Title
Investigating the role of the O-GlcNAc modification during the neural differentiation of human embryonic stem cells

Author
Andres, Lissette

Publication Date
2014

Peer reviewed|Thesis/dissertation

Investigating the role of the O-GlcNAc modification during the neural differentiation of human embryonic stem cells

By

Lissette Andres

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carolyn R. Bertozzi, Chair
Professor Michael Rape
Professor Ellen Robey
Professor David V. Schaffer

Fall 2014

Investigating the role of the O-GlcNAc modification during the neural differentiation of human embryonic stem cells

by Lissette Andres

Abstract

Investigating the role of the O-GlcNAc modification during the neural differentiation of human embryonic stem cells

by

Lissette Andres

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Carolyn R. Bertozzi, Chair

Human embryonic stem cells (hESCs) have the ability of propagating indefinitely and can be induced to differentiate into specialized cell types; however, the molecular mechanisms that govern the self-renewal and conversion of hESCs into a variety of cell types are not well understood. In order to understand the molecular mechanisms involved in the modulation of these processes it is necessary to look beyond what is encoded in the genome, and look into other forms of cellular regulation such as post-translational modifications. The addition of a single monosaccharide, β-*N*-acetylglucosamine (O-GlcNAc), to the hydroxy side chain of serine or threonine residues of protein is an intracellular, post-translational modification that shares qualities with phosphorylation. The enzymes involved in the addition and removal of O-GlcNAc, O-GlcNAc transferase and O-GlcNAcase, respectively, have been shown to target key transcriptional and epigenetic regulators. O-GlcNAc is believed to play a very important role in ESC biology, as this modification is required for cell viability and perturbations to the regulation of O-GlcNAc have been associated with abnormal development.

A great deal of interest is currently devoted towards deciphering the functional role of O-GlcNAc in stem cell maintenance and development. Chapter 1 summarizes the state of knowledge in the area of O-GlcNAc as it pertains to transcriptional regulation and development while also addressing how these processes might be regulated by the interplay between O-GlcNAcylation and phosphorylation. Moreover, it explains how chemical and biochemical tools have advanced our understanding of the functional significance of the O-GlcNAc modification. Chapter 2 describes the application of one of these chemical tools, known as metabolic labeling, in the identification of O-GlcNAcylated proteins in undifferentiated hESCs. The use of this tool coupled with biotin affinity purification and mass spectrometry analysis allowed for the identification of different O-GlcNAcylated proteins, including transcription factors, metabolic enzymes, and histones. The results presented in this chapter represent the first comprehensive proteomic characterization of protein O-GlcNAcylation in hESCs.

Similar to metabolic labeling, lectin weak affinity chromatography (LWAC) is another tool used for the enrichment of O-GlcNAcylated proteins. The O-GlcNAc modification has been found on proteins important for neuronal plasticity and brain

development, and the enzymes responsible for the modification are expressed highest in the brain. Chapter 3 focuses on characterizing the different O-GlcNAcylated proteins present during the neural differentiation of hESCs using mass spectrometry analysis and LWAC. The results discussed here provide fundamental knowledge of stage-dependent protein modification by O-GlcNAc and will help further elucidate the roles of O-GlcNAc in the development of the nervous system.

Finally, chapter 4 describes the effects of perturbing O-GlcNAcylation during the neural induction of hESCs. Inhibition of OGT induced the early expression of neuronal proteins and accelerated the conversion of hESCs into neural stem cells, suggesting a regulatory role of O-GlcNAc in maintaining proper brain development. The results presented in this chapter will help define the molecular behavior of stem cells during neuronal development so that they can be used effectively and reliably for the treatment of neurodegenerative disorders.

This dissertation is dedicated to my husband and my parents for always believing in me.

Investigating the role of the O-GlcNAc modification during the neural differentiation of human embryonic stem cells

Table of contents

Chapter 2. Metabolic incorporation of unnatural monosaccharides by human embryonic stem cells

Chapter 3. Proteomic profile of the O-GlcNAc modification on hESCs undergoing neural differentiation

List of Figures

List of Tables

Acknowledgments

Graduate school has been one of the most challenging experiences of my life. For this reason, I would like to start by thanking the people that helped me the most during this process, my family, specifically my husband Marco. I really wouldn't have been able to finish graduate school without his love and support. He was also very instrumental in helping me write this dissertation by keeping me fed and sane, which I'm sure my mom is also very thankful for. My parents, Luis and Amparo, are the reason I chose to go to graduate school. They encouraged me to always challenge myself and to excel in whatever I decided to do. They also believed in me even when they had no idea as to why I had to spend so much time changing media over the weekends. I would also like to thank my brother Luis and my sister Giselle for their love and support.

Obviously, I wouldn't have been able to do any of what I describe in this dissertation without the support and guidance of my advisor Carolyn Bertozzi. I have learned a lot from her and I appreciate all the time she devoted to me during my many years in graduate school. At the same time I would like to thank Phung Gip for being a great mentor and friend, and for being my unofficial therapist. You have always been there for me when I needed your help with experiments, quals, and nervous breakdowns.

I would also like to express my sincere gratitude to all members of the Bertozzi lab for being supportive and helpful. I have enjoyed the time I've spent with all of you. Special thanks to my fellow 'boricua' Olga Martinez. You have made my time spent in lab more enjoyable. Thank you for being my second unofficial therapist, and for being a great friend. I thank Neil Rumachik and Ian Blong for being the best collaborators, friends and proofreaders. Neil, I'm still amazed by your food selection for lunch; I have learned a lot about what not to eat if I ever decide to become a vegan. Ian, from the first day I met you, I thought you had the most contagious smile. Thank you for putting up with me, especially during the weekends and holidays. I would also like to thank my labmates, Ioannis Mountziaris and Brendan Beahm, for their encouragement and for making 816 a happier place. Special thanks to David Spiciarich for being a great proofreader and friend; and my personal supplier of dark chocolate. Thanks to former undergrad Evan Adams for assisting me in my work, and helping me become a better mentor. Thank you to lab members who have helped me with experiments and/or have provided advice: Michael Boyce, Pamela Thompson, Shana Topp, Kanna Palaniappan, Brian Smart, Sarah Hubbard, Ellen Sletten, Brian Belardi, Jason Tsai, Peter Robinson, Frances Rodriguez, Lauren Wagner, Megumi Ito, and Douglas Fox. I am also grateful to Asia Avelino, Sia Kruschke, Cheryl McVaugh, and Karen Carkhuff for their immense help, and because without them I would have been lost more than usual.

Many of the experiments described in this dissertation wouldn't have been possible without the help of great collaborators: Nam Pham and Jennifer Kohler at UT Southwestern, Sam Myers at UC San Francisco, and Teppei Yamaguchi at UC Berkeley. Special thanks to Minyong Chung for helping me with RNA sequencing analysis.

Part of my graduate school training wouldn't have been possible without the guidance and support of my thesis committee members Michael Rape, David Schaffer and Ellen Robey, and former committee member Sharon Amacher. I would also like to

thank Lily Mirels for allowing me to be part of the CIRM scholars program. I am also grateful for the guidance and support provided by Berta Parra from the MCB department, and Audrey Knowlton from the AMGEN summer program. I was supported by a Kirschstein-NRSA predoctoral fellowship (F31) from the National Institute of Health.

Finally, I thank my amazing friends, Geovannie Ojeda, Laurimer Kuilan, Natalie Del Hoyo, Lance Toyoshima, and Marcelo Baquero for always putting up with me and for being great cheerleaders.

Chapter 1

O-GlcNAc cycling: A novel mechanism for the regulation of stem cell pluripotency and differentiation

Chapter 1. O-GlcNAc cycling: A novel mechanism for the regulation of stem cell pluripotency and differentiation

1.1. Introduction

Pluripotent cells, such as human embryonic stem cells (hESCs), have the capacity of differentiating into any cell type in our body. The signaling networks, transcription factors and epigenetic regulators involved in establishing and maintaining specific cell fates are tightly controlled through multiple mechanisms, including post-translational modifications (PTMs). The post-translational modification of proteins provides additional diversity to extend protein function beyond what is genetically encoded [1]. Glycosylation is one of the most common forms of protein PTM, and involves the addition of monosaccharides and complex oligosaccharides, termed glycans, to proteins. The most abundant forms of protein glycosylation are N- and O-linked glycans. The N-linked glycans are attached to the amide nitrogen of asparagine residues, while the O-linked glycans are attached to the hydroxyl group of serine or threonine protein residues (Fig. 1.1). While typical N- and O-glycans are attached to proteins by glycosyltransferases with their active sites within the lumen compartments of the endoplasmic reticulum (ER) and Golgi apparatus (Golgi), there is a special type of glycosylation that occurs within the cytoplasm and nucleus, O-Linked β-*N*-acetylglucosamine commonly known as O-GlcNAc [2].

O-GlcNAc is a dynamic form of O-linked glycosylation that involves the enzymatic transfer of a monosaccharide (GlcNAc) onto serine or threonine residues of nuclear and cytoplasmic proteins from a UDP-GlcNAc sugar donor (Fig. 1.2). In contrast to the classic protein glycosylation found on cell surface and secreted proteins, O-GlcNAcylation of intracellular proteins is dynamic and reversible, often on the minute or second time-scale [3-5]. In this way, protein O-GlcNAcylation is analogous to phosphorylation cycling, in that a small, covalent post-translational modification is added or removed by dedicated enzymes as part of a signaling cascade that affects substrates proteins' localization, function or stability. Furthermore, several studies have suggested a dynamic interplay between O-GlcNAcylation and phosphorylation in regulating cellular signaling. Proteins from many different functional classes have been identified as being O-GlcNAcylated, including key transcriptional and epigenetic regulators such as RNA polymerase II, histones, transcription factors, TET proteins, and polycomb group proteins. Additionally, recent studies suggest that O-GlcNAc is an important regulator of neuronal development by modifying proteins important for neuronal signaling and synaptic plasticity, and contributing to the development of neurodegenerative diseases [6]. In this chapter, we discuss our current knowledge of the O-GlcNAc modification, focusing on how it is modulated by phosphorylation, how it functions as a regulator of transcription and development, and how the chemical and biochemical tools available have accelerated our understanding of O-GlcNAc.

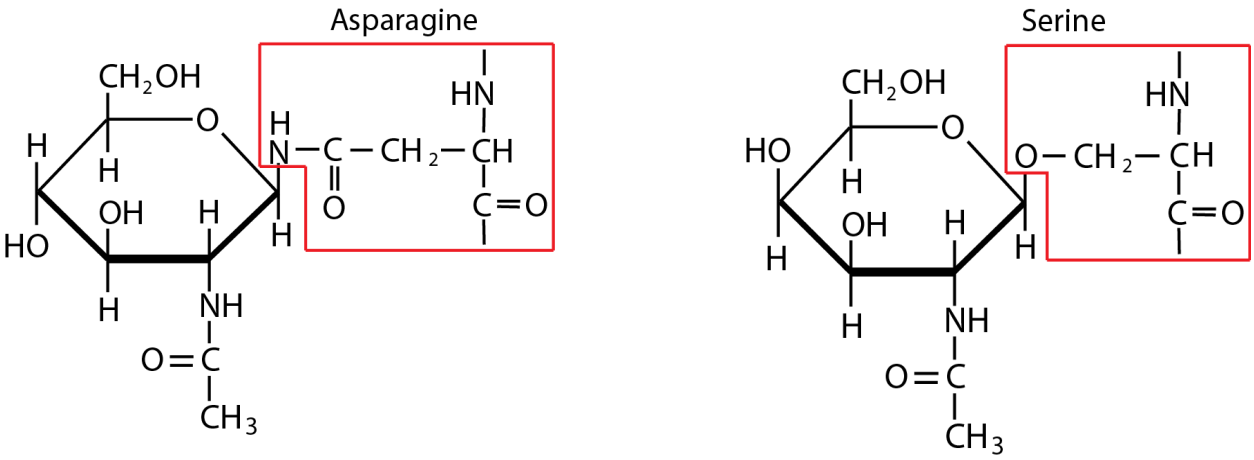


Figure 1.1. Examples of N-linked and O-linked glycosylation. (Left) N-linked glycans are attached to the amide nitrogen of asparagine protein residues. (Right) Meanwhile, O-linked glycans are attached to the hydroxyl of serine (shown here) or threonine proteins residues.

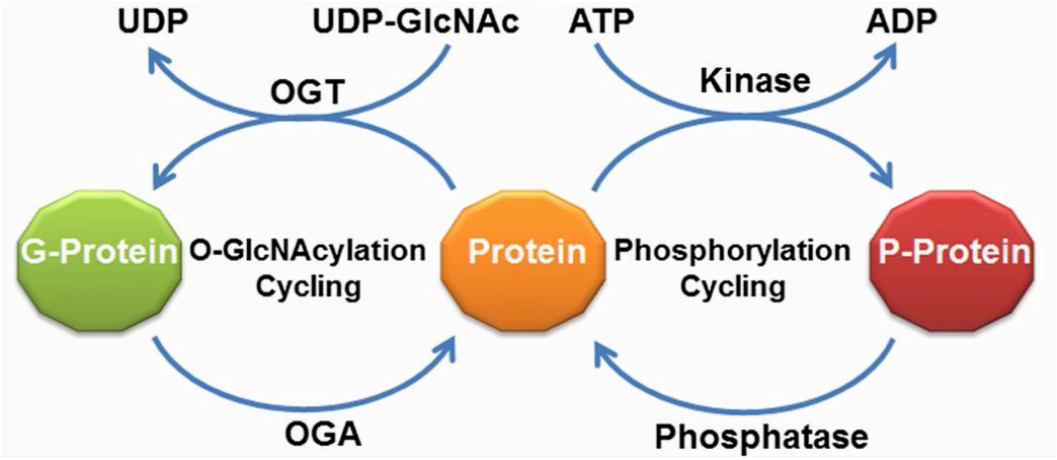


Figure 1.2. O-GlcNAc cycling is similar to phosphorylation. Nuclear and cytoplasmic proteins are post-translationally modified by the addition of β-*N*-acetyl-glucosamine (O-GlcNAc) onto serine or threonine residues. O-GlcNAc transferase (OGT) and O-GlcNAcase (OGA) catalyze the reversible addition and removal of O-GlcNAc, similar to the reversible addition and removal of a phosphate by kinases and phosphatases, respectively. UDP-GlcNAc is the nucleotide sugar donor of OGT, used to transfer GlcNAc onto substrates. (Modified from Gong et al. [7])

1.2. Principles of the O-GlcNAc modification

In recent years, the dynamic, reversible modification of intracellular proteins with O-GlcNAc, has emerged as a ubiquitous and indispensable regulator of diverse cellular processes [8, 9]. Unlike phosphorylation, which has hundreds of known enzymes called kinases and phosphatases that add and remove phosphate from proteins, respectively, only two enzymes catalyze the addition and removal of O-GlcNAc from proteins in mammals; O-GlcNAc transferase (OGT) installs O-GlcNAc, while O-GlcNAcase (OGA) removes it (Figure 1.2) [10-12]. O-GlcNAc has been found in all multicellular organisms investigated so far. However, O-GlcNAc and the enzymes that control it's cycling have not been found in yeast. Many of the O-GlcNAcylated proteins are transcription factors or translation regulatory factors, and the modification also appears to be particularly abundant on proteins involved in signaling, stress response, and energy metabolism [2]. The significance of this modification is emphasized by the fact that it is required for cell viability, and perturbations to the regulation of O-GlcNAc have been associated with the development of a number of diseases, such as cancer, Alzheimer's, cardiovascular disease, diabetes, and glucose toxicity [13, 14].

OGT, which catalyzes the addition of O-GlcNAc, is present in all cells, but is most abundant in the glucose-sensing cells of the pancreas and the brain. It has two distinct domains separated by a putative nuclear localization sequence. The amino terminus contains tetratricopeptide repeats (TRPs) that serve as protein-protein interactions domains and represents the sites to which most of the OGT targeting/regulatory proteins bind. The number of TPRs ranges from 3 to 13 repeats depending on the alternatively spliced isoform of OGT. The longest isoform, termed nucleocytoplasmic OGT (ncOGT), and the short isoform (sOGT) are localized in the nucleus and the cytoplasm. An intermediate isoform, termed mitochondrial OGT (mOGT) is localized in the mitochondria. The carboxy-terminal domain of all isoforms contains the uridine diphosphate N-acetylglucosamine (UDP-GlcNAc; nucleotide sugar donor) binding and catalytic site of the enzyme. Only when OGT is bound by the nucleotide sugar, can it catalyze the transfer of the GlcNAc onto proteins, thus making the UDP-GlcNAc the limiting substrate for protein O-GlcNAcylation. Additionally, OGT is post-translationally modified by tyrosine phosphorylation, O-GlcNAc modification [15], and Cys-nitrosylation [16], which could be another way of regulating its function.

O-GlcNAcase, the enzyme that catalyzes the removal of O-GlcNAc moieties from serine and threonine residues, is found in both the nucleus and cytoplasm. Like OGT, OGA is expressed at the highest levels in pancreas, brain, and thymus, with lesser amounts in other tissues [13]. Different from other β-N-acetylhexosaminidases, which hydrolyze terminal β-N-acetylhexosamine units from glycoproteins, it catalyzes the removal of O-GlcNAc but not of β-linked terminal N-acetylgalactosamine (O-GalNAc). The gene that encodes for OGA in mammals, meningioma expressed antigen 5 (MGEA5), is alternatively spliced to encode two major isoforms that are differentiated by their C-terminus. Both isoforms contain the same amino-terminus encoding the glycosidase domain, whereas only the longer isoform has a domain in the C-terminus that shows homology to the histone acetyl transferase (HAT) family, which specifically acetylates histones to activate gene expression. OGA is also O-GlcNAc modified, although the functional significance of this modification is not known.

