,3-GalNAc) on gangliosides.	GSL	[78] (Kojima, 1994) [79] (Sturgill, 2012)
	ST3GAL3	Lactosamine Type1 (Galβ1-3GlcNAc) >Type2 (Galβ1-4GlcNAc)	N, O, > GSL	[34] (Schnaar, 2014)
	ST3GAL4	Lactosamine Type2 (Galβ1-4GlcNAc) > Type1 (Galβ1-3GlcNAc)	N, O, >GSL	[33] (Schnaar, 2014)
	ST3GAL5	"GM3 Synthase"	GSL	[80] (Inokuchi, 2014)
	ST3GAL6	Type2 NeoLacto specific: nLc4, nLc6, contributes to Sialyl-LeX.	GSL	[81] (Okajima, 2014)
ST6GAL1 Family				
	ST6GAL1	Lactosamine Type2 (Galβ1-4GlcNAc), prefers galactose	N	[82] (Kitazume, 2014) [83]

		residue at an α1,3-Man arm of N-linked glycans.		
	ST6GAL2	LacDiNAc (GalNAcβ1-4GlcNAc)	N	[84] (Takashima, 2014)
ST6-GALNAC Family				
	ST6GALNAC 1	The sialyl-Tn antigen, (Neu5Acα2-6GalNAc-O-Ser/Thr).	O	[85]
	ST6GALNAC 2	Similar to ST6GALNAC1	O	[85]
	ST6GALNAC 3	Targets the NeuAcα2-3Galβ1-3GalNAc. Synthesizes the ganglioside GD1α from GM1b. Also O-glycans.	GSL > O	[86]
	ST6GALNAC 4	Similar specificity to ST#GALNAC3, but O-links in humans.	O	[87]
	ST6GALNAC 5	Synthesizes the ganglioside GD1α from GM1b	GSL	[88]
	ST6GALNAC 6	Accepts GM1b, GD1α and GT1b. Also synthesizes globo-series disialyl Gb5. And lacto-series disialyl-Lewis A.	GSL	[88] [36] (Senda, 2007)
ST8Sia Family				
	ST8SIA1	"GD3 Synthase" GM3 → GD3	GSL	[89]
	ST8SIA2	(STX) Poly- sialylation of NCAM1 and other glycoproteins. (Ectodermal tissues, esp. fetal brain)	N, O	[90] [91]
	ST8SIA3	GD3→GT3 Strong affinity to sialyl-Lc4. (Neu5Acalpha2 --> 8Neu5Acalpha2 --> 3Gal) on N-links	GSL, N	[92] [93]
	ST8SIA4	(PST) Poly-sialylation of NCAM1. (Mesodermal and Endodermal tissues)	N	[94] [91]
	ST8SIA5	Gangliosides GD3→GT3 Also, GD1c, GT1a, GQ1b.	GSL	[95]
	ST8SIA6	Mucin O-glycans.	O	[96]

Stem Cells and Development:

Stem Cells are canonically defined by their ability to self-renew and to generate differentiated cell types, where one cell may generate a whole lineage of cell types. Indeed, their name "Stem" cells comes from the prolific German biologist, Ernst Haeckel, and his study of phylogenetic lineages. In 1898 Haeckel likened how the single cell of an embryo generates all of the many lineages of cells of an adult organism to phylogenetic family trees, or Stammbaume "stem trees" in German [97]. In the intervening hundred or so years many cell types have been called "stem cells", varying across a spectrum of "Potency" defined as the breadth of lineages such a cell can make. The initial fertilized embryo, the single cell zygote is Totipotent. It will give rise to the entirety of the tissues of the developing organism including supportive extra-embryonic tissues such as the placenta. On the other end of the spectrum, more differentiated and fate limited stem cells such as adult tissue resident stem cells (e.g. hematopoietic stem cells) along with many fetal progenitor cells are known as Multipotent and can differentiate into a limited number of tissue specific cells. In development the first step down from totipotency is the lineage restriction of the trophectoderm, which will give rise to the extra-embryonic structures supporting the developing organism, and the inner cell mass (ICM) of the 64 cell blastocyst. The cells of the ICM are Pluripotent as they can give rise to all of the cells and tissues required for the development of the adult organism [98]. It is from these cells that the first pluripotent "embryonic stem cells" were isolated from mouse embryos (mESCs) by Gail Martin in 1981 [99]. This was followed by the isolation of human embryonic stem cells (hESCs) by James Thompson in 1998 [100].

Human embryonic stem cells, with their ability to give rise to virtually all cells of the human body, have become important tools in the fields of regenerative medicine, developmental biology, and the study of human disease[101-103]. This was further revolutionized by work of Shinya Yamanaka[104] and others[105] in 2007 with the creation of induced pluripotent stem cell (iPSC) technology allowing the induction of adult somatic cells into an embryonic stem cell-like state, allowing for cells and tissues to be grown from and for individual human patients and without the need for human embryos.

One of the key abilities of pluripotent stem cells, both embryonic and induced, is their ability to recreate the processes of embryonic development. During embryogenesis, early development proceeds with gastrulation producing the three main "germ layer" lineages that will give rise to the adult tissues: Ectoderm, which gives rise to the nervous system, neural crest cells and the epidermis; Mesoderm, which forms the axial skeleton, musculature, cartilage and connective tissues; and Endoderm from which the gastrointestinal, respiratory, and vascular system form [106].

With the ability to grow human tissues in a dish, pluripotent stem cells are useful tools for studying human development and diseases. They are of particular utility in the study of biological process that would be otherwise invasive or impossible to observe in human subjects, such as test the cellular function in specific tissues without the need for biopsy or in examining early events in human development, such as neural differentiation. Important and continuing developmental biology studies[107] have led to the greater understanding of neural development and the creation of methods for highly controllable in-vitro neural differentiations. Distinct neuronal cell populations can be

generated, such as superficial cortical neurons [108], inhibitory interneurons[109], or brain stem dopaminergic neurons [110]. Complex heterogeneous differentiations, such as cerebral organoids recreate complex structures and autonomous organization that mimics cortical development in-vivo, allowing for the study of neural networks and cellular interactions[111].

Thus with these developments patient derived iPSCs and genetically modified ESCs have been used to study many diverse neurodevelopmental diseases, from schizophrenia[112, 113] to a host of Autism spectrum Syndromes like Fragile X [114], Rett Syndrome[115], and Timothy Syndrome[116]. Syndromic disorders with high penetrant chromosomal or genetic disruption are rare minority of Autism Spectrum Disorders (ASD). Future developments and understanding will depend on understanding the many minor genetic correlations and environmental factors that contribute to idiopathic ASD. Models of neural development and function using ASD patient derived iPSC cultures will allow for novel investigations into this poorly understood disorder.

Autism Spectrum Disorder and Neuroligins:

Autism is a poorly understood neurodevelopmental disorder with complex and heterogeneous clinical presentation. It is clinically categorized in the DSM-V by core impairments in social-communicative abilities and restricted, repetitive patterns of behavior and interests. Autism Spectrum Disorders are estimated to have a worldwide prevalence near 1 in 100, and can vary in degree of morbidity, with 45% of ASD individuals having an intellectual disability, 32% undergoing childhood regression of social and communicative skills at the mean age of 1.78 yo. Other comorbid condition

often seen in ASD patients include epilepsy, gastrointestinal problems, immune
dysregulation, and many other psychiatric conditions[117]. Current therapies for ASD
are supportive only, with pharmaceutical approaches still playing a minor or
experimental role. Further understanding of the biological causes and mechanisms is
required for more effective diagnosis, support, and treatment of individuals with ASD.

Autism Genetics

Twin and family studies have shown high concordance rates of ASD and
demonstrated high heritability of ASD risk genes and Autistic behavioral traits[118].
While the etiology of Autism appears to be largely genetic, no single gene accounts for
more than 1-2% of Autism cases. Several dozen highly penetrant gene mutations and
chromosomal abnormalities have been shown to reproducibly cause ASD, while many
other genes identified as risk factors likely interact with other genes, environmental
factors, or epigenetics[119]. Known Autism associated genes play roles in a wide range
of biological process that may impact neural development and neuronal function. Genes
and processes implicated include MECP2 in epigenetic regulation[120], FMR1 in RNA
trafficking and translation[102], DIA1 and DIA1R in the trafficking of proteins through the
ER, enzymes involved in glycosylation[121], UBE3A and regulators of protein
turnover[122], as well as many of the proteins involved in synaptic signal
transduction[123]. Perhaps most highly implicated are synaptic adhesion proteins
including Contactins[124], CNTNAP2[125], Neurexins and the Neuroligins 3 and 4[1].
See Chen et al. (2014)[126] for a review and more comprehensive list.

Neuroligins

Neuroligins are a family of post-synaptic cell adhesion glycoproteins that have been shown to play key roles in neuronal synaptic formation and function and as such have received attention for their role in ASD. In humans, there are 4 major NLGNs, all with the conserved structural components of a cytoplasmic C-terminal domain capable of binding to post-synaptic scaffolding through a PDZ domain, a single pass trans-membrane domain, a glycosylated linker domain, and an extracellular cholinesterase-like domain which mediates extracellular binding[127]. Neuroligins were first discovered for their role in mediating synapse formation and specification by binding to the LGN laminin-type domain of neurexins. This Neuroligin-Neurexin adhesion aligns the developing synaptic connection and recruits the post synaptic receptor machinery (**Figure1.4**). Different neuroligins regulate the excitatory/inhibitory specification of the synapse by binding of different splice-forms of neurexin (α/β) and intracellular association of receptor specific post-synaptic scaffolding such as PSD-95 or Gephyrin[128]. The major excitatory form, NLGN1, selectively binds β-neurexin and PSD-95 on the intracellular side, assembling the scaffolding complex that presents glutamate receptors to the synapse. The inhibitory synapse neuroligin, NLGN2, recognizes α-neurexin and GABAergic synapse scaffolding such as gephyrin and collybistin. NLGN3 was found to heterodimerize with either NLGN1 or 2 and co-localize to both excitatory and inhibitory synapses.
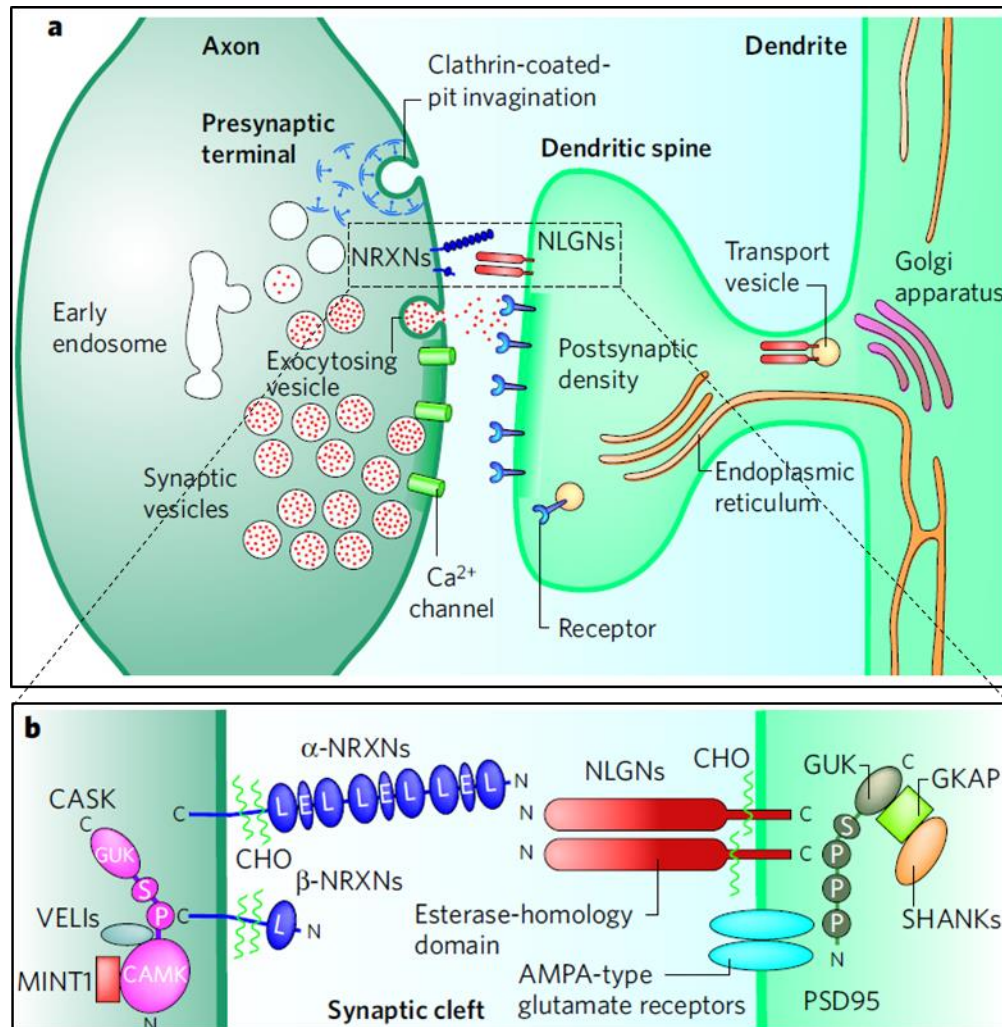
Figure 1.4 Neuroligins as Synaptic Adhesion Molecules
Neuroligins were first characterized by the Sudhof group for their role in neuronal synapse formation and specification.
A) At the neuronal synapse, pre-synaptic neurotransmitter release must be aligned with post-synaptic receptor machinery.
B) Post-synaptic Neuroligins (NLGNs) adhere to pre-synaptic Neurexins to physically stabilize the synapse. The intracellular domain of NLGNs contains PDZ domains that interact with synaptic scaffolding proteins, such as PSD95, necessary for proper localization of neurotransmitter receptors.
(Adapted from Sudhof, 2008)[1]

The fourth member of the neuroligin family, neuroligin-4, is the most implicated in Autism with 11 known human mutations in NLGN4x, 1 in NLGN3, and none in NLGN1 or NLGN2. Despite its human relevance, understanding of its function has been impeded due to low homology in mice of only 59% compared to NLGN1 (98%), NLGN2(98%), and NLGN3(98%) [129]. Human neuroligin-4 is X-linked, designated as NLGN4x to distinguish it from rodent NLGN4 which some categorize as a completely different protein[127]. NLGN4x has been described as a pseudo-autosomal gene, as a 97% identical NLGN4y is located on the Y-chromosome[130] and NLGN4x has recently been shown to be one of the few genes that escapes X-inactivation in female mosaicism[131]. Consistent with these genetic differences, studies of mouse NLGN4[132] and human NLGN4x[133] protein expressed in mouse neurons have produced conflicting reports of localization and function. The divergence of NLGN4x from mouse NLGN4 and the other NLGNs, and its robust connection to Autism, indicate that NLGN4x has a unique role in human neural development. NLGN4x expression and function in human neural development needs to be explored, yet no studies have characterized endogenous NLGN4x protein localization.

Summary and Scope of Dissertation:

The goal of this work is to better understand the mechanisms by which cellular glycosylation is regulated in developing human cells. Utilizing advanced transcriptomics combined with novel mass spectrometric glycome analysis, we analyzed hESC derived differentiated cells and tissues across multiple stages of development and from distinct developmental lineages. By identifying patterns in the transcriptional regulation of glycosylation we aim to advance the knowledge of how glycosylation is regulated in

development and with this, better understand glycomic alterations observed in developmental diseases such as Autism Spectrum Disorders.

In Chapter 2, quantitative N-glycan structural profiling of these hESC derived cell types identifies novel cell-type unique structure and patterns of structures. Transcriptomic analysis identifies biosynthetic steps correlating with the observed glycan profiles indicating loci of transcriptional control, as well as structure/transcript relationships indicating multiple levels of control.

In Chapter 3, GSL profiles of these same cells are analyzed. Comparisons with GSL biosynthetic enzyme transcripts again identifies points of correlation and distinguishing characteristics of the different cell lines. By utilizing both N-glycan and GSL analysis on multiple developmental types and stages, some trends previously thought to be cell specific are found to be more universal trends of cellular differentiation and loss of pluripotency.

Chapter 4 reviews the key and novel findings, discuss the limitations of the approaches used and outlines strategies for further investigation.

Works Cited:

1.   Südhof, T.C., *Neuroligins and Neurexins Link Synaptic Function to Cognitive Disease.* Nature, 2008. **455**(7215): p. 903-911.

2.   Schnaar, R.L., R. Gerardy-Schahn, and H. Hildebrandt, *Sialic Acids in the Brain: Gangliosides and Polysialic Acid in Nervous System Development, Stability, Disease, and Regeneration.* Physiological Reviews, 2014. **94**(2): p. 461-518.

3.   Norihiko, S., et al., *LacdiNAc (GalNAcβ1-4GlcNAc) Contributes to Self-Renewal of Mouse Embryonic Stem Cells by Regulating Leukemia Inhibitory Factor/STAT3 Signaling.* STEM CELLS, 2011. **29**(4): p. 641-650.

4.   Wang, Y.-C., et al., *Glycosyltransferase ST6GAL1 contributes to the regulation of pluripotency in human pluripotent stem cells.* Scientific Reports, 2015. **5**: p. 13317.

5.   Lu, J., et al., *Expression of N-acetylglucosaminyltransferase III suppresses α2,3 sialylation and its distinctive functions in cell migration are attributed to α2,6 sialylation levels.* Journal of Biological Chemistry, 2016.

6.   Wang, X., et al., *Dysregulation of TGF-β1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(44): p. 15791-15796.

7.   Scott, H. and V.M. Panin, *N-glycosylation in Regulation of the Nervous System.* Advances in neurobiology, 2014. **9**: p. 367-394.

8.   Freeze, H.H. and B.G. Ng, *Golgi Glycosylation and Human Inherited Diseases.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(9).

9.   Stanley, P., N. Taniguchi, and M. Aebi, *N-Glycans*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.: Cold Spring Harbor (NY). p. 99-111.

10.   Allam, H., et al., *Glycomic Analysis of Membrane Glycoproteins with Bisecting Glycosylation from Ovarian Cancer Tissues Reveals Novel Structures and Functions.* Journal of Proteome Research, 2015. **14**(1): p. 434-446.

11.   De Jaco, A., et al., *Neuroligin Trafficking Deficiencies Arising from Mutations in the α/β-Hydrolase Fold Protein Family.* The Journal of Biological Chemistry, 2010. **285**(37): p. 28674-28682.

12.   Vaidyanathan, K., S. Durning, and L. Wells, *Functional O-GlcNAc modifications: Implications in molecular regulation and pathophysiology.* Critical reviews in biochemistry and molecular biology, 2014. **49**(2): p. 140-163.

13.   Sheikh, M.O., S.M. Halmo, and L. Wells, *Recent advancements in understanding mammalian O-mannosylation.* Glycobiology, 2017. **27**(9): p. 806-819.

14.   Praissman, J.L., et al., *B4GAT1 is the priming enzyme for the LARGE-dependent functional glycosylation of α-dystroglycan.* eLife, 2014. **3**: p. e03943.

15. Dobson, C.M., et al., *O-Mannosylation and human disease.* Cellular and Molecular Life Sciences, 2013. **70**(16): p. 2849-2857.

16. Dwyer, C.A., et al., *Neurons and Glia Modify Receptor Protein-tyrosine Phosphatase ζ (RPTPζ)/Phosphacan with Cell-specific O-Mannosyl Glycans in the Developing Brain.* Journal of Biological Chemistry, 2015. **290**(16): p. 10256-10273.

17. Schnaar, R.L. and T. Kinoshita, *Glycosphingolipids*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY). p. 125-135.

18. Sezgin, E., et al., *The mystery of membrane organization: composition, regulation and physiological relevance of lipid rafts.* Nature reviews. Molecular cell biology, 2017. **18**(6): p. 361-374.

19. Sych, T., Y. Mély, and W. Römer, *Lipid self-assembly and lectin-induced reorganization of the plasma membrane.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2018. **373**(1747).

20. Nechipurenko, I.V., D.B. Doroquez, and P. Sengupta, *Primary Cilia and Dendritic Spines: Different but Similar Signaling Compartments.* Molecules and Cells, 2013. **36**(4): p. 288-303.

21. Nairn, A.V., et al., *Regulation of Glycan Structures in Murine Embryonic Stem Cells: COMBINED TRANSCRIPT PROFILING OF GLYCAN-RELATED GENES AND GLYCAN STRUCTURAL ANALYSIS.* Journal of Biological Chemistry, 2012. **287**(45): p. 37835-37856.

22. Parodi A, C.R., Aebi M., *Glycans in Glycoprotein Quality Control*, in *Essentials of Glycobiology* C.R. Varki A, Esko JD, et al., editors., Editor. 2017, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.: [Internet].

23. Kizuka, Y. and N. Taniguchi, *Enzymes for N-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer.* Biomolecules, 2016. **6**(2): p. 25.

24. Inamori, K.-i., M. Pierce, and N. Taniguchi, *Mannosyl (Alpha-1,6-)-Glycoprotein Beta-1,6-N-Acetyl-Glucosaminyltransferase, Isozyme B (MGAT5B)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 247-255.

25.   Calderon, A.D., et al., *Substrate specificity of FUT8 and chemoenzymatic synthesis of core-fucosylated asymmetric N-glycans.* Organic & Biomolecular Chemistry, 2016. **14**(17): p. 4027-4031.

26.   Svennerholm, L., *The gangliosides.* Journal of Lipid Research, 1964. **5**(2): p. 145-155.

27.   Lee, P.L., J.J. Kohler, and S.R. Pfeffer, *Association of β-1,3-N-acetylglucosaminyltransferase 1 and β-1,4-galactosyltransferase 1, trans-Golgi enzymes involved in coupled poly-N-acetyllactosamine synthesis.* Glycobiology, 2009. **19**(6): p. 655-664.

28.   Ujita, M., et al., *Poly-N-acetyllactosamine Synthesis in BranchedN-Glycans Is Controlled by Complemental Branch Specificity of i-Extension Enzyme and β1,4-Galactosyltransferase I.* Journal of Biological Chemistry, 1999. **274**(24): p. 16717-16726.

29.   Seko, A. and K. Yamashita, *Characterization of a novel galactose β1,3-N-acetylglucosaminyltransferase (β3Gn-T8): the complex formation of β3Gn-T2 and β3Gn-T8 enhances enzymatic activity.* Glycobiology, 2005. **15**(10): p. 943-951.

30.   Seko, A. and K. Yamashita, *Activation of β1,3-N-Acetylglucosaminyltransferase-2 (β3Gn-T2) by β3Gn-T8: POSSIBLE INVOLVEMENT OF β3Gn-T8 IN INCREASING POLY-N-ACETYLLACTOSAMINE CHAINS IN DIFFERENTIATED HL-60 CELLS.* Journal of Biological Chemistry, 2008. **283**(48): p. 33094-33100.

31.   Ishida, H., et al., *A novel beta1,3-N-acetylglucosaminyltransferase (beta3Gn-T8), which synthesizes poly-N-acetyllactosamine, is dramatically upregulated in colon cancer.* FEBS Lett, 2005. **579**(1): p. 71-8.

32.   Mika, K., et al., *α1,3-Fucoslytransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX.* FEBS Letters, 1999. **452**(3): p. 237-242.

33.   Schnaar, R.L., *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 4 (ST3GAL4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 667-674.

34.   Schnaar, R.L., *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 3 (ST3GAL3)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 657-665.

35. Sato, T., et al., *Correlated Gene Expression between β-1,4-Galactosyltransferase V and N-Acetylglucosaminyltransferase V in Human Cancer Cell Lines.* Biochemical and Biophysical Research Communications, 2000. **276**(3): p. 1019-1023.

36. Senda, M., et al., *Identification and expression of a sialyltransferase responsible for the synthesis of disialylgalactosylgloboside in normal and malignant kidney cells: downregulation of ST6GalNAc VI in renal cancers.* Biochemical Journal, 2007. **402**(3): p. 459-470.

37. Souady, J., et al., *Differences in CD75s- and iso-CD75s-ganglioside content and altered mRNA expression of sialyltransferases ST6GAL1 and ST3GAL6 in human hepatocellular carcinomas and nontumoral liver tissues.* Glycobiology, 2011. **21**(5): p. 584-594.

38. Li, W., et al., *Core Fucosylation of μ Heavy Chains Regulates Assembly and Intracellular Signaling of Precursor B Cell Receptors.* Journal of Biological Chemistry, 2012. **287**(4): p. 2500-2508.

39. Osumi, D., et al., *Core fucosylation of E-cadherin enhances cell–cell adhesion in human colon carcinoma WiDr cells.* Cancer Science, 2009. **100**(5): p. 888-895.

40. Grabenhorst, E., et al., *In Vivo Specificity of Human α1,3/4-Fucosyltransferases III-VII in the Biosynthesis of LewisX and Sialyl LewisX Motifs on Complex-type N-Glycans: COEXPRESSION STUDIES FROM BHK-21 CELLS TOGETHER WITH HUMAN β-TRACE PROTEIN.* Journal of Biological Chemistry, 1998. **273**(47): p. 30985-30994.

41. Brito, C., et al., *Human fucosyltransferase IX: Specificity towards N-linked glycoproteins and relevance of the cytoplasmic domain in intra-Golgi localization.* Biochimie, 2008. **90**(9): p. 1279-1290.

42. Shimoda, Y., et al., *Pax6 Controls the Expression of Lewis x Epitope in the Embryonic Forebrain by Regulating α1,3-Fucosyltransferase IX Expression.* Journal of Biological Chemistry, 2002. **277**(3): p. 2033-2039.

43. Gouveia, R., et al., *Expression of glycogenes in differentiating human NT2N neurons. Downregulation of fucosyltransferase 9 leads to decreased Lewisx levels and impaired neurite outgrowth.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2012. **1820**(12): p. 2007-2019.

44. Kumar, A., et al., *The Lewis X-related α1,3-Fucosyltransferase, Fut10, Is Required for the Maintenance of Stem Cell Populations.* Journal of Biological Chemistry, 2013. **288**(40): p. 28859-28868.

45. Sato, T., S. Guo, and K. Furukawa, *Occurrence of poly-N-acetyllactosamine synthesis in Sf-9 cells upon transfection of individual human β-1,4-galactosyltransferase I, II, III, IV, V and VI cDNAs.* Biochimie, 2001. **83**(8): p. 719-725.

46. Ito, H., et al., *Strategy for the fine characterization of glycosyltransferase specificity using isotopomer assembly*. Vol. 4. 2007. 577-82.

47. Qasba, P.K., B. Ramakrishnan, and E. Boeggeman, *Structure and Function of β-1,4-Galactosyltransferase.* Current drug targets, 2008. **9**(4): p. 292-309.

48. Ujita, M., et al., *Poly-N-acetyllactosamine Extension inN-Glycans and Core 2- and Core 4-branchedO-Glycans Is Differentially Controlled by i-Extension Enzyme and Different Members of the β1,4-Galactosyltransferase Gene Family.* Journal of Biological Chemistry, 2000. **275**(21): p. 15868-15875.

49. Stanley, P., *What Have We Learned from Glycosyltransferase Knockouts in Mice?* Journal of Molecular Biology, 2016. **428**(16): p. 3166-3182.

50. Furukawa, K., H. Clausen, and T. Sato, *UDP-Gal: BetaGlcNAc Beta 1,4-Galactosyltransferase, Polypeptide 2-6; Xylosylprotein Beta 1,4-Galactosyltransferase, Polypeptide 7 (Galactosyltransferase I) (B4GALT2-7)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 63-72.

51. Schwientek, T., et al., *Cloning of a Novel Member of the UDP-Galactose:β-N-Acetylglucosamine β1,4-Galactosyltransferase Family, β4Gal-T4, Involved in Glycosphingolipid Biosynthesis.* Journal of Biological Chemistry, 1998. **273**(45): p. 29331-29340.

52. Yamaji, T. and K. Hanada, *Establishment of HeLa Cell Mutants Deficient in Sphingolipid-Related Genes Using TALENs.* PLOS ONE, 2014. **9**(2): p. e88124.

53. Amado, M., et al., *A Family of Human β3-Galactosyltransferases: CHARACTERIZATION OF FOUR MEMBERS OF A UDP-GALACTOSE:β-N-ACETYL-GLUCOSAMINE/β-NACETYL-GALACTOSAMINE β-1,3-GALACTOSYLTRANSFERASE FAMILY.* Journal of Biological Chemistry, 1998. **273**(21): p. 12770-12778.

54. Vester-Christensen, M.B., L. Hansen, and H. Clausen, *UDP-Gal: BetaGlcNAc Beta 1,3-Galactosyltransferase, Polypeptide 1,2 (B3GALT1,2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 73-80.

55. Okajima, T., et al., *Expression Cloning of Human Globoside Synthase cDNAs: IDENTIFICATION OF β3Gal-T3 AS UDP-N-ACETYLGALACTOSAMINE:GLOBOTRIAOSYLCERAMIDE β1,3-N-ACETYLGALACTOSAMINYLTRANSFERASE.* Journal of Biological Chemistry, 2000. **275**(51): p. 40498-40503.

56. Furukawa, K., Y. Ohmi, and K. Furukawa, *UDP-Gal: BetaGlcNAc Beta 1,3-Galactosyltransferase, Polypeptide 4 (B3GALT4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 81-88.

57. Zhou, D., et al., *The β1,3-Galactosyltransferase β3GalT-V Is a Stage-specific Embryonic Antigen-3 (SSEA-3) Synthase.* Journal of Biological Chemistry, 2000. **275**(30): p. 22631-22634.

58. Zhou, D., E.G. Berger, and T. Hennet, *Molecular cloning of a human UDP-galactose:GlcNAcbeta1,3GalNAc beta1,3 galactosyltransferase gene encoding an O-linked core3-elongation enzyme.* European Journal of Biochemistry, 1999. **263**(2): p. 571-576.

59. Togayachi, A. and H. Narimatsu, *UDP-Gal: BetaGlcNAc Beta 1,3-Galactosyltransferase, Polypeptide 5 (B3GALT5)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 89-99.

60. Shiraishi, N., et al., *Identification and Characterization of Three Novel β1,3-N-Acetylglucosaminyltransferases Structurally Related to the β1,3-Galactosyltransferase Family.* Journal of Biological Chemistry, 2001. **276**(5): p. 3498-3507.

61. Yeh, J.-C., et al., *Novel Sulfated Lymphocyte Homing Receptors and Their Control by a Core1 Extension β1,3-N-Acetylglucosaminyltransferase.* Cell, 2001. **105**(7): p. 957-969.

62. Togayachi, A. and H. Narimatsu, *UDP-GlcNAc: BetaGal Beta-1,3-N-Acetylglucosaminyltransferase 2 (B3GNT2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 283-294.

63. Biellmann, F., et al., *The Lc3-synthase gene B3gnt5is essential to pre-implantation development of the murine embryo.* BMC Developmental Biology, 2008. **8**(1): p. 109.

64. Takamiya, K., et al., *Mice with disrupted GM2/GD2 synthase gene lack complex gangliosides but exhibit only subtle defects in their nervous system.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(20): p. 10662-10667.

65. Furukawa, K., et al., *Beta-1,4 N-Acetylgalactosaminyltransferase 1,2 (B4GALNT1,2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 417-428.

66. Montiel, M.D., et al., *Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2-3Galbeta-R beta1,4-N-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: evidence for an unusual extended cytoplasmic domain.* Biochem J, 2003. **373**(Pt 2): p. 369-79.

67. Baenziger, J.U., *Beta1,4-N-Acetylgalactosaminyltransferase-3 (B4GALNT3) and Beta1,4-N-Acetylgalactosaminyltransferase-4 (B4GALNT4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 429-437.

68. Liang, Y.J., et al., *Changes in glycosphingolipid composition during differentiation of human embryonic stem cells to ectodermal or endodermal lineages.* Stem Cells, 2011. **29**(12): p. 1995-2004.

69. Satomaa, T., et al., *The N-glycome of human embryonic stem cells.* BMC Cell Biology, 2009. **10**(1): p. 42.

70. Dupuy, F., et al., *Structure/functional study of Lewis α3- and α 3/4-fucosyltransferases: The α1,4 fucosylation requires and aromatic residue in the acceptor-binding domain.* Vol. 14. 2004. 347-56.

71. Kudo, T. and H. Narimatsu, *Fucosyltransferase 3. GDP-Fucose Lactosamine α1,3/4-Fucosyltransferase. Lea and Leb Histo-Blood Groups (FUT3, Lewis Enzyme)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 531-539.

72. Kudo, T. and H. Narimatsu, *Fucosyltransferase 4. GDP-Fucose Lactosamine α1,3-Fucosyltransferase. Myeloid Specific (FUT4)*, in *Handbook of*

Glycosyltransferases and Related Genes, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 541-547.

73. Lauc, G., et al., *Genomics Meets Glycomics—The First GWAS Study of Human N-Glycome Identifies HNF1α as a Master Regulator of Plasma Protein Fucosylation.* PLOS Genetics, 2010. **6**(12): p. e1001256.

74. Kudo, T. and H. Narimatsu, *Fucosyltransferase 7. GDP-Fucose Lactosamine α1,3-Fucosyltransferase. Sialyl-Lex Specific (FUT7)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 573-580.

75. Ihara, H., et al., *Fucosyltransferase 8. GDP-Fucose N-Glycan Core α6-Fucosyltransferase (FUT8)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 581-596.

76. Nishihara, S., et al., *α1,3-Fucosyltransferase 9 (FUT9; Fuc-TIX) preferentially fucosylates the distal GlcNAc residue of polylactosamine chain while the other four α1,3FUT members preferentially fucosylate the inner GlcNAc residue.* FEBS Letters, 1999. **462**(3): p. 289-294.

77. Hakomori, S.-i. and M. Palcic, *Histo-Blood Group A and B Transferases, Their Gene Structures, and Common O Group Gene Structures*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 463-477.

78. Tsuji, S. and S. Takashima, *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 2 (ST3GAL2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 645-656.

79. Sturgill, E.R., et al., *Biosynthesis of the major brain gangliosides GD1a and GT1b.* Glycobiology, 2012. **22**(10): p. 1289-1301.

80. Inokuchi, J.-i. and S. Uemura, *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 5 (ST3GAL5)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 675-686.

81. Okajima, T. and K. Furukawa, *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 6 (ST3GAL6)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 687-691.

82. Kitazume, S., *ST6 Beta-Galactoside Alpha-2,6-Sialyltranferase 1 (ST6GAL1)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 693-703.

83. Joziasse, D.H., et al., *Branch specificity of bovine colostrum CMP-sialic acid: N-acetyllactosaminide alpha 2----6-sialyltransferase. Interaction with biantennary oligosaccharides and glycopeptides of N-glycosylproteins.* Journal of Biological Chemistry, 1985. **260**(2): p. 714-719.

84. Takashima, S. and S. Tsuji, *ST6 Beta-Galactoside Alpha-2,6-Sialyltranferase 2 (ST6GAL2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 705-714.

85. Tsuji, S. and S. Takashima, *ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 1 (ST6GALNAC1)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 715-725.

86. Takashima, S. and S. Tsuji, *ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 3 (ST6GALNAC3)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 737-748.

87. Takashima, S. and S. Tsuji, *ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 4 (ST6GALNAC4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 749-758.

88. Furukawa, K., et al., *ST6 N-Acetylgalactosaminide Alpha-2,6-Sialyltransferase 5,6 (ST6GALNAC5,6)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 759-765.

89. Bobowski, M., A. Harduin-Lepers, and P. Delannoy, *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 1 (ST8SIA1)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 767-780.

90. Sato, C., *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 2 (ST8SIA2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 781-795.

91.	Berger, R.P., et al., *ST8SIA4-Dependent Polysialylation is Part of a Developmental Program Required for Germ Layer Formation from Human Pluripotent Stem Cells.* Stem Cells, 2016. **34**(7): p. 1742-52.

92.	Tsuji, S. and S. Takashima, *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 3 (ST8SIA3)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 797-804.

93.	Sato, C., T. Matsuda, and K. Kitajima, *Neuronal Differentiation-dependent Expression of the Disialic Acid Epitope on CD166 and Its Involvement in Neurite Formation in Neuro2A Cells.* Journal of Biological Chemistry, 2002. **277**(47): p. 45299-45305.

94.	Angata, K. and M. Fukuda, *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 4 (ST8SIA4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 805-812.

95.	Tsuji, S. and S. Takashima, *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 5 (ST8SIA5)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 813-821.

96.	Takashima, S. and S. Tsuji, *ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 6 (ST8SIA6)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 823-832.

97.	Ramalho-Santos, M. and H. Willenbring, *On the origin of the term "stem cell".* Cell Stem Cell, 2007. **1**(1): p. 35-8.

98.	Posfai, E., O.H. Tam, and J. Rossant, *Chapter One - Mechanisms of Pluripotency In Vivo and In Vitro*, in *Current Topics in Developmental Biology*, M. Rendl, Editor. 2014, Academic Press. p. 1-37.

99.	Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proceedings of the National Academy of Sciences of the United States of America, 1981. **78**(12): p. 7634-7638.

100.	Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.

101.	Wang, H. and L.C. Doering, *Induced Pluripotent Stem Cells to Model and Treat Neurogenetic Disorders.* Neural Plasticity, 2012. **2012**: p. 346053.

102. Marchetto, M.C.N., et al., *A model for neural development and treatment of Rett Syndrome using human induced pluripotent stem cells.* Cell, 2010. **143**(4): p. 527-539.

103. Drury-Stewart, D., et al., *Highly efficient differentiation of neural precursors from human embryonic stem cells and benefits of transplantation after ischemic stroke in mice.* Stem Cell Research & Therapy, 2013. **4**(4): p. 93-93.

104. Takahashi, K., et al., *Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.* Cell, 2007. **131**(5): p. 861-872.

105. Yu, J., et al., *Induced pluripotent stem cell lines derived from human somatic cells.* Science, 2007. **318**(5858): p. 1917-20.

106. Kiecker, C., T. Bates, and E. Bell, *Molecular specification of germ layers in vertebrate embryos.* Cellular and Molecular Life Sciences, 2015. **73**(5): p. 923-947.

107. Abranches, E., et al., *Neural Differentiation of Embryonic Stem Cells In Vitro: A Road Map to Neurogenesis in the Embryo.* PLOS ONE, 2009. **4**(7): p. e6286.

108. Boissart, C., et al., *Differentiation from human pluripotent stem cells of cortical neurons of the superficial layers amenable to psychiatric disease modeling and high-throughput drug screening.* Transl Psychiatry, 2013. **3**: p. e294.

109. Kim, T.-G., et al., *Efficient specification of interneurons from human Pluripotent stem cells by dorsoventral and rostrocaudal modulation.* Stem cells (Dayton, Ohio), 2014. **32**(7): p. 1789-1804.

110. Kriks, S., et al., *Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson/'s disease.* Nature, 2011. **480**(7378): p. 547-551.

111. Mariani, J., et al., *Modeling human cortical development in vitro using induced pluripotent stem cells.* Proceedings of the National Academy of Sciences, 2012. **109**(31): p. 12770-12775.

112. Brennand, K., et al., *Modeling schizophrenia using hiPSC neurons.* Nature, 2011. **473**(7346): p. 221-225.

113. Brennand, K., et al., *Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia.* Molecular Psychiatry, 2015. **20**(3): p. 361-368.

114. Bar-Nur, O., I. Caspi, and N. Benvenisty, *Molecular analysis of FMR1 reactivation in fragile-X induced pluripotent stem cells and their neuronal derivatives.* Journal of Molecular Cell Biology, 2012. **4**(3): p. 180-183.

115. Ananiev, G., et al., *Isogenic Pairs of Wild Type and Mutant Induced Pluripotent Stem Cell (iPSC) Lines from Rett Syndrome Patients as In Vitro Disease Model.* PLoS ONE, 2011. **6**(9): p. e25255.

116. Paşca, S.P., et al., *Using iPS cell-derived neurons to uncover cellular phenotypes associated with Timothy Syndrome.* Nature medicine, 2011. **17**(12): p. 1657-1662.

117. Lai, M.-C., M.V. Lombardo, and S. Baron-Cohen, *Autism.* The Lancet. **383**(9920): p. 896-910.

118. Ronald, A. and R.A. Hoekstra, *Autism spectrum disorders and autistic traits: A decade of new twin studies.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2011. **156**(3): p. 255-274.

119. Geschwind, D.H., *Autism: Many Genes, Common Pathways?* Cell, 2008. **135**(3): p. 391-395.

120. Urbach, A., et al., *Differential modeling of Fragile X syndrome by human embryonic stem cells and induced-pluripotent stem cells.* Cell Stem Cell, 2010. **6**(5): p. 407-411.

121. Aziz, A., S.P. Harrop, and N.E. Bishop, *DIA1R Is an X-Linked Gene Related to Deleted In Autism-1.* PLoS ONE, 2011. **6**(1): p. e14534.

122. Yi, J.J., et al., *An autism-linked mutation disables phosphorylation control of UBE3A.* Cell, 2015. **162**(4): p. 795-807.

123. Gilman, Sarah R., et al., *Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses.* Neuron, 2011. **70**(5): p. 898-907.

124. Zuko, A., et al., *Contactins in the neurobiology of autism.* European Journal of Pharmacology, 2013. **719**(1–3): p. 63-74.

125. Poot, M., *Connecting the CNTNAP2 Networks with Neurodevelopmental Disorders.* Molecular Syndromology, 2015. **6**(1): p. 7-22.

126.   Chen, J., et al., *Synaptic proteins and receptors defects in autism spectrum disorders.* Frontiers in Cellular Neuroscience, 2014. **8**: p. 276.

127.   Bemben, M.A., et al., *The cellular and molecular landscape of neuroligins.* Trends in Neurosciences, 2015. **38**(8): p. 496-505.

128.   Krueger, D.D., et al., *The role of neurexins and neuroligins in the formation, maturation, and function of vertebrate synapses.* Current Opinion in Neurobiology, 2012. **22**(3): p. 412-422.

129.   Bolliger, M.F., et al., *Unusually rapid evolution of Neuroligin-4 in mice.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(17): p. 6421-6426.

130.   BOLLIGER, M.F., et al., *Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression.* Biochemical Journal, 2001. **356**(2): p. 581-588.

131.   Bellott, D.W., et al., *Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators.* Nature, 2014. **508**(7497): p. 494-499.

132.   Hoon, M., et al., *Neuroligin-4 is localized to glycinergic postsynapses and regulates inhibition in the retina.* Proceedings of the National Academy of Sciences, 2011. **108**(7): p. 3053-3058.

133.   Bemben, M.A., et al., *Autism-associated mutation inhibits protein kinase C-mediated neuroligin-4X enhancement of excitatory synapses.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(8): p. 2551-2556.

Chapter 2:

Transcriptional Regulation of N-Glycan Biosynthesis

in Human Stem Cell Differentiation

Introduction:

Glycoconjugates mediate cell adhesion and cell signaling events that are essential for normal tissue development [1, 2]. Therefore, understanding the cellular repertoires of glycan structures and the dynamic changes associated with cellular differentiation will lead to insights into the mechanisms governing specific developmental processes. For example, one of our surprising findings was that polysialylation enzyme ST8SIA4 is highly upregulated in differentiated smooth muscle.

Human embryonic stem cells and induced pluripotent stem cells (iPSCs), with their ability to give rise to virtually all cells of the human body, have become important tools in the fields of regenerative medicine, developmental biology, and the study of human disease [3, 4]. Many of the most useful markers for characterizing pluripotency and subsequent commitment to specific cell fates, such as SSEA3/4, Tra-1-81, etc., are glycan epitopes [5]. Given their ubiquitous and abundant presence on the surface of all eukaryotic cells, it is reasonable to expect that glycomic shifts associated with cell differentiation may provide valuable targets for defining cell-specific markers and for understanding cell-cell interactions that underlie tissue development.

A variety of enzymatic, antibody/lectin, and mass spectrometry approaches have been applied to characterize pluripotent cells and the differentiated cells derived from these pluripotent cell sources[6-8]. These methods have sought to identify the full profile of stem cell glycans and to identify glycans or glycan profiles unique to pluripotent and differentiated cell types. Each of these approaches had individual strengths and limitations, but generally were limited in scope to one or two cell types. In this study we advance the understanding of cell specific glycosylation by combing advanced glycomic and transcriptomic analysis of multiple cell types across the spectrum of development.

We have undertaken comprehensive glycomic and glyco-transcriptomic analysis of human ES and multiple derived cell types in order to investigate lineage-specific regulation of protein glycosylation and to assess the concordance of gene transcription with glycan structural elaboration. In this chapter, we report detailed N-glycan profiles, comprehensive RNAseq and targeted qRT-PCR transcriptomics of glycosylation processing genes in human embryonic stem cells and in cell types differentiated from these ES cells into lineages representative of all three germ layers and multiple stages of maturation. Thus, allowing a more dynamic view to examine the developmental regulation of cell specific protein glycosylation.

We find that combined analysis of transcriptomes and glycomes of these cells revealed shifts in biosynthetic pathways between pluripotent, multipotent, and differentiated cells leading to the generation of glycan structural profiles unique to the different cell types.

Experimental Procedures:

<u>Stem Cell Culture and Differentiation</u>:

Samples for glycomic and transcriptomic analyses were generated from human embryonic stem cells (H9 cell line) and differentiated cell types designated NC, WT1, SM, Liv. For each of the five cell types, four independent biological replicates were collected (n=4). Sample material from each replicate was split for paired glycomic and RNA transcript analysis. Stem cell culture and differentiation conditions utilized feeder-free and serum-free defined media conditions to restrict the introduction of exogenous or non-human glycans. All stem cell and differentiated cultures were grown in 37°C/5%CO2 incubator conditions. The following stem cell culture and sample collection was performed by Michael Kulick and collaborators in the Steven Dalton laboratory.

H9 hESCs were maintained in defined and feeder free conditions on Geltrex matrix coated plates. Defined Media consisted of: DMEM/F12 without glutamine (Cellgro), with the final concentration of following factors: 2% Probumin, Life Science Grade (Millipore), 1x NEAA (Cellgro) 2mM GlutaGro (Cellgro), 1x Pen/Strep (Cellgro), 10ug/ml Human Transferrin, (Athens Research & Tech),1x Trace elements A, B, & C (Cellgro), 50ug/ml L-Ascorbic acid 2-phosphate sesqimagnesium salt hydrate (Sigma), and 10uL/mL 2β-mercaptoethanol (Sigma). For final stem cell medium the following factors were added to the following final concentrations: 10ng/ml Heregulin β (Peprotech), 10ng/ml Activin A (R&D Systems), 200ng/ml Long R3-IGF1 (SAFC/Sigma), 8ng/ml bFGF (RnD Systems). For further stem cell culture details see [9].

Neural crest stem cells (NC) were differentiated from H9 hESCs as previously described[9, 10]. Briefly, hESCs were single-cell dissociated and seeded at 9×10^4 cells/cm^2 on Geltrex in stem cell media. Careful control of cell seeding density is important as overly dense cultures may develop a mixture of neural stem cells and neural crest stem cells[11]. Stem cell media was changed 24hrs after plating to NC induction media by the removal of Activin-A, and the addition of GSK3 inhibitor IX (BIO) and ALK inhibitor SB431542.

WT1 antigen positive multipotent mesothelial precursors (WT1) were differentiated from H9 hESCs through an Isl1-positive mesodermal intermediate (protocol to be published in manuscript by Colunga, et al. 2018). These multipotent WT1 cells are capable of further differentiation into smooth muscle, fibroblasts, and endothelial cells[12]. From these WT1 cells terminally differentiated, functional smooth muscle (SM) cell cultures were generated.

Functionally mature hepatocytes (Liv) were generated from H9 hESCs via a protocol adapted from the Keller Lab [13]. Dissociated H9 hESCs in stem cell media supplemented with 10nM Y27632 were aggregated in suspension culture on a rotary shaker and induced to definitive endoderm (DE) precursors in Defined Media with 100ng/ml Activin, 8ng/ml bFGF and 3uM CHIR. DE precursors in aggregate suspension were then directed through hepatic linage specification and maturation, via extended culture (36 days) through a series of media conditions.

<u>Gene Selection and Primer Design:</u>

Human glycan related genes were selected based on previous studies of the transcriptomics of glycan biosynthetic pathways in murine tissues and stem cells [14-

16]. Human gene sequence primers were designed and validated as described in[17].

The following sample RNA processing and gene expression analysis via qRT-PCR and

RNA-seq was performed by Alison Nairn and collaborators in the Kelley Moremen

laboratory.

RNA Isolation and cDNA Synthesis:

As all primer sets for qRT-PCR were designed within a single exon, and to

ensure that mRNA transcripts were not contaminated, it is important that no genomic

DNA (gDNA) remained in the RNA isolates. RNA was isolated from harvested cell

pellets using the RNeasy Plus Mini RNA isolation kit (Qiagen), which contains a column

to remove gDNA. Further screening for gDNA contamination was conducted along with

the cDNA synthesis protocol.

For RT-PCR, cDNA was synthesized from 1µg RNA using the SuperScript III First

Strand Synthesis kit (Invitrogen) according to the manufacturer's protocol, with the

modification that both the both oligo(dT)and random-primers (1:1) were included in the

cDNA synthesis reactions. cDNA reaction products (20µL) were diluted 1:20 to be used

as templates for qRT-PCR reactions. Control reactions for each sample were prepared

without reverse transcriptase ("no-RT") and analyzed for lack of amplification to ensure

the total removal of gDNA.

Ion Torrent Library Preparation and Whole Transcriptome Sequencing

Total RNA was isolated from flash frozen cell pellets (10^{6-7} cells) provided by the

Dalton lab using the RNeasy® Plus Mini Kit (Qiagen). Total RNA quality was checked

using an Agilent® 2200 Tape Station for a RIN>7 and an ERCC RNA Spike-In Control Mix

was added. Dynabeads® mRNA DIRECT™ Micro Kit (ThermoFisher) was used for

isolating poly(A) mRNA from 5μg of total RNA. RNAse III was used to fragment the mRNA and the size distribution was assessed using an Agilent® 2100 Bioanalyzer®. The Ion Total RNA-Seq Kit v2.0 (ThermoFisher) was used to create strand-specific sequencing libraries following manufacturer's protocol. Two libraries, from biological replicates, for hES, WT1, SM and Liver samples were sequenced on separate Ion PI™ V2 chips (ThermoFisher). For NC samples, one biological replicate was run on a single chip and a barcoded library of 4 NC biological replicate samples were run on another chip.

RNA-Seq Data Alignment and Normalization

Sequences in fastq format were subjected to quality filtering (Q20) and trimming (Discard bases after 150, Minimum read length 35) in the Partek® Flow® software. Sequences were aligned to hg19 using STAR v2.5.3a and Bowtie2 v2.2.5 (local mode) in a 2-step alignment process [Sun, Y.A. *et al*. (2013) Two-step alignments for Ion Proton™ sequencer RNA-seq analysis. *Life Technologies White Paper]*. Basically, reads were aligned using the STAR algorithm and the unaligned read were subjected to alignment by Bowtie2, then the aligned reads from both STAR and Bowtie2 were combined into a single file. The combined aligned reads were quantified to the RefSeq transcriptome (8/2/17 release) using the Partek E/M method [18]. Normalized counts (RPKM with 0.01 offset) were used for histogram plots or analyzed for differential expression with the Gene-specific analysis (GSA) method within the Partek Flow software package [version 7.0.18, Copyright © 2018, Partek Inc., St. Louis, MO, USA].

qRT-PCR:

Quantitative RT-PCR analysis methods used were as described in [16]. Reactions were set up in a 96-well plate format using a 20-μl reaction volume (Bio-Rad

MyIQ) consisting of 50% iQ™ SYBR® Green Supermix (Bio-Rad), 25% diluted cDNA template, and 25% primer pair mix (500 nM each primer, 125 nM final concentration; Eurofins MWG Operon). Transcript abundance for each primer pair was analyzed in technical triplicate on four biological replicate samples for each cell type, and a normalization control gene was included on each plate. Relative transcript abundance was calculated using the ΔΔCt method[19] using ribosomal protein gene Rpl4 for normalization as described in detail previously [17].

Extraction and Preparation of N-linked glycans:

Cell pellets were homogenized on ice in an extraction mixture of chloroform/methanol/water (C:M:W, 4:8:3). Isolation of lipid and glycoprotein fractions, selective release of N-linked glycans by PNGase-F, and their purification were performed as previously described[20]. Following CMW extraction and centrifugation, the combined liquid phase containing Glycolipids was dried down on under N^2 flow/37°C and stored desiccated at -20° awaiting further analysis (See chapter 3: Stem Cell GSLs). The precipitated protein pellet underwent further acetone washes at -20° to remove free sugar polymers and other contaminants was then thoroughly dried and stored desiccated at -20°.

As part of a method for absolute quantification of nmol glycan per mg protein, recently developed by our lab [21], careful and consistent handling of protein yields was necessary. Approximately 2mg of dry protein powder from each sample was well suspended via vortexing and sonication in trypsin buffer(0.1 M Tris-HCl, pH 8.2, 1mM CaCl2) and digested overnight with 2mg/mL trypsin and 2mg/mL chymotrypsin. Following digestion, the total volume of soluble glycoprotein digest was recorded and

10μL taken for protein quantification by BCA reaction (BCA Kit, Pierce, cat 23225). Glycopeptide processing continued as described in [20], with post Trypsin C_{18} column clean up, PNGase-F enzymatic release of N-glycans, final C_{18} column purification of released N-Glycans, eluted and dried by lyophilization. For mass spectrometry analysis, N-glycans were permethylated in solution of anhydrous NaOH in DMSO with Iodomethane [22]. In our hands, carefully applying this protocol as detailed in [23] greatly increases permethylation efficiency.

Mass Spectrometry Analysis:

Permethylated N-glycans from equivalent amounts of protein were analyzed by direct infusion mass spectrometry. Total permethylated N-glycans were resuspended in a volume of 100% methanol such that 10uL would be equivalent to 150mg starting protein. This normalized amount was dissolved in a final infusion buffer of 50ul 50% methanol/50% water with 1mM sodium hydroxide, and 10pmol of internal standard, heavy C_{13} permethylated maltotetrose (STD-DP4). This standard has been shown to demonstrate a linear quantitative relationship over relevant concentrations(30nM to 3μM) and mass range (m/z 899- 2831)[21].

Nano-Spray Ionization (NSI) direct infusion was performed on a Thermo-Fischer NSI-LTQ analyzer, with tandem Fourier-transform Orbitrap (FTMS) and Linear Ion Trap (ITMS) modes. Various scan and fragmentation modes described below were utilized for identification, interrogation and quantification of glycan structures. All scans were performed in positive ion mode. MS^n fragmentation was collected in ITMS centroid mode using Collision Induced Dissociation (CID) with settings of Collision Energy at 35-40%, Activation Q of 0.250, and Activation Time of 30.0ms.

For absolute quantitation of N-glycan m/z peak intensities, direct infusion samples with the DP4 internal standard were analyzed in FTMS profile mode. Spectra were recorded over mass range of 600-2000m/z, and peak intensities recorded as the average of 3 scans of 10 microscans each.

<u>Glycan Structure Annotation and Quantification</u>:

Glycan structure assignments for detected m/z peaks were based on the calculated masses of the predicted intact glycan molecule, the presence of diagnostic MS^n fragments representing unique N-glycan structural features, and the limitations imposed on N-Glycan structural diversity by known glycan biosynthetic pathways. Public access glycomics analysis platforms such as the GRITS Toolbox (<u>www.grits-toolbox.org</u>), also contain curated glycan structure databases and automated tools for the annotation of glycomics mass spectrometry data. GlycoWorkbench v.2 was used for calculation of intact glycan masses and manual review of possible fragmentation products[24, 25].
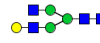
Initially, spectra were annotated by GRITS glycomics toolbox, identifying 97 masses as N-glycans by characteristic MS2 fragmentation. GRITS MS^2 fragmentation analysis revealed that many m/z values included multiple isomeric forms, such as GlcNAc branch placement and the presence of mixtures of diagnostic fragments for multiple isotypes of fucose placement (e.g. 474m/z chitobiose core α1,6-linked fucose or 660m/z external arm Lewis or H-type α1,2/3-linked fucose). **Table 2.1** displays the glycan masses detected along with the observed structural isomers for each peak.
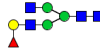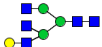
Quantification of glycan structures is complicated by the presence of isomeric and, in the case of NSI MS, multiple charge state m/z peaks. To address this, MS

Table 2.1: **Identified N-Glycan Masses and Detected Isomeric Structures**
Detected masses arranged from smallest to largest. Theoretical non-sodiated
(z=0) masses displayed, along with the number of structural isomers identified
and their presumptive structures based on fragmentation data and known N-
glycan biosynthetic rules. CID MSn fragmentation confirmed amount of core
branching and composition of terminal antennae, though exact linkages and
placement of arms was not interrogated. In some samples, external fucose
structures contained a mix of LeX α1,3-GlcNAc linked fucose and H-antigen
α1,2-Gal linked fucose.

Mass (z=0)	Isomers	Structure 1	Structure 2	Structure 3	Structure 4	Structure 5	Structure 6
944.48	single						
1118.58	single						
1148.58	single						
1322.67	single						
1352.68	single						
1393.72	single						
1556.78	single						
1567.81	single						

1597.82	2					
1638.85	2					
1730.87	single					
1760.88	single					
1771.90	3					
1801.91	2					
1812.93	2					
1842.95	4					
1883.97	single					
1945.99	single					
1958.99	single					

1964.98	single						
1976.00	3						
2006.01	single						
2017.04	6						
2047.04	4						
2058.05	2						
2088.06	2						
2133.07	single						
2150.09	single						
2163.09	single						
2169.08	single						
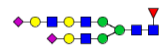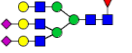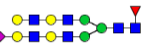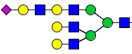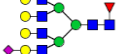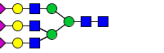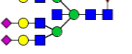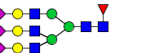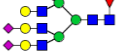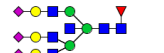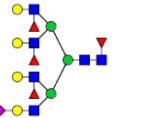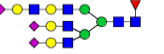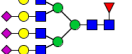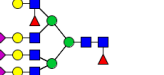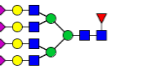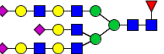
2180. 10	2					
2191. 12	2					
2204. 12	2					
2221. 13	6					
2251. 15	2					
2262. 15	6					
2292. 17	3					
2303. 18	x					
2337. 17	single					
2367. 18	single					
2373. 18	single					

2378.21	3						
2395.22	4						
2408.22	2						
2425.23	single						
2436.24	single						
2449.25	single						
2466.25	4						
2496.27	2						
2541.28	single						
2569.30	2						
2577.29	single						

2582.30	4						
2599.32	single						
2610.34	single						
2612.32	single						
2623.34	5						
2640.35	4						
2653.35	2						
2664.36	single						
2670.36	4						
2756.39	single						
2769.38	single						

2797.42	2						
2814.43	2						
2827.43	4						
2838.45	single						
2844.44	2						
2857.44	2						
2868.45	single						
2915.48	3						
2930.49	single						
2943.47	single						
2984.51	single						

3001.52	single						
3014.52	single						
3025.54	single						
3031.54	4						
3119.59	3						
3188.60	2						
3205.62	3						
3218.62	2						
3276.66	3						
3304.66	2						
3379.71	single						

3392.71	2						
3480.76	3						
3579.79	single						
3637.84	2						
3753.87	single						
3841.93	3						
3999.00	single						
4003.02	single						
4203.10	2						
4379.17	single						
4564.27	single						
4652.33	single						

proteomics peak deconvolution program xTract (Thermo) was modified for use with glycan oligosaccharide structures. xTract detects isotopic peaks based on theoretical distribution of constituent atoms with default "Averagine" values of ($C_{4.94}$ $H_{7.76}$ $N_{1.36}$ $O_{1.48}$ $S_{0.042}$) based on average amino acid composition. For N-glycan oligosaccharide structures, these parameters were changed to an average-saccharide unit "Averagose" ($C_{6.00}$ $H_{9.81}$ $N_{0.37}$ $O_{4.34}$ $S_{0.0}$) [26]. These settings increased the peak fit% to 99.7% providing more accurate glycan peak selection and isotopic peak integration.

Monoisotopic xTract peak intensities were assigned to the glycan structures identified for that peak by GRITS or manual MS^n analysis. For peaks with multiple possible isomeric structures, the ratios of mutually exclusive diagnostic fragments were used to assign the relative amounts of that peak's intensity. For some masses, multiple levels of selection and fragmentation were required to quantitatively distinguish the ratios of the structures contained in that m/z peak.

We were able to distinguish overlapping isomeric glycan structures by utilizing MS^n CID fragmentation to determine the presence, absence, or relative ratios of mutually exclusive and structurally diagnostic ion fragments. The sensitivity of our Ion Trap allowed for fragmentation levels of up to MS^8 and allowed for the resolution key structural differences such as the placement of external fucose motifs, core structure signatures that separate bisecting GlcNAc from terminal arm GlcNAc, and the elongation of Poly-Lactosamine repeats versus increased branching. For example, in **Figure 2.1** the presence of the 444m/z peak rather than a 458m/z peak confirms a bisecting core structure (similar to the bisecting/branching analysis in [27]). Manual MS^n annotation relied on consistent fragmentation behavior of permethylated N-glycans in
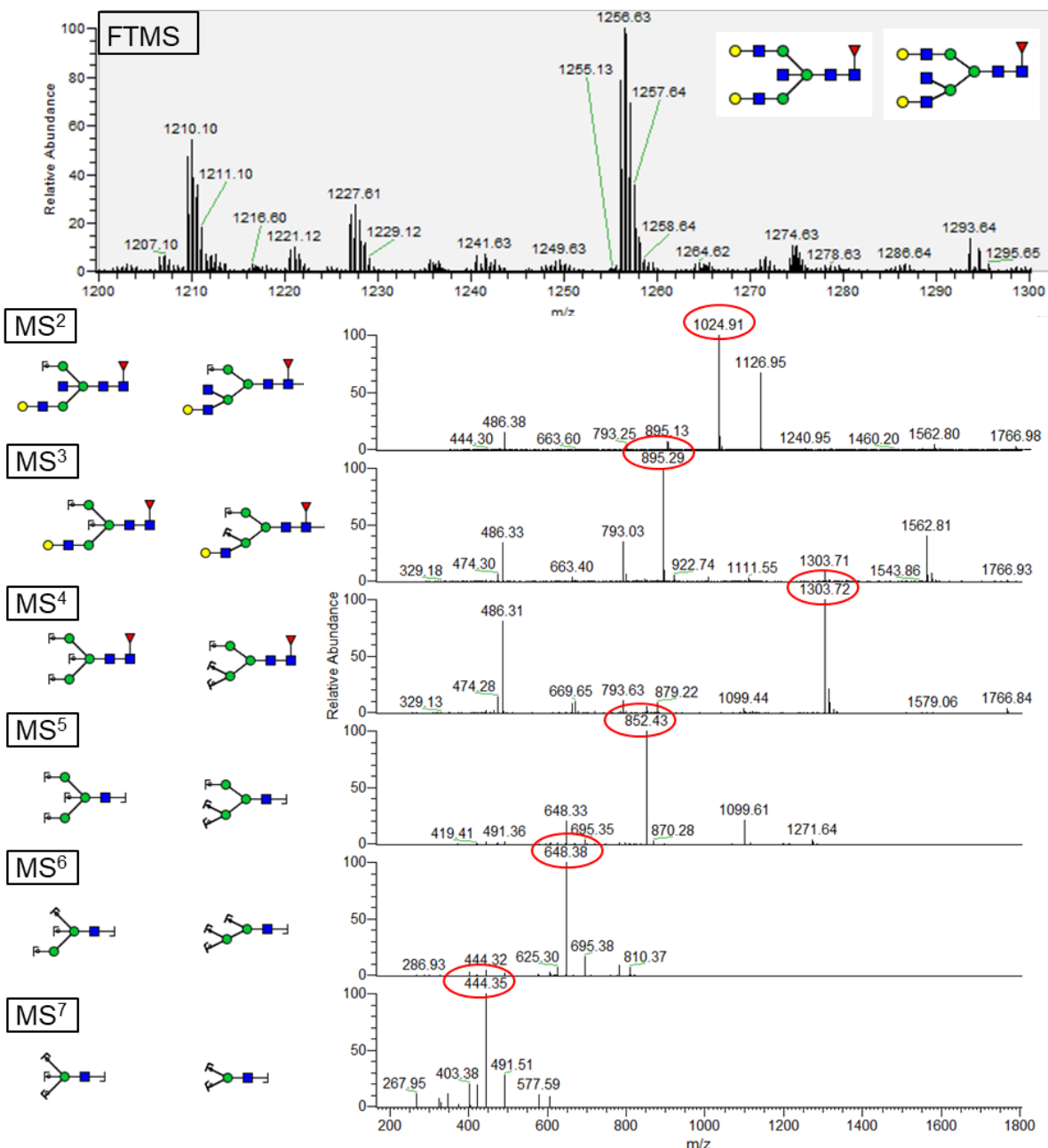
Figure 2.1: Sequential MS^n fragmentation pattern to determine placement of terminal GlcNAc. **FTMS**: The 1256 m/z peak matches the mass of a doubly charged (z=2) glycan with multiple possible structures. **MS^2** (1256.6): Y-ion loss of external lactosamine at 1024.5m/z (z=2) and corresponding singly charged (z=1) 486.2m/z fragment peak. **MS^3** (1024.9): Y-ion loss of GlcNAc at 894.9m/z (z=2). **MS^4** (895): Loss of lactosamine resulting in the singly charged core structure at 1303.6. **MS^5** (1303.6): Loss of core fucosylated reducing end GlcNAc and fucose at 852.4m/z (z=1). **MS^6** (852) →**MS^5** (648): fragmentation to either bisecting core fragment at 444.2m/z or branching core fragment 458.2m/z.