1.3. Regulation of O-GlcNAc cycling

O-GlcNAcylation is the final step of a metabolic process known as the hexosamine biosynthetic pathway (HBP). HBP links glucose metabolism with the generation of UDP-GlcNAc (Fig. 1.3). UDP-GlcNAc is utilized not only for complex glycan structures in the ER and Golgi but also as the nucleotide sugar donor of OGT. The fluctuating levels of UDP-GlcNAc are a reflection of a cell's nutrient status, since its biosynthesis is affected and regulated by nearly every metabolic pathway in the cell [13]. UDP-GlcNAc is also the limiting substrate for O-GlcNAcylation, thus it has been proposed that O-GlcNAcylation serves as a nutrient sensor by modulating cellular signaling in response to nutrient status. For example, increased levels of intracellular UDP-GlcNAc enhance the activity of OGT, leading to an increase in cellular O-GlcNAcylation, as was seen recently in a study with breast cancer cells [17, 18]. Similarly, treatment of mesangial cells with high glucose (25 mM) increased cellular O-GlcNAc, compared to cells with normal glucose (5.5 mM) [19]. Conversely, in a study in cardiomyocytes glucose deprivation produced an increase in O-GlcNAc levels in a time dependent manner. Glucose deprivation was also associated with decreased OGA and calcium dependency [20]. These results suggest that the regulation of O-GlcNAc, by way of the regulation of UDP-GlcNAc production via the HBP, is complex and is sensitive to nutrient alterations.

O-GlcNAcylation has been shown to participate in a very extensive crosstalk with protein phosphorylation. Both modifications share common features: they are involved in the regulation of many cellular processes, such as transcription/translation, cell cycle and stress response, and are rapid and dynamic PTMs [13]. Recently, the Burlingame lab performed a proteomic characterization of O-GlcNAcylation and phosphorylation in the murine synapse and found that proteins found to be extensively O-GlcNAcylated were almost always phosphorylated to a similar or greater extent, indicating that the O-GlcNAc enzymes are specifically targeting a subset of the proteome that is also phosphorylated [21]. Site mapping studies have shown that there are at least three different types of crosstalk between O-GlcNAc and O-phosphate: 1) Competitive occupancy for the same site, e.g. Thr-41 of β-catenin [22]; 2) Competitive occupancy at proximal sites, e.g. O-GlcNAc at Thr-562 and Ser-576, and O-phosphate at Ser-566 synapsin I [23]; 3) Simultaneous modifications but on different sites, e.g. O-GlcNAc at Ser-40 and O-phosphate at Ser-133 in cyclic AMP-response element binding protein (CREB) [24]. For many OGT substrates, changes in O-GlcNAc level causes a reciprocal change in the phosphorylation state of the modified site. For example, increased O-GlcNAcylation decreased the phosphorylation of dynamin-related protein 1 (DRP1) at Ser-637, thus inducing its translocation from the cytoplasm to mitochondria [25]. Similarly, crosstalk between O-GlcNAcylation and phosphorylation of a given protein can trigger changes in their function and interaction with other proteins. The C-terminal domain (CTD) of RNA polymerase II is O-GlcNAcylated and phosphorylated at the same and adjacent sites, thus cycling of O-GlcNAc and phosphorylation was found to be important for the assembly of the pre-initiation complex and start of transcription [26].

Moreover, protein kinases were found to be more extensively O-GlcNAcylated in comparison to other protein classes, indicating the potential for interplay of phosphorylation with O-GlcNAcylation via regulation of enzymatic activity [21]. Recently,

Dias and colleagues used a human protein array and identified 42 kinases that were O-GlcNAcylated *in vitro*, and confirmed the presence of O-GlcNAc in 3 kinases *in vivo* [27]. Similarly, the catalytic subunit of casein kinase II (CK2) was found to be O-GlcNAcylated at Ser-37, antagonizing phosphorylation at Thr-344, and modulating the substrate selectivity of CK2 [28]. Another study found that O-GlcNAcylation of protein kinase B, also known as Akt, inhibited its phosphorylation by disrupting the interaction between Akt and PDK1 [29]. Finally, O-GlcNAcylation sites located in the active site of calcium/calmodulin-dependent kinase IV (CaMKIV) were shown to modulate its phosphorylation at Thr-200 and its activity toward cAMP-response element-binding transcription factor [30].

O-GlcNAcylation has also been shown to regulate several cellular processes by directly working in coordination with phosphorylation. In a recent study, Bullen and colleagues demonstrated that activation of AMP-activated protein kinase (AMPK) resulted in the alteration of the substrate selectivity of OGT in several cell lines and in the nuclear localization of OGT in C2C12 skeletal muscle myotubes. Moreover, O-GlcNAcylation of AMPK increased with AMPK activity, and acute inhibition of O-GlcNAc cycling disrupted activation of AMPK in muscle cells [31]. In another study, AMPK's phosphorylation of OGT at Thr-444 inhibited OGT-chromatin association, histone O-GlcNAcylation and gene transcription [32]. Interplay between O-GlcNAcylation and phosphorylation is also important for cell cycle progression and cytokinesis, as OGT and OGA interact with Aurora kinase B and protein phosphatase 1 to regulate the posttranslational status of the cytoskeletal protein vimentin [33]. Similarly, perturbations to global O-GlcNAcylation altered the phosphorylation of key proteins associated with the mitotic spindle and midbody. Overexpression of OGT increased the inhibitory phosphorylation of cyclin-dependent kinase 1 (CDK1) and reduced the phosphorylation of CDK1 substrates [34]. Finally, O-GlcNAcylation was found to work coordinately with phosphorylation to fine-tune circadian clock through the reciprocal regulation of OGT and glycogen synthase kinase 3β (GSK-3β) [35].

Recent studies have also suggested a potential interplay between O-GlcNAc and ubiquitination. Ubiquitination is a PTM in which ubiquitin is attached to a protein substrate by forming a covalent attachment between the last amino acid of ubiquitin (glycine 76) and a lysine residue of the substrate. Ubiquitin is best known for targeting proteins for proteasomal degradation [36]. Both ubiquitin and O-GlcNAc levels are equally increased or decreased when perturbing the HBP flux or O-GlcNAc cycling, through the use of chemical inhibitors and RNA interference [37]. Additionally, the ubiquitin-activating enzyme E1 was found to be O-GlcNAcylated, and its GlcNAcylation influenced its interaction with the heat shock protein, Hsp70. In another study, O-GlcNAcylation of the tumor suppressor p53 at Ser-149 stabilized p53 by blocking ubiquitin-dependent proteolysis [38]. Conversely, Fujiki and colleagues recently demonstrated that O-GlcNAcylation of histone H2B at Ser-112 promoted K-120 monoubiquitination, in which the GlcNAc moiety could serve as an anchor for a H2B ubiquitin ligase [39]. Overall, these results suggest a possible cross-talk between O-GlcNAcylation and ubiquitination in regulating a protein's function, stability, and interactions.

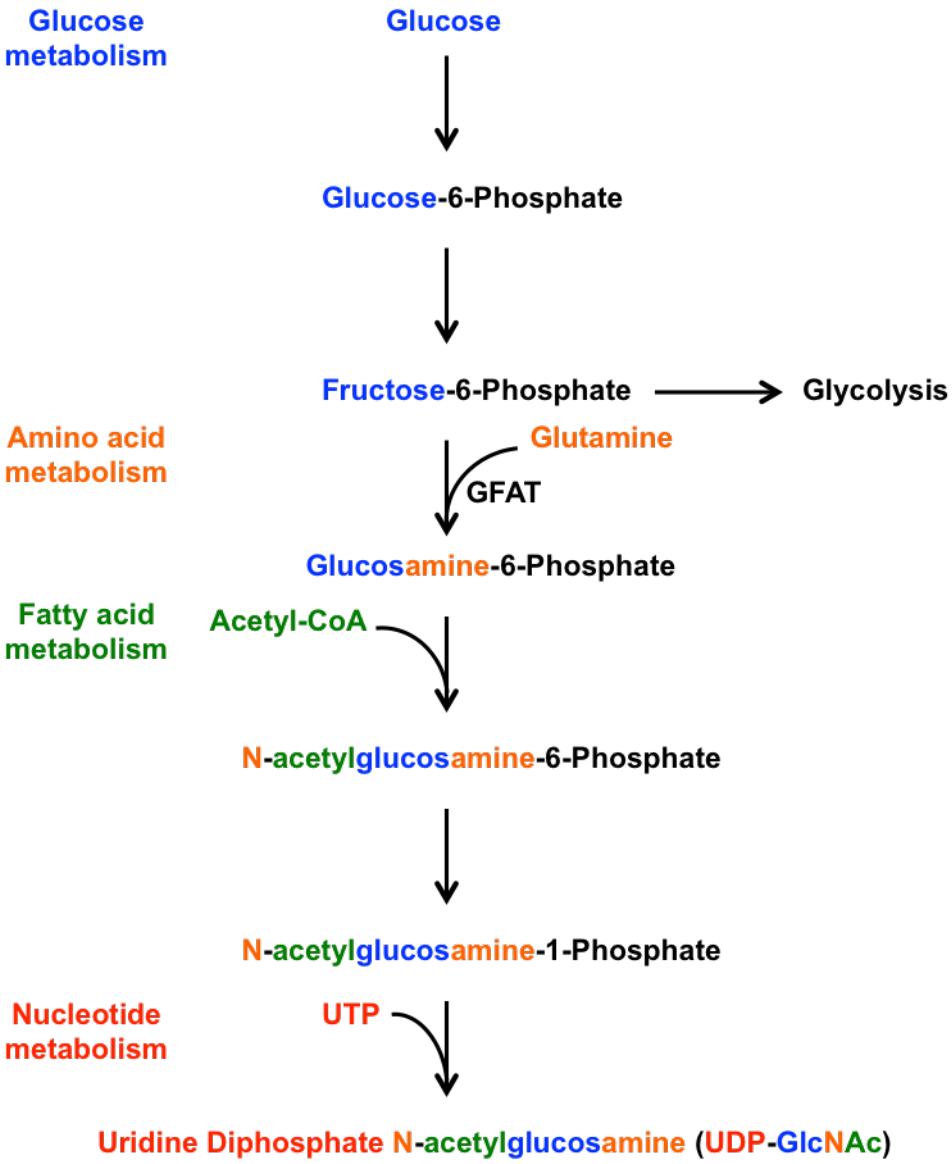


Figure 1.3. The hexosamine biosynthetic pathway provides the sugar substrate for O-GlcNAcylation. When glucose enters into the cell, a small percentage is channeled directly into the hexosamine biosynthetic pathway, where it is converted into Uridine Diphosphate-N-acetylglucosamine (UDP-GlcNAc). The biosynthesis of UDP-GlcNAc is also affected and regulated by amino acid, fatty acid and nucleotide metabolism. The enzyme glucose:fructose 6-phosphate amidotransferase (GFAT) catalyzes the rate-limiting step.

1.4. Roles of O-GlcNAc in transcription regulation and development

Transcription regulation

Transcription factors are some of the most highly O-GlcNAcylated proteins in cells. O-GlcNAcylation influences their transcriptional activity, DNA binding, localization, stability, and interaction with other proteins. Jackson and Tjian discovered the first O-GlcNAc-modified transcription factor, Sp1, in 1998 [40]. Sp1 is a ubiquitous transcription factor that activates transcription of genes that contain GC-rich sequences at their promoter regions. Shortly after the identification of the transcription factor, Yang and colleagues discovered that O-GlcNAcylation of Sp1 decreased its transcriptional activity both *in vitro* and *in vivo* [41]. Additionally, O-GlcNAcylation of Sp1 has been shown to regulate the factor's nuclear localization, protein stability, and its interaction with other transcriptional regulators [42-48]. OGT is also known to interact with transcriptional repressors, such as Sin3A [43]. Sin3A is a critical regulator of transcription networks required for proliferation and development [49]. OGT cooperates and modifies the co-repressors Sin3A and histone deacetylation complex (HDAC) to repress transcription. The same study found that both the TPR and the C-terminal catalytic domain of OGT participate in gene repression. Similarly, another O-GlcNAcylated factor, Host cell factor 1 (HCF-1), also interacts with OGT, Sin3A and HDAC to regulate the cell cycle [50-52]. OGT O-GlcNAcylates and promotes cleavage and activation of HCF-1. Recent studies elegantly demonstrated that conversion of the cleavage site glutamate into serine converted a HCF-1 proteolytic repeat into a glycosylation substrate, thus cleavage and glycosylation occurred in the same active site [52]. HCF-1 was also shown to recruit OGT to O-GlcNAcylate the peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC-1α), thus protecting PGC-1α from degradation and promoting gluconeogenesis [53]. OGT is also known to modify and interact with the nuclear receptor co-repressor proteins 1 and 2 (NCoR1 and NCoR2), which also form a repressor complex with Sin3A [54-56]. Moreover, increased O-GlcNAc levels correlate with the overexpression of genes that encode for NCoR, suggesting two mechanisms by which OGT may enhance transcriptional repression through NCoR [56].

O-GlcNAc modification of the C-term domain (CTD) of RNA polymerase II (RNA Pol II) is also important for the regulation of gene expression [57]. Reciprocal modification of CTD by O-GlcNAc and phosphorylation was found to modulate transcription initiation. Ranuncolo and colleagues demonstrated that inhibition of OGT and OGA blocks transcription during pre-initiation complex assembly [26]. OGT is also known to interact with RNA Pol II, and OGT is a component of the pre-initiation complex, thus providing further evidence of the role of O-GlcNAc in regulating transcription.

O-GlcNAc has also been shown to be important for polycomb repression. Polycomb group (PcG) proteins effect gene repression through epigenetic modifications and are required for the maintenance of both embryonic and adult stem cells [58]. Recently, two groups identified the *Drosophila* homologue of OGT (*sxc*) as a polycomb group protein critical for homeotic gene repression and embryonic development [59, 60]. Similarly, one of the PcG members, polyhomeotic, is glycosylated by OGT in *Drosophila*. Importantly, human OGT fully complemented the biochemical and

developmental phenotypes of *sxc* mutant flies, implying that OGT may play a similar regulatory role in polycomb-mediated gene silencing in mammals, a crucial process for human embryonic stem cell (hESC) self-renewal and differentiation [59]. Myers and colleagues identified O-GlcNAcylation of polyhomeotic homolog 3 in mouse embryonic stem cells (mESCs) [54]. The same study showed that the polycomb repressive complex 2 (PRC2), one of the two classes of PcG, was important for maintaining normal levels of OGT in mESCs. Recently, the E3 ubiquitin-protein ligase Ring1B, a member of PRC1, the second class of PcG, was found to interact with OGT in ESCs; however, the functional significance of this interaction remains unknown [61]. Another group of proteins involved in the epigenetic regulation of ESC pluripotency and differentiation, known as TET proteins, were recently identified as substrates and interacting partners of OGT [62-67]. Ten-eleven translocation (TET) proteins oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in DNA [68]. OGT modifies and interacts with all the TET proteins, TET1, TET2 and TET3. Interestingly, though, sites of O-GlcNAc modification have only been mapped for TET1 [54]. Additionally, OGT was involved in regulating the enzymatic activity of TET1 and TET3, and the subcellular localization of TET3. Furthermore, all three TET proteins were shown to play critical roles in targeting OGT to chromatin and enhancing OGT enzymatic activity. For example, down-regulation of TET2 reduced the amount of histone 2B Ser-112 O-GlcNAcylation [65]. Altogether, these results suggest that the TET-OGT complexes might be involved in regulating transcriptional activation in ESCs.

Histone O-GlcNAc sites have also been identified in the core sequences of histone 2A (H2A) and H4, including sites known to be phosphorylated [17, 69]. O-GlcNAc modification has also been found on histone H3, where it also competes with phosphorylation for site occupancy [70]. Interestingly, O-GlcNAcylation of Ser-10 at H3 was associated with both active (H3K4trimethylation) and repressed (H3K9 trimethylation) chromatin. Furthermore, histone O-GlcNAcylation changed during mitosis, particularly through interplay with phosphorylation. Taken together, these results show that O-GlcNAcylation of histones is dynamic, and demonstrate the role of O-GlcNAc in the epigenetic regulation of gene expression.

Development

Several lines of evidence have demonstrated that the modification of proteins with *O*-GlcNAc plays a pivotal role in embryonic development. For example, the complete absence of OGT function is lethal to mouse ESCs and embryos [71], and experiments with conditional alleles revealed that OGT is required in a tissue specific manner at later stages of development [72]. In addition, OGT activity is required for oocyte maturation, cell survival, and epiboly movements during gastrulation in zebrafish embryos [73]. In *C. elegans*, dysregulation of O-GlcNAc cycling led to changes in transcription, insulin signaling, and other signaling pathways [74-76]. Moreover, deletion of OGT and OGA significantly altered their lifespan.

O-GlcNAcylation has also been shown to both enhance and suppress activity of transcription factors important for ESC pluripotency and differentiation. Recently, Jang and colleagues demonstrated that loss of OGT reduced proliferation and self-renewal of

ES cells, as well as decreased reprogramming efficiency of induced pluripotent stem cells (iPSCs) [77]. Conversely, high O-GlcNAcylation increased reprogramming efficiency and decreased differentiation. The study also showed that O-GlcNAcylation of pluripotency factors, OCT4 and SOX2, was necessary for maintaining ESC pluripotency, and that their modification decreased following ESC differentiation. Previous studies had already demonstrated the interaction of OCT4 with OGT in mouse ESCs [78, 79]. One of the two studies also suggested an interaction between OGT and Dax1, Esrrb, and Tcfcp2l1 [79], all important in embryonic development. These results correlate with a recent study that demonstrated that excess O-GlcNAcylation, caused by inhibition of OGA, impaired differentiation of mouse ESCs without affecting naïve-to-primed cell transition [80]. Conversely, a study using human pluripotent cells (hPSCs) found that up-regulating of O-GlcNAc with OGA inhibitors had no effect on either the self-renewal capabilities of hPSCs or on their differentiation potential [81]. However, excess O-GlcNAcylation did alter the expression of specific lineage markers. Still, the authors of this study did not demonstrate if knockdown of OGT had any effects in hPSC pluripotency, like Jang and colleagues demonstrated in mESCs.

Proper O-GlcNAcylation is also important for the development of several tissues. Dysregulation of O-GlcNAcylation blocks spontaneous differentiation of ESCs to cardiomyocytes [82]. Similarly, increased O-GlcNAcylation negatively regulated differentiation of skeletal myogenesis [83]. Conversely, an increase in O-GlcNAcylation and OGT expression were important for proper adipocyte formation [84] and chondrogenic differentiation [85]. O-GlcNAcylation has also been shown to play a significant role in brain development (Figure 1.4). The enzymes responsible for the modification are highly expressed in the brain [15, 86]. Moreover, neuron-specific OGT deletion caused neuronal dysfunction that resulted in neonatal death [72]. Accumulation of protein O-GlcNAcylation in the mouse hippocampus and in mouse neural precursor cells coincided with neuronal apoptosis [87, 88]. Additionally, overexpression of OGA was found to increase the percentage of neurons exhibiting axon branching [89]. The modification has been found on proteins important for neuronal plasticity, and brain development [21, 90-94], including cAMP-responsive element binding protein (CREB) [95, 96], β-amyloid precursor protein (APP) [97], tubulin [98], and tau [99]. Moreover, the interplay of O-GlcNAc with phosphorylation has also been involved in modulating neuronal plasticity and disease. Changes in O-GlcNAc levels have been shown to regulate long-term potentiation through increased ERK1/2 and CaM kinase II activation-specific phosphorylation, along with increased synapsin phosphorylation [100]. Manipulation of phosphorylation in cerebellar neurons resulted an inverse relationship between phosphorylation and O-GlcNAcylation [101]. Inducing O-GlcNAcylation of tau negatively regulated its hyperphosphorylation [99, 102-104], a hallmark of Alzheimer's disease [105]. Similarly, APP, which forms the β-amyloid plaques characteristic of Alzheimer's disease, was both O-GlcNAcylated and phosphorylated [97]. Thus, O-GlcNAc may regulate aspects of neuronal development and degeneration in concert with phosphorylation.

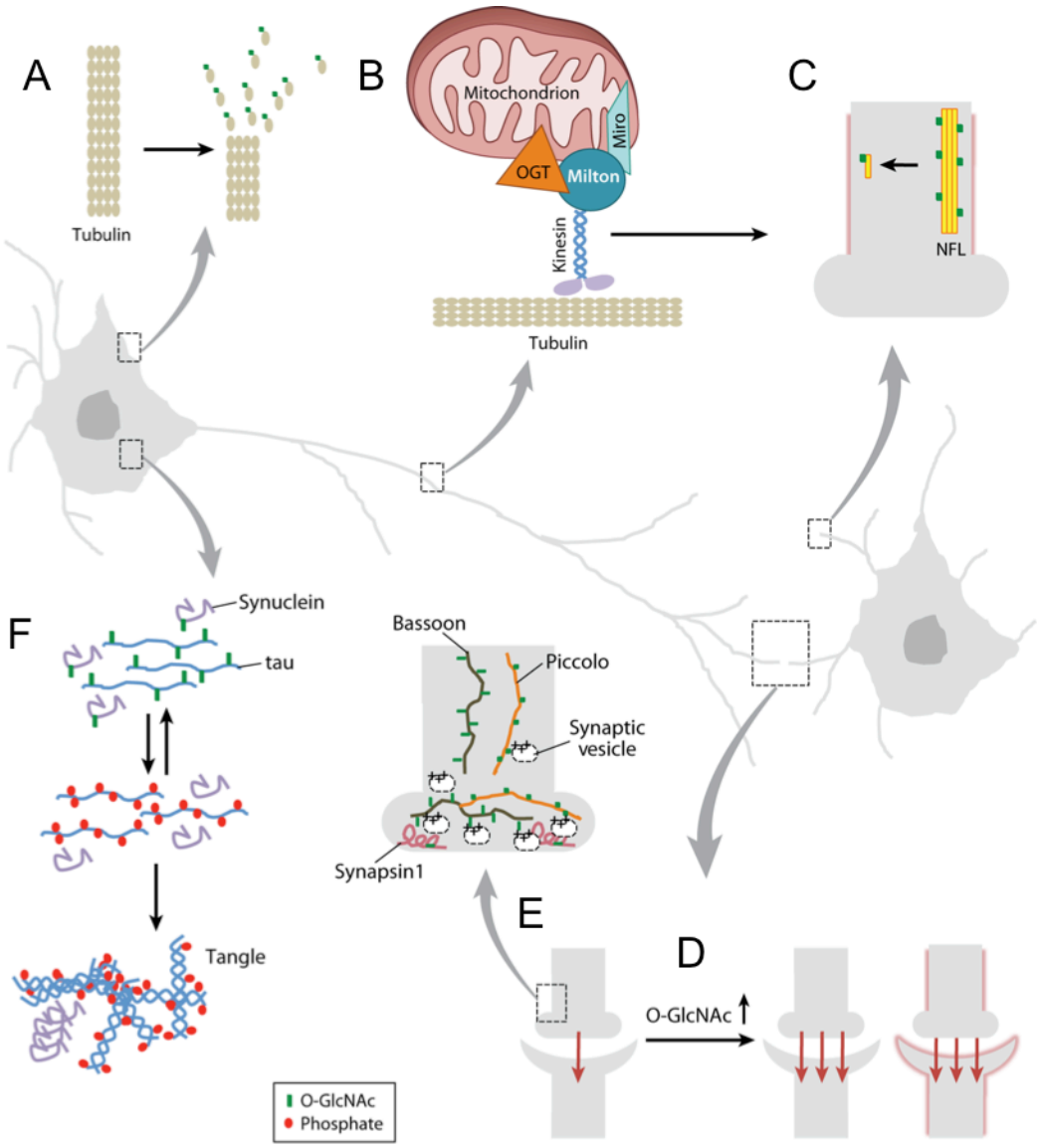


Figure 1.4. O-GlcNAc is important for proper neuronal function. (A) O-GlcNAcylation of tubulin is important in regulating microtubule formation [98]. (B) The protein milton, required for the axonal transport of mitochondria throughout the nervous system, is associated with OGT and miro, and is responsible for recruiting kinesin to the mitochondrial surface [106]. (C) O-GlcNAcylation of neurofilament proteins is important for filament assembly and network formation [107, 108]. (D) Changes in O-GlcNAc levels have been shown to regulate long-term potentiation. (E) Many proteins involved in neurotransmitter release, such as bassoon, synapsin-1, and piccolo, are heavily modified by O-GlcNAc in the synapse. (F) Adult brains are extensively O-GlcNAcylated and not significantly phosphorylated. However, in the brains of patients with Alzheimer's disease, tau becomes less O-GlcNAcylated and more extensively phosphorylated, leading to the formation of intraneuronal "tangles", which are characteristic of this disease [2]. (Adopted from Hart et al. [13])

1.5. Tools for studying and detecting O-GlcNAc

Chemical inhibitors

Several small-molecule inhibitors have been developed for the study of OGT and OGA activity. Currently, chemical inhibitors of OGT include alloxan, 5-thioglucosamine (5S-GlcNAc), and inhibitors with a oxobenzo[d]oxazole core structure. Alloxan is the most common inhibitor of OGT largely due to its commercial availability. It is an analog of uracil, and thus shows multiple nonspecific effects [109]. Gross and co-workers also identified several OGT inhibitors through a library screening that have been shown to work in different systems [18, 110-112]. Recently, Gloster and colleagues developed an OGT-specific inhibitor that relies on 'hijacking' the biosynthetic pathway of UDP-GlcNAc [113]. The sugar analogue of GlcNAc, per-O-acetylated 2-acetamido-2-deoxy-5-thio-D-glucopyranose (Ac-5SGlcNAc) is converted through the HBP into UDP-5SGlcNAc, which binds to OGT and inhibits its activity, thus decreasing O-GlcNAc levels in cells. Since its publication, this inhibitor has been tested on a range of cell lines, including HeLa [114], HepG2 [115] and pancreatic cancer cells [116].

Interestingly, there are more inhibitors available for OGA than for OGT. These include PUGNAc, Thiamet-G, GlcNAcstatin C, streptozotocin, NAG-thiazoline and NButGT. One of the inhibitors that was identified earliest is streptozotocin (STZ), which has been used for studying hyperglycemia models [117]. STZ functions as a GlcNAc analog and has several off-target effects. Another commonly used inhibitor of OGA is O-(2-acetamido-2-deoxy-D-glucopyranosylidene)amino-N-phenylcarbamate (PUGNAc). It functions as a transition-state analog for OGA [118, 119], but it also inhibits lysosomal hexosaminidases [120]. Yet, it has been used in several studies to increase cellular O-GlcNAc levels [121]. Dorfmueller and co-workers designed a nagstatin derivative based on the crystal structure of bacterial OGA, and called it GlcNAcstatin-C [122]. This inhibitor showed a high selectivity towards the bacterial hydrolase compared to the lysosomal hexosaminidases, but it showed weaker inhibition toward human OGA than PUGNAc. Finally, Vocadlo et al. developed the two OGA inhibitors, Thiamet-G and NAG-thiazoline, with the knowledge that the enzyme uses a two-step catalytic mechanism involving substrate-assisted catalysis. NAG-thiazoline resembles the oxazoline intermediate, making it a potent inhibitor of OGA [123]; however, it also showed strong binding for lysosomal hexosaminidases. Increasing bulky groups on the thiazoline ring led to the development of 1,2-dideoxy-2′-propyl-α-D-glucopyranoso-[2,1-D]-Δ2′-thiazoline (NButGT) with an increased selectivity for OGA but with loss of potency [118]. Finally, Vocadlo and co-workers developed an aminothiazoline (Thiamet-G), which showed higher inhibition of OGA than any of the other inhibitors, and was found to cross the blood-brain barrier in rats [124]. Similarly, oral treatment of JNPL3 tauopathy mice with Thiamet-G increased tau O-GlcNAc, hindered formation of tau aggregates and decreased neuronal cell loss, which demonstrates that Thiamet-G is a powerful tool for probing the functional role of O-GlcNAc [125].

Enrichment and detection strategies

Traditional methods for detecting and enriching O-GlcNAc glycosylation include the use of wheat germ agglutinin (WGA) lectin, O-GlcNAc-specific antibodies, chemoenzymatic tagging and radioactive labeling using β-1,4-galactosylatransferase (GalT), and metabolic labeling. Several decades ago, O-GlcNAc was first detected using tritiated UDP-galactose (UDP-[^{3}H]-galactose) [126, 127]. GalT transfers [^{3}H]-galactose to the GlcNAc residue attached onto the target protein, allowing detection by autoradiography. Although it greatly facilitated detection of O-GlcNAc-modified proteins at the time, this method is not specific to only O-GlcNAcylated substrates, as the GalT will attach [^{3}H]-galactose to any glycan with a terminal GlcNAc residue. So, it is important that samples are pre-treated with peptide: N-glycosidase F (PNGase F) before performing the GalT transfer reaction to remove N-linked GlcNAc glycosylated proteins. Additionally, tritium labeling suffers from low sensitivity, which requires lengthy exposure times, often ranging from several days to weeks [128]. Chemoenzymatic tagging is a similar technique that uses a mutant GalT to transfer unnatural substrates onto terminal GlcNAc [129]. This mutant GalT has a larger active site pocket, making it possible to fit an unnatural substrate containing either an azide or a ketone moiety at the carbon 2 position of UDP-galactose [6]. These chemically reactive tags then allow the attachment of a biotin derivative for highly selective enrichment of O-GlcNAcylated proteins or peptides. This approach has been widely used in the enrichment of O-GlcNAcylated peptides and proteins in conjunction with mass spectrometry analysis for proteomic profile of O-GlcNAc substrates [24, 34, 91, 94, 130-132]. It has also been used in the detection of O-GlcNAc-modified proteins by in-gel fluorescence [133], and for analysis of O-GlcNAc dynamics [24, 91, 96, 134, 135].

The WGA lectin is another useful tool for studying O-GlcNAc. It binds to GlcNAc residues, and has much higher affinity for heavily GlcNAc-containing proteins. Vosseller and colleagues developed a lectin weak affinity chromatography (LWAC) strategy that takes advantage of WGA's weak affinity to proteins with a small number of GlcNAc residues [92]. Although this low affinity makes LWAC technically challenging, several studies have used it to identify several O-GlcNAcylated proteins [21, 54, 93, 136]. Succinylation of WGA increases its specificity for GlcNAc, but it also reduces WGA's affinity for the sugar. Nevertheless, this other form of WGA is routinely used for the detection of the O-GlcNAc modification, usually coupled with western blotting [64, 77, 135]. O-GlcNAc antibodies are also commonly used for fast detection of O-GlcNAcylated proteins, but suffer from low affinity and specificity, making them inefficient for immunoprecipitation. They preferentially bind proteins with high abundance of O-GlcNAc residues. Teo et al. developed a panel of pan-specific O-GlcNAc monoclonal antibodies, but these antibodies have not been widely used [137].

The Bertozzi lab pioneered a metabolic labeling strategy for the study of protein glycosylation [138]. This strategy allows covalent labeling of glycoproteins in living cells and organisms, and can be harnessed for affinity capture or visualization of glycoproteins. It involves unnatural monosaccharides (azido sugars), which are very similar to natural sugars but bear a reactive chemical functionality, typically an azide. These azido sugars are introduced to cells, where they are processed and incorporated into glycoproteins in the place of natural sugars [139, 140]. Azides are small functional

groups that are metabolically stable, essentially inert in biological systems, and can selectively react with phosphines and alkynes via the Staudinger ligation and the [3+2] cycloaddition, respectively [141]. The metabolic labeling method offers an advantage over existing labeling methodologies because the glycoproteins that are labeled represent a population that has been newly synthesized, rather than those proteins present at steady-state levels. Vocadlo et al. first used this technology for the detection of O-GlcNAc in Jurkat cells [142]. Since then, the technology has been used for the detection and enrichment of O-GlcNAcylated proteins in many different cell lines [143-151].

Proteomics and site-mapping

Many of the tools discussed above have been used in conjunction with mass spectrometry (MS) methods to identify O-GlcNAcylated proteins and map their sites of glycosylation. Early MS approaches used conventional fragmentation methods such as collision-induced dissociation (CID). Though widely used for the identification of peptides and proteins, this fragmentation method poses a significant problem when attempting to analyze glycoproteins: CID utilizes high-energy collisions that often cause the cleavage of the glycosidic bond, as this bond is quite labile relative to the bonds in the peptide backbone. Nevertheless, many O-GlcNAcylated proteins were identified by coupling CID with methods such as β-elimination followed by Michael addition with dithiothreitol (DTT) or BEMAD. In BEMAD, the GlcNAc can be removed by mild β-elimination with DTT, thus converting the modified serine and threonine residues into their dehydrated equivalents [152]. The site is then determined by the loss of a water molecule [18 mass/charge (m/z)] within the sequence. Given that other modifications of serine and threonine can be eliminated, this approach can produce false-positives, making validation of the glycosylation necessary. A density-labeled DTT also allows for the comparative quantification of site occupancy of a given O-GlcNAcylated peptide between different samples [13]. Most importantly, the modified peptide has better fragmentation in CID relative to many other fragmentation methods. Recently, a higher energy form of collisional fragmentation was discovered, termed higher-energy collisional induced dissociation (HCD) [153]. HCD fragmentation is a type of CID technique specific to the orbitrap mass spectrometer, which allows for low-mass cutoff, high resolution ion detection, and increased ion fragments resulting in higher quality tandem MS spectra [154]. HCD tends to generate ions characteristic of N-acetylhexosamines (GlcNAc or GalNAc) oxonium ions at a low m/z region, which can serve as diagnostic tools for glycosylation [155], and it has therefore largely replaced CID when HCD technology is available.

Other fragmentation methods, such as electron-capture dissociation (ECD) and electron transfer dissociation (ETD), have greatly advanced our capability to detect and map O-GlcNAc sites. These fragmentation methods hold unique utility for the analysis of glycoproteins and glycopeptides, as these methods frequently conserve the labile glycosidic bond. This is accomplished by fragmenting proteins and peptides by the transfer of a single electron to induce fragmentation, avoiding the high energies associated with collisional dissociation. More than 1000 sites of O-GlcNAc modification

have been mapped by coupling these “softer” fragmentation methods with the enrichment methods discussed above [21, 54, 93, 94, 132, 136, 155, 156].

1.6. Conclusion

O-GlcNAc is a dynamic and reversible post-translational modification that has gained a lot of interest in recent years due to its influence in many of the biological processes in the cell. Over the past few years, the development of new tools for the study of O-GlcNAc has accelerated our understanding of the function of this modification in regulating diverse cellular processes, such as transcription and neuronal development. The intrinsic difficulties associated with the study of this PTM pose significant obstacles to investigators around the world; however, the continuation of technological innovation and the development of chemical tools have catalyzed research towards understanding this vital modification.

Human embryonic stem cells can be changed into virtually any cell type in the adult body, and thus, have the potential to cure a vast majority of existing human disorders. Studies of the O-GlcNAc modification may lead to a greatly increased understanding of how stem cells retain their ability to be changed into other cell types, and also how the fate of stem cells is decided upon differentiation. Both are critical areas that need to be explored to enable modern regenerative medicine to realize its full potential as tool for the treatment of human diseases, such as Alzheimer’s, Parkinson, and Batten disease.

1.7. References

1. Wang, Y.-C., S.E. Peterson, and J.F. Loring, *Protein post-translational modifications and regulation of pluripotency in human stem cells.* Cell Res, 2014. **24**(2): p. 143-160.

2. Hart, G.W. and Y. Akimoto, *The O-GlcNAc Modification*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009, Cold Spring Harbor Laboratory Press The Consortium of Glycobiology Editors, La Jolla, California: Cold Spring Harbor (NY).

3. Kearse, K.P. and G.W. Hart, *Lymphocyte activation induces rapid changes in nuclear and cytoplasmic glycoproteins.* Proc Natl Acad Sci U S A, 1991. **88**(5): p. 1701-5.