CID fragmentation, such as the preferred production of b- and y-ion fragments[28]. The MSn CID fragmentation produced stereotyped patterns of structure deconstruction, generally beginning with the distal ends of antennae and through iterated rounds of fragmentation deconstructing down to the N-glycan core structure of mannose linked to the chitobiose GlcNAc-GlcNAc-R reducing end. The same designated structure-specific fragments were used for all samples. See **Supplemental Table 2.1** for all structures and diagnostic fragmentations patterns.

Absolute quantification (nmol glycan per mg protein) of the identified glycan structures was calculated using glycan m/z peak intensities, DP4-STD peak intensity, known concentration of the internal DP4 standard, the quantified amount of initial protein assayed, and the total assayed volume [21]. Relative abundances were calculated as percentage of total profile (%Total Profile) by normalizing to the totalsum of identified glycan structures.

Data Analysis and Statistics:

Transcript abundances from qRT-PCR data were calculated as average relative abundances as described previously [17]. Relative transcript abundances were plotted as histograms of the mean of four sample replicates (n=4) with error bars indicating Standard Error (S.E.) and displayed on a Log_{10} scale due to the wide dynamic range of gene expression levels. The -fold change of transcript abundances were calculated by dividing the average transcript abundance of a gene from a differentiated cell type (NC, WT1, Liv, SM) from the corresponding value in the undifferentiated H9 cells. Average -Fold change values were plotted as histograms with error bars displaying S.E. on a

linear scale. To equally display increases and decreases, genes with decreased transcript abundances (ratio<1) were displayed as a negative reciprocal.

For Glycan analysis absolute abundances (nmol glycan per mg protein) or relative abundances (%Total Profile) for individual glycan structures, and grouped structural classes, were plotted as histograms of the mean of four sample replicates (n=4) with error bars indicating Standard Error (S.E.) and plotted on a linear scale.

Hierarchical clustering heat maps and dendrograms were generated using the R statistical programing language (version 3.1.1), implementing Euclidean distance methods as previously described [29, 30]. Fold difference values relative to the undifferentiated H9 mean values were calculated for each sample replicate. Variation in N-glycan structure relative abundances and glycogene transcript abundances was calculated by clustering Log_2 fold difference values across all sample replicates.

Results:

Stem Cell Differentiation of the Three Main Germ Layer Embryonic Lineages.

Human embryonic stem cells (H9 line) were differentiated into ectodermal neural crest stem cells (NC), mesodermal WT1-antigen positive mesothelial precursors (WT1) and terminally differentiated smooth muscle (SM), and mature endodermal hepatocytes (Liv). These derived cell targets were chosen because they provide cells representative of all three germ line lineages and are also varied along an axis of potency-differentiation corresponding to pluripotent (H9 ESC), multipotent (NC and WT1), and terminally differentiated (Liv and SM) (**Figure 2.2a**). The identity and homogeneity of each derived population was confirmed by methods including RNAseq. Transcript levels
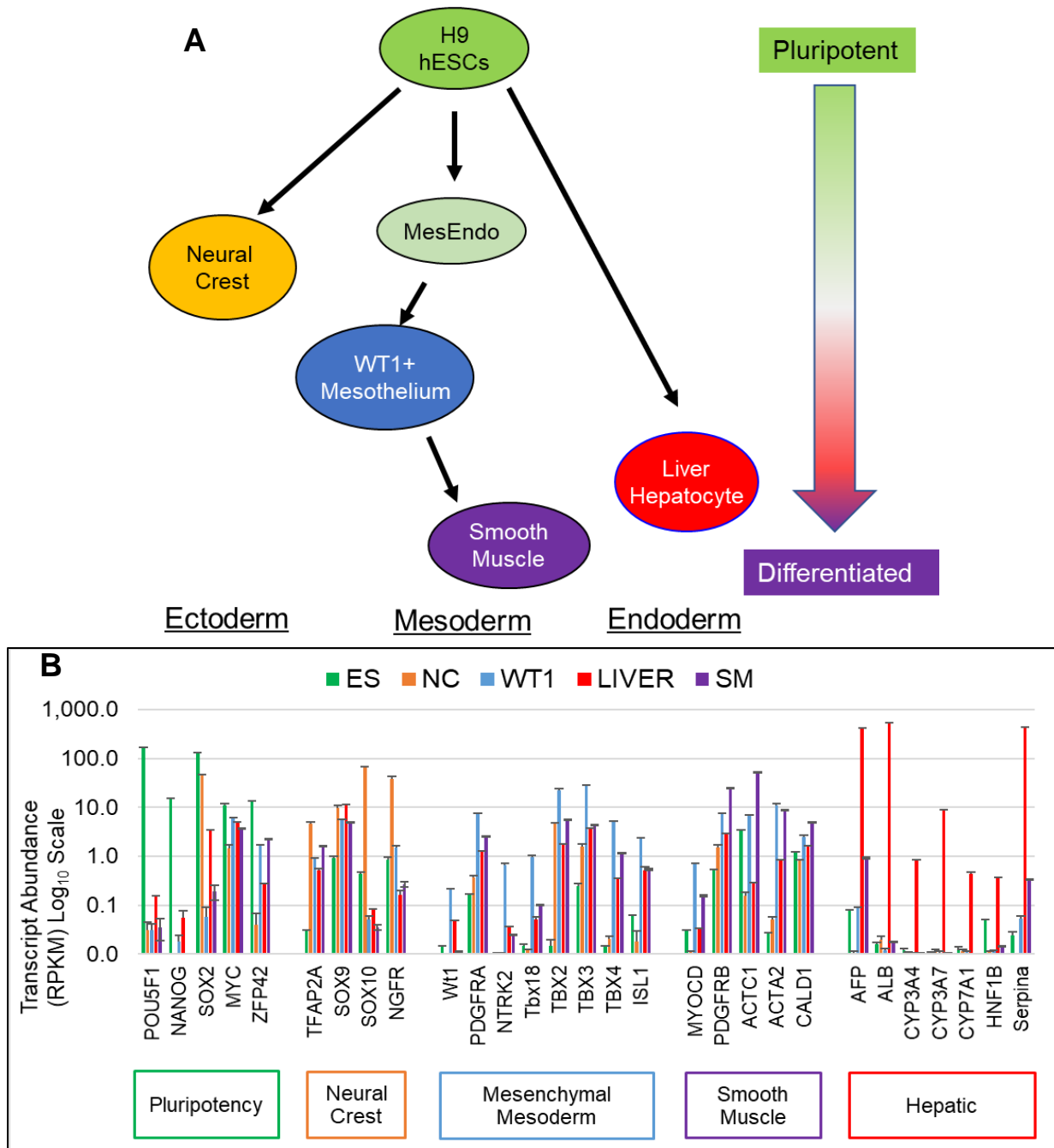
Figure 2.2: Stem Cell Differentiation into Cell lines Representative of Embryonic Development. A) H9 hESCs were differentiated into cell lines representative of the three embryonic germ layers, Ectoderm (NC), Mesoderm (WT1 and SM), and Endoderm (Liv). **B)** RNAseq gene expression confirms the identities of the different cell types with distinct elevations of developmentally regulated genes specific for each cell type. Transcript abundance calculated as reads per kilobase million (RPKM) plotted on a Log_{10} scale. Error bars are SEM, n=2 for each cell type.

of known pluripotency and cell lineage specific genes show restricted expression in the

expected cell types. H9 ESCs have a 1100- to 5400-fold elevation of POU5F1(OCT3/4).

Neural crest stem cells (NC) have 150- to 1800-fold elevation of SOX10. Expression of

WT1 gene is up to 14-fold increased in WT1+ mesenchymal mesoderm, along with fold-

increases up to 44x in other mesodermal progenitor markers such as PDGFRA. SM

smooth muscle cells showed increased expression of smooth, vascular, and cardiac

muscle related genes PDGFRB, ACTC1, and ACTC2 up to 49-, 331-, and 319-fold. Liv

hepatocytes had massive and specific fold-increases in secreted liver protein genes

ALB (28,000 – 43,000x) and AFB (444 - 5250x), as well as liver specific drug

metabolizing CYP enzymes (**Figure 2.2b**). Overall the greatest differences in gene

expression were seen between undifferentiated and differentiated cells and across

lineage boundaries, while mesodermal WT1 and SM clustered closer together and H9

and NC shared more similar expression of neuroectodermal markers such as SOX2.

Broad Characterization of N-linked Glycomic Features Distinguishes Pluripotent and
Differentiated Cells.
        NSI-MSn analysis identified and quantified 179 N-glycan structures

(**Supplemental Table 2.1**). We quantified absolute glycan abundances relative to a

permethylated glycan standard and detected significant and unexpected increases in

glycan amount relative to protein amount as stem cells differentiate. The most

differentiated cell types analyzed (*Liv* and *SM*) exhibited the greatest increase in total

glycan abundance, whereas the cell populations with greater multipotency (*NC* and *WT*)

exhibited lesser increases compared to the *H9* pluripotent stem cells (**Figure 2.3b**). The

cell type with the highest total glycosylation (SM) shows increases in all hybrid and
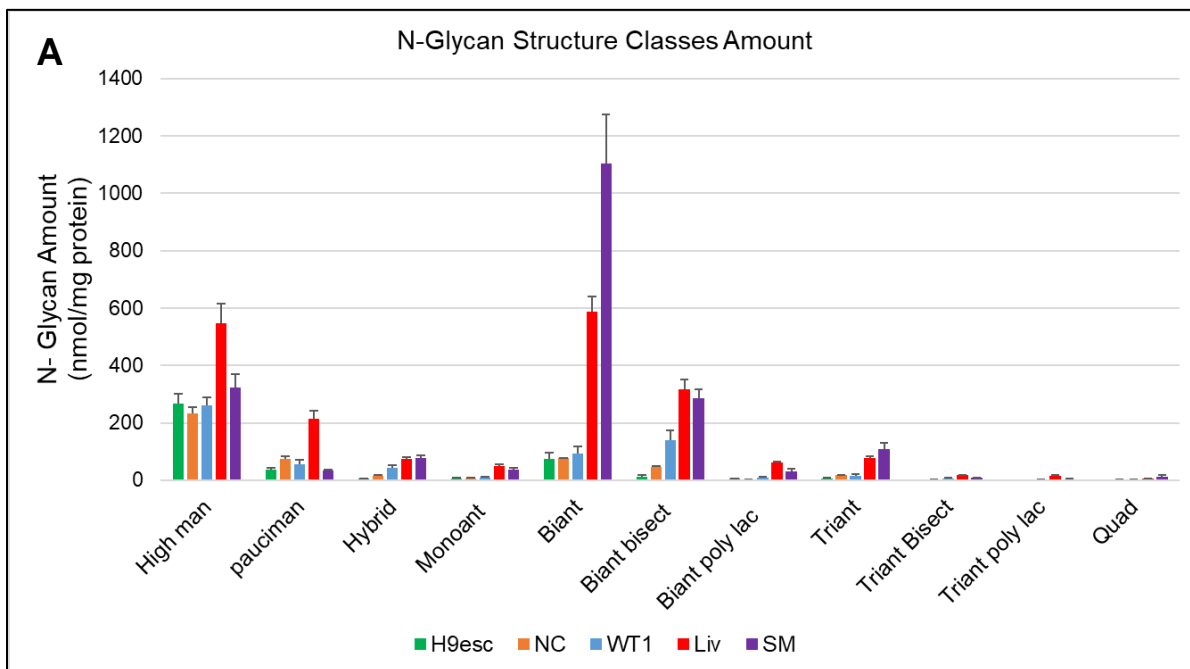
Figure 2.3: Quantified N-Glycan Abundance
A) Glycan abundance of structural classes. **B)** Glycan abundance increases with cellular differentiation. The most differentiated cell types (*Liv, SM*) derived from pluripotent ES cells (*H9*) produce between 3- and 4-fold more glycan per mg protein than less differentiated cells (H9, *NC, WT*). Error bars SEM, n=4
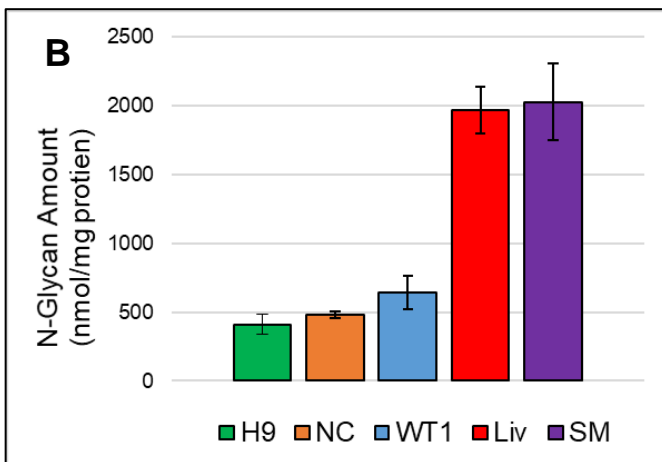
complex structures, indicating that the glycoproteins expressed by the most differentiated cell types are, 1) carrying more N-glycans per protein molecule, and 2) are more efficiently processed to complex type structures as multipotency decreases (**Figure 2.3a**).

Quantifying the relative abundance of structural classes as the percent of the total profile contributed by each class of glycan highlights overlapping and characteristic biosynthetic capacities of pluripotent, multipotent, and differentiated cell types (**Figure 2.4**). The percent total profile parameter, similar to glycan amount, demonstrates that pluripotent (*H9*) and multipotent cell types (*NC, WT*) are enriched in high-Man and less complex glycans, while more differentiated cell types (*Liv, SM*) are enriched in more highly processed Complex-type glycans. Additionally, it reveals nuanced, graded changes in glycan classes, such as the consistently downward trend in the relative abundance of high-mannose glycans as multipotency is lost and the unexpected enrichment in bisected biantennary glycans in WT1 cells.

<u>Pluripotent and Differentiated Cell Types Present Distinct Glycan and Transcript Profiles</u>

Changes in the N-linked glycome and glycogene transcriptome were evaluated by calculating the fold-change in the abundance of individual glycans and transcripts between each of the 4 derived cell lines and the baseline abundance in H9 ES cells. The variance among the cell types and individual replicates was visualized using Hierarchical clustering with these fold-change values for each glycan and gene. Both glycan and transcript abundances independently segregated each of the derived cell types (**Figure 2.5**). Thus, each of the ES-derived cell types presents a signature set of
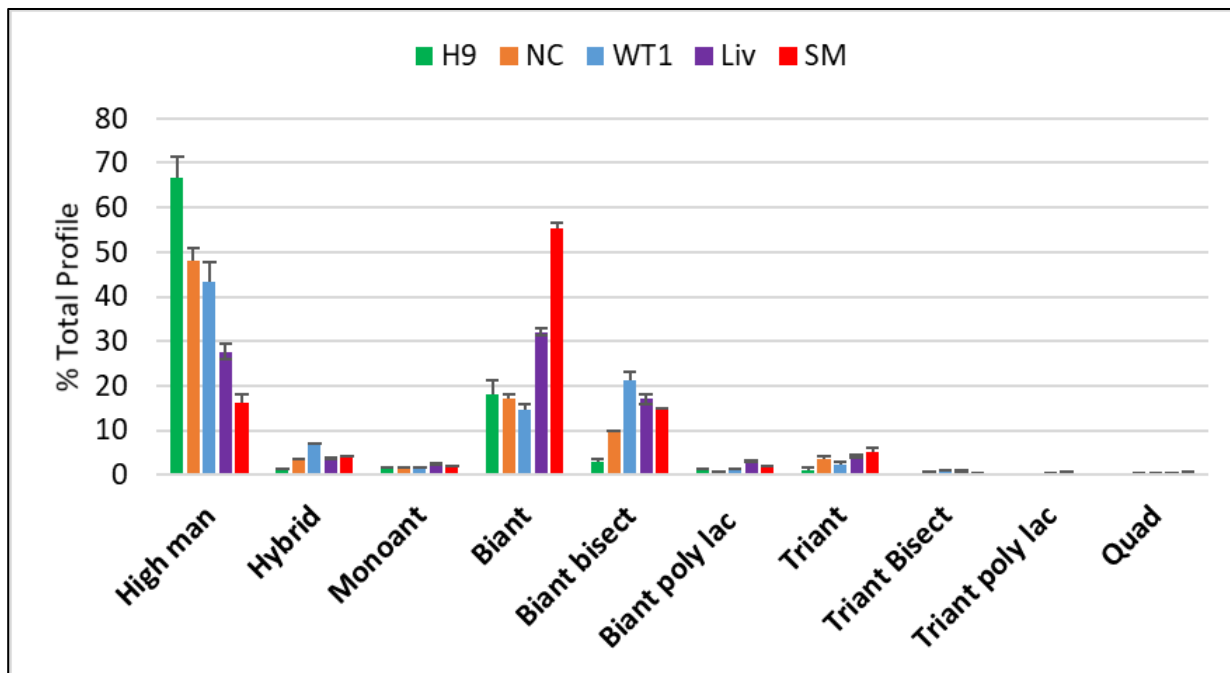
Figure 2.4: Glycan Structure Relative Abundances
Distribution of glycan structural features in stem cells and derived cell lines.
Quantification of the relative abundance of specific glycan structural classes and
features highlights shifts in the biosynthetic capacity of pluripotent, multipotent, and
differentiated cell types. Pluripotent (H9) and multipotent cell types (NC, WT) are
enriched in high-Man and less complex glycans, while more differentiated cell types
(Liv, SM) are enriched in more highly processed glycans. Values for each cell line
presented as average(n=4) percent of total profile, with error bars SEM.
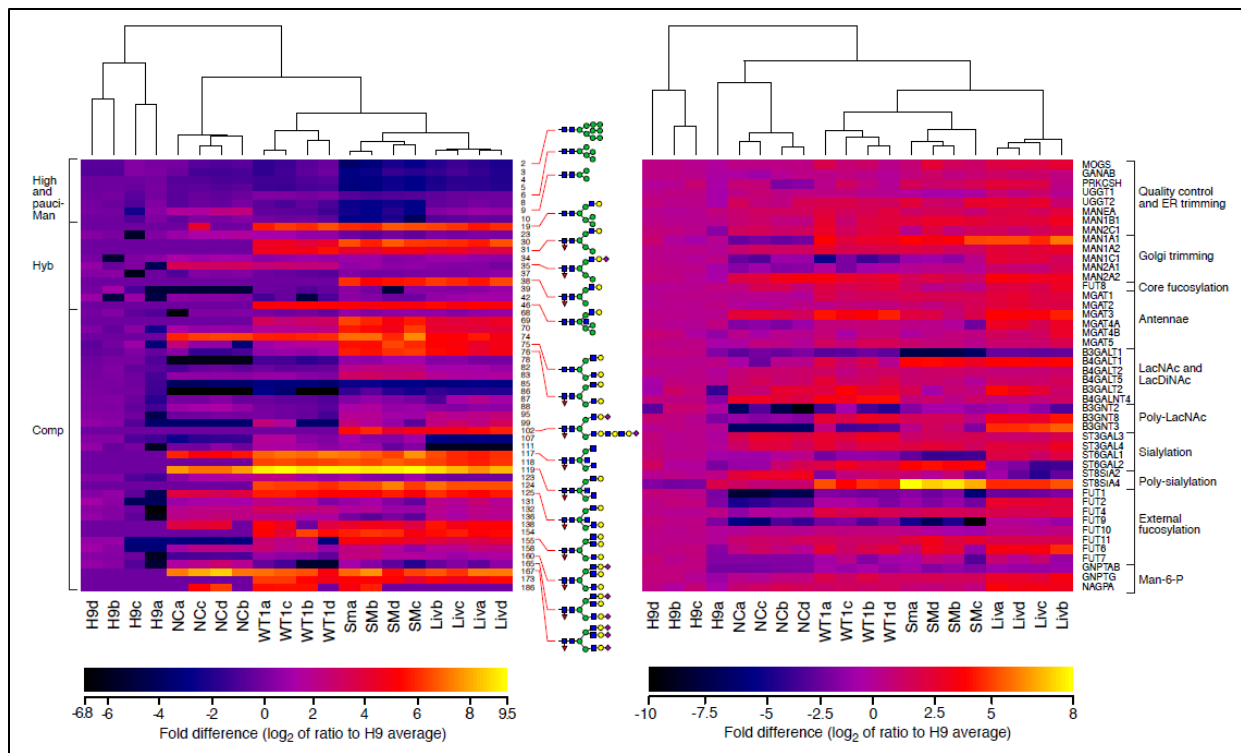
Figure 2.5: Hierarchical clustering heat maps of the fold change relative to the H9 average for the individual replicates of each cell type. **Left)** N-Glycan structures, organized by structure class and with some example structures shown. **Right)** N-linked Glycoprotein biosynthetic enzymes grouped by function. Both N-glycan structures and transcripts show distinct segregation of cell types. Dendrograms also show higher variation in H9 undifferentiated cells compared to the well segregated differentiated types, indicating that differentiated cell types coalesce into distinct structure and biosynthetic patterns.

glycan and transcript abundances. Many of these cell-specific glycan structural features correlate with the expression of relevant biosynthetic genes, as further described below.

<u>Early N-Glycan Trimming and Branching Reflects Degree of Differentiation</u>

ER α-glucosidase-I (MOGS) transcript levels in the Liv and SM cell lines were 2 to 3-fold higher than those in H9esc, NC and WT1 lines (**Figure 2.6b**). These transcript levels correlate inversely with decreased prevalence of high-Man glycan structures as a class in Liv and SM (**Figure 2.4**), as well as with a decrease in the prevalence of the individual $Glc_1Man_9GlcNAc_2$ structure (**Figure 2.6d**). Further trimming of the high-Man structures from $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$ is accomplished by ER and early-Golgi mannosidases. Processing from Man9 to Man8 was increased in differentiated cells as ER MAN1B1 and Golgi MANEA transcripts were increased ~5-fold in SM and Liv. Golgi mannosidases MAN1A1, MAN1A2, and MAN1C1 cleave the remaining mannose residues to the $Man_5GlcNAc_2$ form that becomes the substrate for further elaboration into Hybrid or Complex Structures. Transcript levels for these enzymes were elevated over H9esc levels in all differentiated lines, with fold changes ranging from 1.5 to 33 (**Figure 2.6b,c**).

<u>$Man_5GlcNAc_2$ Branching and Trimming Enzymes Generate Hybrid Structure Diversity Found in Structure Profiles</u>

The addition of β2-linked GlcNAc to the Man(α1-3) arm by MGAT1 produces hybrid-type structures, which can be further trimmed by MAN2A1 and MAN2A2 mannosidases and capped by MGAT2 with β2-linked GlcNAc on the Man(α1-6) arm to produce biantennary complex-type structures. Transcript levels of MGAT1 were elevated in WT1, SM, and Liv over H9, with fold changes of 2.5, 3.0, and 4.3,

Figure 2.6: Early ER and Golgi N-Glycan processing, Trimming and Branching.

A) Biosynthetic schematic of early N-glycan processing. The initial co-translationally added $Glc_3Man_9GlcNAc_2$ undergoes trimming and branching to generate hybrid and complex type structures. Once the high mannose structures are trimmed down to $Man_5GlcNAc_2$ they can be extended into hybrid structures or extended into complex type structures if both arms are trimmed. Further branching with various linkages of GlcNAc. Bisecting β1-4 GlcNAc can be added to either Hybrid or Complex structures and can inhibit further branching or modification (denoted by red asterisks). Black asterisks indicate GlcNAc that may be further elongated as shown in **Figure 2.8**.

B) qRT-PCR expression of glycosylases and glycotransferases responsible for the enzymatic steps denoted with corresponding number in part A. Mean gene expression for each cell type displayed on a log_{10} scale, w/ error bars SEM.

C) Fold Changes in gene expression between each differentiated cell type and the undifferentiated H9. Displayed on a linear scale as a log_2 ratio of the mean transcript abundance of each line over the H9 mean. Error bars SEM, n=4.

D) Relative amounts of the individual high mannose glycan structures detected, Glc_1Man_9 to Man_5 corresponding to steps 2-7 in part A, along with the summed amounts of structural classes generated by further modification of the Man_5 N-glycan core. Average % Total Profile of each structure/ structural class, error bars SEM, n=4.

Figure 2.6
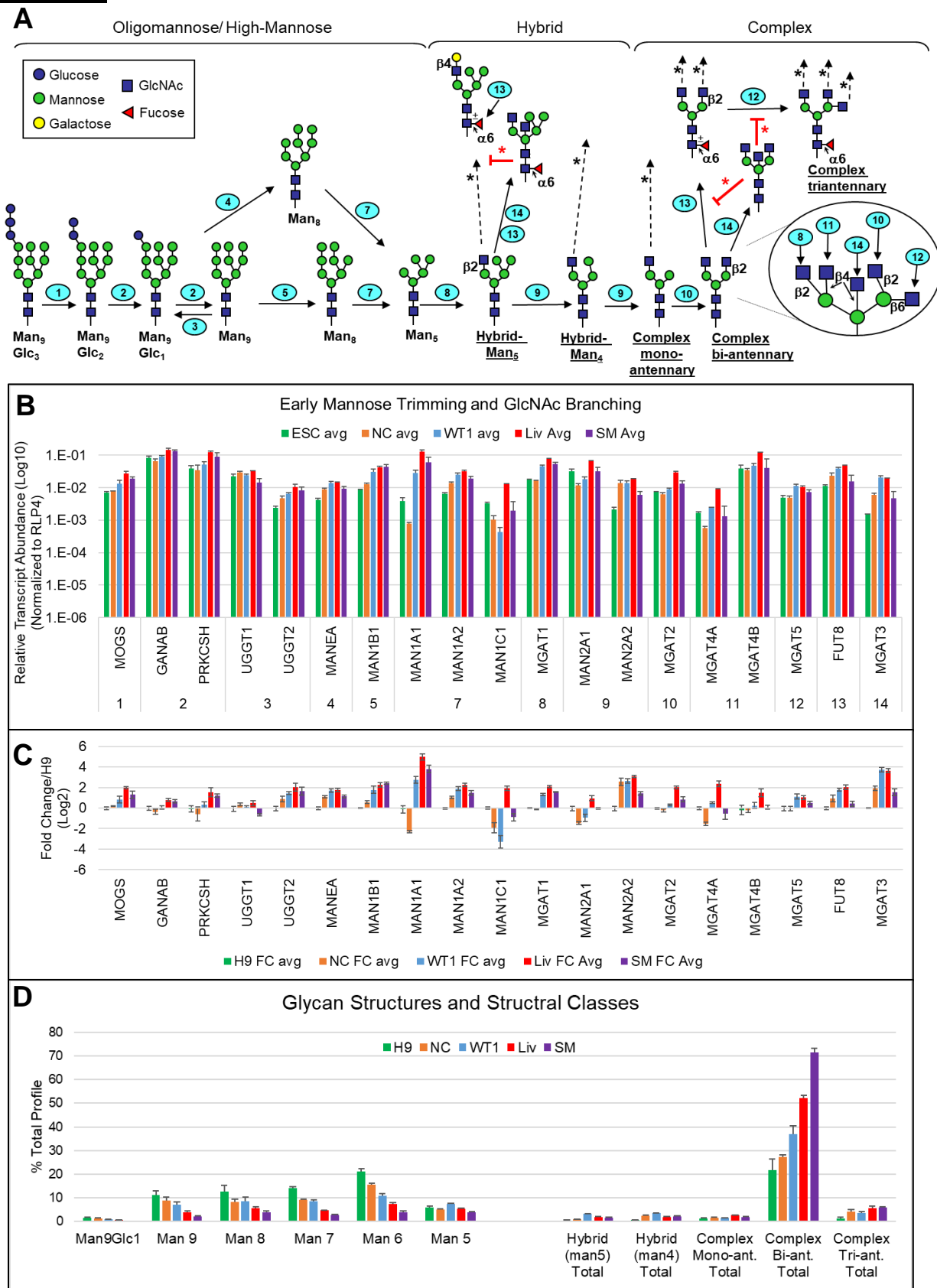
respectively. MGAT2 transcript levels are also elevated in SM (1.9 fold) and Liv (4.1 fold) (**Figure2.6c**). These transcriptional changes are consistent with the glycan profile shifts towards increased complex type structures observed in SM and Liv cells.

While all differentiated cell types had an increased proportion of hybrid structures, WT1 was 2-fold elevated above the rest and 6-fold higher than undifferentiated H9 (**Figure 2.7a**). Flux through the biosynthetic pathway may explain this: with entry into hybrid type synthesis initiated by the addition of the first β2-linked GlcNAc to the 3-arm mannose by MGAT1, mannose trimming by MAN2A1, and exit to complex type by MGAT2 addition of GlcNAc to the now exposed 6-arm mannose. WT1 has the highest ratio of MGAT1 to MGAT2 expression and a low level of MAN2A1 expression (**Figure 2.6b**). While NC also has low MAN2A1 expression it also has lower MGAT1 expression. Further supporting the role of flux through mannose trimming in hybrid structure synthesis, transcription of MAN2A1 also correlates with the relative amounts of Man5-hybrid, Man4-hybrid, and monoantennary complex structures. Cell types with the lowest MAN2A1, NC and WT1, have more untrimmed hybrid than fully trimmed mono-antennary (**Figure 2.7b**).

MGAT3 Expression and Bisecting GlcNAc versus MGAT4 and MGAT5 Branching

Both Hybrid and Complex type structures may also be modified with a "bisecting" GlcNAc added by MGAT3 to the β-linked core mannose. Unique from the other GlcNAc branches, the MGAT3 bisecting branch is not extended and its presence blocks the action of many other glycosyltransferases including MGAT4, MGAT5, FUT8, and α2,3-sialyl transferases [31, 32]. Transcript levels demonstrate MGAT3 is increased in all differentiated lines, 3-4-fold in NC and SM, and up to ~13-fold in WT1 and Liv
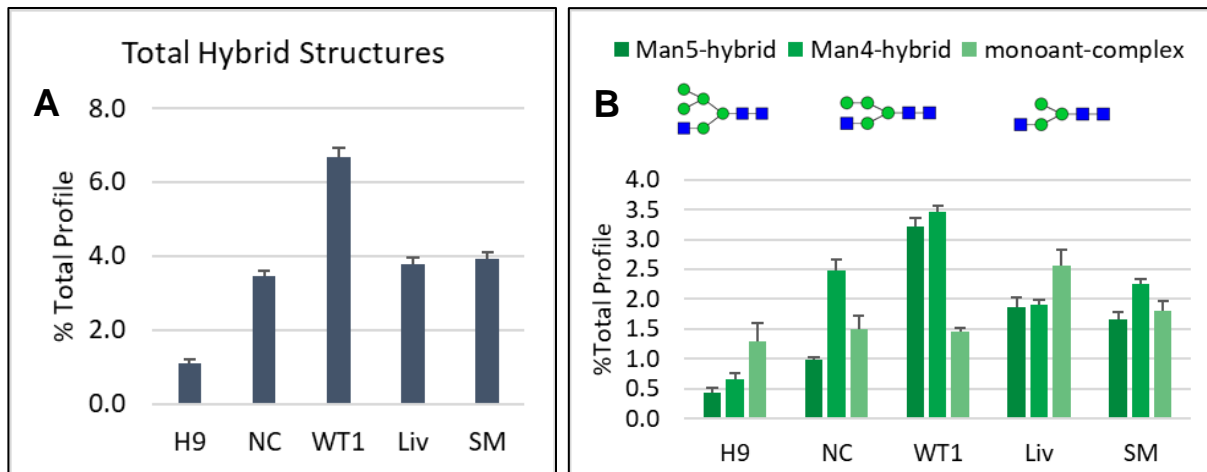
Figure 2.7: A) Total% Profile levels of Hybrid structures. **B)** Relative amounts of hybrid structures based on the core structures: Man5, Man4 and monoantennary complex. The progression from Man5 to mono-complex is mediated by trimming of α2-mannoses by MAN2A1(See Fig.4). Cell types with higher expression of MAN2A1 (H9, Liv, SM) have a higher relative amount of more processed monoantennary-complex, while cells with lower expression (WT1, NC) have higher relative mounts of Man5-hybrid and Man4-hybrid. All graphs average values (n=4), with error bars displaying SEM.

(**Figure2.6**). Consistent with the increase in MGAT3 transcript, WT1 and Liv exhibit the highest relative abundance of bisected glycans (**Figure 2.4**).

All differentiated cell types had increased relative abundance of tri- and tetra-antennary structures, SM with the highest proportion, then Liv, NC, and WT1 (**Figure 2.3**). The MGAT4b isozyme is the more highly expressed form of MGAT4 with PCR transcript levels higher in all differentiated cell types compared to MGAT4a (as well as MGAT5). Both MGAT4a and MGAT4b are elevated 3- to 5-fold in Liv compared to other cell types (**Figure 2.6**). MGAT5 transcript levels are similar in H9, NC, and SM, but are elevated 2-fold in WT1 and Liv (**Figure 2.6**).

Modification with Core Fucose by FUT8

Transcript levels of FUT8 were elevated in differentiated cells, with fold changes of 1.4 (SM), 2.1 (NC), 3.5 (WT1), and 4.2 (Liv) compared to H9 ESC. With the exception of SM, elevated FUT8 transcription correlated with the relative proportion of core-fucosylated, which increased upon differentiation such that H9<NC<WT1<Liv<SM (**Figure 2.8a**). Non-transcriptional regulation may explain this lack of correlation in SM with its high amount of core fucose relative to its lower expression of FUT8. As further explored in the discussion section, the substrate specificity of FUT8 has been well studied, with stronger affinity towards biantennary structures [32]. As SM has the highest relative amount of biantennary structures (see **Figure 2.4**), this may be a likely explanation. Comparing the percent of core-fucosylation present among N-glycan structure classes finds that, across all cell types, biantennary structures and biantennary-bisecting structures are most thoroughly core-fucosylated (**Figure 2.8b**).
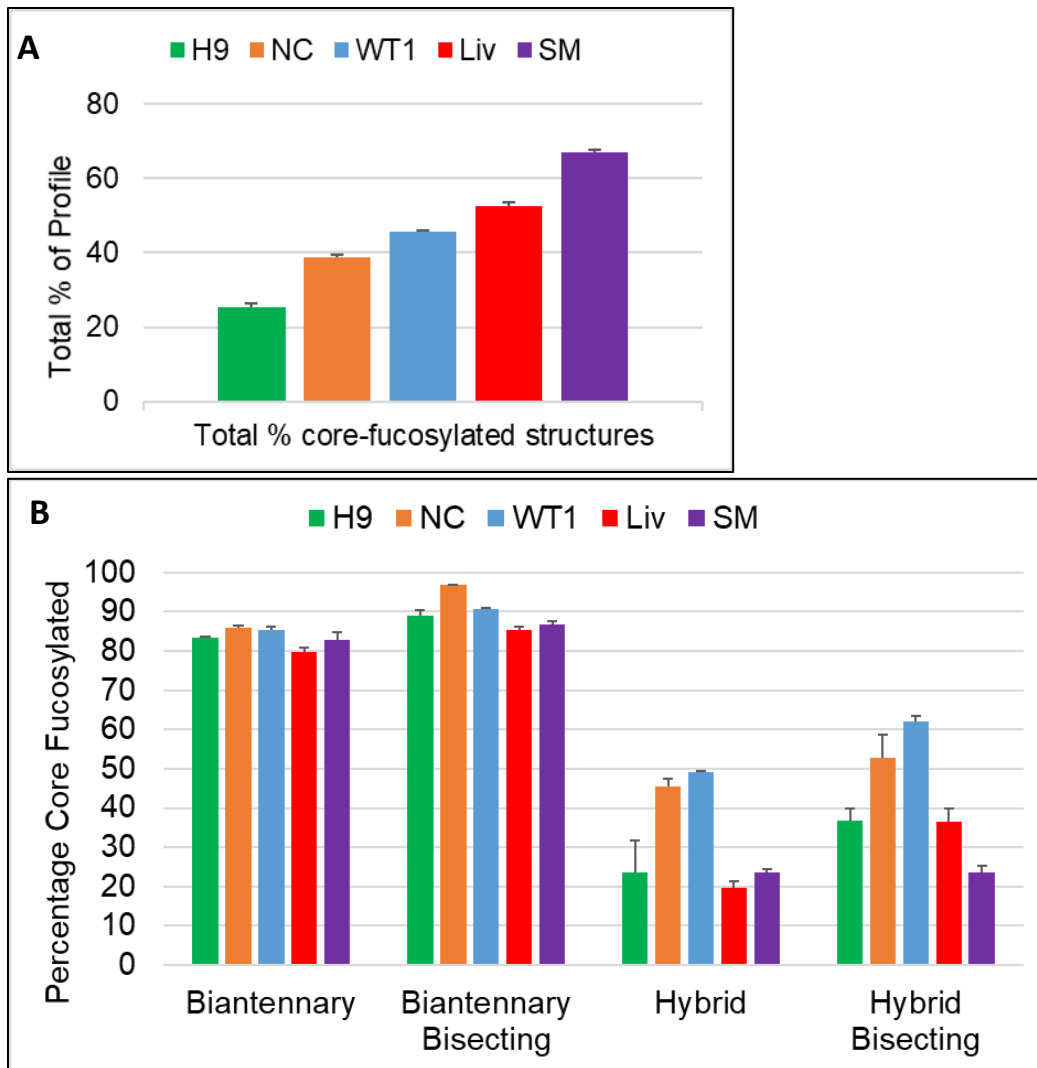
Figure 2.8: **A)** Percentage of core fucosylated structures increases with differentiation. The average sum %Profile made up by core fucose containing structures in the different cell lines, error bars displaying SEM, n=4. **B)** The percentage of core-fucosylated structures within different N-glycan structural classes. Complex biantennary have a high degree of core-fucosylation across all cell types, while hybrid structures had much lower rates of core-fucosylation.

Extension of N-glycan Branches Generates Structure Diversity

Following the addition of a terminal GlcNAc, complex and hybrid N-glycan may be further extended with β3-linked Gal to generate Type-1 lactosamine/LacNAc structures, with β4-Gal to generated Type-2 lactosamine, or by β4-linked GalNAc to create LacDiNAc structures. These branches may then become substrates for elaboration with additional sugars such as fucose or for further extension by poly-lactosamine or sialic acid repeats (**Figure 2.9**).

While the relative amounts of hybrid and complex glycan structures varied between cell types, in all cell types the majority of available terminal GlcNAc were extended, with 92 percent of available GlcNAcs extended in H9, 86% in NC, 77% in WT1, 87% in Liv, and 85% in SM. The majority of these structures were lactosamine branch extensions. We did not interrogate MSn spectra for cross-ring fragments that distinguish between β3- and β4-linked Gal in our routine MS analysis. However, transcripts for β4/Type-2 GalT enzymes were detected to be more highly-expressed in all cell types compared to the β3/Type-1 GalT enzymes. B4GALT1, the main N-glycan type-2 lactosamine extension enzyme, shows a highly significant increase in expression in differentiated cells, up 2.5-fold in WT1, 10.8-fold in SM, and 12-fold in Liv compared to undifferentiated H9 cells (**Figure 2.9c**). While NC cells exhibited a slight decrease in B4GALT1 expression, other tissue expression and KO studies have shown B4GALT2 is the neural β4GalT enzyme [1]. H9escs also exhibited higher B4GALT2 expression than B4GALT1, but B4GALT2 is expressed nearly 2-fold higher in NC and the other differentiated cells. The combined pattern of these two enzymes strongly correlates with

Figure 2.9: N-Glycan Extension and Elaboration Pathways

A) Biosynthetic schematic of extension and elaboration of non-reducing end GlcNAc branches. Elongation of N-glycan antennae begins with addition of LacDiNAc, Type-1 Lactosamine, or Type-2 Lactosamine structures.

B) qRT-PCR expression of glycosylases and glycotransferases responsible for each step in the elongation pathway. Log10 scale, error bars SEM. Some enzymes have substrate specificities limiting them to one pathway, such as the preference of ST6GAL2 for LacDiNAc structures, while other glycosyltransferases such as FUT1 may accept either Type1 or Type2 structures.

C) Fold Changes in gene expression between each differentiated cell type and the undifferentiated H9 ESCs. Displayed on a linear scale as a $\log_2$ ratio of the mean transcript abundance of each line over the H9 mean. Error bars SEM, n=4.

D) The total percent of extended structures in each cell type, designated as N-glycans containing non-reducing end GlcNAc branches with further elaboration.

E) The relative percent of structures with extended GlcNAc branches out of all structures containing extendable GlcNAc branches (i.e., not a bisecting core linked GlcNAc).

F) Percentage of total profile of LacDiNAc structures, along with the distribution of Hybrid versus Biantennary complex structures carrying LacDiNAc motifs. Error bars SEM, n=4. Wt1 cells are most enriched in LAcDiNAc, uniquely so in hybrid forms.

G) Ratio of Biantennary Polylactosamine structures to Triantennary structures, and ratio of Triantennary Polylactosamine structures to Quadrantennary structures.

Figure 2.9   **A**
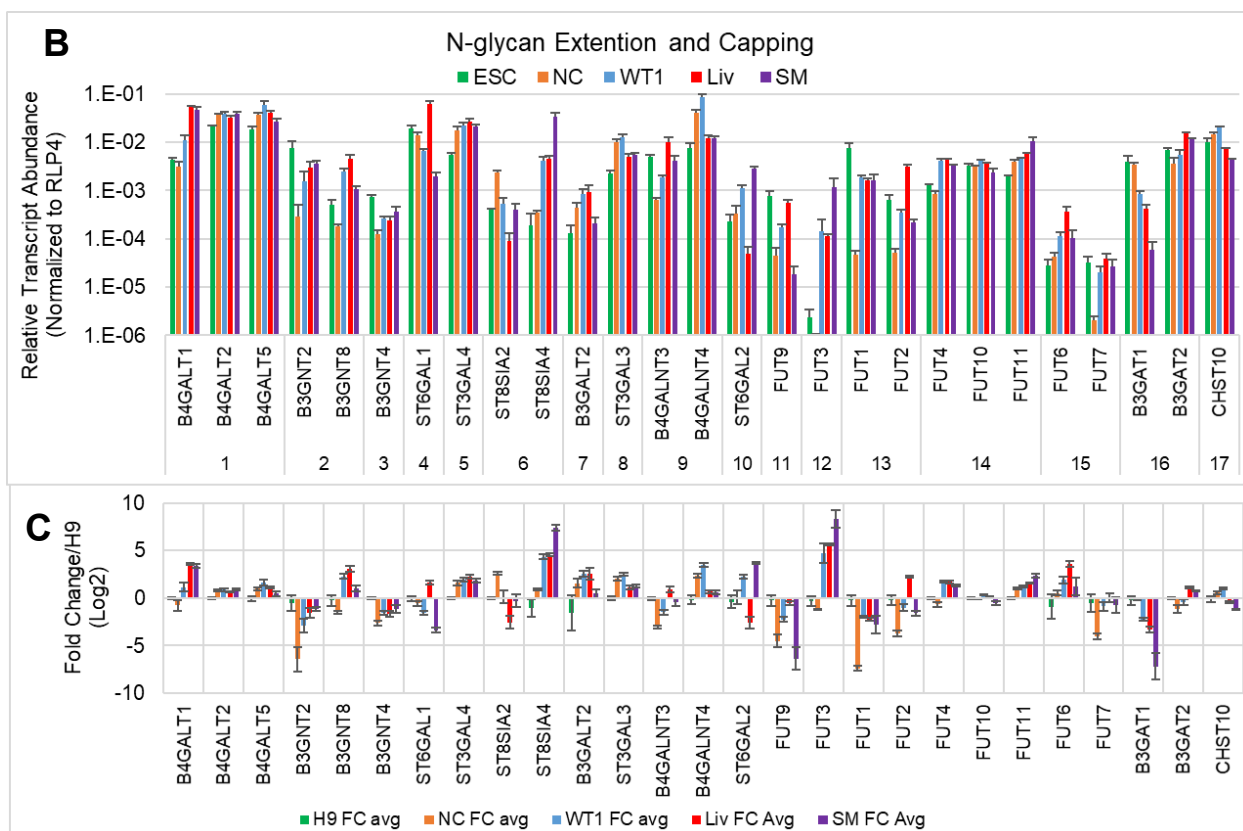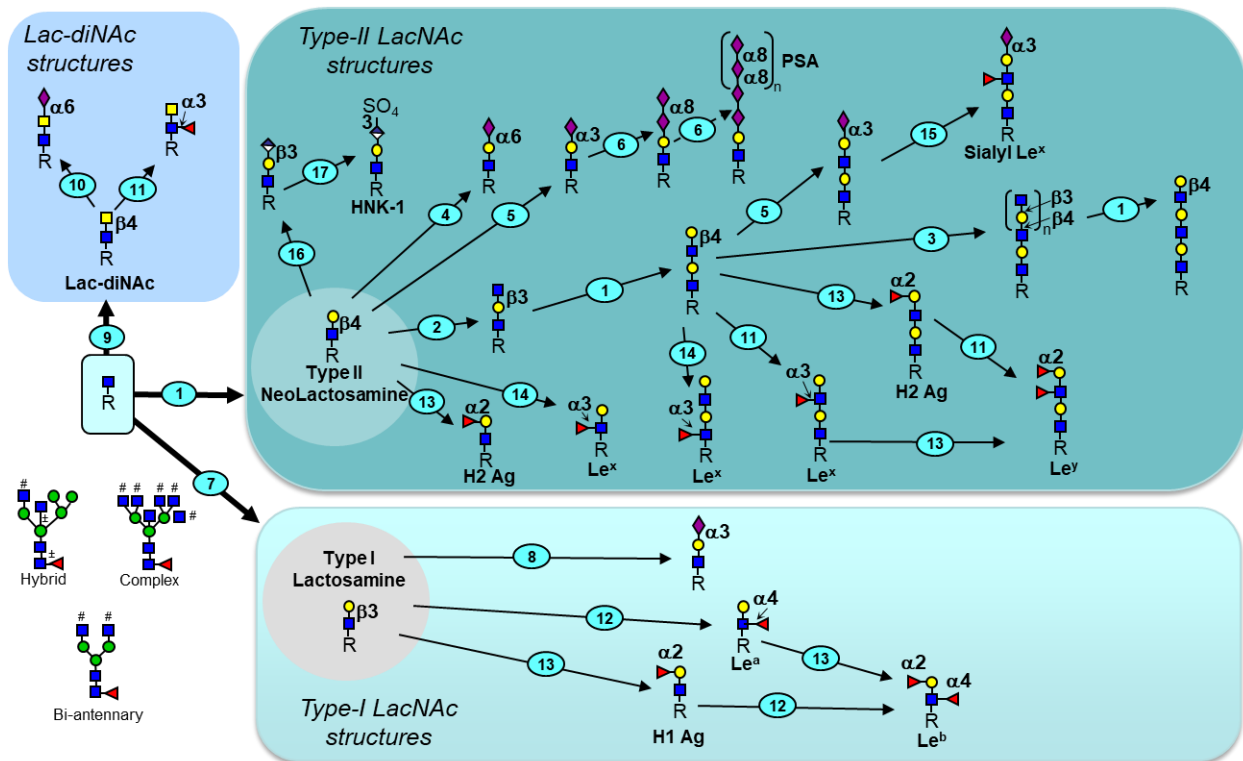
Figure 2.9
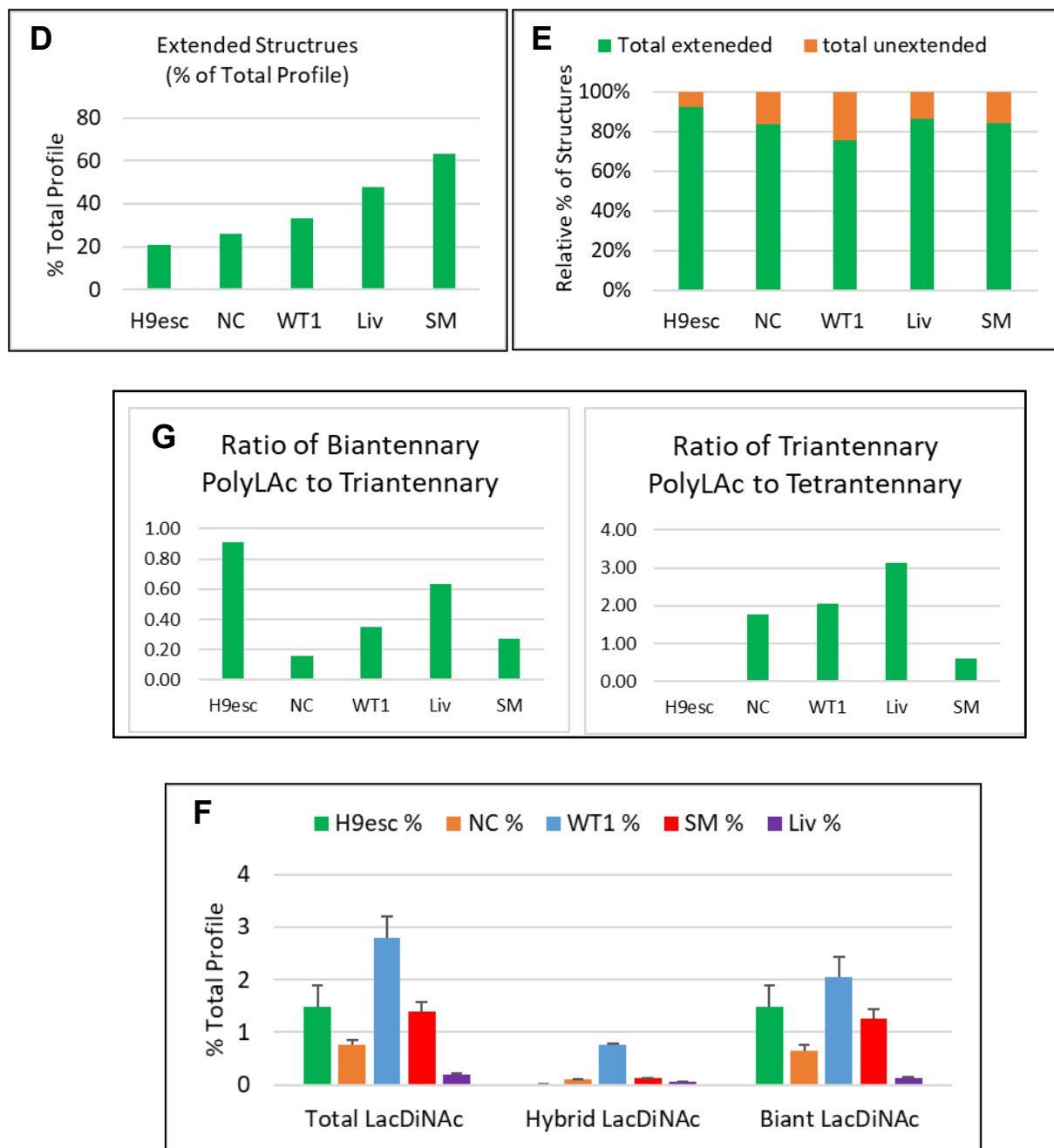
the pattern of total percent of extended hybrid and complex structures across the different cell lines (H9<NC<WT1<Liv<SM) (**Figure 2.9d**).

LacDiNAc structures were found to be present in the glycan profiles, with WT1 having the highest proportion (**Figure 2.9f**). Transcript levels for the LacDiNAc GalNAc transferase B4GalNT4 were overall higher than B4GalNT3, and highly elevated in WT1 and NC. This matches the relative expression of LacDiNAc structures in Wt1 but remains incongruous with the other cell types. B4GalNT3 and B4GalNt4 have been shown to have different tissue distributions, and to have selective specificities for individual protein substrates [33]. Thus, factors such as differential substrate protein expression may explain inconsistencies in enzyme-product correlation.

LacNAc arms can be extended by the iterative action of the B3GNT-family of GlcNAc transferases and B4GALTs. Interestingly, the undifferentiated H9escs showed the highest expression of B3GNT2, at least two-fold over the differentiated cell types, followed by SM. Liv and WT1 expressed the highest levels of B3GNT8 by at 2.5- to 4.5-fold. NC has very little expression (**Figure 2.9b,c**). These transcript changes correlated well to the levels and types of poly-LacNAc detected (**Figure 2.9g**). H9 and Liv cultures had the highest percent profile of biantennary extended structures as would be elongated by B3GNT2. The percent profile of triantennary polylactosamine extended structures, likely elongated via the action of MGAT5 and B3GNT8, was highest in Liv and WT1 cultures (**Figure 2.9g**). The ratio of extended versus branched structures correlated with the expression levels, highest in H9, Liv and WT1, and lowest in NC.

<u>Terminal Capping and modification</u>

Sialic Acid can be α3-linked to Gal by ST3Gal sialyltransferases, α6-linked to Gal by ST6Gal sialyltransferases, α6-linked to GalNAc by ST6GalNAc sialyltransferase, or α8-linked to an underlying sialic acid by St8sia polysialic acid (PSA) transferases. The prominent α3 enzymes for N-linked structures are ST3Gal3 which prefers Type-1 glycans, ST3Gal4 which prefers Type-2 N linked structures, and ST3Gal6 which can modify Type-2 N-linked structures but prefers the Galβ1-4GlcNAc-R structures of glycosphingolipids [34]. ST3Gal3 was elevated in differentiated cells, most markedly in NC and Wt1. ST3Gal4 expression was elevated in all differentiated cells, 3-4x over undifferentiated H9esc. ST3Gal6 expression was lower overall but highest in SM, NC and Liv (**Figure 2.9b**). α6-sialyltransferase ST6Gal1 expression is highest in Liv cultures, 3-fold over H9, though H9 has the next highest expression. ST1Gal1 levels decrease in NC and WT1 and are lowest in SM. H9esc cultures have high levels of ST6GAL1, higher than their expression of ST3Gal genes, indicating that stem cell sialic acid is predominantly α6-linked. Comparing the relative expression of the ST3Gal4 to ST6Gal1 enzymes in each cell type indicates that NC would have increased sialyltransferase expression overall, compared to H9, with a mix of α3 and α6-linked sialic acid, WT1 and SM would be enriched in α3, and Liv which has the highest expression overall would be enriched in α6-linked.

N-glycan poly-sialyl transferases St8sia2(STX) and ST8Sia4(PST) are important developmentally regulated genes, with ST8Sia2 upregulated in ectoderm and ST8Sia4 upregulated in endoderm and mesoderm [35]. Unique to NC, ST8Sia2 is upregulated 6

fold, and ST8Sia4 is upregulated most highly in SM with a 177-fold increase over H9, as well as WT1 (21x) and Liv (23x) (**Figure 2.9c**).

External fucosylation, like sialylation, plays a key role in regulating glycan interactions. Fucosyltransferases, grouped as α2 (FUT1, FUT2) or α3/4 types (FUT3-7,9-11), add fucose to the external Gal, GlcNAc, and sometimes GalNAc residues of N-linked, O-linked, and GSL glycans. External fucosylation can combine with sialylation to form sialyl-Lewis type structures, or it can compete with sialylation as a mutually exclusive modification [36, 37]. External fucose containing structures such as LeX, LeY and H-type2 were identified in the glycan profiles, though no Sialyl LeX or Sialyl LeA. H9 cells were distinctly enriched in external fucose structures, consistent with previous reports that externally fucosylated structures such as SSEA-5 and are markers of undifferentiated stem cells. MSn analysis also indicated that H9 external fucose residues are predominantly H-Type (Fucα2Gal), Lewix X (Galβ4(Fucα3)GlcNAc) and difucosylated LeA (Fucα3Galβ4(Fucα3)GlcNAc). Alpha1,2-Fucosyltransferase FUT1 is highly elevated in undifferentiated H9esc cells, nearly 4-fold over WT1, SM, and Liv, and is negligibly expressed in NC. FUT2 is also expressed in H9 but has the highest expression in Liv (**Figure2.9**). Of Lewis X alpha1,3-Fucosyltransferases, FUT4, FUT10, and FUT11 had the highest overall expression, though FUT9 has the most efficient LeX synthesis activity [38]. Further corresponding to the high proportion of external fucose structures in H9 cells, FUT9 is most elevated in H9, followed by Liv and WT1 (**Figure 2.9b**). Sialyl-LeX generating alpha1,3-Fucosyltransferases FUT6 and FUT7 were the lowest expressed, concomitant with a lack of Sialyl-LeX structures identified.

Discussion:

N-Glycans Have Cell Specific Roles in Development and Tissue Functions

N-linked glycans are attached to cell surface receptors and secreted proteins where they play important roles in cell signaling, adhesion, and migration; processes that are important for development[39-41]. Developmental signaling can be modulated by the transcriptionally regulated N-glycosylation of growth factor receptors. Differential N-glycan branching, mediated by MGAT5 and MGAt4 expression, interacts with carbohydrate binding lectins such as galectin-3 to traffic growth factor receptors to the cell surface, stabilize their presence on the cell membrane and regulate activation by dimerization [39, 42]. Cell adhesion and migration mediated by extracellular binding proteins integrins and cadherins is regulated by the competitive expression of MGAT3 and GMAT5 and their mutually exclusive modifications of bisecting or branching N-Glycan structures. The functional importance of N-glycans is further made clear by the severe consequences of loss of function of early-acting N-linked synthesis genes. Animal KO models are often lethal before birth and human congenital disorders of glycosylation result in severe developmental disorders [1, 41, 43, 44].

Cell surface N-glycan structures also serve as identifiers cell identity, as evidenced by the existence of various endogenous and exogenous glycan binding proteins such as a variety of Lectins with specific binding affinities for different glycan structures[45]. Glycan binding lectins of the innate immune system and antibodies from adaptive immune system are a key mechanism of the body's surveillance of Self versus non-Self [46]. Given these roles, biologists have taken advantage of these glycan binding proteins and their target glycan epitopes for use as markers of cell identity.

Many important developmental markers, such as the Stage Specific Embryonic Antigens (SSEAs), CD15, and HNK1, used to identify cancer cell states and stem cell pluripotency, are glycan based [47]. The development of regenerative medicine over the last decade due to hESC and iPSC technology has led to increased demands for stringent characterization of these cells to access their pluripotency, differentiation potential, and tumorigenicity [48, 49]. Unrecognized or out of place Xeno-glycan epitopes such as Gal-(1,3)-Gal or Neu5Gc sialic acid, pathogen associated mannose antigens, and ABO blood group antigens cause immune responses that can lead to graft rejection[50].

Given these known roles for N-glycans of cell function and identity, recent studies have attempted to profile the N-glycomes of human pluripotent stem cells [6-8, 51-53] to identify stem cell specific glycan traits or markers. Others have also sought to characterize differentiated cell types generated from stem cells, such as cardiomyocytes[54, 55] and hepatocytes[56]. This previous literature as well as previous studies from our lab have demonstrated that specific tissues[14] and cell types[16] have distinct profiles of glycan structures. Despite the important roles N-glcyan structures play, a comprehensive understanding of the biosynthetic control of developmental changes in protein glycosylation is currently lacking [40].

The New Developments of this Study

The control of N-glycan biosynthesis driving this diversity is likely the product of a complex system involving the expression, localization, and function of glycosyltransferases; the substrate specificities and trafficking of glycoproteins through the Golgi; and the availability of sugar nucleotide donors[57]. Many of these factors are

yet unknown, so a full and accurate model of N-glycan biosynthesis requires comprehensive information on these components. This current study advances this understanding N-glycan biosynthesis by utilizing comprehensive transcriptomics combined with methods for unprecedented breadth of N-glycan structure identification, to examine the transcriptional regulation of glycosylation in multiple distinct cell types from all three somatic germ layer lineages and across multiple stages of differentiation.

Using transcriptomics methods developed by collaborators in the Moreman lab, RNAseq data allowed for broad analysis of both direct effectors of N-glycan biosynthesis such as glycosyltransferase expression, but also ancillary genes involved in sugar precursor metabolism, synthesis and transport. Using expert knowledge of glycan biosynthetic pathways[58] and technical developments in quantitative RT-PCR[16, 17], we were able to target and quantify a large number of key glycosyltransferase responsible for N-glycan synthesis. As seen in **Figure 2.5**, these transcript profiles demonstrated distinct transcriptional differences that clearly segregated the cell types analyzed. By taking a broad transcriptomics approach in parallel to the glycomic analysis, rather than one following it, we were able to make unexpected findings such as the high expression of poly-sialylation enzyme ST8SIA4 unique to smooth muscle (SM).

By combining multiple advances in MS analytical glycomics, our novel Direct Infusion (DI) MS-MS^n method allowed us to identify and quantify N-glycan structures utilizing a single method, increasing assignment confidence and generating a broader profile (170+ structures) than previously achieved. The use of an external standard to quantify all glycan m/z peaks in a single scan minimized scan-scan variability, as might

be seen in chromatographic separation methods such as LC-MS. The DI-MS limitation

of overlapping isometric structures was overcome utilizing MS^n serial fragmentation to

interrogate the presence, and relative ratios, of unique diagnostic fragment ions for

overlapping structures. Other groups, such as the Reinhold Lab, have used Ion Trap

CID fragmentation to MS^n levels deeper than 2-3, to show with known standards and

biological samples that consistent fragmentation patterns and ratios can be used to

distinguish relative measures of isotopic mixtures[59-61]. By utilizing sensitive orbitrap

and ion trap analyzers we were able reach fragmentation levels of MS^7 and MS^8. Thus,

while this current "one-pot" method was still limited in its ability to consistently determine

exact linkages (e.g. α2,3- vs α2,6- sialic acid), it was able to resolve structural

differences that may have remained opaque to previous studies. For example, the

placement of terminal HexNAc as either an unextended branching GlcNAc antennae, a

core bisecting GlcNAc, or a terminal GalNAc of a LacDiNAc arm (**Figure 2.1**). We were

also able to distinguish the placement of LacNAc groups as either additional branches

or as extended poly-lactosamine arms, finding a previously undescribed relative

enrichment of extended biantennary structures in stem cells (H9) (**Figure 2.9g**) that

positively correlated with the level biantennary specific LacNAc extension enzyme

B3GNT2 (**Figure 2.9b**).

Previous studies examined the profiles of stem cells alone[52, 62], stem cells

differentiated into non-specific mixed populations such as EBs[8, 16], one or two cell

types[6, 7, 51, 54-56], or many unrelated samples[53]. For a summary comparison of

their findings see **Table 2.2** and further discussion below. In this study, we examine

multiple ESC differentiated cell types across different developmental lineages and

timepoints. This allows us to examine the regulation of glycosylation through development and by examining diverse cell types at once we better understand which glycome shifts are truly cell specific and which are part of more general developmental changes seen across multiple lineages.

<u>Comparisons to the Previous Literature</u>

One of the first comprehensive N-glycan profilings of hESCs was by (Satomaa et al 2009)[8]. They compared hESC to embryoid body(EB) based differentiations and found an abundance of high mannose structures, complex externally fucosylated LeX and H-type-2 structures, and complex structures with combinations of terminal sialyl groups, external fucose and extra LacNAc. Our data agrees with these findings and would conclude based on our qRT-PCR data that the sialic acid would be α2,6-linked, and based on our structural data that the extra LacNAc groups would be extended poly-lactosamine arms. Other groups confirmed the high abundance of externally fucosylated structures, though limitations in different techniques left some structures unresolved. Fujitani et al. used MALDI-TOF and found externally fucosylated, triantennary/poly-lactosamine, and bisecting/LacDiNAc structures [53]. Our H9 hESC data agrees and would confirm the latter structures as LacDiNAc, of which a few externally fucosylated LacDiNAc structures were found to be unique to H9 cells in our data.

Table 2.2: Comparison of Previous Stem Cell Glycomics Studies.