4. Kneass, Z.T. and R.B. Marchase, *Neutrophils exhibit rapid agonist-induced increases in protein-associated O-GlcNAc.* J Biol Chem, 2004. **279**(44): p. 45759-65.

5. Yang, X.Y., et al., *Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance.* Nature, 2008. **451**(7181): p. 964-U1.

6. Rexach, J.E., P.M. Clark, and L.C. Hsieh-Wilson, *Chemical approaches to understanding O-GlcNAc glycosylation in the brain.* Nat Chem Biol, 2008. **4**(2): p. 97-106.

7. Gong, C.X., F. Liu, and K. Iqbal, *O-GlcNAc cycling modulates neurodegeneration.* Proc Natl Acad Sci U S A, 2012. **109**(43): p. 17319-20.

8. Hanover, J.A., M.W. Krause, and D.C. Love, *The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine.* Biochimica et Biophysica Acta, 2009(0006-3002): p. ahead of print.

9. Butkinaree, C., K. Park, and G.W. Hart, *O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with\ phosphorylation to regulate signaling and transcription in response to nutrients \ and stress.* Biochimica et Biophysica Acta, 2009(0006-3002): p. ahead of print.

10. Hanover, J.A., M.W. Krause, and D.C. Love, *The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine.* Biochim Biophys Acta, 2010. **1800**(2): p. 80-95.

11. Love, D.C., M.W. Krause, and J.A. Hanover, *O-GlcNAc cycling: emerging roles in development and epigenetics.* Semin Cell Dev Biol, 2010. **21**(6): p. 646-54.

12. Hu, P., S. Shimoji, and G.W. Hart, *Site-specific interplay between O-GlcNAcylation and phosphorylation in cellular regulation.* FEBS Lett, 2010. **584**(12): p. 2526-38.

13. Hart, G.W., et al., *Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease.* Annu Rev Biochem, 2011. **80**: p. 825-58.

14. Bond, M.R. and J.A. Hanover, *O-GlcNAc cycling: a link between metabolism and chronic disease.* Annu Rev Nutr, 2013. **33**: p. 205-29.

15. Kreppel, L.K., M.A. Blomberg, and G.W. Hart, *Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats.* J Biol Chem, 1997. **272**(14): p. 9308-15.

16. Ryu, I.-H. and S.-I. Do, *Denitrosylation of S-nitrosylated OGT is triggered in LPS-stimulated innate immune response.* Biochemical and Biophysical Research Communications, 2011. **408**(1): p. 52-57.

17. Hanover, J.A., M.W. Krause, and D.C. Love, *Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation.* Nat Rev Mol Cell Biol, 2012. **13**(5): p. 312-21.

18. Caldwell, S.A., et al., *Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1.* Oncogene, 2010. **29**(19): p. 2831-42.

19. Park, M.J., et al., *High glucose-induced O-GlcNAcylated carbohydrate response element-binding protein (ChREBP) mediates mesangial cell lipogenesis and fibrosis: the possible role in the development of diabetic nephropathy.* J Biol Chem, 2014. **289**(19): p. 13519-30.

20. Zou, L., et al., *Glucose deprivation induced increase in protein O-GlcNAcylation in cardiomyocytes is calcium dependent.* Journal of Biological Chemistry, 2012.

21. Trinidad, J.C., et al., *Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse.* Mol Cell Proteomics, 2012. **11**(8): p. 215-29.

22. Olivier-Van Stichelen, S., et al., *O-GlcNAcylation stabilizes beta-catenin through direct competition with phosphorylation at threonine 41.* Faseb j, 2014.

23. Cole, R.N. and G.W. Hart, *Glycosylation Sites Flank Phosphorylation Sites on Synapsin I.* Journal of Neurochemistry, 1999. **73**(1): p. 418-428.

24. Rexach, J.E., et al., *Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags.* Nat Chem Biol, 2010. **6**(9): p. 645-51.

25. Gawlowski, T., et al., *Modulation of dynamin-related protein 1 (DRP1) function by increased O-linked-beta-N-acetylglucosamine modification (O-GlcNAc) in cardiac myocytes.* J Biol Chem, 2012. **287**(35): p. 30024-34.

26. Ranuncolo, S.M., et al., *Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo.* J Biol Chem, 2012. **287**(28): p. 23549-61.

27. Dias, W.B., W.D. Cheung, and G.W. Hart, *O-GlcNAcylation of kinases.* Biochem Biophys Res Commun, 2012. **422**(2): p. 224-8.

28. Tarrant, M.K., et al., *Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis.* Nat Chem Biol, 2012. **8**(3): p. 262-9.

29. Wang, S., et al., *Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling.* PLoS One, 2012. **7**(5): p. e37427.

30. Dias, W.B., et al., *Regulation of calcium/calmodulin-dependent kinase IV by O-GlcNAc modification.* J Biol Chem, 2009. **284**(32): p. 21327-37.

31. Bullen, J.W., et al., *Cross-talk between two essential nutrient-sensitive enzymes: O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK).* J Biol Chem, 2014. **289**(15): p. 10592-606.

32. Xu, Q., et al., *AMPK regulates histone H2B O-GlcNAcylation.* Nucleic Acids Res, 2014. **42**(9): p. 5594-604.

33. Slawson, C., et al., *A mitotic GlcNAcylation/phosphorylation signaling complex alters the posttranslational state of the cytoskeletal protein vimentin.* Mol Biol Cell, 2008. **19**(10): p. 4130-40.

34. Wang, Z., et al., *Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis.* Sci Signal, 2010. **3**(104): p. ra2.

35. Kaasik, K., et al., *Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock.* Cell Metab, 2013. **17**(2): p. 291-302.

36. Sun, L. and Z.J. Chen, *The novel functions of ubiquitination in signaling.* Current Opinion in Cell Biology, 2004. **16**(2): p. 119-126.

37. Guinez, C., et al., *Protein ubiquitination is modulated by O-GlcNAc glycosylation.* Faseb j, 2008. **22**(8): p. 2901-11.

38. Yang, W.H., et al., *Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability.* Nat Cell Biol, 2006. **8**(10): p. 1074-1083.

39. Fujiki, R., et al., *GlcNAcylation of histone H2B facilitates its monoubiquitination.* Nature, 2011. **480**(7378): p. 557-60.

40. Jackson, S.P. and R. Tjian, *O-glycosylation of eukaryotic transcription factors: Implications for mechanisms of transcriptional regulation.* Cell, 1988. **55**(1): p. 125-133.

41. Yang, X., et al., *O-linkage of N-acetylglucosamine to Sp1 activation domain inhibits its transcriptional capability.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6611-6.

42. Majumdar, G., et al., *Insulin dynamically regulates calmodulin gene expression by sequential o-glycosylation and phosphorylation of sp1 and its subcellular compartmentalization in liver cells.* J Biol Chem, 2006. **281**(6): p. 3642-50.

43. Yang, X., F. Zhang, and J.E. Kudlow, *Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression.* Cell, 2002. **110**(1): p. 69-80.

44. Han, I. and J.E. Kudlow, *Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility.* Mol Cell Biol, 1997. **17**(5): p. 2550-8.

45. Roos, M.D., et al., *O glycosylation of an Sp1-derived peptide blocks known Sp1 protein interactions.* Mol Cell Biol, 1997. **17**(11): p. 6472-80.

46. Lim, K. and H.I. Chang, *O-GlcNAcylation of Sp1 interrupts Sp1 interaction with NF-Y.* Biochem Biophys Res Commun, 2009. **382**(3): p. 593-7.

47. Lim, K. and H.I. Chang, *O-GlcNAc inhibits interaction between Sp1 and Elf-1 transcription factors.* Biochem Biophys Res Commun, 2009. **380**(3): p. 569-74.

48. Lim, K. and H.I. Chang, *O-GlcNAc modification of Sp1 inhibits the functional interaction between Sp1 and Oct1.* FEBS Lett, 2009. **583**(3): p. 512-20.

49. McDonel, P., et al., *Sin3a is essential for the genome integrity and viability of pluripotent cells.* Developmental Biology, 2012. **363**(1): p. 62-73.

50. Daou, S., et al., *Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2747-52.

51. Capotosti, F., et al., *O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1.* Cell, 2011. **144**(3): p. 376-88.

52. Lazarus, M.B., et al., *HCF-1 is cleaved in the active site of O-GlcNAc transferase.* Science, 2013. **342**(6163): p. 1235-9.

53. Ruan, H.B., et al., *O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability.* Cell Metab, 2012. **16**(2): p. 226-37.

54. Myers, S.A., B. Panning, and A.L. Burlingame, *Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9490-5.

55. McDonel, P., I. Costello, and B. Hendrich, *Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development.* Int J Biochem Cell Biol, 2009. **41**(1): p. 108-16.

56. Bowe, D.B., et al., *O-GlcNAc integrates the proteasome and transcriptome to regulate nuclear hormone receptors.* Mol Cell Biol, 2006. **26**(22): p. 8539-50.

57. Kelly, W.G., M.E. Dahmus, and G.W. Hart, *RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc.* J Biol Chem, 1993. **268**(14): p. 10416-24.

58. Di Croce, L. and K. Helin, *Transcriptional regulation by Polycomb group proteins.* Nat Struct Mol Biol, 2013. **20**(10): p. 1147-55.

59. Sinclair, D.A.R., et al., *Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc).* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(32): p. 13427-13432.

60. Gambetta, M.C., K. Oktaba, and J. Muller, *Essential role of the glycosyltransferase sxc/Ogt in polycomb repression.* Science, 2009. **325**(5936): p. 93-6.

61. Illingworth, R.S., et al., *PRC1 and PRC2 are not required for targeting of H2A.Z to developmental genes in embryonic stem cells.* PLoS One, 2012. **7**(4): p. e34848.

62. Shi, F.T., et al., *Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells.* J Biol Chem, 2013. **288**(29): p. 20776-84.

63. Deplus, R., et al., *TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS.* Embo j, 2013. **32**(5): p. 645-55.

64. Vella, P., et al., *Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.* Mol Cell, 2013. **49**(4): p. 645-56.

65. Chen, Q., et al., *TET2 promotes histone O-GlcNAcylation during gene transcription.* Nature, 2013. **493**(7433): p. 561-4.

66. Ito, R., et al., *TET3-OGT interaction increases the stability and the presence of OGT in chromatin.* Genes Cells, 2014. **19**(1): p. 52-65.

67. Zhang, Q., et al., *Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT).* J Biol Chem, 2014. **289**(9): p. 5986-96.

68. Pastor, W.A., L. Aravind, and A. Rao, *TETonic shift: biological roles of TET proteins in DNA demethylation and transcription.* Nat Rev Mol Cell Biol, 2013. **14**(6): p. 341-356.

69. Sakabe, K., Z. Wang, and G.W. Hart, *Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code.* Proc Natl Acad Sci U S A, 2010. **107**(46): p. 19915-20.

70. Fong, J.J., et al., *beta-N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3.* J Biol Chem, 2012. **287**(15): p. 12195-203.

71. Shafi, R., et al., *The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny.* Proceedings of the National Academy of Sciences, 2000. **97**(11): p. 5735-5739.

72. O'Donnell, N., et al., *Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability.* Mol Cell Biol, 2004. **24**(4): p. 1680-90.

73. Webster, D.M., et al., *O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development.* BMC Dev Biol, 2009. **9**: p. 28.

74. Love, D.C., et al., *Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7413-8.

75. Forsythe, M.E., et al., *Caenorhabditis elegans ortholog of a diabetes susceptibility locus: oga-1 (O-GlcNAcase) knockout impacts O-GlcNAc cycling, metabolism, and dauer.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11952-7.

76. Hanover, J.A., et al., *A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11266-71.

77. Jang, H., et al., *O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network.* Cell Stem Cell, 2012. **11**(1): p. 62-74.

78. van den Berg, D.L., et al., *An Oct4-centered protein interaction network in embryonic stem cells.* Cell Stem Cell, 2010. **6**(4): p. 369-81.

79. Pardo, M., et al., *An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease.* Cell Stem Cell, 2010. **6**(4): p. 382-395.

80. Speakman, C.M., et al., *Elevated O-GlcNAc levels activate epigenetically repressed genes and delay mouse ES cell differentiation without affecting naïve to primed cell transition.* STEM CELLS, 2014: p. n/a-n/a.

81. Maury, J.J., et al., *Excess of O-linked N-acetylglucosamine modifies human pluripotent stem cell differentiation.* Stem Cell Res, 2013. **11**(2): p. 926-37.

82. Kim, H.S., et al., *Excessive O-GlcNAcylation of proteins suppresses spontaneous cardiogenesis in ES cells.* FEBS Lett, 2009. **583**(15): p. 2474-8.

83. Mizuguchi-Hata, C., et al., *Quantitative regulation of nuclear pore complex proteins by O-GlcNAcylation.* Biochim Biophys Acta, 2013. **1833**(12): p. 2682-9.

84. Ishihara, K., et al., *Characteristic increase in nucleocytoplasmic protein glycosylation by O-GlcNAc in 3T3-L1 adipocyte differentiation.* Biochem Biophys Res Commun, 2010. **398**(3): p. 489-94.

85. Andres-Bergos, J., et al., *The increase in O-linked N-acetylglucosamine protein modification stimulates chondrogenic differentiation both in vitro and in vivo.* J Biol Chem, 2012. **287**(40): p. 33615-28.

86. Gao, Y., et al., *Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain.* J Biol Chem, 2001. **276**(13): p. 9838-45.

87. Liu, K., et al., *Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism.* Journal of Neurochemistry, 2004. **89**(4): p. 1044-1055.

88. Yanagisawa, M. and R.K. Yu, *O-linked beta-N-acetylglucosaminylation in mouse embryonic neural precursor cells.* J Neurosci Res, 2009. **87**(16): p. 3535-45.

89. Francisco, H., et al., *O-GLcNAc post-translational modifications regulate the entry of neurons into an axon branching program.* Dev Neurobiol, 2009. **69**(2-3): p. 162-73.

90. Khidekel, N., et al., *Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(36): p. 13132-13137.

91. Khidekel, N., et al., *Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics.* Nat Chem Biol, 2007. **3**(6): p. 339-48.

92. Vosseller, K., et al., *O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry.* Mol Cell Proteomics, 2006. **5**(5): p. 923-34.

93. Chalkley, R.J., et al., *Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides.* Proc Natl Acad Sci U S A, 2009. **106**(22): p. 8894-9.

94. Alfaro, J.F., et al., *Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets.* Proc Natl Acad Sci U S A, 2012. **109**(19): p. 7280-5.

95. Lamarre-Vincent, N. and L.C. Hsieh-Wilson, *Dynamic Glycosylation of the Transcription Factor CREB: A Potential Role in Gene Regulation.* Journal of the American Chemical Society, 2003. **125**(22): p. 6612-6613.

96. Rexach, J.E., et al., *Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation.* Nat Chem Biol, 2012. **8**(3): p. 253-61.

97. Griffith, L.S., M. Mathes, and B. Schmitz, *β-Amyloid precursor protein is modified with O-linked N-acetylglucosamine.* Journal of Neuroscience Research, 1995. **41**(2): p. 270-278.

98. Ji, S., et al., *O-GlcNAcylation of tubulin inhibits its polymerization.* Amino Acids, 2011. **40**(3): p. 809-18.

99. Liu, F., et al., *O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(29): p. 10804-10809.

100. Tallent, M.K., et al., *In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation.* J Biol Chem, 2009. **284**(1): p. 174-81.

101. Griffith, L.S. and B. Schmitz, *O-linked N-acetylglucosamine levels in cerebellar neurons respond reciprocally to pertubations of phosphorylation.* Eur J Biochem, 1999. **262**(3): p. 824-31.

102. Robertson, L.A., K.L. Moya, and K.C. Breen, *The potential role of tau protein O-glycosylation in Alzheimer's disease.* J Alzheimers Dis, 2004. **6**(5): p. 489-95.

103. Lefebvre, T., et al., *Evidence of a balance between phosphorylation and O-GlcNAc glycosylation of Tau proteins--a role in nuclear localization.* Biochim Biophys Acta, 2003. **1619**(2): p. 167-76.

104. Lefebvre, T., et al., *Evidence of a balance between phosphorylation and O-GlcNAc glycosylation of Tau proteins—a role in nuclear localization.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2003. **1619**(2): p. 167-176.

105. Ballatore, C., V.M.Y. Lee, and J.Q. Trojanowski, *Tau-mediated neurodegeneration in Alzheimer's disease and related disorders.* Nat Rev Neurosci, 2007. **8**(9): p. 663-672.

106. Glater, E.E., et al., *Axonal transport of mitochondria requires milton to recruit kinesin heavy chain and is light chain independent.* J Cell Biol, 2006. **173**(4): p. 545-57.

107. Dong, D.L., et al., *Glycosylation of mammalian neurofilaments. Localization of multiple O-linked N-acetylglucosamine moieties on neurofilament polypeptides L and M.* J Biol Chem, 1993. **268**(22): p. 16679-87.

108. Dong, D.L., et al., *Cytoplasmic O-GlcNAc modification of the head domain and the KSP repeat motif of the neurofilament protein neurofilament-H.* J Biol Chem, 1996. **271**(34): p. 20845-52.

109. Ostrowski, A. and D.M. van Aalten, *Chemical tools to probe cellular O-GlcNAc signalling.* Biochem J, 2013. **456**(1): p. 1-12.

110. Gross, B.J., B.C. Kraybill, and S. Walker, *Discovery of O-GlcNAc transferase inhibitors.* J Am Chem Soc, 2005. **127**(42): p. 14588-9.

111. Dehennaut, V., et al., *O-linked N-acetylglucosaminyltransferase inhibition prevents G2/M transition in Xenopus laevis oocytes.* J Biol Chem, 2007. **282**(17): p. 12527-36.

112. Itkonen, H.M., et al., *O-GlcNAc transferase integrates metabolic pathways to regulate the stability of c-MYC in human prostate cancer cells.* Cancer Res, 2013. **73**(16): p. 5277-87.