Study	Stem Cell Type	Method	Stem Cell Associated Findings
Satomaa et al., 2009 Ref.[8]	•ESCs vs •Embryoid Body (EB) differentiation	• MALDI-TOF MS, • NMR, • Glycosidase digestion, • Lectin binding	• # N-glycans identified: 40 • High Mannose structures most abundant. • $(Hex)_5(HexNAc)_4(Fuc)_2$ • $(Neu5Ac)_1(Hex)_5(HexNAc)_4(Fuc)_2$ • $(Neu5Ac)_1(Hex)_5(HexNAc)_4(Fuc)_3$ • Complex external fucosylated structures, isomeric mixture ~50:50 Le^X (α1,3-GlcNAc) and H-type2 (α1,2-Gal). • All LacNAc units type2.
Tateno et al., 2011 Ref.[6]	•Somatic cells •iPSCs •ESCs	• Lectin Array • Microarray gene expression	• # N-glycans identified: N/A • Increased (α2,6)-sialic acid • Increased (α1,2)-fucose. • Decreased (α1,6)-fucose. • Increased Type-1 LacNAc with decreased Type-2 LacNAc. • Decreased bisecting GlcNAc • Decreased tetra-antennary N-glycans • Decreased High Mannose N-glycans.
Hasehira et al., 2012 Ref.[7]	•Fibroblast somatic cells •iPSCs	• HPLC • MALDI-TOF MS • Glycosidase digestion	• # N-glycans identified: 44 After FB→ iPSC reprogramming: • High Mannose structures most abundant and increased. • Reduced Complex type structures. • Increased exposed terminal GlcNAc. • Increased (α2,6)-sialic acid. • Increased (α1,2)-fucose. • Decreased (α1,6) core fucose • Increased Type-1 LacNAc
An et al., 2012 Ref.**[51]**	•Fibroblast somatic cells •ESCs •Breast cancer cell line	• Cell membrane enrichment • FT-ICR MS • nanoLC • Lectin binding	• # N-glycans identified: 39 masses quantified, and 136 structures identified • High Mannose structures most abundant. (74-85%) »Man8 and Man9 most abundant.
Fujitani et al., 2013 Ref.**[53]**	•ESCs, iPSCs, and a variety of human cancer, fetal, and somatic cell lines	• MALDI-TOF MS	• # N-glycans identified: 93 • High Mannose structures most abundant in all cell types, though highest in pluripotent cells. • Reduced Pauci-mannose glycans. • Hex6HexNAc5Fuc2NeuAc1 • multiply fucosylated type (N-30, -80, -65) » (Hex)1 (HexNAc)2 (Fuc)2 + (Man)3(GlcNAc)2 » (Hex)4 (HexNAc)4 (Fuc)1 (NeuAc)3 + (Man)3(GlcNAc)2 »(Hex)2 (HexNAc)2 (Fuc)2 (NeuAc)1 + (Man)3(GlcNAc)2

			• neutral triantennary type (N-41), »(Hex)3 (HexNAc)3 + (Man)3(GlcNAc)2 • bisect and/or LacdiNAc type (N-22, N-42) » (HexNAc)3 + (Man)3(GlcNAc)2 » (Hex)1(HexNAc)4 (Fuc)1 + (Man)3(GlcNAc)2
Thiesler et al., 2016 Ref.**[62]**	• ESCs • iPSCs • iPSCs generated from PMM2-CDG patient.	• N-glycans detected by (xCGE-LIF): multiplexed Capillary Gel Electrophoresis with Laser Induced Fluorescence detection. • Lectin binding. • Glycosidase digestion	• # N-glycans identified: 61 » 22 quantified. • High Mannose structures most abundant, Man9-Man5: (~70%) • Hybrid and Complex low abundance. • Exclusively (α2,6)-linked sialic acid.
Stadlmann et al., 2017 Ref.**[52]**	• Mouse mESCs • Human hESCs	• nLC-ESI-MS/MS Glycoproteomics	• # N-glycans identified: • High Mannose structures most abundant (~70%). • Complex type (~30%). • Complex type mainly terminated with sialic acid. »2059 Da: »2205 Da: »2350 Da: • Identified glycan carrying proteins(236). • LAMP1 carrier of Le^X external fucose biantenttary complex glycans.
Kawamura et al., 2015 Ref.**[54]**	• iPSc • Cardio-myogenic differentiation. • 1° Human cardiomyocytes (hCMCs)	• HPLC and MALDI-TOF MS • RT-PCR •	• # N-glycans identified: 52 isolated, 38 identifed. • High Mannose structures were predominant in iPSCs (~70%), but no change in differentiated cells. • Increased external fucose. • Increased terminal GlcNAc.
Montacir et al., 2017 Ref.**[56]**	• ESCs • Definitive Endoderm (DE) • Hepatocyte-like cells (HCLs)	• Cell surface N-glycans collected by trypsination. • MALDI-TOF MS • Glycosidase digestion •	• # N-glycans identified: • High Mannose structures most abundant. H9N2--H3N2 (~39%) • Complex type (~11%) • Hybrid type (~3%) •
Konze et al., 2017 Ref.**[55]**	• iPSCs • Cardio-vascular progenitors d7. • Cardiomyocytes (CM) d15.	• N-glycans detected by (xCGE-LIF) • Glycosidase digestion	• High Mannose structures most abundant (~50%), but no significant difference between iPSCs and differentiated CM. • Exclusively (α2,6)-linked sialic acid. • Presence of (β1,3)-linked terminal Gal, Type1 LacNAc. • Mono-sialylated and externally fucosylated complex structures decreased upon differentiation. »SA1Gal2Fuc1N2M3N2F

Hasehira et al.[7] and Tateno[6] also agreed upon α2,6-linked sialic acid, and external fucose, but indicated that the major LacNAc form was Type1.

While our MS methods were not able to reliably differentiate Type1 and Type2 lactosamine, our transcriptomic data would disagree. As these latter two studies were conducted using lectin arrays, it is possible that these Type1 structures were glycosphingolipid based (as will be discussed in Chapter 3) or O-linked. In general, our N-glycan profiling of pluripotent stem cells agrees with previous work but expands on the number of structures identified and transcripts examined, including the confirmation of extended lactosamine structures.

Glycome profiling of differentiated cell types has also sought cell specific markers. Pluripotent stem cells differentiated into cardiomyocyte-like [55, 63] and hepatic cell types[56] have been examine for N-glycan profile changes. Cardiomyocyte-like cells compared to hiPSCs after a 15 day differentiation had N-glycan profile decreases in high mannose structures, α2,6-linked sialic acid, and external fucose, with increased biantennary, bisecting and α2,3-linked sialic acid structures[55]. In a study by Moerkamp, et al., glycan binding antibodies were utilized to isolate what they described as cardiac progenitor cells and upon differentiation of these found increased glycotransferase gene expression of ST3GAL6 and mannosidases[63]. In both cases, the results mirrored the general trends in our data of changes due to increasing differentiation, exemplified by SM cell glycan and transcriptomic profiles. In fact in the latter study[63], the differentiated cells designated as "cardiomyocytes" may have been more akin to our smooth muscle(SM) cells. These differentiated cells developed tubular

aSMA-positive structures reminiscent of vascular smooth muscle and the antibody used to isolate the "cardiac progenitor cells" (CPCs) also colocalized with WT1+ cells.

Montcacir et al. differentiated hESCs through DE to "hepatocyte like cells HLCs" over 16 days and examined the top 30 cell surface N-glycan structures by MALDI-TOF. Similar to our Liv cell findings, they found a decrease in high mannose structures in their more differentiated HLCs, with the two highest abundance structures being core-fucosylated biantennary complex with zero or one sialic acid (of compositions H5N4F1 and S1H5N4F1) [56]. While the above cases claimed to have characterized cell specific characteristics, our analysis of multiple differentiated cell types found that these trends of decreasing high mannose and predominant biantennary complex structures were not cell-specific trends, but rather general features of highly differentiated cells. Our findings caution against the use of one or even two cell types for the purposes of discovering unique glycan structure expression and underscore the need for a broad examination of glycosylation in many diverse cell types and developmental stages.

Summary of the comparative characteristics of ES and derived cell lines

Comparative glycomics and glycogene transcriptional analysis of the ES and multiple differentiated cell types presented here revealed signatures for pluripotency and for differentiation toward specific cell identifies. Some changes are shared across cell types and some are specific to a particular type (**Table 2.3**).

Key Findings and Directions for Further Study

By combining glycomic and transcriptomic analysis we sought to investigate the contribution of transcriptional control in glycan biosynthetic pathway. For example, as seen by An et al.[51] and other studies of stem cell differentiation, as stem cells

Table 2.3: Cell Line Specific Glycan Profile Features and Transcript Correlations

Line	Lineage	Glycan Profile Features	Transcript Correlation
H9 ESC	Pluripotent	•Predominantly High Mannose(HM), highest percent total profile of HM. (~66%) •High percentage of external fucose. Predominantly of the a1,2-Gal H-type 2 form. •Unique Structures: externally fucosylated structures	•Low Golgi mannosidases MAN1A1, MAN1A2, and MAN1C1. •High (α1,2) Fucosyltransferase FUT1 and (α1,3) FUT9
Neural Crest Stem Cell (NC)	Ectoderm	•High % Paucimannose (PM) •Low external fucose •High Sialylation •Unique structures: Sialyl-sialyl and disialylayted antennae	•High HEXDC GlcNAcase. •Low expression of FUT genes •ST3Gal6 •ST8SIA2 •
Mesothelium (WT1)	Mesoderm	•High Hybrid structures •High percentage of Bisecting structures. •LacDiNAc •Unique Structures: LacDiNAc hybrid structrues.	•MAN2A1 vs MGAT2 levels •MGAT3 •B4GALNT4
Smooth Muscle (SM)	Mesoderm	•Lowest HM •Highest Complex •Overwhelmingly Biantennary	•High levels of mannosidases and MGAT1, MGAT2
Hepatocytes (Liv)	Endoderm	•High HM and Complex type •Paucimannose (by absolute amount) •Poly-Lactosamine, especially Tri-ant	•High overall transcript levels •Increased catabolic enzymes. •Elevated B3GNT8 Lysosomal Man-6-P High MGAT4a, GLUT2 RNAseq

differentiate their predominantly high mannose N-glycan profiles switch to increasing

percentages of hybrid and complex structures. This pattern was also seen in our data

(**Figure 2.4**), and some straight forward gene-glycan correlations may explain certain

aspects, such as increases in early mannose trimming enzymes MANEA, MAN1B1,

MAN1A1, MAN1A2, and MAN1C1 (**Figure 2.6**) resulting in less high mannose and

more Man5GlcNAc2 for processing into hybrid or complex structures. Yet other more

complex relationships between flux through multiple enzymes was evident, as seen in

the relative amount of hybrid structures and the interplay of mannose trimming enzyme

MAN2A1 and the GlcNAc transferases MGAT1 and MGAT2 (**Figure 2.7**).

The relationship between MAN2A1, MGAT1/MGAT2, and hybrid structure levels

was most apparent in the WT1 cell type. While no glycomic analysis of mesothelial

progenitor cells could be found in the literature, there are striking resemblances

between the presently analyzed WT1 mesothelial mesenchyme and tumor cells

undergoing epithelial-mesenchymal transition (EMT). In hepatocellular carcinoma, N-

glycan profiles were shifted to increased hybrid structures and decreased tri- and tetra-

antennary structures [64], and in bladder epithelial tumor cells undergoing EMT a similar

shift was seen accompanied by reduced expression of MAN2A1 [65]. Thus, the

accumulated evidence points to hybrid structure abundance being controlled by the

availability of these three enzymes at the transcriptional level.

Other structure-transcript relationships may be impacted by post transcriptional

factors such as enzyme substrate specificity. FUT8 transcript levels correlated with total

percent of core-fucosylated structures in the differentiated cell types, except for SM.

While SM had the highest percentage of core-fucosylated structures, it had relatively

low FUT8 expression compared to the other differentiated cell type. Calderon et al., demonstrated that FUT8 has the highest substrate affinity for biantennary glycans without terminal sialylation or fucosylation [32]. Thus, in the SM cells which have the highest percentage of biantennary structures un-capped by sialic acid or fucose, the high amount of core-fucosylation may be regulated by enzyme-substrate affinity.

Golgi trafficking of developing glycoproteins and the localization of glycosyltransferases to certain pathways or compartments is another factor affecting the processing of glycan structures [40, 66, 67]. Similar to findings from our previous studies with mouse ESCs [16], the prevalence of core-fucose varied by glycan structural class. In this study core-fucosylation was significantly lowest in Hybrid structures (**Figure 2.8**). Also, while MGAT3 addition of bisecting GlcNac has been show to block FUT8 action[68], complex bisecting structures had the highest percentage of core-fucosylation—around 90% in all cell types. This structure specific prevalence indicates that FUT8 mediated core-fucosylation can also be mediated by position or path taken through the Golgi processing system. Thus, it appears that hybrid structure synthesis occurs earlier in Golgi processing stream than FUT8, while MGAT3 acts later.

Taken together, our N-glycan structure profile and biosynthetic pathway transcriptomic data have expanded the scope of identified N-glycan structures on a previously examined cell type, undifferentiated pluripotent stem cells (H9), and broadened the array of human cell types profiled. By examining multiple cell types from divergent developmental lineages and stages, we were able to learn that some "cell specific" glycan structures are shared across diverse linages as well identify structure-transcript trends that correlate across levels of differentiation—indicating consistent

transcriptional control across development. While we were able to identify some cell type specific structures and were certainly able to discriminate cell types based on total profiles, many aspects of both the structural compositions and biosynthetic regulation remain opaque. Further studies utilizing high throughput analytical glycan characterization capable of distinguishing specific glyosidic bond linkages are necessary to confidently identify all structures and make sense of differences in linkage-varying glycotransferases.

Literature Cited:

1. Stanley, P., *What Have We Learned from Glycosyltransferase Knockouts in Mice?* Journal of Molecular Biology, 2016. **428**(16): p. 3166-3182.

2. Scott, H. and V.M. Panin, *N-glycosylation in Regulation of the Nervous System.* Advances in neurobiology, 2014. **9**: p. 367-394.

3. Takahashi, K., et al., *Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors.* Cell, 2007. **131**(5): p. 861-872.

4. Ardhanareeswaran, K., G. Coppola, and F. Vaccarino, *The Use of Stem Cells to Study Autism Spectrum Disorder.* The Yale Journal of Biology and Medicine, 2015. **88**(1): p. 5-16.

5. P., B.R., et al., *Glycosylation and stem cells: Regulatory roles and application of iPSCs in the study of glycosylation-related disorders.* BioEssays, 2016. **38**(12): p. 1255-1265.

6. Tateno, H., et al., *Glycome Diagnosis of Human Induced Pluripotent Stem Cells Using Lectin Microarray.* Journal of Biological Chemistry, 2011. **286**(23): p. 20345-20353.

7. Hasehira, K., et al., *Structural and Quantitative Evidence for Dynamic Glycome Shift on Production of Induced Pluripotent Stem Cells.* Molecular & Cellular Proteomics, 2012. **11**(12): p. 1913-1923.

8. Satomaa, T., et al., *The N-glycome of human embryonic stem cells.* BMC Cell Biology, 2009. **10**(1): p. 42.

9.   Avery, J. and S. Dalton, *Methods for Derivation of Multipotent Neural Crest Cells Derived from Human Pluripotent Stem Cells*, in *Embryonic Stem Cell Protocols*, K. Turksen, Editor. 2016, Springer New York: New York, NY. p. 197-208.

10. Menendez, L., et al., *Directed differentiation of human pluripotent cells to neural crest stem cells.* Nature Protocols, 2013. **8**: p. 203.

11. Chambers, S.M., et al., *Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling.* Nature biotechnology, 2009. **27**(3): p. 275-280.

12. Rinkevich, Y., et al., *Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells & fibroblasts for mammalian internal organs, and their vasculature*. Vol. 14. 2012.

13. Ogawa, S., et al., *Three-dimensional culture and cAMP signaling promote the maturation of human pluripotent stem cell-derived hepatocytes.* Development, 2013. **140**(15): p. 3285-3296.

14. Nairn, A.V., et al., *Regulation of Glycan Structures in Animal Tissues: TRANSCRIPT PROFILING OF GLYCAN-RELATED GENES.* Journal of Biological Chemistry, 2008. **283**(25): p. 17298-17313.

15. Nairn, A.V., M. dela Rosa, and K.W. Moremen, *Chapter Four - Transcript Analysis of Stem Cells*, in *Methods in Enzymology*, M. Fukuda, Editor. 2010, Academic Press. p. 73-91.

16. Nairn, A.V., et al., *Regulation of Glycan Structures in Murine Embryonic Stem Cells: COMBINED TRANSCRIPT PROFILING OF GLYCAN-RELATED GENES AND GLYCAN STRUCTURAL ANALYSIS.* Journal of Biological Chemistry, 2012. **287**(45): p. 37835-37856.

17. Nairn, A.V., M.d. Rosa, and K.W. Moremen, *Transcript Analysis of Stem Cells.* Methods in enzymology, 2010. **479**: p. 73-91.

18. Xing, Y., et al., *An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs.* Nucleic acids research, 2006. **34**(10): p. 3150-3160.

19. Livak, K.J. and T.D. Schmittgen, *Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method.* Methods, 2001. **25**(4): p. 402-408.

20. Aoki, K., et al., *Dynamic Developmental Elaboration of N-Linked Glycan Complexity in the Drosophila melanogaster Embryo.* Journal of Biological Chemistry, 2007. **282**(12): p. 9127-9142.

21. Mehta, N., et al., *Mass Spectrometric Quantification of N-Linked Glycans by Reference to Exogenous Standards.* Journal of Proteome Research, 2016. **15**(9): p. 2969-2980.

22. Anumula, K.R. and P.B. Taylor, *A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates.* Analytical Biochemistry, 1992. **203**(1): p. 101-108.

23. Nix, D.B., et al., *Improved In-gel Reductive β-Elimination for Comprehensive O-linked and Sulfo-glycomics by Mass Spectrometry.* Journal of visualized experiments : JoVE, 2014(93): p. e51840-e51840.

24. Damerell, D., et al., *Annotation of Glycomics MS and MS/MS Spectra Using the GlycoWorkbench Software Tool*, in *Glycoinformatics*, T. Lütteke and M. Frank, Editors. 2015, Springer New York: New York, NY. p. 3-15.

25. Ceroni, A., et al., *GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans.* Journal of Proteome Research, 2008. **7**(4): p. 1650-1659.

26. R., K.S., et al., *The glycolyzer: Automated glycan annotation software for high performance mass spectrometry and its application to ovarian cancer glycan biomarker discovery.* PROTEOMICS, 2012. **12**(15-16): p. 2523-2538.

27. Allam, H., et al., *Glycomic Analysis of Membrane Glycoproteins with Bisecting Glycosylation from Ovarian Cancer Tissues Reveals Novel Structures and Functions.* Journal of Proteome Research, 2015. **14**(1): p. 434-446.

28. Domon, B. and C.E. Costello, *A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates.* Glycoconjugate Journal, 1988. **5**(4): p. 397-409.

29. Ferguson, M.W., K.L. Wycoff, and A.R. Ayers, *Use of cluster analysis with monoclonal antibodies for taxonomic differentiation of phytopathogenic fungi and for screening and clustering antibodies.* Current Microbiology, 1988. **17**(3): p. 127-132.

30. Pattathil, S., et al., *A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies.* Plant Physiol, 2010. **153**(2): p. 514-25.

31. Kizuka, Y. and N. Taniguchi, *Enzymes for N-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer.* Biomolecules, 2016. **6**(2): p. 25.

32. Calderon, A.D., et al., *Substrate specificity of FUT8 and chemoenzymatic synthesis of core-fucosylated asymmetric N-glycans.* Organic & Biomolecular Chemistry, 2016. **14**(17): p. 4027-4031.

33. Baenziger, J.U., *Beta1,4-N-Acetylgalactosaminyltransferase-3 (B4GALNT3) and Beta1,4-N-Acetylgalactosaminyltransferase-4 (B4GALNT4)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 429-437.

34. Chung, C.-Y., et al., *Assessment of the coordinated role of ST3GAL3, ST3GAL4 and ST3GAL6 on the α2,3 sialylation linkage of mammalian glycoproteins.* Biochemical and Biophysical Research Communications, 2015. **463**(3): p. 211-215.

35. Berger, R.P., et al., *ST8SIA4 dependent polysialylation is part of a developmental program required for germ layer formation from human pluripotent stem cells.* Stem cells (Dayton, Ohio), 2016. **34**(7): p. 1742-1752.

36. Grabenhorst, E., et al., *In Vivo Specificity of Human α1,3/4-Fucosyltransferases III-VII in the Biosynthesis of LewisX and Sialyl LewisX Motifs on Complex-type N-Glycans: COEXPRESSION STUDIES FROM BHK-21 CELLS TOGETHER WITH HUMAN β-TRACE PROTEIN.* Journal of Biological Chemistry, 1998. **273**(47): p. 30985-30994.

37. Brito, C., et al., *Human fucosyltransferase IX: Specificity towards N-linked glycoproteins and relevance of the cytoplasmic domain in intra-Golgi localization.* Biochimie, 2008. **90**(9): p. 1279-1290.

38. Mika, K., et al., *α1,3-Fucoslytransferase IX (Fuc-TIX) is very highly conserved between human and mouse; molecular cloning, characterization and tissue distribution of human Fuc-TIX.* FEBS Letters, 1999. **452**(3): p. 237-242.

39. Dennis, J.W., et al., *Adaptive regulation at the cell surface by N-glycosylation.* Traffic, 2009. **10**(11): p. 1569-78.

40. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nature Reviews Molecular Cell Biology, 2012. **13**: p. 448.

41. Haltiwanger, R.S. and J.B. Lowe, *Role of Glycosylation in Development.* Annual Review of Biochemistry, 2004. **73**(1): p. 491-537.

42. Vaiana, C.A., T. Kurcon, and L.K. Mahal, *MicroRNA-424 Predicts a Role for β-1,4 Branched Glycosylation in Cell Cycle Progression.* Journal of Biological Chemistry, 2016. **291**(3): p. 1529-1537.

43.   Freeze HH, B.L., Varki A. , *Glycans in Systemic Physiology*, in *Essentials of Glycobiology* C.R. Varki A, Esko JD, et al., editors., Editor. 2017, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.: [Internet].

44.   Freeze HH, S.H., Kinoshita T., *Genetic Disorders of Glycosylation*, in *Essentials of Glycobiology* C.R. Varki A, Esko JD, et al., editors., Editor. 2017, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.: [Internet].

45.   Taylor, M.E., et al., *Discovery and Classification of Glycan-Binding Proteins*, in *Essentials of Glycobiology*, rd, et al., Editors. 2015, Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017.: Cold Spring Harbor (NY). p. 361-372.

46.   Maverakis, E., et al., *Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: A critical review.* Journal of Autoimmunity, 2015. **57**: p. 1-13.

47.   Furukawa, J.-i., K. Okada, and Y. Shinohara, *Glycomics of human embryonic stem cells and human induced pluripotent stem cells.* Glycoconjugate Journal, 2016. **33**(5): p. 707-715.

48.   The International Stem Cell, I., *Characterization of human embryonic stem cell lines by the International Stem Cell Initiative.* Nature Biotechnology, 2007. **25**: p. 803.

49.   Jun, H., et al., *A Novel Probe as Surface Glycan Marker of Pluripotent Stem Cells: Research Outcomes and Application to Regenerative Medicine.* Advanced Healthcare Materials, 2015. **4**(16): p. 2520-2529.

50.   Cobb, B.A., *Is donor glycosylation the gatekeeper for xenotransplantation?* Am J Transplant, 2014. **14**(4): p. 745-7.

51.   An, H.J., et al., *Extensive Determination of Glycan Heterogeneity Reveals an Unusual Abundance of High Mannose Glycans in Enriched Plasma Membranes of Human Embryonic Stem Cells.* Molecular & Cellular Proteomics, 2012. **11**(4).

52.   Stadlmann, J., et al., *Comparative glycoproteomics of stem cells identifies new players in ricin toxicity.* Nature, 2017. **549**: p. 538.

53.   Fujitani, N., et al., *Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers.* Proceedings of the National Academy of Sciences, 2013. **110**(6): p. 2105-2110.

54.   Kawamura, T., et al., *Structural Changes in N-Glycans on Induced Pluripotent Stem Cells Differentiating Toward Cardiomyocytes.* Stem Cells Transl Med, 2015. **4**(11): p. 1258-64.
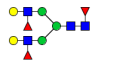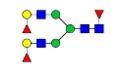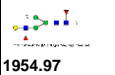
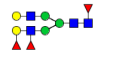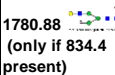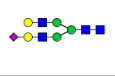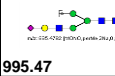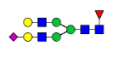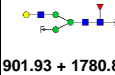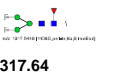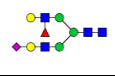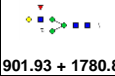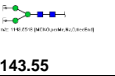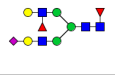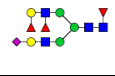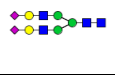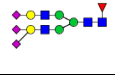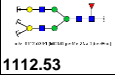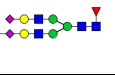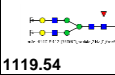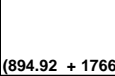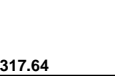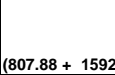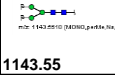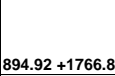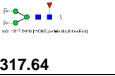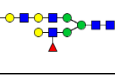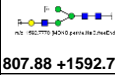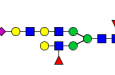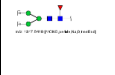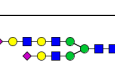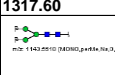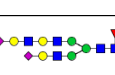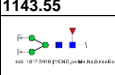55.   Konze, S.A., et al., *Quantitative Assessment of Sialo-Glycoproteins and N-Glycans during Cardiomyogenic Differentiation of Human Induced Pluripotent Stem Cells.* Chembiochem, 2017. **18**(13): p. 1317-1331.

56.   Montacir, H., et al., *The Cell-Surface N-Glycome of Human Embryonic Stem Cells and Differentiated Hepatic Cells thereof.* Chembiochem, 2017. **18**(13): p. 1234-1241.

57.   Neelamegham, S. and L.K. Mahal, *Multi-level regulation of cellular glycosylation: from genes to transcript to enzyme to structure.* Current Opinion in Structural Biology, 2016. **40**: p. 145-152.

58.   Nairn, A. and K. Moremen, *Chapter 5 - Glycotranscriptomics A2 - Cummings, Richard D*, in *Handbook of Glycomics*, J.M. Pierce, Editor. 2010, Academic Press: San Diego. p. 95-135.

59.   Ashline, D.J., et al., *Structural Documentation of Glycan Epitopes: Sequential Mass Spectrometry and Spectral Matching.* Journal of The American Society for Mass Spectrometry, 2014. **25**(3): p. 444-453.

60.   Ashline, D.J., et al., *Structural Characterization by Multistage Mass Spectrometry (MSn) of Human Milk Glycans Recognized by Human Rotaviruses.* Molecular & Cellular Proteomics, 2014. **13**(11): p. 2961-2974.

61.   Ashline, D.J., H. Zhang, and V.N. Reinhold, *Isomeric complexity of glycosylation documented by MSn.* Analytical and Bioanalytical Chemistry, 2017. **409**(2): p. 439-451.

62.   Thiesler, C.T., et al., *Glycomic Characterization of Induced Pluripotent Stem Cells Derived from a Patient Suffering from Phosphomannomutase 2 Congenital Disorder of Glycosylation (PMM2-CDG).* Molecular & Cellular Proteomics, 2016. **15**(4): p. 1435-1452.

63.   Moerkamp, A.T., et al., *Glycosylated Cell Surface Markers for the Isolation of Human Cardiac Progenitors.* Stem Cells Dev, 2017. **26**(21): p. 1552-1565.

64.   Goldman, R., et al., *Detection of Hepatocellular Carcinoma Using Glycomic Analysis.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2009. **15**(5): p. 1808-1813.

65.   Guo, J., et al., *Alteration of N-glycans and Expression of Their Related Glycogenes in the Epithelial-Mesenchymal Transition of HCV29 Bladder Epithelial Cells.* Molecules, 2014. **19**(12): p. 20073.
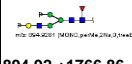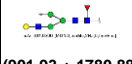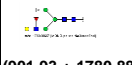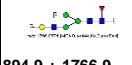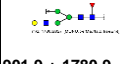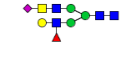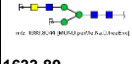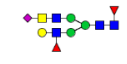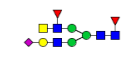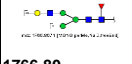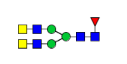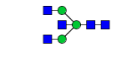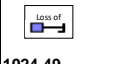
66.   Tu, L. and D.K. Banfield, *Localization of Golgi-resident glycosyltransferases.* Cellular and Molecular Life Sciences, 2010. **67**(1): p. 29-41.

67.   Banfield, D.K., *Mechanisms of Protein Retention in the Golgi.* Cold Spring Harbor Perspectives in Biology, 2011. **3**(8).

68.   Longmore, G.D. and H. Schachter, *Product-identification and substrate-specificity studies of the GDP-l-fucose: 2-acetamido-2-deoxy-β-d-glucoside (fuc→asn-linked GlcNAc) 6-α-l-fucosyltransferase in a golgi-rich fraction from porcine liver.* Carbohydrate Research, 1982. **100**(1): p. 365-392.

Supplemental Figures for Chapter 2:

Final order	Structure	Alt. Structure	Single Form or Isome	steps	1st step	1st Dx fragment	2nd Step	2nd Dx fragment	3nd Step	3nd Dx fragment
1			single	0						
2			single	0						
3			single	0						
4			single	0						
5			single	0						
6			single	0						
7			single	0						
8			single	0						
9			single	0						
10			single	0						
11			single	0						
12			single	0						
13			iso1/2	1		**794.37 + 1565.76**				
14			iso1/3	1		**(881.42+ 1739.85)**				
15			iso1/4	1	MS2	**794.37 + 1565.76**				
16			iso1/6	1	Ms2	**881.42 + 1739.85**				
17			iso4	1	MS2	**881.4 + 1739.85**				
18			iso4	1	Ms2 (1126) >MS3	**881.42 + 1739.85**				
19			single	0						

20			iso2	1	MS2(916.94)> MS3	692.32 + 1361.66				
21			iso1/2	1		(881.42+ 1739.85)				
22			iso2/2	1		(794.37 + 1565.76)				
23			single	0						
24				0						
25			iso1/2	1	ms2	692.32 + 1361.6				
26			iso1/3	1		779.37 + 1535.75				
27			iso1/4	1	MS2	692.32 + 1361.66				
28			iso1/6	1	MS2	(779.37 + 1535.75)				
29			iso3	1	MS2(1024)> MS3:	779.37 + 1535.75				
30			iso2/2	1		692.32 + 1361.66				
31			iso2/3	1		779.37 + 1535.75				
32			iso3/3	1		692.32 + 1361.66				
33			single	0						
34			single	0						
35			single	0						
36			single	0						
37			single	0						
38			iso2/2	1	ms2	590.27 + 1157.56				
39			iso2/3	1		677.32 + 1331.65				

#			iso	n	MS method	value 1	MS3 method	value 2		
40			iso3/3	1		**590.27 + 1157.56**				
41			single	0						
42			single	0						
43			single	0						
44			iso2/4	2	MS2	Loss of **916.94 + 1810.89**	ms3 917	**787.3 + 1551.7**		
45			iso2/6	2	Ms2	Loss of **1003.90**	Ms3(1003.9) >MS4	**874.4 + 1725.8 vs (1521.7)**		
46			iso2	1	MS2(916.94)>MS3	**685.32 + 1347.65**				
47			single	0						
48				0						
49			single	0						
50			iso2/4	2	MS2	**814.89 + 1606.79**	MS3(814.9)> MS4:	**685.32 + 1347.65**		
51			iso2/6	2	MS2	**(901.93 + 1780.88)**	MS3 901.9	**(772.36 + 1521.74)**		
52			iso3/4	2	MS2	**(814.9 + 1606.8)**	ms3 (1606.8)	**685.32 + 1347.65**		
53			iso3/6	2	Ms2	Loss of **901.93 + 1780.88**	Ms3(902.03) >MS4	**772.3 + 1521.7**		
54			iso4/6	2	Ms2	Loss of **1003.90**	Ms3(1003.9) >MS4	**685.3 + 1347.6 vs (1521.7)**		
55			iso4	2	MS2	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1521.74**		
56			iso1/2	1	Ms2	**1097.52**				
57			iso4	2	Ms2 (1126) >MS3	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1521.74**		
58			iso3/3	1	MS2(937.45)> MS3(807.88)> MS4:	**1333.63**				
59			iso1/2	1		**444.00**				

#	Structure	Structure 2	iso	n	MS	Value 1	MS col	Value 2	MS col	Value 3	Final
60			iso1/2	1	Ms2(799.88) > Ms3(670.31) > Ms4(1317.64) > Ms5(866.39) > Ms6(662) > Ms7	**444.20**					
61			iso3/4	3	MS2	**814.89 + 1606.79**	MS3 (814.9)> MS4:	**583.27 + 1143.55**	MS3(1143.5) >MSn		**444.20**
62			iso3/6	3	MS2	**(901.93 + 1780.88)**	MS3 901.9	**(670.31 + 1317.64)**	MSn (1317.75) >866 >662 >		**444.20**
63			iso4/6	3	MS2	**(901.93 + 1780.88)**	MS3 901.9	**(583.27 + 1143.55)**	MSn (1143.55) >866 >662 >		**444.20**
64			iso1/2	1							
65			iso1/2	1	Msn	**444.00**					
66			iso3	2	MS2(1024)> MS3:	**799.88 +1576.78**		**444.20**			
67			iso2/2	1		**458.00**					
68			iso2/2	1	Ms3(670.31) > Ms4(1317.64) > Ms5(866.39) > Ms6(662) > Ms7	**458.20**					
69			iso4/4	3	MS2	**814.89 + 1606.79**	MS3(814.9)> MS4:	**583.27 + 1143.55**	MS3(1143.5) >MSn		**458.20**
70			iso5/6	3	MS2	**(901.93 + 1780.88)**	MS3 901.9	**(670.31 + 1317.64)**	MSn (1317.75) >866 >662 >		**458.20**
71			iso6/6	3	MS2	**(901.93 + 1780.88)**	MS3 901.9	**(583.27 + 1143.55)**	MSn (1143.55) >866 >662 >		**458.20**
72			iso2/2	1							
73			iso2/2	1	Msn	**458.00**					
74			iso3	2	MS2(1024)> MS3:	**799.88 +1576.78**		**458.20**			
75			iso4/4	2	MS2	**(814.9 + 1606.8)**	ms3 (1606.8)	**583.27 + 1143.55**			
76			iso5/6	2	Ms2	Loss of **901.93 + 1780.88**	Ms3(902.03) >MS4	**670.3 + 1317.6**			
77			iso6/6	2	Ms2	Loss of **901.93 + 1780.88**	Ms3(902.03) >MS4	**583.2 + 1143.5**			
78			iso4	2	MS2	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1317.64**			
79			iso4	2	MS2	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1143.55**			

#	Structure	Structure 2	iso	n	MS method	Value			
80			iso2	1	Ms2	**1954.97**			
81			iso2	1	Ms2	**1780.88 (only if 834.4 present)**			
82			iso2/2	1	Ms2	**995.47**			
83			iso4	2	Ms2 (1126) >MS3	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1317.64**	
84			iso4	2	Ms2 (1126) >MS3	**901.93 + 1780.88**	Ms2 (901.9 or 1780.8) >Ms3	**1143.55**	
85			single	0					
86			single	0					
87			single	0					
88			single	0					
89			iso1/2	1	Ms2(999.47) > Ms3(1300.12) > ms4:	**1112.53**			
90			iso2/2	1	Ms2(999.47) > Ms3(1300.12) > ms4:	**1119.54**			
91			iso4	2	Ms2 (1127) >ms3	**(894.92 + 1766.86)**	MS3 (894.92) > Ms4:	**1317.64**	
92			iso4	1	Ms2 (1127) >ms3	**(807.88 + 1592.77)**			
93			iso2						
94			iso2	1	MS2(1264.11) > MS3(1032.49)> MS4(808.16)	**1143.55**			
95			iso2/4	2	Ms2(1351) >Ms3(1126.54) > Ms4	**894.92 +1766.86**	MS4(894.9 or 1766.8) >MS5	**1317.64**	
96			iso4/4	1	Ms2(1351) >Ms3(1126.54) > Ms4	**807.88 +1592.77**			
97			iso3/3	1	Ms2(1438) >Ms3(1213) >Ms4(1767) >Ms5	**1317.60**			
98			iso2	1	MS3(1257.11)> MS4(1032.49)> MS5(807.88)> MS6:	**1143.55**			
99			iso2/2	1	Ms3 (1344.15)> Ms4(1119.54)> Ms5 (894.92or 1766.86> Ms6:	**1317.60**			

#	Structure	Structure	Iso	n	Method	Value	Method	Value	Extra	Extra
100			iso3	1	MS2(1351.16) > MS3(1119.54) > MS4	**894.92 +1766.86**				
101			iso3	2	MS2(1575.77) > MS3(1344.15) > MS4(1119.5)> MS5:	**894.92 + 1766.86**	MS5(887.9 or 1766.86) > MS6:	**1317.64**		
102			iso3	1	MS3(1568.76) > MS4(1344.15)> Ms5(1119.54)> MS6	**894.92 +1766.86**				
103			iso6	2	Ms2	**(901.93 + 1780.88)**	Ms3 (1780.9)	**1317.64**		
104			iso6	2	Ms2	**(901.93 + 1780.88)**	Ms3 (1780.9)	**1143.55**		
105										
106			iso1/5	1	Ms2(1147) >Ms3	**894.9 + 1766.9**				
107			iso2/5	1	Ms2(1147) >Ms3	**901.9 + 1780.9**				
108			iso3/5	1	Ms2(1147) >Ms3	**1633.80**				
109			iso2/2	1	Ms2 1234 >Ms3	**1807.80**				
110			iso1/2	1	Ms2 1234 >Ms3	**1766.80**				
111			single	0						
112										
113			single	0						
114			single	0						
115										
116			single	0						
117			iso2	1	MSn	**444.18**				
118			iso1/2	1		**444.18**				
119			iso6	3	Ms2	**1024.49**	Ms2(1024.5) >Ms3(894.9) >Ms5:	**1303.60**	Msn 1303	**444.20**

#	Structure	Structure 2	iso	n	MS method	Mass		MSn method	Mass		Value
120			iso6	3	Ms2	1024.49	Loss of	Ms2(1024.5) >Ms3(894.9) >Ms5:	1129.53	Msn 1129	444.20
121			single	0							
122			single	0							
123			iso4/5	2	Ms2(1147) >Ms3	1017.49		>Ms3(1017) >Ms4(888) >Ms5(1303) >MSn (444)	444.20		
124			iso1/3	2	MS2(937.45)> MS3(807.88)> MS4:	1129.53		MSn (1129.5)	444.20		
125			iso1/4	2	Ms2(1126.54)> Ms3(894.9)> MS4	663.3 + 1303.62		Msn 1303	444.18		
126			iso2/4	2	Ms2(1126.54)> Ms3(894.9)> MS4	576.26 + 1129.53		Msn 1129	444.18		
127			iso1/4	2	MS2	1111.54		MS2(1111.5)> MSn	444.20		
128			iso2/4	2	MS2	1024.49		MS2(1024.5)> MSn	444.20		
129			iso2								
130			iso2								
131			iso1/2	1	1162.06 > 1032.49 > 1592.77 > 1129.53 > 925.43 > 648.28	444.00					
132			iso4	2	MS2(1249.1) > MS3(1119.9)> MS4:	887.92 + 1752.85		Msn	444.20		
133			iso4	2	MS2(1249.1) > MS3(1119.9)> MS4:	800.87 + 1578.76		Msn	444.20		
134			single	0							
135			single	0							
136			iso1/2	1	> 1017.49> 1562.76> 1303.62> 852.38> 648.28>	444.18					
137			iso3	1	MS3(1017.49 or 1119.5)> MS4(887.92)> MS5:	663.3 + 1303.62					
138			iso3	1	Ms2 (1473.72) > Ms3(1242.1) > Ms4(1112.53) > Ms5:	887.92 + 1752.85					
139			iso1	1		1303.60					

#	Structure	Structure 2	iso	n	MS method	mass	MS method 2	mass 2	MS method 3	mass 3
140			iso2	1	MSn	**458.19**				
141			iso2/2	1		**458.19**				
142			iso6	3	Ms2	**1024.49**	Ms2(1024.5) >Ms3(894.9) >Ms5:	**1303.60**	Msn 1303	**458.20**
143			iso6	3	Ms2	**1024.49**	Ms2(1024.5) >Ms3(894.9) >Ms5:	**1129.53**	Msn 1129	**458.20**
144			iso5/5	2	Ms2(1147) >Ms3	**1017.49**	>Ms3(1017) >Ms4(888) >Ms5(1303) >MSn (458)	**458.20**		
145			iso2/3	2	MS2(937.45)> MS3(807.88)> MS4:	**1129.53**	MSn (1129.5)	**458.20**		
146			iso3/4	2	Ms2(1126.54)> Ms3(894.9)> MS4	**663.3 + 1303.62**	Msn 1303	**458.19**		
147			iso4/4	2	Ms2(1126.54)> Ms3(894.9)> MS4	**576.26 + 1129.53**	Msn 1129	**458.19**		
148			iso3/4	2	MS2	**1111.54**	MS2(1111.5)> MSn	**458.20**		
149			iso4/4	2	MS2	**1024.49**	MS2(1024.5)> MSn	**458.20**		
150			iso2/2	1	1162.06 > 1032.49 > 1592.77 > 1129.53 > 925.43 > 648.28	**458.00**				
151			iso4	2	MS2(1249.1) > MS3(1119.9)> MS4:	**887.92 + 1752.85**	Msn	**458.20**		
152			iso4	2	MS2(1249.1) > MS3(1119.9)> MS4:	**800.87 + 1578.76**	Msn	**458.20**		
153			iso2/2	1	> 1017.49> 1562.76> 1303.62> 852.38> 648.28>	**458.19**				
154				0						
155			iso4	2	Ms2 (1127) >ms3	**(894.92 + 1766.86)**	MS3 (894.92) > Ms4:	**1303.62**		
156			iso4	2	Ms2 (1127) >ms3	**(894.92 + 1766.86)**	MS3 (894.92) > Ms4:	**1129.53**		
157			iso2							
158			iso2	1	MS2(1264.11) > MS3(1032.49)> MS4(808.16)	**1129.53**				
159			iso2	1	MS3(1257.11)> MS4(1032.49)> MS5(807.88)> MS6:	**1129.53**				

#	Structure		iso		MS workflow	Fragment 1	MS workflow 2	Fragment 2		
160	MS5: 1303, MS6: 852		iso1/4	2	Ms2(1351) >Ms3(1126.54) > Ms4	894.92 +1766.86	MS4(894.9 or 1766.8) >MS5	1303.62		
161			iso3/4	2	Ms2(1351) >Ms3(1126.54) > Ms4	894.92 +1766.86	MS4(894.9 or 1766.8) >MS5	1129.53		
162			iso1/3	1	Ms2(1438) >Ms3(1213) >Ms4(1767) >Ms5	1303.60				
163			single	0						
164			iso2/3	1	Ms2(1438) >Ms3(1213) >Ms4(1767) >Ms5	1129.50				
165			iso1/2	1	Ms3 (1344.15)> Ms4(1119.54)> Ms5 (894.92or 1766.86> Ms6:	1303.60				
166			single	1						
167			single	0						
168			iso3	2	MS2(1351.16) > MS3(1119.54) > MS4	887.92 + 1752.85	MS4(887.9 or 1752.8) > MS5(1301.61) > MS6	852.38		
169			iso3	2	MS2(1575.77) > MS3(1344.15) > MS4(1119.5)> MS5:	894.92 + 1766.86	MS5(887.9 or 1766.86) > MS6:	1303.62		
170			iso3	2	MS3(1568.76) > MS4(1344.15)> Ms5(1119.54)> MS6	887.92 + 1752.85	MS6(887.9)	1303.62		
171			iso2	1	MS3(1173.89)> MS4(1561.75)> MS5(1337.14)> MS6(1112.54)>	1303.62				
172										
173			iso3	1	MS3(1017.49 or 1119.5)> MS4(887.92)> MS5:	656.29 + 1289.61				
174			iso3	1	MS3(1017.49 or 1119.5)> MS4(887.92)> MS5:	569.25 + 1115.52				
175			iso3	1	Ms2 (1473.72) > Ms3(1242.1) > Ms4(1112.53) > Ms5:	880.91 + 1738.83				
176			iso3	1	Ms2 (1473.72) > Ms3(1242.1) > Ms4(1112.53) > Ms5:	793.86 + 1564.74				
177			iso2	1		1289.60				
178			single	0						
179			single							

	Structure			Type		MS path	Image		Image	Image		
180				iso3	2	MS2(1351.16) > MS3(1119.54) > MS4	**887.92 + 1752.85**	MS4(887.9 or 1752.8) > MS5(1301.61) > MS6	**838.36**		
181				iso3	1	MS2(1575.77) > MS3(1344.15) > MS4(1119.5)> MS5:	**887.92 + 1752.85**				
182				single	0						
183				iso3	2	MS2(1178.57) > MS3(1568.76) > MS4(1344.15)> Ms5(1119.54)> MS6	**887.92 + 1752.85**	MS6(887.9)	**1289.61**		
184				iso2	1	MS2(1296.96)> MS3(1173.89)> MS4(1561.75)> MS5(1337.14)> MS6(1112.54)> MS7(887.92)>	**1289.61**				
185				single	0						
186				single	0						

CHAPTER 3

REGULATION OF GLYCOSPHINGOLIPID EXPRESSION IN HUMAN STEM CELL DIFFERENTIATION

Introduction:

Glycosphingolipids (GSLs) are glycan bearing lipid structures that are a significant component of the cell membrane. As with cell surface glycoproteins, GSL mediate a variety of cell surface interactions and cell signal transduction events via extracellularly presented glycan structures interacting with carbohydrate binding molecules of adjacent cells or through cis-regulatory interactions with molecules on their own cell's surface[1, 2].

While the lipid surface of the cell membrane is in constant flux, one role of GSLs is to organize stabilized cell membrane signaling domains known as lipid rafts[3]. Within these domains, cell surface receptor signaling is modulated by different GSL glycoforms, influencing cellular adhesion and signaling through the formation of a "glycosynapse" [4]. GSL membrane domains also regulate innate immune system inflammatory responses via the recognition of pathogen-associated molecular patterns(PAMPs) by stabilizing Toll-like Receptors(TRLs)[5]. Various tissue specific GSLs modulate signaling involved cellular growth and differentiation[6], and neuronal synaptic transmission[7] by stabilizing growth factor receptors at the cell surface.

Glycosphingolipids are developmentally regulated and show cell specific expression. State Specific Embryonic Antigens (SSEAs), markers of cell developmental staging and stem cell pluripotency, such as SSEA1, SSEA3/4, and SSEA5 are all GSL epitopes[8, 9]. Given these known roles, we seek to better understand the regulation of GSLs in human cellular development and differentiation. In this study we combined GSL structure profiling with biosynthetic pathway transcriptomics in human embryonic stem cells differentiated into multiple cell lines representative of the three embryonic germ layer lineages and different states of differentiation. We identified distinct glycan profile trends between undifferentiated and differentiated cell types, indicating a shift in GSL biosynthetic pathways upon differentiation, and resulting in cell-type specific GSL profiles.

Experimental Procedures:

This section details the analysis of the glycolipid fractions isolated from the H9, NC, WT1, SM and Liv cell samples as described in Chapter 2. The methods for cell culture, the preparation of RNA samples for qRT-PCR and RNAseq, and analysis of transcriptomics data were as described therein.

Isolation and Preparation of Glycosphingolipids:

All organic solvents used were of HPLC certified purity, and water ultra-purified by Milli-Q reverse osmosis (EMD-Millipore). All glass collection tubes and glassware used for sample handling were thoroughly washed with water, 50/50 water-methanol, and 100% methanol prior to sample processing in order to remove any trace or detergent contaminants.

As described in the Chapter2- Experimental Procedures, section: "Extraction and preparation of N-linked glycans", total cell samples were homogenized and lipid fractions isolated from the protein fraction by means of a 4:8:3 ratio of chloroform/methanol/water solvent mixture optimized for the extraction of total glycolipids from complex biological samples [10, 11]. Three sequential solvent extractions of the homogenized cell samples were combined, dried down under gentle nitrogen flow in parallel with total protein collection, and stored at -20°C after drying. Isolation of the glycosphingolipids from the mixture of total cellular lipids proceeded as previously described in [12, 13]. Lipid extracts were saponified in 500uL 0.5M NaOH in methanol/water(95:5) at 37°C overnight to remove glycerophospholipids. Following saponification, samples solutions were neutralized with 5% AcOH on ice, adjusting to 1mL 50:50 methanol/water. Sample pH values were tested to ensure the proper acidification(<7pH) required for C18 column binding. All neutralized samples measured between pH of 5.5-6.0. The sample solutions were loaded via glass syringe onto a Sep-Pak tC18 columns (Waters, Sep-Pak Vac 1cc, 100mg resin), pre-washed with methanol and equilibrated with water. The column filtrate was collected and re-applied to ensure complete recovery. The loaded column was then washed with a total volume of 30mL water to remove non-lipid contaminants. GSLs were eluted in 4mL methanol and dried under nitrogen stream, resuspended in 1mL Acetone and dried again two more times to fully remove any remaining water. To remove free-fatty acids released by saponification, dried GSLs were washed with 1mL ice-cold hexanes (ACS Grade) while kept over ice, centrifuged at 4°C. This was repeated a total of 5 times, collecting the first two hexane washes for later testing. Purified GSLs and hexane washes were dried under nitrogen.

Glycosphingolipid Analysis by Thin Layer Chromatography (TLC):

For initial characterization, GSL samples from the five cell lines were analyzed by TLC. Samples were suspended in a 2:2:0.1 Chloroform/Methanol/Water solution, with volumes normalized by total sample protein yields. 20μL of each sample was set aside for MS analysis. 6uL were spotted onto oven-dried high-performance HPTLC silica plates, alongside standard mixtures (STD1) of sialylated ganglio-tetraosylceramides and (STD2) neutral glycosphingolipids (Matreya, cat# 1511 and 1505), and allowed to dry. TLC plates were run with a developing solution of 60:40:10 C:M:W for 9 minutes, and dried thoroughly[13]. Application of Orcinol–H2SO4 sugar staining reagent was used for visualization of GSLs. TLC analysis of dried hexane fractions was performed in the same manner and was found to mainly consist of fast migrating non-GSL bands, with minimal signal in the GSL range (**Figure 3.1**).

Mass Spectrometric Analysis of GSL Structures:

The 20μL of sample suspension solutions collected and normalized as described above were thoroughly dried down prior to permethylation for MS analysis. Similar to the protocol described in Chapter 2, native GSLs were permethylated in a solution of anhydrous NaOH in DMSO with Iodomethane [14].

For positive mode, direct infusion MS (DI-MS), the following GSL infusion buffer was used: 1mM NaOH in a 16:3:3:2 ratio mixture of methanol, 2-propanal, 1-propanol, and 12mM aqueous NaOH. Permethylated samples were resuspended in 100μL methanol and final infusion solution of 60μL consisted of 10uL sample solution, 45μL GSL(+) infusion buffer, and 5μL of 2pmol/μL DP4-STD.
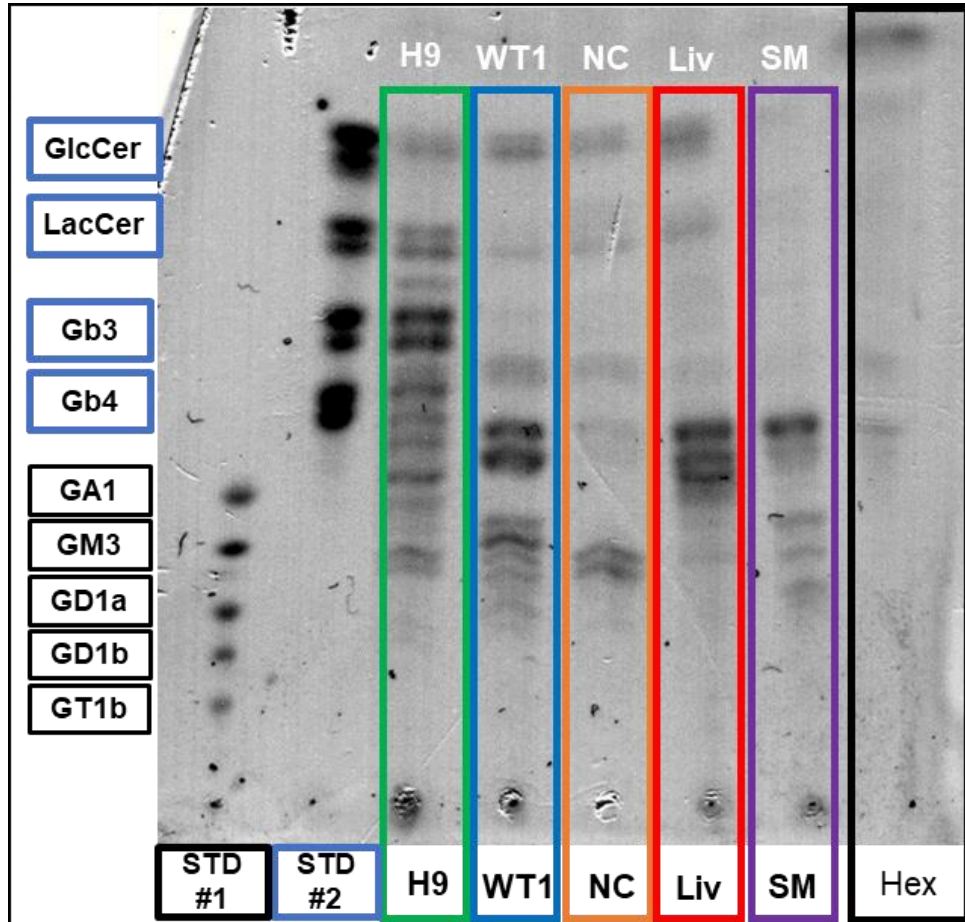
Figure 3.1: Thin layer chromatography separation of GSL structures.
(STD #1) Ganglioside standards, **(STD#2)** Neutral GSL standards, **(H9)** H9 ESC, **(WT1)** WT1 mesothelial precursor, **(NC)** Neural Crest Stem Cell, **(Liv)** mature hepatocyte, **(SM)** terminally differentiated smooth muscle, and **(Hex)** pooled Hexane wash negative control.

Mass spectrometry analysis was performed on an NSI-LTQ Obritrap-Ion Trap analyzer (Thermo Fischer). Nanospray ionization, direct infusion analysis of permethylated GSL samples was conducted in positive ion mode with a syringe flow rate of 0.40 µL/min and capillary temperature set to 210°C. For peak quantification, spectra were collected in FTMS profile mode over a 600-2000m/z range as an average of 5 scans each of 10 microscans.

Detection and identification of individual glycolipid structures was accomplished using ITMS-centroid mode MS/MS fragmentation, using the total ion mapping (TIM) and neutral loss scan (NL scan) functionality of the Xcalibur software package version 2.0 (Thermo Fisher Scientific) as previously described [11]. CID fragmentation settings were set to 35% Collision Energy, Activation Q of 0.250, and Activation Time of 30.0ms.

For the data-independent TIM fragmentation method, the m/z range from 600 to 2000 was automatically scanned with MS^2 fragmentation in successive steps of 2.2 mass unit windows, with a window-to-window overlap of 0.2 mass units, designed to capture the summed signal of the isotopic mass peaks of each glycolipid species to maximize detection sensitivity.

For the data-dependent NL method, an MS workflow was defined in which the highest intensity peaks detected by full FTMS were subjected to CID fragmentation. Previous studies [13] and preliminary analyses demonstrated that the major fragment ions in CID MS/MS scans of glycolipid preparations correspond to the neutral loss of the ceramide moiety, leaving intact glycolipid oligosaccharide ions. Thus, in the NL method the selection of peaks in the MS^2 level was determined by the presence of an ion with m/z equivalent to loss of the most prevalent ceramide moiety: in this case

permethylated mass of ceramide d18:1/C16:0, m/z of 547.54 (at +1 Na⁺). If detected, the presumptive GLS glycan was subjected to further MS^3 fragmentation.

<u>GSL Data Analysis</u>:

The manual analysis of GSL spectra can be complicated. A single glycoform may be attached to different ceramides, differing by FA chain length and degree of saturation, and present as multiple sets of peaks for each. Additionally, most glycolipids components were identified as singly and doubly sodiated species (M+Na, M+2Na) forms, so a single glycoform may consist of up to a dozen individual m/z peak values. Therefore, our MS workflows were valuable for the identification of unknown GSL glycan structures from complex biological samples. Data-dependent NL scans identified and confirmed multiple GSL glycan structures. Data-independent TIM scans allowed for identification of the multiple ceramide conjugates and charge states of each glycan. From this analysis glycan structure abundances were quantified by the addition of these peak intensities as measured in the FTMS scan. By utilizing this tandem approach, and quantifying peak intensities against an internal standard in a single scan we generated consistent quantification by eliminating possible variation due to spray conditions or other interruptions over the course of a DI experiment.

Results:

When GSL glycan abundances were quantified relative to protein amount of the sample the highest total abundance of GSLs was found in the undifferentiated H9 cells, followed by the multipotent cell types of WT1 and NC, and the lowest abundance found in the most differentiated cell types of Liv and SM (**Figure 3.2a**). Interestingly, this
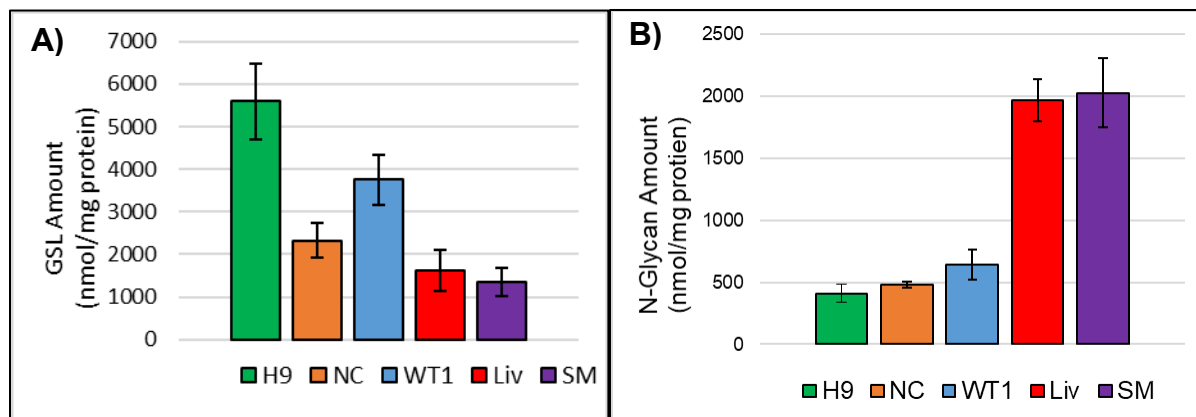
Figure 3.2: Total Amount of GSL to Total Amount of N-Glycan
A) Total GSL amounts for each line. **B)** Total N-linked glycans for each line.
Both are N=4, error bars are SEM. The amount of total GSL measured from each line followed an approximately inverse pattern to the total N-Glycan amounts.

pluripotent > multipotent > differentiated abundance pattern was the inverse of the of

glycoprotein N-glycan abundances observed in Chapter 2, where total amount of

glycan/protein increased with differentiation (**Figure 3.2a**).

Biosynthetic Pathways Responsible for GSL Core Types Distinguish Pluripotent and
Undifferentiated Cell Types.

Examining the relative distribution of GSL glycan structural types revealed that

undifferentiated H9 cells were predominantly of the globoside type with relatively little

ganglioside structures present (**Figure 3.3**). As cell types differentiated, this ratio

switched to a high percentage of gangliosides. This broad increase in ganglioside

structures following differentiation demonstrated some cell lineage specific features

between the different differentiated cells. Ectodermal NC profiles consisted mainly of the

GD3 derived b-series gangliosides, segregating it from the mesodermal and

endodermal cell profiles consisting mainly of GM3 and derived a-series gangliosides.

Transcript levels of the relevant biosynthetic genes correspond to the observed

glycan changes. Accompanying the abundance of globoside structures in H9, the Gb3

synthase α4-Gal transferase A4GALT is elevated in H9 cells, while the expression of

the ganglioside generating ST3GAL5 is low (**Figure 3.4**). The NC cells, high in GD2

ganglioside, have the highest expression of ST3GAL5 and ST8SIA1. WT1, Liv and SM

have high ST3GAL5 as well, with lesser amounts of ST8SIA1, thus producing a high

percentage of ganglioside GM3.

The percentage of Lacto-series structures is elevated in differentiated lines

(**Figure 3.3**), and the initial enzyme in that pathway B3GNT5 is up ~6- to 8 fold in NC,
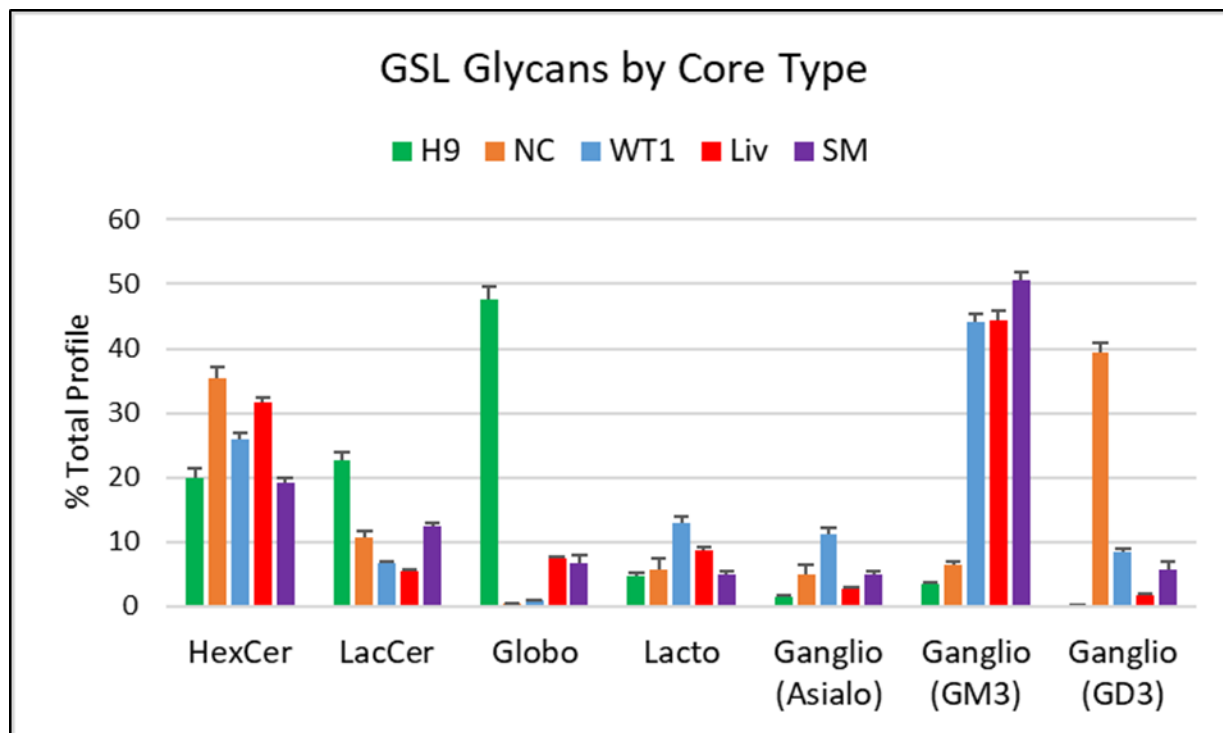
Figure 3.3: Relative Amounts of Core GSL Classes Between Cell Types
Relative abundance of GSL precursors HexCer and LacCer along with GSL structures grouped by class type. Some group assignments are overlapping due to isomeric structures that were unable to be resolved (e.g Asaialo GM1b vs IV-sialyl-Lc4). Obvious shifts between the distribution of Globoside structures and ganglioside structures differentiate undifferentiated versus differentiated cell types. N=4, error bars are SEM.
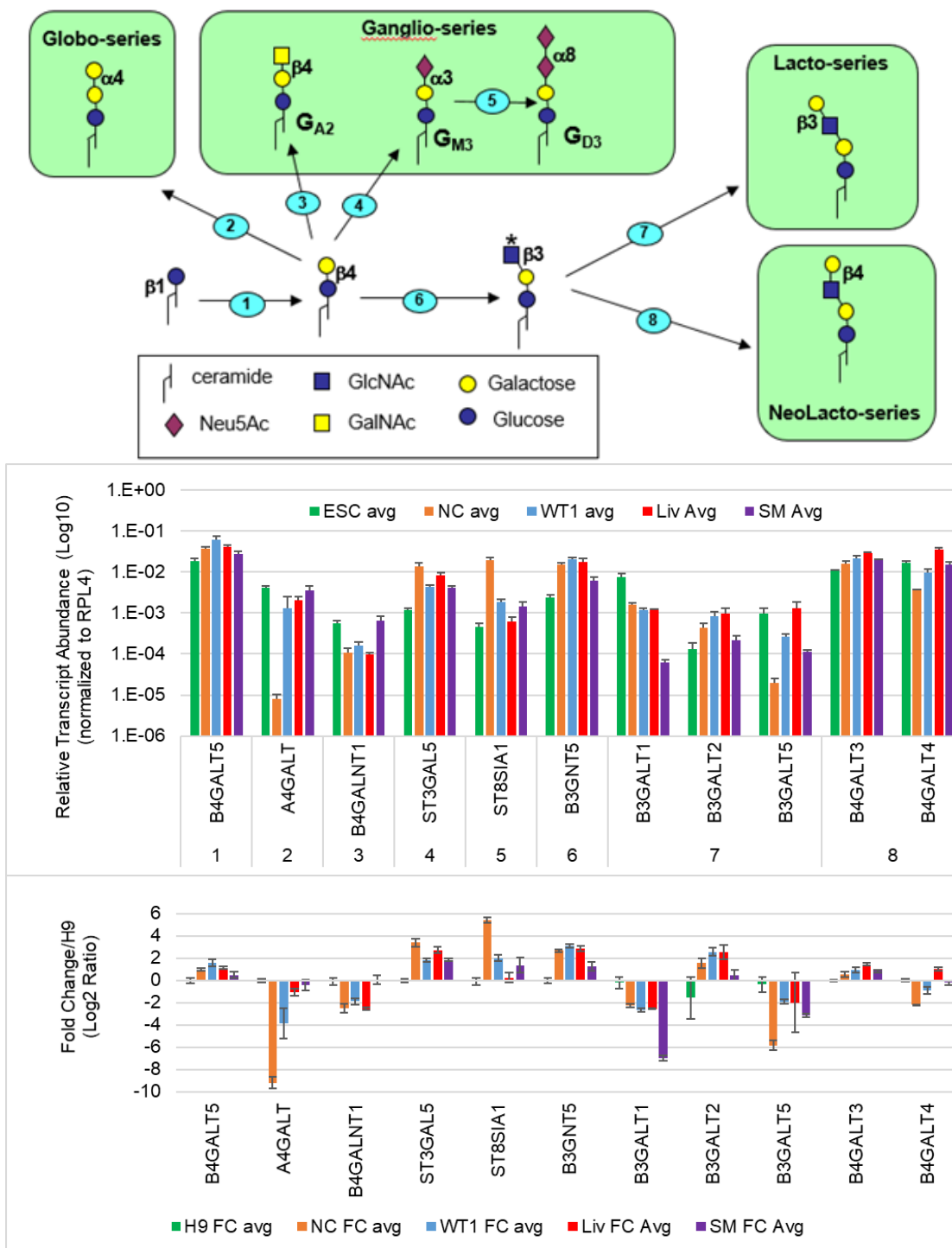
Figure 3.4: **Top)** Biosynthetic pathway diagram showing the generation of the main GSL core structures from shared precursor of Lactosylceramide. **Middle)** Corresponding qRT-PCR transcript abundances for the biosynthetic steps, $\log_{10}$ scale, n=4, error bars are SEM. **Bottom)** $\log_2$ fold changes of these genes relative to H9 hESCs, linear scale.