113. Gloster, T.M., et al., *Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells.* Nat Chem Biol, 2011. **7**(3): p. 174-81.

114. Olivier-Van Stichelen, S., et al., *Serum-stimulated cell cycle entry promotes ncOGT synthesis required for cyclin D expression.* Oncogenesis, 2012. **1**: p. e36.

115. Perez-Cervera, Y., et al., *Insulin signaling controls the expression of O-GlcNAc transferase and its interaction with lipid microdomains.* Faseb j, 2013. **27**(9): p. 3478-86.

116. Ma, Z., D.J. Vocadlo, and K. Vosseller, *Hyper-O-GlcNAcylation is anti-apoptotic and maintains constitutive NF-kappaB activity in pancreatic cancer cells.* J Biol Chem, 2013. **288**(21): p. 15121-30.

117. Rakieten, N., M.L. Rakieten, and M.V. Nadkarni, *Studies on the diabetogenic action of streptozotocin (NSC-37917).* Cancer Chemother Rep, 1963. **29**: p. 91-8.

118. Whitworth, G.E., et al., *Analysis of PUGNAc and NAG-thiazoline as transition state analogues for human O-GlcNAcase: mechanistic and structural insights into inhibitor selectivity and transition state poise.* J Am Chem Soc, 2007. **129**(3): p. 635-44.

119. Lameira, J., et al., *Quantum mechanical/molecular mechanical molecular dynamics simulation of wild-type and seven mutants of CpNagJ in complex with PUGNAc.* J Phys Chem B, 2010. **114**(20): p. 7029-36.

120. Macauley, M.S., et al., *O-GlcNAcase uses substrate-assisted catalysis: kinetic analysis and development of highly selective mechanism-inspired inhibitors.* J Biol Chem, 2005. **280**(27): p. 25313-22.

121. Banerjee, P.S., G.W. Hart, and J.W. Cho, *Chemical approaches to study O-GlcNAcylation.* Chem Soc Rev, 2013. **42**(10): p. 4345-57.

122. Dorfmueller, H.C., et al., *GlcNAcstatin: a picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-glcNAcylation levels.* J Am Chem Soc, 2006. **128**(51): p. 16484-5.

123. Mark, B.L., et al., *Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase.* J Biol Chem, 2001. **276**(13): p. 10330-7.

124. Yuzwa, S.A., et al., *A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo.* Nat Chem Biol, 2008. **4**(8): p. 483-90.

125. Yuzwa, S.A., et al., *Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation.* Nat Chem Biol, 2012. **8**(4): p. 393-9.

126. Holt, G.D. and G.W. Hart, *The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage, O-linked GlcNAc.* J Biol Chem, 1986. **261**(17): p. 8049-57.

127. Torres, C.R. and G.W. Hart, *Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc.* J Biol Chem, 1984. **259**(5): p. 3308-17.

128. Ma, J. and G.W. Hart, *O-GlcNAc profiling: from proteins to proteomes.* Clin Proteomics, 2014. **11**(1): p. 8.

129. Khidekel, N., et al., *A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications.* J Am Chem Soc, 2003. **125**(52): p. 16162-3.

130. Tai, H.C., et al., *Parallel identification of O-GlcNAc-modified proteins from cell lysates.* J Am Chem Soc, 2004. **126**(34): p. 10500-1.

131. Wang, Z., et al., *Site-specific GlcNAcylation of human erythrocyte proteins: potential biomarker(s) for diabetes.* Diabetes, 2009. **58**(2): p. 309-17.

132. Wang, Z., et al., *Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry.* Mol Cell Proteomics, 2010. **9**(1): p. 153-60.

133. Clark, P.M., et al., *Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins.* J Am Chem Soc, 2008. **130**(35): p. 11576-7.

134. Yi, W., et al., *Phosphofructokinase 1 glycosylation regulates cell growth and metabolism.* Science, 2012. **337**(6097): p. 975-80.

135. Ramakrishnan, P., et al., *Activation of the transcriptional function of the NF-kappaB protein c-Rel by O-GlcNAc glycosylation.* Sci Signal, 2013. **6**(290): p. ra75.

136. Nagel, A.K., et al., *Identification of O-linked N-acetylglucosamine (O-GlcNAc)-modified osteoblast proteins by electron transfer dissociation tandem mass spectrometry reveals proteins critical for bone formation.* Mol Cell Proteomics, 2013. **12**(4): p. 945-55.

137. Teo, C.F., et al., *Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc.* Nat Chem Biol, 2010. **6**(5): p. 338-43.

138. Mahal, L.K., K.J. Yarema, and C.R. Bertozzi, *Engineering Chemical Reactivity on Cell Surfaces Through Oligosaccharide Biosynthesis.* Science, 1997. **276**(5315): p. 1125-1128.

139. Saxon, E. and C.R. Bertozzi, *Cell Surface Engineering by a Modified Staudinger Reaction.* Science, 2000. **287**(5460): p. 2007-2010.

140. Laughlin, S.T. and C.R. Bertozzi, *Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation.* Nat. Protocols, 2007. **2**(11): p. 2930-2944.

141. Prescher, J.A. and C.R. Bertozzi, *Chemical Technologies for Probing Glycans.* Cell, 2006. **126**(5): p. 851-854.

142. Vocadlo, D.J., et al., *A chemical approach for identifying O-GlcNAc-modified proteins in cells.* Proceedings of the National Academy of Sciences, 2003. **100**(16): p. 9116-9121.

143. Zaro, B.W., et al., *Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8146-51.

144. Boyce, M., et al., *Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway.* Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3141-6.

145. Yu, S.-H., et al., *Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners.* Proceedings of the National Academy of Sciences, 2012. **109**(13): p. 4834-4839.

146. Palaniappan, K.K., et al., *A chemical glycoproteomics platform reveals O-GlcNAcylation of mitochondrial voltage-dependent anion channel 2.* Cell Rep, 2013. **5**(2): p. 546-52.

147. Sprung, R., et al., *Tagging-via-substrate strategy for probing O-GlcNAc modified proteins.* J Proteome Res, 2005. **4**(3): p. 950-7.

148. Nandi, A., et al., *Global identification of O-GlcNAc-modified proteins.* Anal Chem, 2006. **78**(2): p. 452-8.

149. Gurcel, C., et al., *Identification of new O-GlcNAc modified proteins using a click-chemistry-based tagging.* Anal Bioanal Chem, 2008. **390**(8): p. 2089-97.

150. Hahne, H., et al., *Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry.* J Proteome Res, 2013. **12**(2): p. 927-36.

151. Teo, C.F. and L. Wells, *Monitoring Protein O-GlcNAc Status via Metabolic Labeling and Copper-free Click Chemistry.* Anal Biochem, 2014.

152. Greis, K.D., et al., *Selective detection and site-analysis of O-GlcNAc-modified glycopeptides by beta-elimination and tandem electrospray mass spectrometry.* Anal Biochem, 1996. **234**(1): p. 38-49.

153. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.

154. Jedrychowski, M.P., et al., *Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics.* Mol Cell Proteomics, 2011. **10**(12): p. M111.009910.

155. Zhao, P., et al., *Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment.* J Proteome Res, 2011. **10**(9): p. 4088-104.

156. Myers, S.A., et al., *Electron transfer dissociation (ETD): the mass spectrometric breakthrough essential for O-GlcNAc protein site assignments-a study of the O-GlcNAcylated protein host cell factor C1.* Proteomics, 2013. **13**(6): p. 982-91.

Chapter 2

Metabolic incorporation of unnatural monosaccharides by human embryonic stem cells

Chapter 2. Metabolic incorporation of unnatural monosaccharides by human embryonic stem cells[1]

2.1. Introduction

Changes in glycosylation are thought to accompany the molecular transformations necessary for the development of human embryonic stem cells (hESCs). Several of the markers used to characterize hESC's pluripotency and differentiation are glycosylated, such as stage-specific embryonic antigens, SSEA-3 and SSEA-4, and polysialylated neural cell adhesion molecule (PSA-NCAM) [1, 2]. Recent studies suggested that cell surface glycosylation might play an important role in hESC biology [3]. However, a more dynamic form of glycosylation, known as the O-GlcNAc modification, has also been shown to be important for the maintenance and differentiation of hESCs [4-6]. *N*-acetylglucosamine (GlcNAc) is reversibly attached by the enzyme O-GlcNAc transferase (OGT) to serine and threonine residues of nuclear and cytoplasmic proteins. Although, several pluripotency transcription factors and developmental regulators have been shown to be O-GlcNAcylated, a comprehensive proteomic study of O-GlcNAcylated proteins has not been carried out in hESCs.

The Bertozzi lab has developed a metabolic labeling strategy for the study of glycoproteins *in vivo* and *in vitro* [7-13]. Metabolic labeling relies on the use of unnatural monosaccharides (azido sugars), which are structurally very similar to natural sugars but bear a reactive chemical functionality (the azide). Azido sugars are taken up by cells where they are processed and incorporated into glycoproteins in the place of natural sugars (Figure 2.1) [14, 15]. The azide functionality can then be used to selectively label glycoproteins with affinity tags, by the addition of reactive partners bearing a phosphine (e.g. phosphine-FLAG) or an alkyne (e.g. alkyne-biotin). This methodology allows covalent labeling of glycoproteins in living cells and organisms with minimal perturbation to physiological processes, and has been used by the Bertozzi lab and others to study O-GlcNAc [10, 12, 16-20].

In this chapter, we used the metabolic labeling strategy, as a means for enrichment of O-GlcNAc modified proteins, coupled with mass spectrometry (MS) analysis, to identify O-GlcNAcylated proteins in hESCs. We used a peracetylated azido analog of *N*-acetylgalactosamine (GalNAc) termed Ac_4GalNAz to metabolically label O-GlcNAcylated proteins. In prior work, we showed that metabolic cross-talk between O-GlcNAcylation and the GalNAc salvage pathway could be used to label O-GlcNAcylated proteins with Ac_4GalNAz [12, 17]. We compared several affinity purification strategies for the enrichment of nuclear and cytoplasmic O-GlcNAcylated substrates, and found that affinity purification with biotin resulted in the best capture of the metabolically labeled O-GlcNAcylated proteins. We were able to identify more than 150 O-GlcNAcylated proteins, with several already known to be O-GlcNAc modified. Amongst the O-GlcNAcylated proteins identified included transcription factors, nuclear pore complex proteins, and histones. Altogether, our results demonstrate the utility of this approach for studying O-GlcNAc signaling in hESCs.

[1] David Spiciarich, Brian Smart, Phung Gip, and Sarah Hubbard contributed to the work described in this chapter.

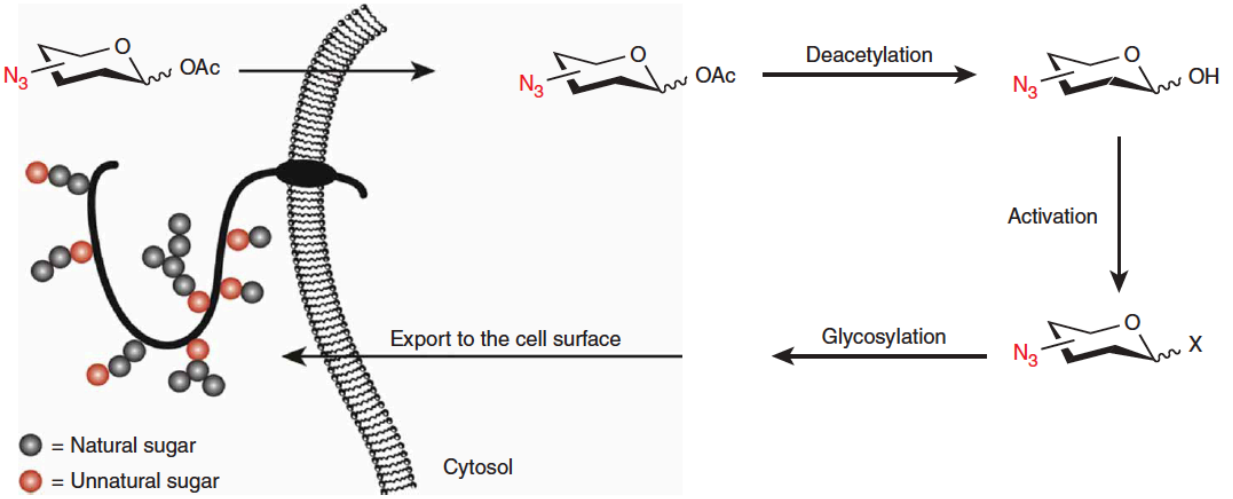


Figure 2.1. Metabolic labeling with unnatural sugars. Per-O-acetylated azido sugars passively diffuse across the cell membrane, and once in the cytosol, the acetyl groups are removed by intracellular esterases. Then, the azido sugar is activated to a nucleotide sugar donor [X = uridine diphosphate for GalNAz (UDP-GalNAz)]. UDP-GalNAz is attached onto glycans in the secretory pathway and exported to the cell surface, or it can be converted into UDP-GlcNAz with the enzyme UDP-galactose-4-epimerase (GALE) (not shown). OGT can accept UDP-GlcNAz as the nucleotide sugar donor, and transfer GlcNAz onto substrates, thereby substituting endogenous GlcNAc with the azido sugar and forming O-GlcNAz. (Adopted from Laughlin et al. [14])

2.2. Results and Discussion

Human embryonic stem cells incorporate per-O-acetylated GalNAz into O-GlcNAz

First, we wanted to test whether hESCs were capable of incorporating the azido sugar, per-O-acetylated GalNAz (Ac_4GalNAz), and substitute it for endogenous nuclear/cytosolic GlcNAc. Acetylation of the azido sugar is necessary for efficient diffusion through the cell membrane; otherwise metabolic incorporation of the azido sugar is minimal. Once inside the cell, GalNAz is phosphorylated followed by activation to the UDP donor sugar. The activated UDP-GalNAz is a substrate for the UDP-galactose-4-epimerase (GALE), which normally catalyzes the interconversion of UDP-GalNAc to UDP-GlcNAc in the cell. UDP-GlcNAz is then accepted by OGT, and used to transfer GlcNAz onto nuclear and cytosolic proteins. We cultured hESCs in the presence of 50 μM Ac_4GalNAz for 72 hours (Figure 2.2). Then, we harvested and lysed the cells, and reacted azido-labeled proteins via the Staudinger ligation, using a phosphine-FLAG probe, which selectively reacts with azides [21]. As shown in Figure 2.3, phosphine-FLAG labeling was dependent on the presence of Ac_4GalNAz, and labeling was depleted after treatment with a recombinant-purified bacterial O-GlcNAcase (OGA) homologue, which selectively removed the GlcNAz from proteins. The remaining FLAG signal in the sample treated with OGA and labeled with

Ac_4GalNAz and phosphine-FLAG was from the labeling of cell surface glycans with Ac_4GalNAz and phosphine-FLAG since the samples were not fractionated into nuclear/cytosolic extracts, and human OGA only hydrolyzes O-GlcNAc glysidic bonds. Additionally, we observed an increase in O-GlcNAc signal in samples that had been treated with Ac_4GalNAz, compared to non-treated samples. It is possible that the addition of the unnatural sugar onto proteins is having an inhibitory effect on human OGA, thus causing a slight increase in O-GlcNAc levels. Nevertheless, the bacterial OGA homologue was able to cleave off both the endogenous and unnatural sugars with comparable efficiencies, as shown by western blot analysis. Overall, these results suggest that GalNAz is being incorporated into O-GlcNAz once it is metabolized inside the cells. In addition, prolonged exposure of hESCs to Ac_4GalNAz had no apparent toxic effects, and the cells appeared normal in all respects, including morphology and growth rate.

Comparing strategies for affinity purification of O-GlcNAzylated proteins

Next, we tested the efficiency of three affinity handles (FLAG, FLAG-His_6 and biotin) in capturing nuclear azido-labeled glycoproteins. We treated hESCs with Ac_4GalNAz for 3 days, isolated nuclear and cytoplasmic fractions, and labeled azidoglycoproteins with a (1) phosphine-FLAG, (2) alkyne-FLAG-His_6, or (3) an alkyne-biotin probe. In the first strategy, we purified the FLAG-labeled azidoglycoproteins using a monoclonal mouse antibody against the epitope, followed by capture on Protein A/G resin. The eluted sample was then resolved by molecular weight using SDS-PAGE. The gels were stained with SimplyBlue SafeStain, and individual bands of approximately 1 - 2 mm in width were cut manually, washed, reduced and alkylated, and digested with trypsin. Each digest was de-salted and concentrated prior to liquid chromatographic separation on C_{18} resin in-line with an LTQ XL mass spectrometer. Following tandem MS analysis and database search, we were able to compare the FLAG affinity purification of hESCs treated with or without Ac_4GalNAz. Based on the database search, we were able to selectively affinity capture 19 azidoglycoproteins. Unfortunately, we identified 149 proteins in the non-Ac_4GalNAz treated hESCs, indicating a poor affinity enrichment of the azidoglycoproteins using phosphine-FLAG (Figure 2.4A). Given that one-step affinity purification resulted in background contamination, we decided to try a two-step purification. So, we performed affinity enrichment of the azidoglycoproteins by reacting Ac_4GalNAz-labeled cells with an alkyne-FLAG-His_6 probe via the [3+2] cycloaddition [22]. We purified the labeled azidoglycoproteins via sequential anti-FLAG, as described above, and immobilized metal affinity chromatographic steps for His_6 purification. Instead of resolving the eluate using SDS-PAGE, we performed in-solution digestion by treating the eluted proteins with trypsin. In-solution digestion was performed to allow the identification of a greater number of purified azidoglycoproteins. Tandem MS analysis and database search of purified azidoglycoproteins using FLAG-His_6 allowed the identification of 8 unique proteins in Ac_4GalNAz-labeled cells, and only 1 unique protein in non-labeled cells (Figure 2.4A). Although, affinity purification of O-GlcNAzylated proteins with FLAG-His_6 reduced background contamination, it was only able to enrich for a very small number of azidoglycoproteins. It is possible that the protein amount used to start the affinity

enrichment needs to be higher in order to purify a higher quantity of azidoglycoproteins with a two-step affinity purification protocol.