WT1, Liver and SM (**Figure 3.4**). Transcript levels of Lc3 extending B3GALT1 and B4GALT3 indicate a shift in type-1 vs type-2 lactosamine chains going from pluripotent H9 to differentiated cell types. B3GALT1, the major GSL β3-Gal transferase is up ~5-6x in H9 over NC, WT1, and Liv, while levels are extremely low in SM. B4GALT3 expression is up over H9 1.5-fold in NC, 2 fold in WT1, 2.7 fold in Liv, and 1.8 fold in SM. This trend indicates that shifts in the lactosamine linkage type correlate with the transition from pluripotency to differentiated state.

Another distinction between undifferentiated and differentiated cell types was the increased presence of sulfated GSL glycans in the former. While detection of sulfated structures was low in positive ion mode analysis, the sulfated galactose structure, Sulfatide, was detected in some of the cell types including H9, NC and Liv, but the largest amount of sulfated GSL detected was sulfo-Lactosylceramide in the undifferentiated H9 (**Figure 3.5**).

Elaboration of GSL Core Structures:

Once the core structure of globoside, gangliosides, and lactosylceramides are synthesized, further elongation and capping generate additional complexity. In the case of gangliosides, the major modification is the addition of Neu5Ac sialic acid residues in various linkages and positions (**Figure 3.6**). Many of the glycotransferase responsible for the elongation of asialo-series, G_{M3} based a-series, and G_{D3} based b-series are shared, further establishing the importance of the series committing enzymes B4GALNT1 (Asialo), ST3GAL5 (G_{M3}), and ST8SIA1 (G_{D3}). GSL structure profiles (**Figure 3.5**) show very little expression of elongated ganglioside structures, with only SM showing appreciable levels of a-series G_{M2}, G_{M1a}, and G_{D1a}. Negligible amounts of
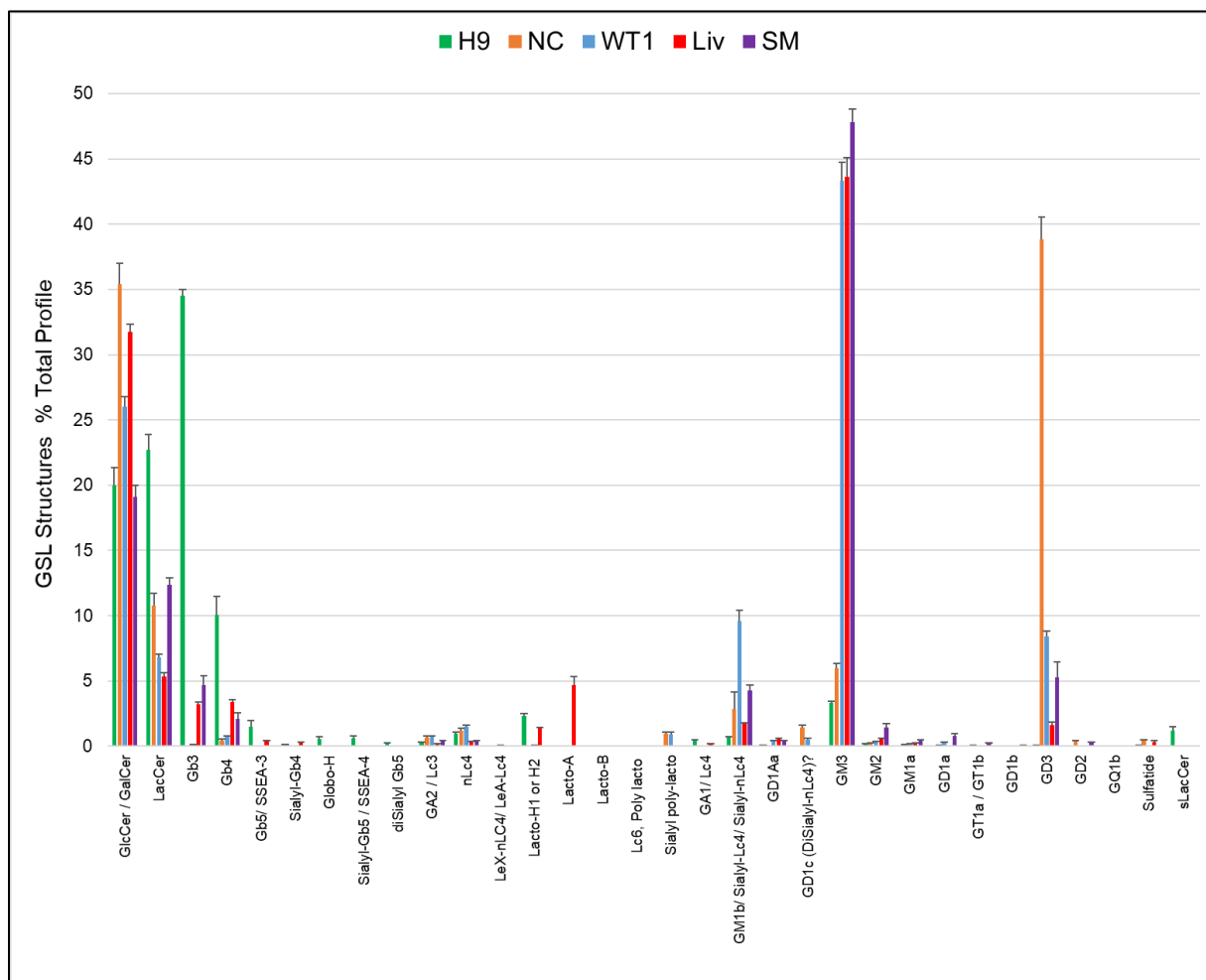
Figure 3.5: **GSL Structure Profile**
Complete profile of GSL glycan structures annotated by MS analysis. Relative abundances, n=4, error bars are SEM. Masses that had more than one possible isomeric structure that was not able to differentiated by MS^n fragmentation analysis have multiple assignments listed, e.g. GM1b/Sialyl-Lc4/Sialyl-nLc4.

Figure 3.6: **Ganglio-Series Biosynthetic Pathway and Enzymes**
Top) Biosynthetic pathway diagram for the elaboration of ganglioside structures. **Middle)** Corresponding qRT-PCR transcript abundances for the biosynthetic steps, $\log_{10}$ scale, n=4, error bars are SEM.
Bottom) $\log_2$ fold changes of these genes relative to H9 hESCs, linear scale.
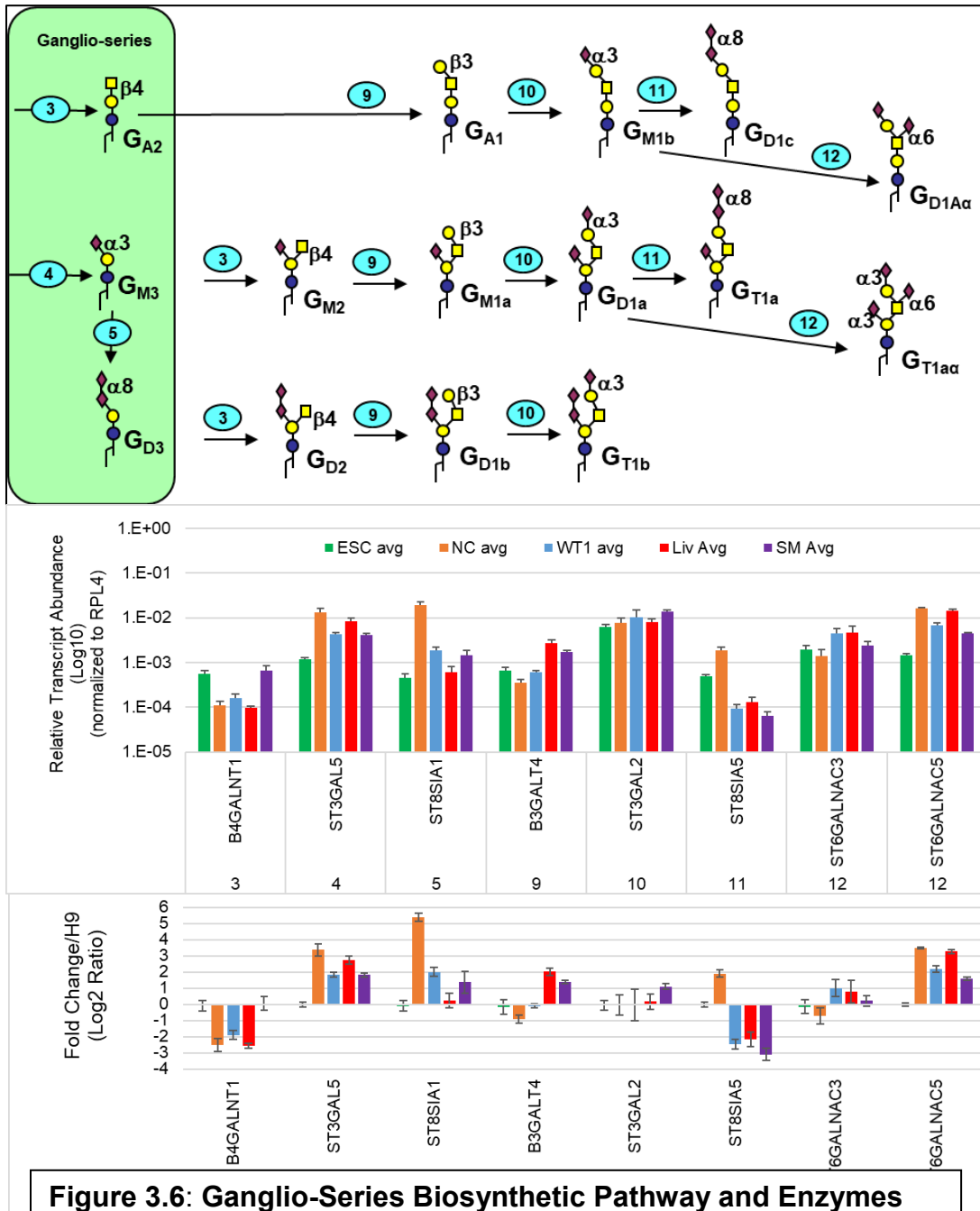
poly-sialyated, GT1, and GQ1 are found. Expression of B4GALNT1, responsible for the initial synthesis of GA2 as well as the first step in elongation of GM3 and GD3, is relatively low in all lines, though H9 and SM show the highest expression. Expression of B3GALT4, responsible for synthesis of GM1a is highest in Liv and SM. GM3 and GD3 account for the majority of ganglioside present in all of the differentiated cell types, and the low levels of further extension are likely due to the low expression of B4GALNT1.

Elaboration of Gobo-series Structures:

Globo-series extension is most prevalent in undifferentiated H9 and generates some of the characteristic stem cell/pluripotency markers. The initial Gb3 structure can be elongated and capped by sialic acid, creating some of the stem cell markers SSEA3 and SSEA4, or fucosylated generating blood group Type-4 antigens (**Figure 3.7**). SM and Liv also maintain some globoside expression, while globoside structures are nearly absent in NC and WT1.

H9 hESCs high expression of A4GALT (**Figure 3.4**) likely is responsible for the elevated amount of globosides in this line. While the largest percentage of globoside structures are Gb3 and Gb4, structures widely reported to be stem cell specific markers such as SSEA-3, Globo-H, SSEA-4 are found (**Figure 3.5**). Other structures such as Sialyl-Gb4, and diSialyl-Gb5 may be newly reported in stem cells, although diSialyl-Gb5 has been reported in malignant kidney cells [15]. That study provides evidence for ST6GALNAC6 being responsible for the synthesis of diSialyl-G_{b5}. Considering transcripts levels in **Figure 3.7**, it appears that sialylating enzymes such as ST3GAL2 and ST6GALNAC6 are highly expressed in all lines, with no elevation seen in H9, and
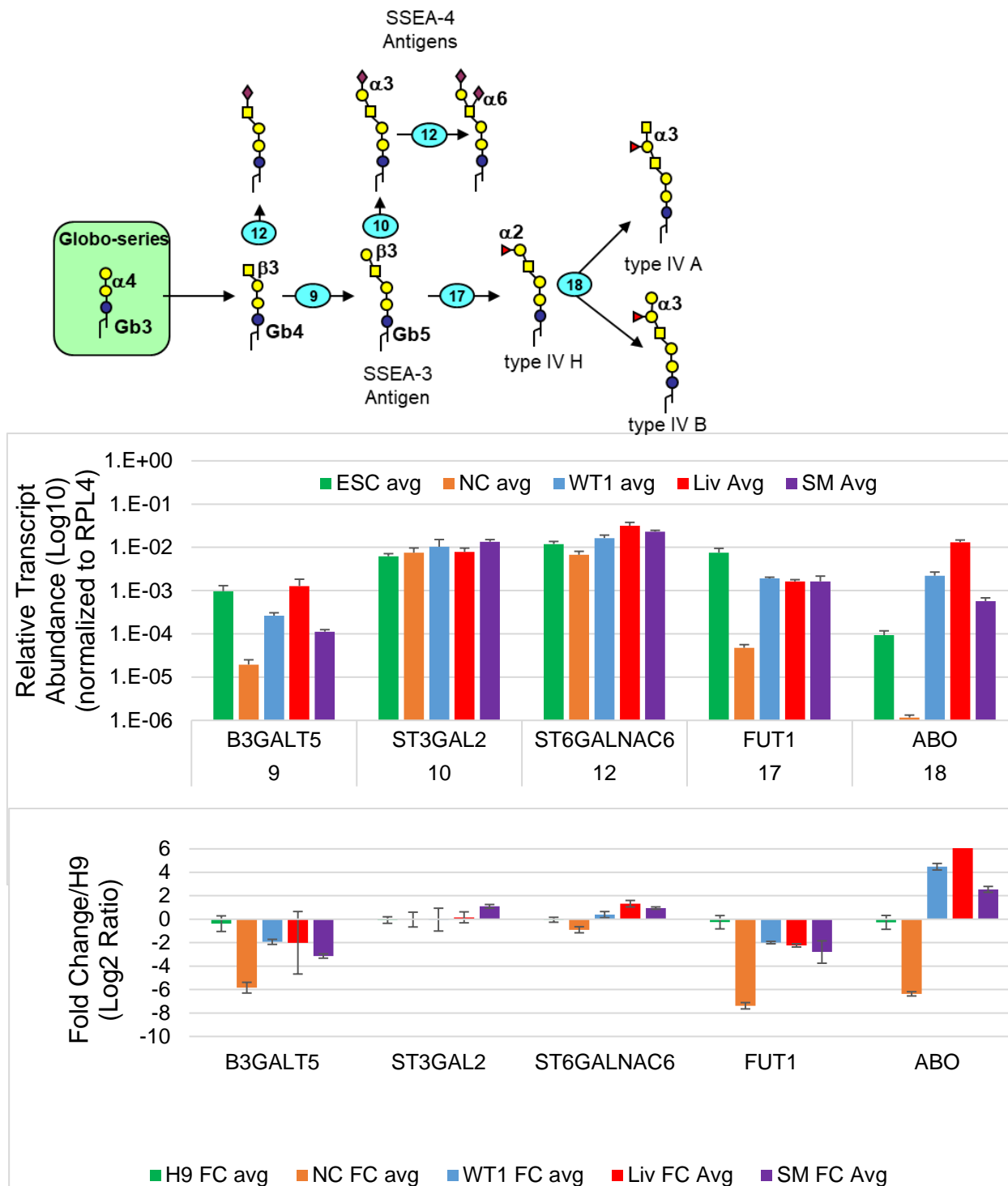
Figure 3.7: Globo-series pathway expansion
Top) Biosynthetic pathway diagram for the elongation and capping of globoside structures. **Middle)** Corresponding qRT-PCR transcript abundances for the biosynthetic steps, $\log_{10}$ scale, n=4, error bars are SEM. **Bottom)** $\log_2$ fold changes of these genes relative to H9 hESCs, linear scale.

probably represent constituently active glycosyltransferase. Instead, the transcripts of the enzymes for the generation of the precursors of the sialylated forms are the ones elevated in H9, with B3GALT5 producing G_{b5}, the precursor to the SSEA4 and diSialyl-G_{b5}. Transcription of FUT1 α1,2-fucosyltransferase is also elevated in H9, as is its glycan product (Globo-H)/ H-antigen Type IV is also elevated in the H9 profile (**Figure3.5**). Interestingly, "stem cell" structures G_{b5}/SSEA3 and sialyl-G_{B4} are also found in Liv which also has elevated levels of B3GALT5 and ST6GALNAC6.

Elaboration of Lacto-series Structures:

Similar to lactosamines structures on N-glycans, the terminal GlcNAc of L_{c3} can be elongated as Type-1 Lacto-series(L_{c4}) or as Type-2 neoLacto-series (nL_{c4}) GSLs. Expression of Type-2 generating β4-Gal transferases was generally an order of a magnitude higher than the Type-1 β3-Gal transferases, with the exception of H9 (**Figure 3.4**). Undifferentiated H9 cells had the highest expression of B3GALT1, nearly to the levels of B4GALT3/4, indicating a uniquely elevated presence of Type-1 GSL structures in undifferentiated cells. Interrogation of the β3/ β4 linkage by MS^n was able to distinguish some, but not all possible structural possibilities. The Type-2 nLc4 was found at higher levels than type-1 Lc4, though Lc4 was present in the H9 structure profile (**Figure 3.5**). Sialyl-Lc4, also known as sialyl-lactotetra, has been described as a stem cell specific marker [16]. This glycan structure shares the same mass (1273.61m/z, +1Na) and similar linkages to GM1b and sialyl-nLc4, making deconvolution of these structures by CID MS^2 difficult. However, each of these structures is sialylated by distinct ST3GAL transferases with substrate specificities unique to each. ST3GAL2 has higher affinity for asialo-type Gal-GalNAc [17], ST3GAL3

has highest affinity for Lc4 Gal-β1,3-GlcNAc [18], and ST3GAL6 specific affinity for type-2 Gal-β1,4-GlcNAc structures such as nLC4[19]. Examining the expression of these enzymes in each line (**Figure 3.8**) shows that, a) the sum expression of these transcripts correlated with the amount of total percent profile of the 1273m/z structure, and b) the relative amounts of each differed between structures, with H9 and SM being predominantly GM1b (ST3GAL2>>ST3GAL3≈ST3GAL6), NC and WT1 having the highest expression the Sialyl-Lc4 (ST3GAL3>ST3GAL2>>ST3GAL6), and Liv showing moderate expression of all three, with slight preference for ST3GAL2.

Other notable Lacto-series expression patterns included the presence of ABO blood group structures Type-2 H antigen in H9 and Liv, and Type-2 A antigen in Liv. No B antigen structures were identified. Unique to NC and WT1, sialylated extended lacotsamine structures were identified (**Figure 3.5**). Transcript levels for the poly-lactosamine extension of nLc4, B3GNT5 and B4GALT4 are elevated in NC and WT1 (**Figure 3.9**), correlating with these structural increases. The levels of blood group antigens also correlate FUT1 elevated in H9, and FUT2 and ABO elevated in Liver (**Figure 3.9**).

Utilizing RNAseq Transcriptomics to Understand GSL Profile Diversity.

Limitations in our MS method did not allow for interrogation of all isomeric structures, though inferences from expression data can help to interpret the likely glycoforms. For example, The Hex-Cer(d18:1/16:0) mass at 806.6 m/z could theoretically be either GalCer or GlcCer. GalCer, also known as Cerebroside or oligodendrocyte marker O4, is major component of the myelin in the nervous system, and a precursor to other myelin glycolipids such as sialyl-GalCer (GM4) and sulfo-
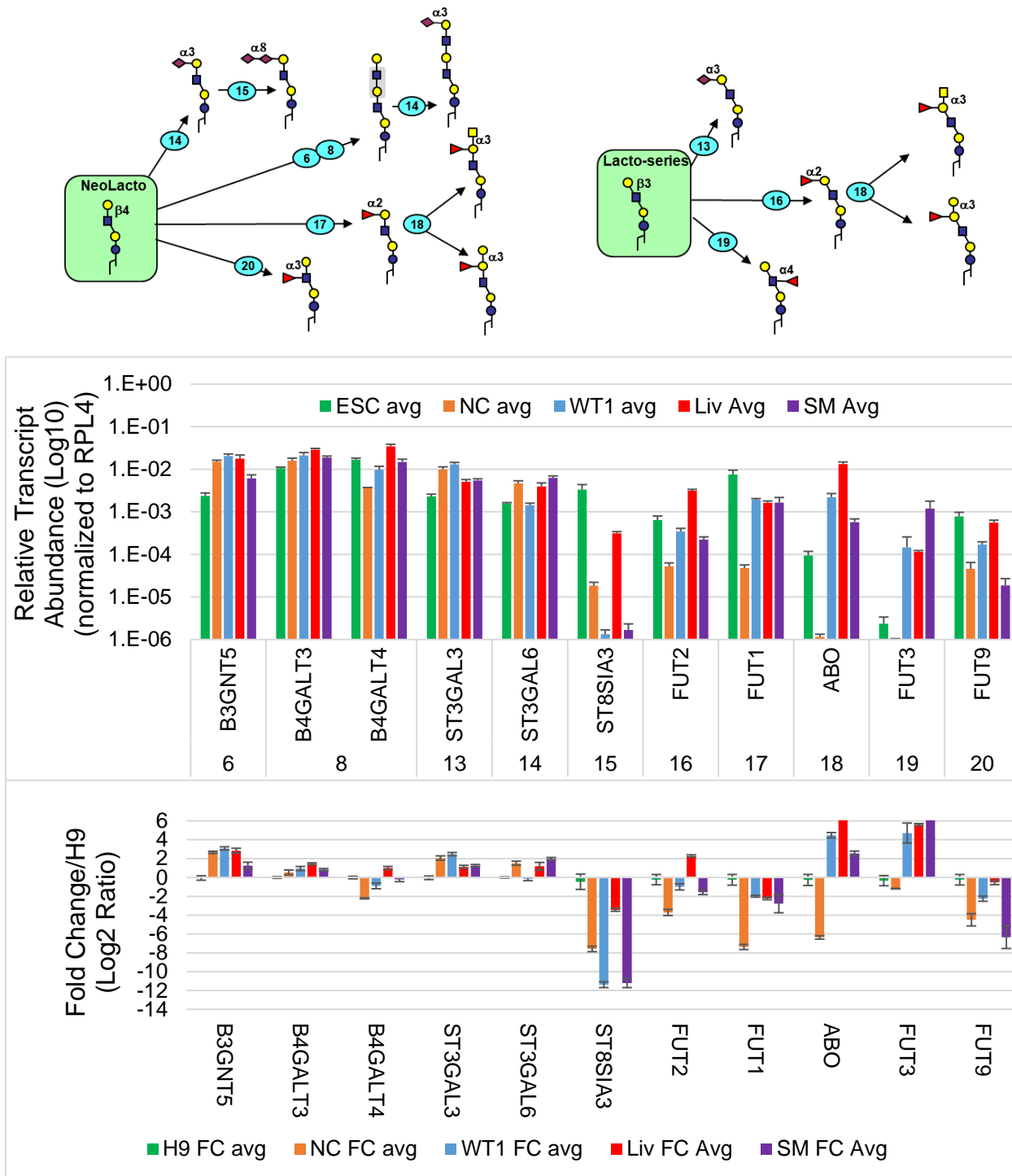
Figure 3.9: Lacto- and NeoLacto Series Expansion Pathways

Top) Biosynthetic pathway diagram for the elongation and capping of neolacto- and lacto-series GSLs. **Middle)** Corresponding qRT-PCR transcript abundances for the biosynthetic steps, $\log_{10}$ scale, n=4, error bars are SEM. **Bottom)** $\log_2$ fold changes of these genes relative to H9 hESCs, Linear scale.

GalCer (Sulfatide) [1]. On the other hand, GlcCer is capable of being extended into the gangliosides, globosides and lacotosylcermaindes that make up the bulk of the structures detected in the cell samples analyzed. As we detected no GM4 and due to the RNAseq expression ratio of GlcCer synthase UGCG to GalCer synthase UGT8 being elevated 4.2-fold in H9, 48-fold in Wt1, 2.9-fold in Liv and 47-fold in SM, we infer that in these lines the HexCer was most likely GlcCer. However, in the NC cells, UGT8 transcript is elevated over UGCG, and NC had the highest percent profile of sulfatide (**Fig. 10**). NC also has the largest percentage of HexCer in it GSL profile, likely due to it being a mixture of GlcCer and GalCer.

Discussion:

GSLs Play Crucial Roles in Development

Human diseases caused by mutations in GSL processing enzymes such as ST3GAL5, cause severe developmental disorders [13]. The role of GSLs is especially important in development, where core GSL synthesis enzymes regulate early processes in embryo development. The loss of Lc3 synthase B3GNT5 disrupts embryo implantation in the uterine wall [20] and cell movements during gastrulation, such as the epithelial-mesenchymal transition(EMT), are regulated by ST3GAL5 and ST6GALNAC5 [21]. These and other examples[22] might put into context our finding that the concentration of total GSLs in highest in stem cells and other multipotent cell types that may resemble early stages of development (H9,WT1, see **Figure 3.2**).

Previous studies of stem cell glycomics compared to differentiated cell types have sought to find stem cell or cell type specific structures. These studies discovered and characterized a number of stem cell specific glycans, such as classical stem cell
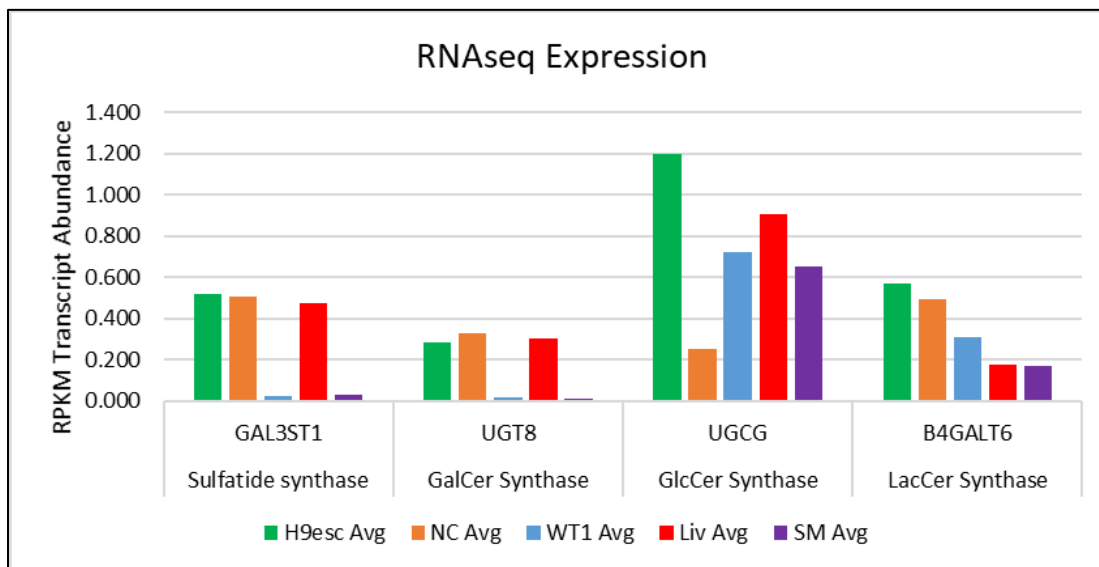
Figure 3.10: GalCer vs GlcCer Synthases
RNAseq examining the relationship between the transcript levels of GalCer synthase and SO_4-GalCer(Sulfatide) synthase responsible for Gala-series structures compared to the enzymes responsible for the production of elongated GSL structures such as gangliosides (GlcCer synthase, LacCer Synthase) the initial substrate. n=2.

epitopes SSEA3 and SSEA4, as well as novel structures like sialylated-Lc4, "Sialyl-lactotetra" [23, 24] .The compared cell GSL profiles underwent clear shifts across differentiation and these changes were attributed to altered expression of a limited number of glycosyltransferase genes examined [23]. However, these studies only examined glycosylation profiles from a few cell types, or non-specific stem cell differentiations such as Embryoid Bodies(EB). In these EB structures that contain a mixture of cells from all three germ layers, they observed switching from globo- and lacto series and to increased ganglio series GSL [23]. Early neural differentiation protocols have also been used and NCS GSL profiles have proven to be similar if not identical to EB GSL profiles[24], possibly due to the fact that neural differentiation predominates in early development/ differentiation[25]. In our data, NC cell types displayed reduced AGALT expression and virtually no globoseries, similar to the finding of previous reports by (Ojima, Umezawa, 2015)[24]. However, our data demonstrated that this is not a general feature of differentiation. A4GALT expression was not significantly reduced in differentiated cells such as Liv and SM, which also contained Globo structures.

To further understand the regulation of cell-type specific glycosylation and to investigate a broader range of cell types, we undertook comprehensive transcriptomic and glycomic analysis on a set of hESCs and hESCs differentiated through all three major germ line linages: Ectoderm (NC), Mesoderm (WT1, SM), and Endoderm (Liv). We were able to demonstrate that the different lineages, and indeed different cell types within these, have unique GSL biosynthesis. Importantly, we also found that some "cell specific structures" were shared between cell types, indicating that some trends in GSL

profile shifts were general features of increasing differentiation rather than cell specific markers.

Specific cellular differentiation lineages invoke unique glycomic expression pathways to generate individual GSL structure diversity.

Our cell specific data, summarized in **Table 1**, demonstrates some of the structure-transcript correlations driving unique features in the different cell types.