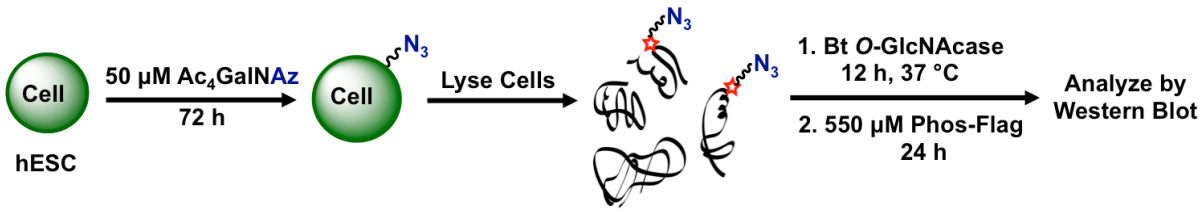


Figure 2.2. Metabolic labeling of human embryonic stem cells. hESCs were treated with 50 µM per-O-acetylated GalNAz (Ac_4GalNAz) for 72 hours to test the cell's ability to incorporate the azido sugar. After lysis, cells were treated with bacterial (Bt) O-GlcNAcase (OGA) and reacted with phosphine-FLAG (Phos-Flag). Metabolic labeling with Ac_4GalNAz was detected by anti-FLAG western blot, and GlcNAc depletion by OGA was detected by anti-O-GlcNAc western blot.

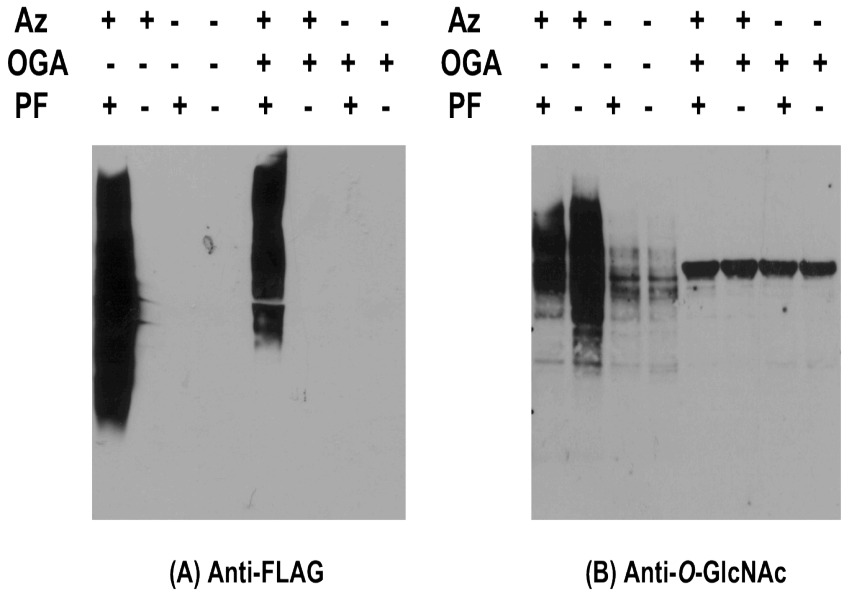
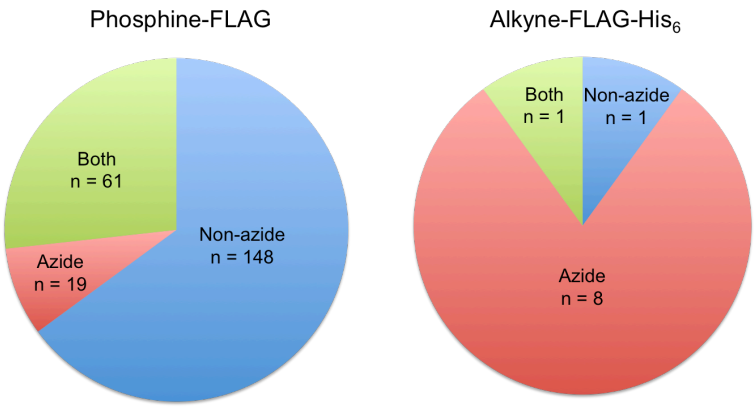


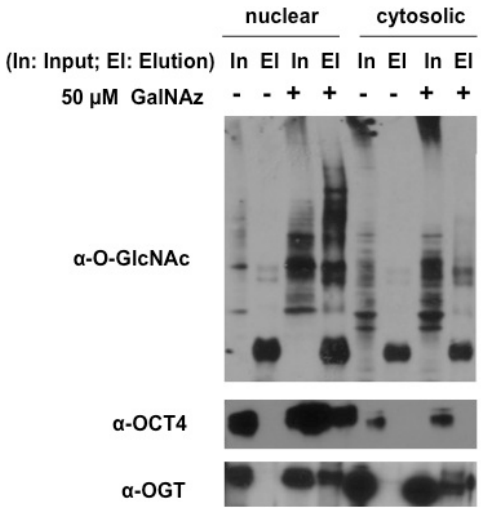
Figure 2.3. hESCs incorporate Ac_4GalNAz into O-GlcNAz. The primary signal observed in anti-FLAG (A) and anti-O-GlcNAc (B) western blots of Ac_4GalNAz-treated hESC whole cell lysates resulted from robust labeling of proteins by Ac_4GalNAz. Only samples reacted with phosphine-FLAG showed labeling in anti-FLAG immunoblot. Treatment of samples with a bacterial OGA homologue resulted in depletion of both anti-FLAG and anti-O-GlcNAc signal. (Az = Ac_4GalNAz, OGA= O-GlcNAcase, PF= phosphine-FLAG)

Finally, we performed affinity purification by labeling Ac_4GalNAz-treated cells with an alkyne-biotin probe. We then enriched the labeled azidoglycoproteins via avidin affinity purification, and performed western blot analysis of the eluted azide-labeled proteins. The interaction between biotin and avidin is of high affinity and specificity, thus allowing for an efficient capture of biotinylated proteins [23]. Western blot analysis showed O-GlcNAc signal only in the eluted azide-labeled glycoproteins, and not in the non-azide labeled glycoproteins, confirming that the enrichment had been specific for the Ac_4GalNAz-labeled proteins (Figure 2.4B). We were also able to enrich for known O-GlcNAcylated proteins, OGT and OCT4, a transcription factor important for hESC pluripotency, in the eluted azidoglycoproteins, but not in the non-azide labeled glycoproteins. Once we were confident that our biotin affinity purification strategy was able to enrich for azidoglycoproteins, we performed on-bead digestion of the capture proteins with trypsin, followed by tandem MS analysis. This allowed us to compare the capture efficiency of azidoglycoproteins using a biotin probe with those captured with FLAG and FLAG-His_6 probes. More than 150 unique O-GlcNAcylated proteins were specifically enriched in the Ac_4GalNAz-treated cells and absent from samples from non-Ac_4GalNAz-treated cells (Figure 2.4C). Importantly, many of the identified proteins had been previously reported to be O-GlcNAcylated, such as several members of the nuclear pore complex (Nup153, 93, 214, 98, 54, 35, and 62) [24], host cell factor 1 (HCF-1) [25], splicing factor 1 (SF1) [26], and RNA binding motif protein 27 (RBM27) [27] (Table 2.1). Additionally, many of the proteins identified are involved in transcription regulation, including Sin3A transcription regulator (Sin3A), Sin3A-associated protein (SAP130), transcription factor Sox-2 (SOX2), Sal-like protein 2 (SALL2), TATA box binding protein (TBP)-associated factor TAF6 (TAF6), cAMP responsive element binding protein 1 (CREB1), SWI/SNF complex subunit SMARCC2 (SMARCC2), and POU class 2 homeobox 1 (POU2F1). Functional and subcellular localization analysis showed that the majority of the identified O-GlcNAcylated proteins were transcription regulators, enzymes, and transporters, and where localized in the nucleus and cytoplasm (Table 2.1). Validation of the O-GlcNAcylated proteins identified in our analysis is necessary to confirm specific labeling of these proteins with Ac_4GalNAz. Nevertheless, these results indicate that metabolic labeling with Ac_4GalNAz, and biotin affinity purification, are useful tools for the discovery of O-GlcNAcylated proteins in mammalian cells.

A.



B.



C.

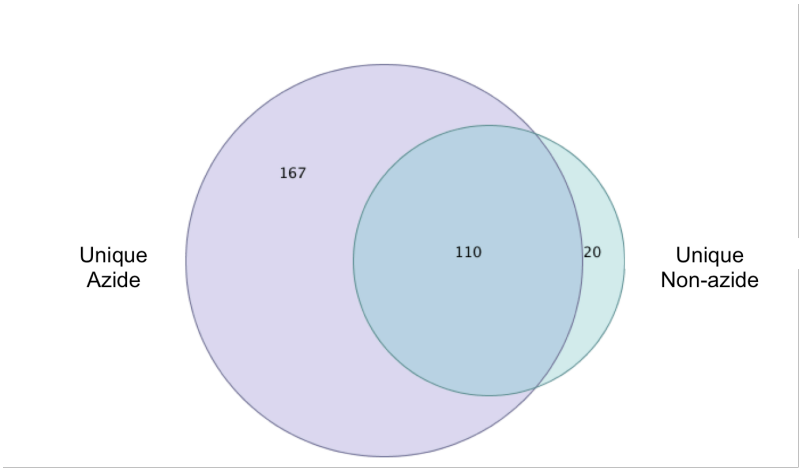


Figure 2.4. Comparison of three enrichment strategies of Ac_4GalNAz-labeled glycoproteins. (A) Enrichment of azidoglycoproteins with alkyne-FLAG-His_6 resulted in less background contamination of non-azide labeled glycoproteins, in comparison to phosphine-FLAG. (B) Ac_4GalNAz-labeled proteins are enriched after biotin purification. Known O-GlcNAcylated proteins were identified in the eluted Ac_4GalNAz-labeled samples. hESCs were treated with or without 50 μM Ac_4GalNAz for 72 h. Nuclear and cytoplasmic extracts were prepared, reacted with alkyne-biotin, affinity purified with avidin beads, and analyzed by immunoblot. Input: sample before enrichment; Elution: sample after enrichment. (C) Mass spectrometry analysis of enriched biotinylated azide-labeled glycoproteins resulted in the identification of 167 unique Ac_4GalNAz-labeled glycoproteins, whereas only 20 unique proteins where identified in the non-azide-labeled sample.

Table 2.1. Proteins enriched from extracts of $Ac_4GalNAz$-treated hESCs

Gene	Description	Location	Family
ACAA2	acetyl-CoA acyltransferase 2	Cytoplasm	enzyme
ACTA1	actin, alpha 1, skeletal muscle	Cytoplasm	other
ACTBL2	actin, beta-like 2	Cytoplasm	other
ACTN1	actinin, alpha 1	Cytoplasm	other
ALDH1B1	aldehyde dehydrogenase 1 family, member B1	Cytoplasm	enzyme
ALDH2	aldehyde dehydrogenase 2 family (mitochondrial)	Cytoplasm	enzyme
ALPL	alkaline phosphatase, liver/bone/kidney	Plasma Membrane	phosphatase
APLP2	amyloid beta (A4) precursor-like protein 2	Extracellular Space	other
ARF4	ADP-ribosylation factor 4	Cytoplasm	enzyme
ARID3A	AT rich interactive domain 3A	Nucleus	transcription regulator
ARID3B	AT rich interactive domain 3B	Nucleus	other
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	Cytoplasm	enzyme
ATP5J2-PTCD1/PTCD1	pentatricopeptide repeat domain 1	Extracellular Space	other
BAX	BCL2-associated X protein	Cytoplasm	transporter
BSG	basigin	Plasma Membrane	transporter
C1QBP	complement component 1, q subcomponent binding protein	Cytoplasm	other
CCAR1	cell division cycle and apoptosis regulator 1	Nucleus	transcription regulator
CCDC134	coiled-coil domain containing 134	Secreted	other
CCT4	chaperonin containing TCP1, subunit 4 (delta)	Cytoplasm	other
CD63	CD63 molecule	Plasma Membrane	other
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	Plasma Membrane	other
CFL1	cofilin 1 (non-muscle)	Nucleus	other
CLU	clusterin	Extracellular Space	other
CREB1	cAMP responsive element binding protein 1	Nucleus	transcription regulator
CSTF2T	cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant	Nucleus	other
CTNNB1	catenin (cadherin-associated protein), beta 1, 88kDa	Nucleus	transcription regulator
CTSD	cathepsin D	Cytoplasm	peptidase
CYP2S1	cytochrome P450, family 2, subfamily S, polypeptide 1	Cytoplasm	enzyme
DBN1	drebrin 1	Cytoplasm	other
DDOST	dolichyl-diphosphooligosaccharide--protein glycosyltransferase	Cytoplasm	enzyme
DSG2	desmoglein 2	Plasma Membrane	other
ECH1	enoyl CoA hydratase 1, peroxisomal	Cytoplasm	enzyme
ELAVL1	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)	Cytoplasm	other
EPB41L2	erythrocyte membrane protein band 4.1-like 2	Plasma Membrane	other
ERLIN2	ER lipid raft associated 2	Plasma Membrane	other
ERP44	endoplasmic reticulum protein 44	Cytoplasm	enzyme

FKBP10	FK506 binding protein 10, 65 kDa	Cytoplasm	enzyme
FLNB	filamin B, beta	Cytoplasm	other
GATAD2A	GATA zinc finger domain containing 2A	Nucleus	transcription regulator
GLG1	golgi glycoprotein 1	Cytoplasm	other
GNAI2	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2	Plasma Membrane	enzyme
GOLIM4	golgi integral membrane protein 4	Cytoplasm	other
GOLM1	golgi membrane protein 1	Cytoplasm	other
GPC4	glypican 4	Plasma Membrane	transmembrane receptor
HCFC1	host cell factor C1 (VP16-accessory protein)	Nucleus	transcription regulator
HIST1H2BN	histone cluster 1, H2bn	Nucleus	other
HIST2H2AA3/HIST2H2AA4	histone cluster 2, H2aa3	Nucleus	other
HK2	hexokinase 2	Cytoplasm	kinase
HM13	histocompatibility (minor) 13	Cytoplasm	peptidase
HNRNPAB	heterogeneous nuclear ribonucleoprotein A/B	Nucleus	enzyme
HNRNPF	heterogeneous nuclear ribonucleoprotein F	Nucleus	other
HNRNPH1	heterogeneous nuclear ribonucleoprotein H1 (H)	Nucleus	other
HNRNPM	heterogeneous nuclear ribonucleoprotein M	Nucleus	other
HSD17B10	hydroxysteroid (17-beta) dehydrogenase 10	Cytoplasm	enzyme
HSP90AA1	heat shock protein 90kDa alpha (cytosolic), class A member 1	Cytoplasm	enzyme
HYOU1	hypoxia up-regulated 1	Cytoplasm	other
IDH2	isocitrate dehydrogenase 2 (NADP+), mitochondrial	Cytoplasm	enzyme
IGF2R	insulin-like growth factor 2 receptor	Plasma Membrane	transmembrane receptor
IPO5	importin 5	Nucleus	transporter
ITGA6	integrin, alpha 6	Plasma Membrane	other
ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	Plasma Membrane	transmembrane receptor
KHSRP	KH-type splicing regulatory protein	Nucleus	enzyme
KPNB1	karyopherin (importin) beta 1	Nucleus	transporter
LMNA	lamin A/C	Nucleus	other
LONP1	lon peptidase 1, mitochondrial	Cytoplasm	peptidase
LRP1	low density lipoprotein receptor-related protein 1	Plasma Membrane	transmembrane receptor
MAN2A1	mannosidase, alpha, class 2A, member 1	Cytoplasm	enzyme
MATR3	matrin 3	Nucleus	other
MCM3AP	minichromosome maintenance complex component 3 associated protein	Nucleus	other
MCM4	minichromosome maintenance complex component 4	Nucleus	enzyme
MFGE8	milk fat globule-EGF factor 8 protein	Extracellular Space	other
mir-1181	microRNA 1181	Cytoplasm	microRNA
mir-1248	microRNA 1248	Cytoplasm	microRNA
MPP6	membrane protein, palmitoylated 6 (MAGUK p55 subfamily member 6)	Plasma Membrane	kinase

MSN	moesin	Plasma Membrane	other
NDUFS1	NADH dehydrogenase (ubiquinone) Fe-S protein 1, 75kDa (NADH-coenzyme Q reductase)	Cytoplasm	enzyme
NDUFS2	NADH dehydrogenase (ubiquinone) Fe-S protein 2, 49kDa (NADH-coenzyme Q reductase)	Cytoplasm	enzyme
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	Nucleus	transcription regulator
NUP153	nucleoporin 153kDa	Nucleus	transporter
NUP214	nucleoporin 214kDa	Nucleus	transporter
NUP35	nucleoporin 35kDa	Nucleus	transporter
NUP54	nucleoporin 54kDa	Nucleus	transporter
NUP62	nucleoporin 62kDa	Nucleus	transporter
NUP93	nucleoporin 93kDa	Nucleus	other
NUP98	nucleoporin 98kDa	Nucleus	transporter
NUPL1	nucleoporin like 1	Nucleus	transporter
OAT	ornithine aminotransferase	Cytoplasm	enzyme
PCBP1	poly(rC) binding protein 1	Nucleus	translation regulator
PCK2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	Cytoplasm	kinase
PCYOX1	prenylcysteine oxidase 1	Cytoplasm	enzyme
PDHB	pyruvate dehydrogenase (lipoamide) beta	Cytoplasm	enzyme
PDIA6	protein disulfide isomerase family A, member 6	Cytoplasm	enzyme
PFN1	profilin 1	Cytoplasm	other
PGRMC1	progesterone receptor membrane component 1	Plasma Membrane	transmembrane receptor
PICALM	phosphatidylinositol binding clathrin assembly protein	Cytoplasm	other
PLD3	phospholipase D family, member 3	Cytoplasm	enzyme
PODXL	podocalyxin-like	Plasma Membrane	kinase
POU2F1	POU class 2 homeobox 1	Nucleus	transcription regulator
PPT1	palmitoyl-protein thioesterase 1	Cytoplasm	enzyme
PRCP	prolylcarboxypeptidase (angiotensinase C)	Cytoplasm	peptidase
PRDX1	peroxiredoxin 1	Cytoplasm	enzyme
PROM1	prominin 1	Plasma Membrane	other
PRRC1	proline-rich coiled-coil 1	Cytoplasm	other
PTK7	PTK7 protein tyrosine kinase 7	Plasma Membrane	kinase
QSER1	glutamine and serine rich 1		other
RAB1A	RAB1A, member RAS oncogene family	Cytoplasm	enzyme
RAB2A	RAB2A, member RAS oncogene family	Cytoplasm	enzyme
RAB7A	RAB7A, member RAS oncogene family	Cytoplasm	enzyme
RANBP2	RAN binding protein 2	Nucleus	enzyme
RANGAP1	Ran GTPase activating protein 1	Nucleus	other
RBM14	RNA binding motif protein 14	Nucleus	transcription regulator
RBM26	RNA binding motif protein 26		other
RBM27	RNA binding motif protein 27	Nucleus	other
REPS1	RALBP1 associated Eps domain containing 1	Plasma Membrane	other

RIF1	RAP1 interacting factor homolog	Nucleus	other
RPN1	ribophorin I	Cytoplasm	enzyme
RPN2	ribophorin II	Cytoplasm	enzyme
RPRD2	regulation of nuclear pre-mRNA domain containing 2		other
SALL2	sal-like 2	Nucleus	transcription regulator
SALL4	sal-like 4	Nucleus	other
SAP130	Sin3A-associated protein, 130kDa	Nucleus	transcription regulator
SEPT2	septin 2	Nucleus	enzyme
SF1	splicing factor 1	Nucleus	transcription regulator
SF3A1	splicing factor 3a, subunit 1, 120kDa	Nucleus	other
SIN3A	SIN3 transcription regulator homolog A	Nucleus	transcription regulator
SLC1A5	solute carrier family 1 (neutral amino acid transporter), member 5	Plasma Membrane	transporter
SLC25A11	solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11	Cytoplasm	transporter
SLC25A13	solute carrier family 25 (aspartate/glutamate carrier), member 13	Cytoplasm	transporter
SLC25A4	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4	Cytoplasm	transporter
SLC2A1	solute carrier family 2 (facilitated glucose transporter), member 1	Plasma Membrane	transporter
SLC38A2	solute carrier family 38, member 2	Plasma Membrane	transporter
SLC39A14	solute carrier family 39 (zinc transporter), member 14	Plasma Membrane	transporter
SLC7A3	solute carrier family 7 (cationic amino acid transporter, y+ system), member 3	Plasma Membrane	transporter
SLC9A3R1	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 1	Plasma Membrane	other
SMARCC2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 2	Nucleus	transcription regulator
SND1	staphylococcal nuclease and tudor domain containing 1	Nucleus	enzyme
SORL1	sortilin-related receptor, L(DLR class) A repeats containing	Plasma Membrane	transporter
SOX2	SRY (sex determining region Y)-box 2	Nucleus	transcription regulator
STOML2	stomatin (EPB72)-like 2	Plasma Membrane	other
TAF6	TAF6 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 80kDa	Nucleus	transcription regulator
TARDBP	TAR DNA binding protein	Nucleus	transcription regulator
TFRC	transferrin receptor (p90, CD71)	Plasma Membrane	transporter
TIMM44	translocase of inner mitochondrial membrane 44 homolog	Cytoplasm	transporter
TIMM50	translocase of inner mitochondrial membrane 50 homolog	Cytoplasm	phosphatase
TMED9	transmembrane emp24 protein transport domain containing 9	Cytoplasm	transporter
TMPO	thymopoietin	Nucleus	other
TOMM40	translocase of outer mitochondrial membrane 40 homolog	Cytoplasm	ion channel
TPM3	tropomyosin 3	Cytoplasm	other
TXNDC5	thioredoxin domain containing 5 (endoplasmic reticulum)	Cytoplasm	enzyme

UBAP2	ubiquitin associated protein 2	Cytoplasm	other
UBAP2L	ubiquitin associated protein 2-like		other
UQCRC2	ubiquinol-cytochrome c reductase core protein II	Cytoplasm	enzyme
VDAC3	voltage-dependent anion channel 3	Cytoplasm	ion channel
XRCC5	X-ray repair complementing defective repair in Chinese hamster cells 5	Nucleus	enzyme
YTHDF3	YTH domain family, member 3	Cytoplasm	other
YWHAE	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	Cytoplasm	other
YWHAQ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide	Cytoplasm	other
YWHAZ	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	Cytoplasm	enzyme
ZFR	zinc finger RNA binding protein	Nucleus	other
ZMYND8	zinc finger, MYND-type containing 8	Nucleus	transcription regulator
ZNF281	zinc finger protein 281	Nucleus	transcription regulator

All listed hits were absent from non-Ac_4GalNAz labeled control cell samples. Subcellular localization and functional analysis were performed with Ingenuity Pathway Analysis software (IPA).

2.3. Conclusion

In this chapter, we demonstrated the utility of metabolic labeling and biotin affinity purification for the identification of O-GlcNAcylated proteins in human embryonic stem cells. Our results show that treating cells with Ac_4GalNAz affords optimal labeling of O-GlcNAcylated proteins, as shown previously by members of the Bertozzi lab. Taken together, the findings presented here will help define the O-GlcNAc-mediated mechanisms involved in the maintenance of pluripotency in hESCs.

2.4. Experimental Methods

General reagents

All chemical reagents were of analytical grade from commercial sources and used without further purification, unless otherwise noted. All chemicals were purchased from Sigma-Aldrich, unless otherwise noted. Ac_4GalNAz, FLAG and FLAG-His_6 peptides were synthesized as described [14, 28]. Alkyne-biotin was purchased from Life technologies (B10185). *Bacteroides* O-GlcNAcase (OGA) was expressed in *Escherichia coli* and purified as described [29]. O-GlcNAcase inhibitor, Thiamet-G, was purchased from Cayman Chemicals.

Human embryonic stem cell (hESC) culture

The NIH approved human embryonic stem cell lines H1 and H9 (WiCell Research Institute) were cultured on 10 cm polystyrene dishes coated with hESC-qualified basement membrane matrix Matrigel (BD Biosciences). Matrigel plates were made according to manufacturer specifications. Cells were grown under standard conditions (37 °C, 5% CO_2) in TeSR2 media (STEMCELL Technologies Inc). Media was exchanged daily after the first 48 h in culture, and cells were passaged every 3 to 5 days using 1 mg/ml of dispase (Gibco) in Knockout DMEM media (Invitrogen) and mechanically removed. Karyotype analysis was routinely performed and indicated that all samples were diploid and had no chromosomal abnormalities.

Treatment of samples with bacterial O-GlcNAcase (OGA) and metabolic labeling

Cells were treated with 50 μM Ac_4GalNAz for 72 hours, and washed twice in 1X PBS at the time of harvest. Whole cell lysates were prepared by resuspending cell pellet in RIPA buffer with protease inhibitors (Roche). Lysates were sonicated (Misonix) and cleared by centrifugation. Samples were then treated with bacterial OGA homologue for 12 h at 37 °C. The following day, each reaction was split in half. One half was reacted with phosphine-FLAG (550 μM) for 24 hours at room temperature. The other half was left at room temperature for the same time period with no further treatment.

Cell extract preparation for affinity purification

hESCs were harvested in the presence of ice-cold 1X PBS (Thermo Scientific), pooled into 50 ml conical tubes and centrifuged for 5 minutes at 1200 rpm at 4 °C. Cell pellets were washed twice with ice-cold 1x PBS and centrifuged for 5 minutes each time. After the second PBS wash, nuclear extracts were made as follows (Lamond's protocol: http://www.lamondlab.com/f7nucleolarprotocol.htm), with protease inhibitors (Roche) and 2 μM Thiamet-G included throughout: Pellets were placed on ice and resuspended with a homogenization buffer consisting of 10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, and 0.5 mM DTT. Cells were swelled for 5 minutes on ice, and broken by Dounce homogenization using a tight pestle (Kimble Chase). Crude nuclei were pelleted by centrifugation (1200 rpm, 10 min at 4 °C). To prepare cytoplasmic extracts, the

nuclei-depleted supernatant was centrifuged at 20,000 x g to pellet insoluble material and the resulting supernatant was saved. To prepare nuclear extracts, the crude nuclear pellet was resuspended in buffer S1 consisting of 250 mM sucrose and 10 mM $MgCl_2$, and layered over an equal volume of buffer S3 consisting of 880 mM sucrose and 0.5 mM $MgCl_2$ and pelleted by centrifugation (3500 rpm, 15 min at 4 °C). This highly purified nuclear pellet was resuspended in 1% Triton X-100, 150 mM NaCl, and 50 mM Tris pH 7.5. Corresponding cytoplasmic extracts were brought to the same Triton X-100, NaCl, and Tris concentrations. All samples were sonicated and cleared by centrifugation.

Affinity purifications and mass spectrometry analysis of Ac_4GalNAz-labeled cells

Enrichment with Phosphine-FLAG and In-gel digestion

Cell extracts were prepared as described above. Samples were reacted with 400 μM phosphine-FLAG overnight at room temperature. Next day, samples were precipitated with cold acetone at -20 °C overnight to remove excess phosphine-FLAG. Samples were spun down, and the resulting pellet was resuspended in RIPA buffer. M2 anti-FLAG antibody (Sigma) was then added in a 1:50 dilution to Ac_4GalNAz-treated and non-treated samples, and the solutions were incubated for 1 hour at room temperature with gentle inversion of the tubes. An appropriate volume of Ultralink Protein A/G resin was equilibrated in lysis buffer and then added to the lysate/antibody mixture. The resin and solution were mixed at room temperature for 1.5 h before centrifugation to pellet the resin and bound protein. The resin was then washed four times with RIPA buffer. Captured proteins were eluted two times with 8 M urea for 15 minutes at room temperature.

For *in-gel digestion*, the eluted sample was resolved by molecular weight using SDS-PAGE. The gels were stained with Simply Blue Safe Stain (Invitrogen) according to the manufacturer's instructions. Bands of approximately 1 mm width were then cut from the control and azide-treated gels, added to low-bind 0.5 mL tubes (PGC Scientifics), and washed with 100 mM ammonium bicarbonate solutions for 5 min. The solution was then removed from the gel pieces, which were then reduced, alkylated, and digested overnight with trypsin (Promega) as described [30]. The digest solutions were then collected, and the gel pieces sonicated with 60% acetonitrile containing 5% formic acid for 10 min. The solutions were then combined with the digest solutions for each sample and evaporated in a speed vacuum (ThermoFisher). Each sample was then desalted using C18 Zip Tips (Millipore) according to the manufacturer's instructions. The samples were then evaporated to dryness using the speed vacuum and stored at -80 °C prior to mass spectrometry analysis.

Nano-LC-MS/MS and data collection [31]

Both mobile phase solvents, ultrapure water as Buffer A (Burdick-Jackson) and HPLC-grade acetonitrile (ACN), Buffer B, contained 0.1% formic acid. Samples were resuspended in water and stored at 4 °C in the temperature-regulated autosampler of

an Agilent 1200 liquid chromatography system. Samples were loaded on to a pressure-packed IntegraFrit capillary trapping column (100 ± 6 µm tip, New Objective) containing 1 cm of C18AQ resin (5 µm, 200 Å; Michrom Bioresources) with 98% water and 2% ACN. Chromatography was achieved using a 13 cm hand-pulled column (50 µm ID, 357.2 µm OD, Polymicro Technologies) pressure-packed with 10 cm of C18AQ (5 µm, 100 Å) resin. Two samples were run per column and trapping column using the following mobile phase gradient: a 10 minute loading step in 2% B, a 35 minute gradient from 2-40% B, a 5 minute gradient from 40-70% B, followed by a 15 minute wash step of 99% B. After the first sample, the column was re-equilibrated to 98% water. A solvent split was used to maintain a flow rate of approximately 400 nL/min at the column tip. The LC gradient program and data acquisition were conducted using Xcalibur 2.0.6 software (ThermoFisher). The data were collected on an LTQ XL mass spectrometer (ThermoFisher) using a column stage built in-house using plans provided by the Yates Laboratory (Scripps Research Institute) with a spray voltage of 1.6 kV, a capillary temperature of 200 °C, capillary voltage of 42.00 V, and a tube lens voltage of 105.00 V. The data were collected in the following manner: 1 full scan (m/z scan range 400-2000) followed by data-dependent scans (minimum signal threshold of 500 counts) from the initial full scan on the ten most abundant precursor ions for subsequent fragmentation and MS/MS analysis. Dynamic exclusion was enabled over a 60 second window. The data were stored as RAW files prior to database searching.

Database searching [31]

The data was processed using BioWorks 3.2 (ThermoFisher). The UniProtKB/Swiss-Prot database was downloaded in FASTA format from the International Protein Index. This comprehensive database was further filtered in BioWorks, using the database Utilities function, to limit entries to those containing the following strings: human, homo, and sapien. This filtered database contained 22734 entries (from 509019 original entries). The database Indexer function was then applied with the following parameters: 1) monoisotopic peptide precursor and fragment mass, 2) trypsin (KR) selected as the protease with fully enzymatic cleavage at both ends of the peptide, 3) up to two missed cleavage sites, 4) peptide tolerance of 3.00 amu and fragment ion tolerance of 2.00 amu, 5) cysteine carboxyamidomethylation and methionine oxidation as differential modifications, and 6) up to three modifications per peptide. The data was searched using the SEQUEST algorithm in BioWorks against the forward and reverse databases and further filtered using the following criteria: 1) Xcorr vs. charge state of 1.8, 2.5, and 3.5 for singly, doubly, and triply charged peptides, respectively; 2) DeltaCN ≥ 0.100, and 3) peptide probability of ≤ 0.05. Single peptide identifications were permitted upon manual inspection of each spectrum and only if the peptide corresponded to a unique protein. The data for each sample set was compiled using the Multiconsensus Report function in Bioworks. The full list of Ac_4GalNAz-labeled proteins was obtained by subtracting the compiled list of background proteins identified from the non-azide-treated samples.

Enrichment with alkyne-FLAG-His$_6$ and in-solution digestion

Control and Ac_4GalNAz-treated cells were fractionated into cell extracts as described above, and 1 mg of protein was reacted with 25 µM alkyne-FLAG-His_6, 100 µM Tris [(1-benzyl-1H-1,2,3-triazol-4-yl) methyl] amine) (TBTA), 1 mM $CuSO_4$ in H_2O, and 1.5 mM sodium ascorbate (prepared fresh in H_2O) for 1 hour at room temperature. Samples were pre-cleared with Protein A/G resin to remove non-specific binding of samples to the resin. Unreacted probe was removed by acetone precipitation overnight at -20 °C. Next day, samples were spun down and supernatant removed. Protein pellet was resuspended in high-salt RIPA buffer consisting of 300 mM NaCl, 20 mM Tris pH 7.4, 1% Triton, and 0.1% SDS. Labeled proteins were enriched overnight at 4 °C using a monoclonal mouse antibody against FLAG (Sigma) at a 1:40 dilution, followed by capture on Protein A/G resin (ThermoFisher) for 2 h rotating at room temperature. Pelleted beads by centrifuging for 2 minutes at 2500 x g, and discarded supernatant. Washed beads a total of 6 times: 2x with high-salt RIPA, and 4x in triton buffer consisting of 300 mM NaCl, 20 mM Tris pH 7.4, and 1% Triton X-100. Proteins were eluted with urea buffer: 8 M urea, 1% Triton X-100,10 mM beta-mercaptoethanol, and 20 mM imidazole in PBS. For secondary capture, Ni-NTA agarose slurry (Qiagen) was added to the eluted proteins and the mixture was incubated for 1 h rotating at room temperature. The beads were washed twice with urea buffer, and four times with 20 mM imidazole in PBS. Proteins were eluted with 250 mM imidazole in PBS.

Started *in-solution digestion* of the eluted samples by adding 6 M urea and adjusting pH to 8.0 with 1 M Tris, pH 8.0. Reduced the samples by adding 20 mM dithiothreitol (DTT), and incubated for 30 minutes at 37 °C. Then, proceeded to alkylate samples by adding 50 mM iodoacetamide, and incubating for 30 minutes in the dark at room temperature. Added 10 mM DTT to scavenge excess alkylating reagent, and incubated for another 30 minutes at room temperature. Diluted the urea to 0.6 M by adding water to the samples, and checked pH to ensure it was 8.0. Finally, added trypsin (Promega) in an enzyme:substrate ratio of 1:50, and incubated overnight at 37 °C. The following day added formic acid to a final concentration of 1% to lower the pH below 3, and inactivate trypsin. The resulting tryptic peptides were de-salted by binding to reversed-phase cartridges containing C_{18} resin (Waters), washing four times with 0.1% formic acid, and eluting with two separate volumes of 80% acetonitrile in 0.1% formic acid. The de-salted peptide samples were dried in a speed-vac and stored in -80 °C prior to mass spectrometry analysis.