Comparisons to the Previous Literature:

H9 ESCs expand the globo-series core of Gb3 into Gb4 and presented many globoside structures that have been shown to be the epitopes of stem cell specific markers. Pluripotency markers such as SSEA3 (Gb5) and SSEA4 (Sialyl Gb5) are uniquely present in the undifferentiated cell types[26]. Sialyl-Gb4, Di-sialyl Gb5, and Globo H are also uniquely expressed or upregulated in H9 ESCs, agreeing with analysis of other human stem cell lines[24, 27, 28]. However other aspects of our findings do not fully align with those of other groups. Barone, et al. claim siayl-Lc4 (sialyl-lactotetra) as stem cell specific marker that disappeared when differentiated into hepatocyte-like cells or cardiomyocyte-like cells[16]. While our MS structural analysis was unable to conclusively determine if the sialylated tri-Hex, mono-HexNAc structure at 1273.61m/z was a type-1 lactosamine, our gene expression data indicates low levels of this GSL structure in H9 and higher levels in differentiated cell types such as WT1. Barone et al. also mentions that this epitope is found on N-linked glycoproteins, and as their identification was based on antibody binding this classification could have been due to non-GSL binding. Therefore, the abundance, type, and specificity of sialylated-lactosylceramide structures remains inconclusive.

Table 3. 1: Cell Line Specific Features and Transcript Correlations

Line	Lineage	GSL Profile Features	Transcript Correlation
H9 ESC	Pluripotent	• Highest amount of total GSL/Protein compared to Diff cells, with high amounts of GlcCer and LacCer. • Globo-series predominant • Low GM3 (which is high in differentiated cells types) • Blood Group H antigens • Sulfated GSL: sLacCer, sulfatide. • Unique structures: "Pluripotency Markers" SSEA3/4, SSEA5, TRA-160, TRA-1-81	• UGCG (GlcCer), B4GALT6(LacCer) • Highest expression of A4GALT • low STGAL5(Ganglio), B3GNT(Lacto) • FUT1 α1,2-fucosyltransferase • GAL3ST1 expressed, but at lower levels than NC and Liv
Neural Crest Stem Cell (NC)	Ectoderm	• Highest GD3, characteristic neural marker. • Absence of Globo series structures.	• Highest ST8sia1, and highest ST3GAL5 • Very low A4GALT
Mesothelium (WT1)	Mesoderm	• High % Lacto/neolacto types • GD1a and Sialyl Lc4 • Absence of Globo-series structures	• Highest B3GNT5 expression • Highest ST3GAL3 expression • A4GALT levels high.
Smooth Muscle (SM)	Mesoderm	• Lowest total GSL amount/protein. • Highest % of Gangliosides, high GM3 with some extended forms.	• High total protein and density of glycoproteins. • High expression of B3GALT4 and highest expression of ST3GAL2.
Hepatocytes (Liv)	Endoderm	• Blood group Antigens, neoLacto Type-2 H and A antigens	• ABO and FUT2, blood group expression • High B4GALT4, NeoLacto.

Distinct features of the Liv cell GSL profile include the presence of blood group structures. Similar to H9, unextended Lacto-H is found in the Liv profile. Liv also express ABO Gal/GalNAc transferase over 100-fold higher than undifferentiated H9 and much higher than any other differentiated cell type (**Figure 3.9**). Accordingly, blood group structure Lacto-A is present only in Liv (**Figure 3.5**). B-type structures were not identified. Interestingly this finding matches the known ABO geneotype of the H9 line (A/O)[29, 30]. Given the potential of pluripotent stem cells in regenerative medicine, previous studies of stem cell glycomics have focused on characterizing profile of immunoreactive surface epitopes, including the Histo-Blood Group ABO system [26, 31]. Similar to our data, they found expression of blood type antigens in some linages (hepatocyte-like cells) and not in others (cardiomyocyte-like cells)[26]. Combined with these studies our data adds credence to the need to characterize stem cell derived cell types used for regenerative medicine.

Key Findings and Directions for Further Study

An exciting finding that warrants follow up is the increased percentage of sulfated structures in the undifferentiated H9 profile. Unpublished data examining stem cell differentiation in mouse cells, also showed a stem cell specific increase in sulfated structures, with many sulfated structures that were identified in mESCs all but disappeared upon generalized EB differentiation. This indicates that the interrogation of sulfated structures in the human stem cells differentiations is warranted. Negative mode ITMS can better detect the negatively charged sulfate structures but was not compatible with the use of our current internal standards. Methods such as solvolytic de-sulfation of

permethylated GSL followed by re-permethylation with isotopic tags may prove to be a valuable approach.

This study added to the depth of understanding of GSL biosynthesis by examining a broad set of cell types. One key finding enabled by our broad transcriptomics analysis was that processing through GSL biosynthetic pathways is controlled by early glycosyltransferases that generate the initial structures of the various structural families. Despite high expression of the glycosyltransferases needed for further elongation, the majority of structures were of the early processed forms. Even in the most terminally differentiated cell type of SM, the major glycoform was GM3, and even with high levels of ST3GAL2 expression there was very little production of di-sialylated gangliosides. This indicates that either in-vitro cells cultures are not developed enough to enable complex ganglioside production (which is contradicted by our other unpublished data of in-vitro neuronal differentiations showing high rates of ganglioside family expansion and sialylation) or that non-transcriptional mechanisms of control determine the degrees of GSL processing in different cell types. Further investigation of ancillary factors such Golgi trafficking, the localization of relevant glycosyltransferases, and enzyme substrate specificities are needed. Within the scope of our work, further improvements in isomeric structure resolution by LC-MS separation or MS^n fragmentation strategies would provide a more precise picture of the structure profiles that could help make sense of the transcriptional data. As Glycosphingolipids are only half glycan, further mining of the RNA-seq transcriptomics data to examine ceramide and sphingosine lipid biosynthetic capacities would be useful.

<u>Works Cited:</u>

1.   Schnaar RL, K.T., *Glycosphingolipids*, in *Essentials of Glycobiology [Internet]. 3rd edition.*, C.R. Varki A, Esko JD, et al., Editor. 2017, Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY).

2.   Sych, T., Y. Mély, and W. Römer, *Lipid self-assembly and lectin-induced reorganization of the plasma membrane.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2018. **373**(1747).

3.   Sezgin, E., et al., *The mystery of membrane organization: composition, regulation and physiological relevance of lipid rafts.* Nature reviews. Molecular cell biology, 2017. **18**(6): p. 361-374.

4.   Todeschini, A.R. and S.-i. Hakomori, *Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains.* Biochimica et biophysica acta, 2008. **1780**(3): p. 421-433.

5.   Nakayama, H., et al., *GSL-Enriched Membrane Microdomains in Innate Immune Responses.* Archivum Immunologiae et Therapiae Experimentalis, 2013. **61**(3): p. 217-228.

6.   Julien, S., et al., *How Do Gangliosides Regulate RTKs Signaling?* Cells, 2013. **2**(4): p. 751-767.

7.   Schnaar, R.L., *Gangliosides of the vertebrate nervous system.* Journal of molecular biology, 2016. **428**(16): p. 3325-3336.

8.   Zhou, D., et al., *The β1,3-Galactosyltransferase β3GalT-V Is a Stage-specific Embryonic Antigen-3 (SSEA-3) Synthase.* Journal of Biological Chemistry, 2000. **275**(30): p. 22631-22634.

9.   Berger, R.P., et al., *Glycosylation and stem cells: Regulatory roles and application of iPSCs in the study of glycosylation-related disorders.* BioEssays : news and reviews in molecular, cellular and developmental biology, 2016. **38**(12): p. 1255-1265.

10.   Antti, S., et al., *Zwitterionic and acidic glycosphingolipids of the Drosophila melanogaster embryo.* European Journal of Biochemistry, 2000. **267**(12): p. 3549-3558.

11.   Aoki, K., et al., *Dynamic Developmental Elaboration of N-Linked Glycan Complexity in the Drosophila melanogaster Embryo.* Journal of Biological Chemistry, 2007. **282**(12): p. 9127-9142.

12. Sturgill, E.R., et al., *Biosynthesis of the major brain gangliosides GD1a and GT1b.* Glycobiology, 2012. **22**(10): p. 1289-1301.

13. Boccuto, L., et al., *A mutation in a ganglioside biosynthetic enzyme, ST3GAL5, results in salt & pepper syndrome, a neurocutaneous disorder with altered glycolipid and glycoprotein glycosylation.* Human Molecular Genetics, 2014. **23**(2): p. 418-433.

14. Anumula, K.R. and P.B. Taylor, *A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates.* Analytical Biochemistry, 1992. **203**(1): p. 101-108.

15. Senda, M., et al., *Identification and expression of a sialyltransferase responsible for the synthesis of disialylgalactosylgloboside in normal and malignant kidney cells: downregulation of ST6GalNAc VI in renal cancers.* Biochemical Journal, 2007. **402**(3): p. 459-470.

16. Barone, A., et al., *Sialyl-lactotetra, a Novel Cell Surface Marker of Undifferentiated Human Pluripotent Stem Cells.* The Journal of Biological Chemistry, 2014. **289**(27): p. 18846-18859.

17. Tsuji, S. and S. Takashima, *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 2 (ST3GAL2)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 645-656.

18. Schnaar, R.L., *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 3 (ST3GAL3)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 657-665.

19. Okajima, T. and K. Furukawa, *ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 6 (ST3GAL6)*, in *Handbook of Glycosyltransferases and Related Genes*, N. Taniguchi, et al., Editors. 2014, Springer Japan: Tokyo. p. 687-691.

20. Biellmann, F., et al., *The Lc3-synthase gene B3gnt5is essential to pre-implantation development of the murine embryo.* BMC Developmental Biology, 2008. **8**(1): p. 109.

21. Kurcon, T., et al., *miRNA proxy approach reveals hidden functions of glycosylation.* Proceedings of the National Academy of Sciences, 2015. **112**(23): p. 7327-7332.

22. Yagi, H. and K. Kato, *Functional roles of glycoconjugates in the maintenance of stemness and differentiation process of neural stem cells.* Glycoconjugate Journal, 2016: p. 1-7.

23. Liang, Y.-J., et al., *Switching of the core structures of glycosphingolipids from globo- and lacto- to ganglio-series upon human embryonic stem cell differentiation.* Proceedings of the National Academy of Sciences, 2010. **107**(52): p. 22564-22569.

24. Ojima, T., et al., *Glycolipid dynamics in generation and differentiation of induced pluripotent stem cells.* Scientific Reports, 2015. **5**: p. 14988.

25. Lupo, G., et al., *From pluripotency to forebrain patterning: an in vitro journey astride embryonic stem cells.* Cellular and Molecular Life Sciences, 2014. **71**(15): p. 2917-2930.

26. Säljö, K., et al., *HLA and Histo-Blood Group Antigen Expression in Human Pluripotent Stem Cells and their Derivatives.* Scientific Reports, 2017. **7**(1): p. 13072.

27. Barone, A., et al., *Structural Complexity of Non-acid Glycosphingolipids in Human Embryonic Stem Cells Grown under Feeder-free Conditions.* The Journal of Biological Chemistry, 2013. **288**(14): p. 10035-10050.

28. Säljö, K., et al., *Comparison of the glycosphingolipids of human-induced pluripotent stem cells and human embryonic stem cells.* Glycobiology, 2017. **27**(4): p. 291-305.

29. Carpenter Melissa, K., et al., *Properties of four human embryonic stem cell lines maintained in a feeder-free culture system.* Developmental Dynamics, 2003. **229**(2): p. 243-258.

30. Chen, Y.-T., et al., *H1 and H9 Human Embryonic Stem Cell Lines Are Heterozygous for the ABO Locus.* Stem Cells and Development, 2008. **17**(5): p. 853-855.

31. Mölne, J., et al., *Blood Group ABO Antigen Expression in Human Embryonic Stem Cells and in Differentiated Hepatocyte- and Cardiomyocyte-Like Cells.* Transplantation, 2008. **86**(10): p. 1407-1413.

Chapter 4:

Conclusions from Current Work and Future Directions

The nature of glycan/glycoconjugate synthesis is complex and yet not fully understood[1]. It is clear from the diversity and high complexity of these carbohydrate structures already known and that they would play important roles in biological functions[2]. This structural complexity is immense and unrivaled by other polymeric biological macromolecules such as nucleic acids and protein amino acid chains. Glycan structural diversity, due to the variety of monosaccharides building blocks and their ability to be joined in a diverse and multitudinous array of linkages, gives cells and organisms a very flexible tool, but it has made the biochemical understanding of these molecules difficult. While the genome has been sequenced[3] and our understanding of proteins has been developing for over a century[4], the total glycomic diversity of cells, tissues and organisms is still being uncovered. This current work adds this endeavor by combining human pluripotent stem cells, highly sensitive mass spectrometry, and comprehensive transcriptomics to take a combined approach to uncovering the diversity of glycan structures and biosynthetic regulation in a diverse set of human cell types.

Context of This Work:

There are many forms of glycan structures and glycoconjugates, including N-linked and O-linked glycoproteins, proteoglycan-GAGs, Hyaluronic-acid disaccharide

polymers, and glycosphingolipids. Within these classes there multiple sub-types classified by the glycosidic linkages with new forms still being discovered, especially in the field of O-linked glycans with linkages such as O-mannose[5] and O-fucose[6]. This work has focused on N-linked glycans, which regulate both the folding and function of cell surface and secreted proteins, and Glycosphingolipids, which are a major component of the cell membrane and regulate the stability and function other proteins on the cell surface such as growth factor receptors.

Multiple approaches have been taken to understanding the diversity of glycan structures, each with unique utility and known blind spots. The key challenges, and ultimate goal, are the combined isolation of all individual glycan structures, the identification of the component monosaccharides, and the definition of the linkages between them.

Biochemical and enzymatic methods are useful for examining linkage analysis via sequential or specific release of monosaccharides that reveals their individual linkages. Full digestion by acid hydrolysis to release all glycosidic bonds can give the compositional distribution of sugars but loses structural information.

Lectins, antibodies, and other tools that physically recognize and bind to defined glycan structures are well suited for identifying specific compositional motifs, such as the presence of the external fucose trisaccharide Lewis-X structure. Lectin and antibodies often have linkage specific binding affinities allowing them to discriminate the presence of linkage variations such as α2,3- or α2,6-linked sialic acid. The very fact that these biological tools, used by cells as well as scientists, bind specific linkages indicates that these seemingly small variations have important biological consequences. These

affinity binding methods, while useful for identifying specific compositional motifs do not specify the complete glycan structure and can bind to glycan motifs on multiple types of glycoconjugates as well.

Mass spectrometry(MS) and Nuclear Magnetic Resonance(NMR) spectroscopy have been long used to define chemical compositions, but only recently have technical[7] and computational[8] developments allowed them to tackle the characterization of glycan structures[9]. These methods rely on mass or chemical bond structure information to determine compositions and linkages. MS approaches also utilize fragmentation methods to looks for diagnostic fragmentation patterns specific to individual glycan structures, and can be coupled to chromatographic methods such as GC-MS or LC-MS. There are a multitude of types of ionization, fragmentation and detection used by different MS approaches. Below I will focus on discussing the strengths and limitations to the method which I used in this study, direct infusion of total glycan/ glycolipid isolates with NSI ionization, CID fragmentation, and detection by Fourier transform orbitrap (FTMS) and linear ion trap (ITMS).

Limitations of Current Study:

While the ultimate goal of this study was the comprehensive analysis of glycomics and transcriptomics, there were notable limitations in this study that restricted the attempt at total characterization. The main three include the technical limitations of the MS method chosen, the lack of analysis of other glycoconjugate classes, and the underutilization of RNA-seq transcriptomics data.

By using a Direct Infusion MS analytical method, all glycan/GSL structures in the sample are analyzed at once. This has some benefits, but the main drawback is the

inability to distinguish or accurate identify structures based on overlaps, either the isotopic/charge state peak overlaps resulting from a complex biological mixture or the literal overlapping of isomeric structures the same m/z. The use and modification of peak deconvolution software xTract, and MS^n fragmentation analysis respectively, attempted to address these issues, but not without faults. It is possible and indeed likely for low abundance peaks, that Xtract missed some peaks resulting in an underestimated total intensity or lack of identification. The MS^n method also had limitations. For MS identification of complete structure linkages, cross-ring fragmentations are need. Due to the nature of CID fragmentation on sodiated ions, which produces mainly B/Y fragments, cross ring fragment are rare until very high MS^n levels are reached and ion abundances low. Thus, we were unable to reliably distinguish certain motifs such as type-1 vs type-2 LacNAc structures and sialic α2,3- or α2,6-linkages. Alternative MS methods may function better in this regard, with LC-MS being able to separate some isomers based on chromatography, and other methods such as ETD fragmentation or the use of Lithium adducts are better able to generate cross ring structures. In future experiments, these or orthogonal methods such as lectin blotting will be needed.

One glaring limitation towards the profiling of "total cellular glycans", was the omission of O-linked analysis. Future work with these samples will have to include such analysis. Also, negatively charged glycan structures such as sulfated glycans were likely under identified in this analysis. The use of the DP4-STD and the desire to quantify all structures from a single scan, precluded analysis in negative mode. Indeed, transcript data and antibody based IF staining indicated that sulfated structures such as

the sulfated glucuronic acid HNK-1 epitope would be present on N-glycans. However, even in NC cells, where HNK-1 is used as an identification marker, no such structures were identified. A solution to this problem would be the use of de-sulfation protocol that leave a characteristic scar/tag that identifies the presence of sulfation while allowing for analysis in positive mode. While the further processing steps required for this method may result in some additional sample loss, the benefits merit its inclusion in future studies.

One limitation that the lack of O-glycan profiling touches on is the difficulty in assigning glycotransferase data to individual biosynthetic steps. While overall, more knowledge of glycotransferase enzyme substrate specificities and kinetics are needed, many enzymes have proposed activity with multiple glycoconjugate types. With an incomplete profile of all glycan structures, this add an unknown amount of variation to the interpretation of glycotransferase levels and their correlation with identified structures. O-linked glycan analysis of these samples will need to be included in future work, as many genes involved in O-linked biosynthesis were found in our RT-PCR and RNA-seq data.

Conclusions from Current Findings:

Despite these limitations, several impactful conclusions can be made from the data acquired. One of the strengths of the direct infusion MS method used was the ability to universally apply an internal standard to all peaks, allowing for absolute quantification. This allowed for the discovery that not only do glycan profiles change through cellular differentiation, but that the amount of N-glycans relative to protein mass also increases as cell become more terminally differentiated. From this we conclude two

possible explanations, 1) that developed cells differentially synthesize complexly glycosylated proteins, with more N-linked glycosylation sites, or 2) that comparable proteins have differential glycan site occupancy and more differentiated cells more efficient process glycoproteins to higher levels of site occupancy and structural elaboration. A combination of both may be true, as evidence supporting both exists from our data: To assess the former, comprehensive RNAseq transcript data was used to identify the top 100 expressed protein coding genes for each cell line and determine the amount of highly expressed glycoproteins bearing N-glycosylation sites (**Figure 4.1**). Mirroring the increase in total glycosylayion observed by quantitative MS, the more differentiated cell types had more glycoproteins among their top expressed genes. When the summed transcript abundances were weighted by the number of N-glycan sites for each protein, this correlation became even stronger. To support the latter hypothesis, the more differentiated cells had higher percentages of complex structures compared to un- or less processed high-mannose structures (as seen in **Figure 2.4**), indicating an increase in glycan processing through the Golgi. Even more specifically, these data indicate that the two highest N-glycan expressing cell types, SM and Liv, have distinct mechanisms for their high abundances, each relevant to their biological roles. Smooth muscle had the highest proportion of its glycan profile as complex type glycans and it had the highest weighted expression of N-glycan carrying proteins in **Figure 4.1c**, indicating more thorough glycan processing and more complex glycosylated glycoproteins. Indeed, measuring the calculating the average number of N-glycan sites per glycoprotein in **Figure 4.1a**, we find the H9 has 1.5, NC has 1.7, WT1 has 1.9, Liv has 2.0 and SM has the highest average amount with 3.0. This fits with

SM's role as a highly differentiated vascular-like smooth muscle that is laying down a complex and mature extracellular network, including complex glycoproteins. In the case of the Liv hepatocyte cells, the mechanism seems to be based on increase protein synthesis flux, befitting the role of hepatocytes as secretory cells. Liv had high amounts of complex type glycans, but also the highest amount of high mannose and paucimannose glycans (**Figure 2.4**), signs of increased glycoprotein biosynthesis and turnover reflective of high protein flux. In **Figure 4.1b**, as well, it also had the highest unweighted expression of glycoproteins, including secreted plasma glycoproteins such as Apolipoprotein-E (APOE) and Fibrinogen(FGA, FGB). Given the novelty and physiological relevance of these findings, further testing of these hypotheses in future experiments is highly warranted.

The use of MS^N fragmentation to deconvolve isometric structure sharing the same mass identified more individual N-glycan structures than any of our previous studies. Key differences in core structure assayed by diagnostic fragment analysis allowed for the identification of more hybrid, bisecting, and polylactosamine structures normally undistinguishable from their biantennary or triantennary branched isomeric structures. This allowed for better grouping of structural class, with less ambiguous structure overlaps, and gave cleaner correlations to relevant transcript levels.

Finally, by examining multiple distinct cell types from different linages and stages of development we were able to draw key observations that may have been missed by previous studies. While different cell types displayed distinct profile features allowing for clear separation by hierarchical clustering (**Figure 2.3**), and cell-type specific structures were identified, many structures previously assigned as stem cell specific markers were

A

| ES		NC		WT1		SM		Liver	
4		**3**		**19**		**24**		**24**	
CALR	1	CALR	1	BAMBI	1	AQP1	1	AFP	1
FTH1	1	FTL	1	CALR	1	BGN	1	APOA1	1
FTL	1	MIF	3	CD248	1	CALR	2	APOE	2
MIF	3			COL1A1	1	CD248	2	CALR	1
				COL1A2	1	COL1A1	1	CD248	2
				COL6A2	5	COL1A2	2	CD63	2
				DIO3	1	COL6A1	2	COL1A1	1
				DLK1	3	COL6A2	1	DLK1	5
				FTH1	1	COL6A3	1	FGA	6
				FTL	1	CTGF	1	FGB	10
				FZD7	2	DLK1	1	FTH1	2
				IGFBP3	3	FKBP10	5	FTL	7
				LUM	4	FN1	6	HLA-C	7
				MFAP4	2	FTH1	10	LUM	3
				MIF	3	FTL	2	MFAP4	2
				OLFML3	2	IGFBP3	3	MIF	1
				SERPINH1	2	LUM	7	NEU1	3
				SPARC	1	MFAP4	7	OLFML3	1
				SST	1	MIF	1	SELENOP	1
						SERPINH1	1	SERPINA1	1
						SLC39A7	3	SLC7A4	1
						SPARC	4	SPARC	1
						TGFBI	2	TIMP1	3
						TIMP1	3	TTR	1
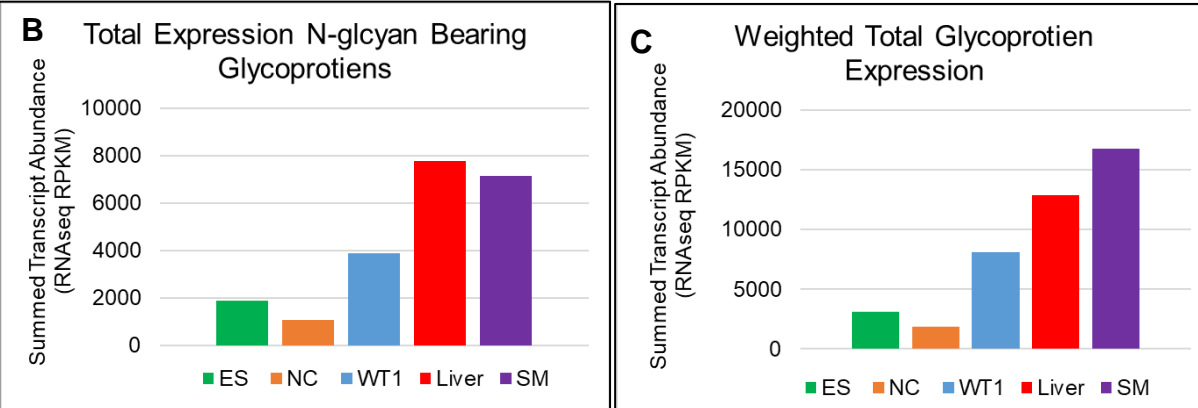
Figure 4.1: N-Glycan Glycoproteins in Top 100 Expressed Transcripts
The top 100 most abundant protein coding genes by RNAseq were catalogued for each line and cross referenced for presence and number of identified N-glycosylation sites according to the NextProt protein database (https://www.nextprot.org). **A)** The glycoprotein gene names and number of N-glycan sites. **B)** Summed transcript abundance of identified glycoproteins for each line, sample mean (n=2). **C)** Summed transcript abundances weighted by the number of N-glycan sites for each protein. Sum of individual transcripts multiplied by # of identified N-glycan sites. Sample mean (n=2).

found in other cell types, like globosides SSEA3 in Liv and sialyl-Lc4 in WT1 (**Figure 3.5, Figure3.8**). We also found that other trends, such as the switch from high-mannose to complex type N-glycans and the switch from Globoside dominated GSL profiles to Gangliosides, were general trends of increasing differentiation to terminally differentiated structures. This data warns of the risks on false positives in biomarker discover if a broad array of unique cell types is not examined.

Directions for Future Investigation:

Towards the goal of understanding the dynamics and biosynthetic control of glycan structures, both certain limitations in the current study and interesting findings, as discussed above, warrant additional investigation in future work.

To address study limitations, the application of orthogonal methods of glycan analysis should be applied to the remaining sample material as well as used in future work. Members of the Tiemeyer and Wells labs are currently working on developing and improving LC-MS glycan separation methods. Combined with continual development of the GRITS spectra annotation toolset, this will likely be the method of choice for glycan profiling moving forward.

One unaccomplished, but ambitious, goal of this study was to combine glycan profile data with transcript levels in a "biosynthetic flux" model. Many of the structure-transcript comparisons hinted at relationships more complex that 1:1 correlations (see **Figure 2.5**). The flux of intermediately processed structures through different enzyme processing steps is likely a more accurate model of glycan biosynthesis. Towards this goal a number of further experiments and developments are needed. As mentioned

above, better information on glycosyltransferase specificities is needed, and the closer

we can get the comprehensive identification of the totality of glycan structures, better

the model fitting will be. Ambiguity in either transcript or profile data increases variation

that make the construction of computation or algorithmic models difficult if not

impossible.

Furthermore, the use of RNA-seq data was underappreciated in this current

analysis. While current limitations include that only 2 replicates were done for most

lines, and the virtually infinite amount of data generated is overwhelming without

appropriate annotation and cataloging. Further use or refinements of this approach

would be useful for understanding the levels of factors that may play a role in the non-

transcriptional regulation of glycan synthesis, such as sugar nucleotide metabolism and

transporters. While not included in the previous chapters, the RNAseq data did provide

indications that this may be the case as in the case of UDP-fucose metabolism being

increased in H9 stem cells possibly contributing to their high proportion of externally

fucosylated structures (**Figure 4.2**).

Lastly, an important goal towards the further understanding of cell and tissue

specific glycosylation is the analysis of further cell types. Cell generated from disease

patient samples provide a natural source of loss of function testing. For example, the

Autism cell line currently under investigation provides a system for examining how

disruptions in signaling leading to glycomic changes. Furthermore, it displays

developmentally staged phenotypes. Full glycomic and transcriptomic analysis of this

cell line through multiple stages of development would yield further insights into the

biosynthetic changes that occur through differentiation, and any alterations of these in

GDP-Mannose --> GDP-Fucose
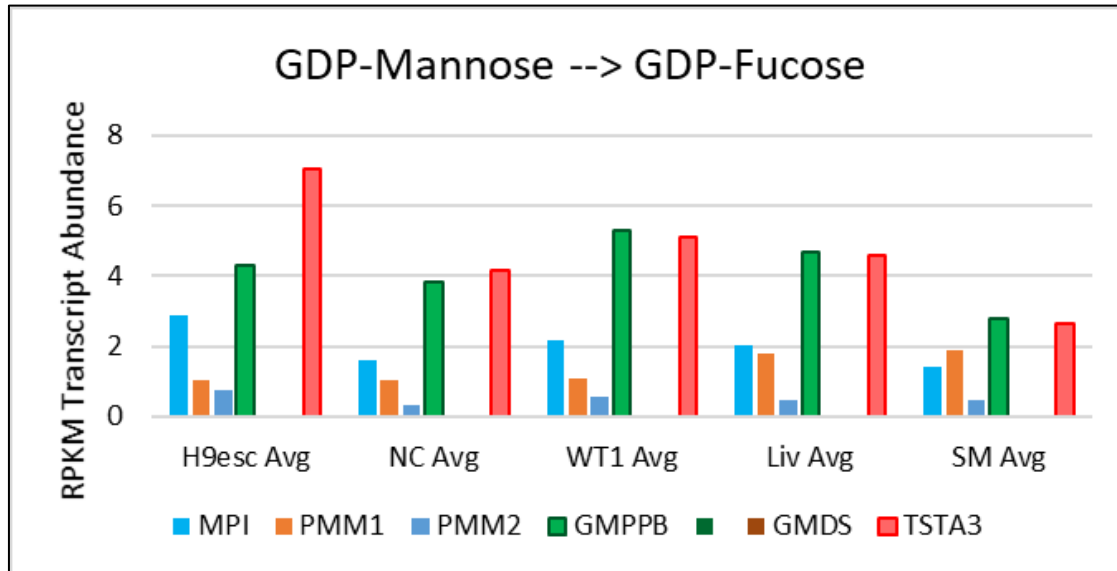
| GDP-Mannose | Fructose-6-P |
| **MPI** |
| Mannose-6-P |
| **PMM1/2** |
| Mannose-1-P |
| **GMPPB** |
| GDP-Mannose |

| GDP-Fucose | **GMDS** |
| GDP-4-dehydro-rhamnose |
| **TSTA3** |
| GDP-L-Fucose |

Figure 4.2: Transcript Expression Sugar Nucleotide Synthesis
RNAseq transcript expression of subset of sugar-nucleotide metabolism genes related to the generation of GDP-Fucose via GDP-Mannose. Transcripts for GDP-Fucose production are highest in H9 cells, which also expressed the highest amount of external fucosylated structures. (n=2, except NC n=4)

the disease cells may lead to a better understanding of the clinical relevance of glycan

regulation. As seen in the example of ST3GAL5 deficiency in the Amish community,

such insights are already leading to clinical treatment trials and option unavailable to

patients before.

Final Conclusions:

This study has led to exciting results and many more examples of future

experiments could be explicated here. To conclude, combined analysis of

transcriptomes and glycomes of these cells revealed shifts in biosynthetic pathways

between pluripotent, multipotent, and differentiated cells leading to the generation of

glycan structural profiles unique to the different cell types. This data indicates a

significant role for the regulation of glycan structures in development. Future studies

following up on these mechanisms and the roles in diseases will lead to an increasing

knowledge base of glycan regulation that, hopefully, will lead to novel interventions to

human diseases.

Works Cited:

1.   Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nature Reviews Molecular Cell Biology, 2012. **13**: p. 448.

2.   Haltiwanger, R.S. and J.B. Lowe, *Role of Glycosylation in Development.* Annual Review of Biochemistry, 2004. **73**(1): p. 491-537.

3.   Venter, J.C., et al., *The Sequence of the Human Genome.* Science, 2001. **291**(5507): p. 1304-1351.

4.   Vickery, H.B. and C.L.A. Schmidt, *The History of the Discovery of the Amino Acids.* Chemical Reviews, 1931. **9**(2): p. 169-318.

5.   Sheikh, M.O., S.M. Halmo, and L. Wells, *Recent advancements in understanding mammalian O-mannosylation.* Glycobiology, 2017. **27**(9): p. 806-819.

6.   Luther, K.B. and R.S. Haltiwanger, *Role of Unusual O-Glycans in Intercellular Signaling.* The international journal of biochemistry & cell biology, 2009. **41**(5): p. 1011-1024.

7.   Haslam, S.M., S.J. North, and A. Dell, *Mass spectrometric analysis of N- and O-glycosylation of tissues and cells.* Current Opinion in Structural Biology, 2006. **16**(5): p. 584-591.

8.   Ceroni, A., et al., *GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans.* Journal of Proteome Research, 2008. **7**(4): p. 1650-1659.

9.   Toukach, F.V. and V.P. Ananikov, *Recent advances in computational predictions of NMR parameters for the structure elucidation of carbohydrates: methods and limitations.* Chemical Society Reviews, 2013. **42**(21): p. 8376-8415.