Nano-LC-MS/MS, data collection, and database searching were performed as described above for phosphine-FLAG enrichment

Enrichment with alkyne-biotin and on-bead digestion

Control and Ac_4GalNAz-treated cells were fractionated into cell extracts as described above, and approx. 1 mg of protein was reacted with 25 µM alkyne-biotin, 100 µM Tris [(1-benzyl-1H-1,2,3-triazol-4-yl) methyl] amine) (TBTA), 1 mM $CuSO_4$ in H_2O, and 1.5 mM sodium ascorbate (prepared fresh in H_2O) for 1 hour at room temperature. PBS and

10% Triton X-100 stock solution was added and samples were rotated for 1 h at room temperature. Samples were then filtered through a disposable PD-10 desalting column (GE Healthcare Life Sciences) and eluted with PBS. 10% SDS stock was added for final concentration of 0.5% SDS and samples were boiled at 90 °C for 8 min. Samples were cooled on ice for 5 min, and additional PBS and avidin slurry (Sigma) were added and rotated at room temperature for 1 h. Beads were concentrated by centrifugation at 1,400 rpm for 3 min at 4 °C, supernatant was removed, and beads were transferred to a washed biospin column (Bio Rad Life Sciences) on a vacuum manifold. Beads were washed twice with PBS, once with 0.2% SDS in PBS, once with fresh 6 M urea, and two more washes with PBS. Freshly prepared 6 M urea and 100 mM tris(2-carboxyethyl)phosphine (TCEP) was delivered and sample was rotated for 30 min at room temperature. Freshly prepared 200 mM iodoacetamide was added and the sample was wrapped in foil and rotated for 30 min at room temperature. Beads were then washed 3 times with PBS and transferred to an RNAse free microcentrifuge tube with 2 washes of PBS. Beads were concentrated by centrifugation at 6000 rpm for 3 min and supernatant was removed. Freshly made 2 M urea was added and *on-bead* trypsin digestion (Promega) was performed overnight in bench top shaker at 37 °C. The next morning, the beads were pelleted at 6,000 rpm for 3 min. The supernatant was collected, and the beads washed with PBS. Formic acid was added and samples were stored at -80 °C.

Multidimensional Protein Identification Technology Analysis (MudPIT) [32]

LTQ XL (Thermo Scientific, San Jose, CA, USA) interfaced at the front end with a quaternary HP 1100 series HPLC pump (Agilent Technology, Santa Clara, CA, USA). The analytical column consisted of a 100 µm diameter fused-silica capillary (J/W Scientific, Agilent Technology, Santa Clara, CA, USA), pulled with a P-2000 laser (Sutter Instrument Co., Novato, CA, USA) and packed with 12 cm of 5 µm C18 resin (Aqua, Phenomenex, Torrence, CA, USA). The biphasic microcapillary trapping column (5 cm of 250 µm diameter) consisted of a fritted capillary with Kasil 1624, packed with 10 cm reversed-phase C18 (Aqua, Phenomenex, Torrence, CA, USA) and 3 cm of strong cation exchange (5 µm Partisphere, Whatman, Maidstone, Kent, UK) packing material. The biphasic column, loaded offline with sample by a pressure pump. In order to split the gradient pump flow to 0.15−0.20 µL/min and supply a spray voltage of 1.8 kV. The LTQ XL was operated via Instrument Method files in the Sequence Setup window of Excalibur. A fully automated 11-cycle chromatographic run was performed on each sample using a three mobile phases system consisting of buffer A (5% acetonitrile (ACN); 0.1% formic acid (FA)) (Sigma Aldrich, San Louis, MO, USA), buffer B (80% ACN, 0.1% FA), and buffer C (500 mM ammonium acetate, 5% ACN, 0.1% FA). For a total time of 10 hours per sample and generating 5 MS files. The first cycle consisted of a 120 min linear gradient from 0 to 100% buffer B. Steps 2−9 show the following profile: 2 min 100% buffer A; 4 min (100 − X)% buffer A, X% buffer C; 45 min from 100% buffer A to 50% buffer A and 50% buffer B; 10 min from 50% to 100% buffer B; 1 min from 100% buffer B to 100% A; 10 min at 100% buffer A. For buffer C, X% was, respectively, 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 100%. Steps 10 and 11 show the following profile: 2 min 100% buffer A; 4 min 10% buffer B, 90% buffer C; 45 min from 100%

buffer A to 50% buffer A and 50% buffer B; 10 min from 50% to 100% buffer B; 1 min from 100% buffer B to 100% A; 10 min at 100% buffer A.

MS Data Analysis by ProLuCID and Ingenuity Pathway Analysis

Mass spectrum data was converted to generate MS1/MS2 files using RawXtractor (currently supports only Thermo Fisher instruments). Files where uploaded to the Integrated Proteomics Pipeline (IP2) webserver, which allows for the identification, quantification, and functionally analysis of proteins in a biological sample using the ProLuCID protein database software (http://fields.scripps.edu/prolucid/).

Subcellular localization and functional annotation of identified proteins, as shown in Table 2.1, was carried out using Ingenuity Pathway Analysis (IPA; Qiagen).

Western blot analysis

Following sample preparation, protein concentration was determined by bicinchoninic acid assay (Pierce). Samples were separated by SDS-PAGE (4-12% Bio-Rad), transferred to nitrocellulose membranes, and blocked with 5% milk (LabScientific, inc.) in PBS with Tween 20 (Sigma). Membranes were incubated with primary antibodies overnight at 4 °C. The following primary antibodies were used to detect the indicated proteins: anti-OGT (1:1000, DM-17, Sigma), anti-OCT4 (1:2500, MAB4401, Millipore), anti-O-GlcNAc (1:1000, RL2, Thermo Scientific), and anti-FLAG-HRP (1:5000, M2, Sigma). The following horseradish peroxidase-linked secondary antibodies were used for detection at a dilution ratio of 1:5000: anti-rabbit HRP-conjugated secondary antibody (4010-05, Southern Biotech) and anti-mouse-HRP-conjugated secondary antibody (1010-05, Southern Biotech).

2.5. References

1. Lanctot, P.M., F.H. Gage, and A.P. Varki, *The glycans of stem cells.* Current Opinion in Chemical Biology, 2007. **11**(4): p. 373-380.

2. Rutishauser, U. and L. Landmesser, *Polysialic acid in the vertebrate nervous system: a promoter of plasticity in cell-cell interactions.* Trends in Neurosciences, 1996. **19**(10): p. 422-427.

3. 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.* Mol Cell Proteomics, 2012. **11**(4): p. M111.010660.

4. Hart, G.W., et al., *Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease.* Annu Rev Biochem, 2011. **80**: p. 825-58.

5. Hanover, J.A., M.W. Krause, and D.C. Love, *Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation.* Nat Rev Mol Cell Biol, 2012. **13**(5): p. 312-21.

6. Jang, H., et al., *O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network.* Cell Stem Cell, 2012. **11**(1): p. 62-74.

7. Saxon, E. and C.R. Bertozzi, *Cell Surface Engineering by a Modified Staudinger Reaction.* Science, 2000. **287**(5460): p. 2007-2010.

8. Prescher, J.A., D.H. Dube, and C.R. Bertozzi, *Chemical remodelling of cell surfaces in living animals.* Nature, 2004. **430**(7002): p. 873-7.

9. Laughlin, S.T., et al., *In vivo imaging of membrane-associated glycans in developing zebrafish.* Science, 2008. **320**(5876): p. 664-7.

10. Vocadlo, D.J., et al., *A chemical approach for identifying O-GlcNAc-modified proteins in cells.* Proceedings of the National Academy of Sciences, 2003. **100**(16): p. 9116-9121.

11. Hang, H.C., et al., *A metabolic labeling approach toward proteomic analysis of mucin-type O-linked glycosylation.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 14846-51.

12. Boyce, M., et al., *Metabolic cross-talk allows labeling of O-linked beta-N-acetylglucosamine-modified proteins via the N-acetylgalactosamine salvage pathway.* Proc Natl Acad Sci U S A, 2011. **108**(8): p. 3141-6.

13. Dube, D.H., et al., *Probing mucin-type O-linked glycosylation in living animals.* Proc Natl Acad Sci U S A, 2006. **103**(13): p. 4819-24.

14. Laughlin, S.T. and C.R. Bertozzi, *Metabolic labeling of glycans with azido sugars and subsequent glycan-profiling and visualization via Staudinger ligation.* Nat. Protocols, 2007. **2**(11): p. 2930-2944.

15. Prescher, J.A. and C.R. Bertozzi, *Chemical Technologies for Probing Glycans.* Cell, 2006. **126**(5): p. 851-854.

16. Yu, S.H., et al., *Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4834-9.

17. Palaniappan, K.K., et al., *A chemical glycoproteomics platform reveals O-GlcNAcylation of mitochondrial voltage-dependent anion channel 2.* Cell Rep, 2013. **5**(2): p. 546-52.

18. Sprung, R., et al., *Tagging-via-substrate strategy for probing O-GlcNAc modified proteins.* J Proteome Res, 2005. **4**(3): p. 950-7.

19. Nandi, A., et al., *Global identification of O-GlcNAc-modified proteins.* Anal Chem, 2006. **78**(2): p. 452-8.

20. Zaro, B.W., et al., *Chemical reporters for fluorescent detection and identification of O-GlcNAc-modified proteins reveal glycosylation of the ubiquitin ligase NEDD4-1.* Proc Natl Acad Sci U S A, 2011. **108**(20): p. 8146-51.

21. Saxon, E., J.I. Armstrong, and C.R. Bertozzi, *A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds.* Org Lett, 2000. **2**(14): p. 2141-3.

22. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.* Angewandte Chemie International Edition, 2001. **40**(11): p. 2004-2021.

23. Rybak, J.N., et al., *Purification of biotinylated proteins on streptavidin resin: a protocol for quantitative elution.* Proteomics, 2004. **4**(8): p. 2296-9.

24. Li, B. and J.J. Kohler, *Glycosylation of the nuclear pore.* Traffic, 2014. **15**(4): p. 347-61.

25. Wysocka, J., et al., *Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1.* Genes Dev, 2003. **17**(7): p. 896-911.

26. Teo, C.F., et al., *Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc.* Nat Chem Biol, 2010. **6**(5): p. 338-43.

27. Wang, Z., et al., *Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis.* Sci Signal, 2010. **3**(104): p. ra2.

28. Kiick, K.L., et al., *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation.* Proceedings of the National Academy of Sciences, 2002. **99**(1): p. 19-24.

29. Dennis, R.J., et al., *Structure and mechanism of a bacterial beta-glucosaminidase having O-GlcNAcase activity.* Nat Struct Mol Biol, 2006. **13**(4): p. 365-71.

30. Rowley, A., et al., *Applications of Protein Mass Spectrometry in Cell Biology.* Methods, 2000. **20**(4): p. 383-397.

31. Hubbard, S.C., et al., *Cell surface glycoproteomic analysis of prostate cancer-derived PC-3 cells.* Bioorganic & Medicinal Chemistry Letters, 2011. **21**(17): p. 4945-4950.

32. Chen, E.I., et al., *Large scale protein profiling by combination of protein fractionation and multidimensional protein identification technology (MudPIT).* Mol Cell Proteomics, 2006. **5**(1): p. 53-6.

Chapter 3

Proteomic profile of the O-GlcNAc modification on hESCs undergoing neural differentiation

Chapter 3. Proteomic profile of the O-GlcNAc modification on hESCs undergoing neural differentiation[1]

3.1. Introduction

The dynamic and reversible modification of intracellular proteins with O-GlcNAc has emerged as a ubiquitous and important regulator of diverse cellular processes. Protein O-GlcNAcylation is analogous to phosphorylation, in that a small, covalent post-translational modification is added or removed by dedicated enzymes, OGT and OGA, respectively, as part of a signaling cascade that affects substrates proteins' localization, function, and stability. To date, more than 4000 proteins have been found to be O-GlcNAc-modifed. Most of these glycoproteins were identified through mass spectrometry (MS) methods [1]. The majority of the O-GlcNAcylated proteins identified are known to be involved in transcription, cellular transport, and translation. Additionally, since the synthesis of UDP-GlcNAc, the substrate donor of OGT, is tightly regulated by multiple metabolic pathways in the cell via the hexosamine biosynthetic pathway (HBP), O-GlcNAc functions to link the nutrient status of the cell with the regulation of signaling pathways [1].

Even though the current understanding of the function of the O-GlcNAc modification of many proteins is incomplete, this post-translational modification has been implicated in many cellular processes. For example, the complete absence of OGT is lethal to mouse ESCs and embryos [2], and experiments with conditional alleles revealed that OGT is required in a tissue specific manner at later stages of development, including neurons [3]. O-GlcNAcylation has also been shown to both enhance and suppress activity of transcription factors important for hESC differentiation (e.g., Sp1 and HCF-1) [4, 5] and maintenance of pluripotency (e.g., SOX2, OCT4 and TET1) [6-8]. The enzymes responsible for the modification, OGT and OGA, are most highly expressed in the brain, including the cerebellar cortex and the hippocampus [9, 10]. Additionally, O-GlcNAc has been found at all stages of embryonic brain development, and aberrant O-GlcNAc cycling has been implicated in the development of neurodegenerative diseases, such as Alzheimer's disease, suggesting the importance of O-GlcNAc within the mammalian brain [11].

A mechanistic understanding of how O-GlcNAc regulates neuronal processes and higher-order brain functions remains unclear. The identification of functionally relevant O-GlcNAcylated proteins in mammalian systems has been made difficult by the genetic lethality of OGT in mice and the difficulty of studying a dynamic post-translational modification such as glycosylation. Additionally, OGT has no clear substrate sequence preference, making substrate or modification site prediction difficult, even using MS. The O-GlcNAc glycosidic bond is labile under traditional mass spectrometry fragmentation methods, such as higher-energy collisional dissociation (HCD), which induce the fragmentation of the post-translational modification (PTM) first from the peptides, followed by fragmentation of the peptide backbone [12]. This initial loss of PTMs obscures any information about the specific site of modification on the peptides. A recently invented fragmentation technique known as electron transfer dissociation (ETD) has greatly enhanced our ability to map O-GlcNAc glycosylation

[1] Neil Rumachik contributed to the work described in this chapter.

sites [13]. ETD fragmentation utilizes a different mechanism of dissociation in which the peptide backbone fragments first, preserving the modification on the appropriate amino acid and thus allowing for specific sequence information. The use of both dissociation methods in tandem allows not only for thorough peptide sequencing for protein identification but also for specific mapping of glycosylation sites, providing an additional level of detail about the glycoproteins. Simultaneously, chemical strategies, antibodies and lectins have been used to enrich O-GlcNAcylated proteins, and have led to the identification of several O-GlcNAc-modifed sites by mass spectrometry. Most recently, the use of lectin weak affinity chromatography (LWAC) and ETD analysis enabled identification of up to 1750 O-GlcNAc sites from 6621 mouse brain synaptosome proteins [14], thus contributing to the better understanding of the role of O-GlcNAc in brain development.

In this study, hESCs were differentiated into neural stem cells (NSCs) to profile the differential expression of O-GlcNAcylated proteins and their sites of glycosylation throughout the differentiation process. To enrich for the O-GlcNAcylated proteins, we used LWAC. Following enrichment of O-GlcNAcylated proteins, ETD and HCD were used to fragment glycoproteins from four different stages of neural differentiation. Using this approach, we identified more than 600 unique O-GlcNAcylated proteins in stage 1 (undifferentiated hESCs) and more than 1000 unique O-GlcNAcylated proteins in stage 4 (NSCs), including 227 unambiguous sites of O-GlcNAcylation. Several of the proteins identified were previously known to be modified by O-GlcNAc, thus helping to validate our method. From the glycoproteins identified, the majority were involved in transcription, transcription regulation, transport and mRNA processing and splicing. The majority of the identified glycoproteins were also localized in the nucleus. Overall, this represents the first comprehensive characterization of protein O-GlcNAcylation in hESCs undergoing differentiation.

3.2. Results and Discussion

Dual inhibition of SMAD signaling converts hESCs to neural stem cells

Human embryonic stem cells were differentiated into neural stem cells (Fig 3.1) using a protocol reported by Chambers et al. [15]. In this protocol, hESCs are treated for 5 days with inhibitors of SMAD signaling: LDN-193189 (LDN) or Noggin, which inhibit bone morphogenic protein (BMP) signaling, and SB431542 (SB), which inhibits transforming growth factor β (TGFβ), Activin and Nodal signaling. The dual-SMAD inhibition protocol was chosen because it produces a pre-rosette neural stem cell in less than two weeks without the need for embryoid body formation or stromal feeder co-culture. These NSCs are capable of further differentiation into dopaminergic and motor neuronal cell fates. Inhibition of both branches of SMAD signaling strongly biases towards the differentiation of neuroectodermal lineage, thereby suppressing alternative embryonic germ layers [15]. During NSC formation, expression of pluripotency marker OCT4 is undetectable by western blotting after four days of neural induction, while the emergence of transcription factor and neuroectodermal marker PAX6 is observed by western blot five days post induction, indicating a successful differentiation of the human embryonic stem cell line H1 to NSCs over 11 days, as reported by Chambers et

al. (Fig 3.2A). Decrease in PAX6 expression after day 10 is indicative of cells differentiating into more mature NSCs. Verification of successful NSC formation after 11 days of induction is also shown by positive staining of PAX6 and TUJ1, an early neuronal marker, via immunofluorescence analysis (Fig 3.2B). Minimal fluorescence staining in the isotype control samples indicates that the staining observed in our experimental samples is specific to the expression of PAX6 and TUJ1, and not to due to non-specific fluorescence.

Global O-GlcNAc profile changes as hESCs differentiate to NSCs

Following verification of successful differentiation of hESCs into NSCs, we wanted to investigate the effects of developmental changes in the cellular state of protein O-GlcNAcylation. We analyzed global O-GlcNAcylation at every day of neural differentiation by western blotting with an O-GlcNAc-specific antibody (RL2) (Fig 3.3). Global O-GlcNAc levels fluctuate during neural induction, slightly decreasing during the first few days and towards the end of differentiation. It is possible that the proteins being O-GlcNAcylated are changing during the differentiation process. Meanwhile, OGT and OGA expression decrease by day 11 of differentiation (Fig 3.3). Similar results were obtained by other groups that looked at levels of O-GlcNAcylation in the rat brain [11], and in mouse embryonic neural precursor cells [16]. Furthermore, treatment of these neural precursor cells with an OGA inhibitor caused an increase in protein O-GlcNAcylation and a reduction in the number of neural precursor cells [16]. Overall, these results suggest that a decrease in O-GlcNAcylation seems to be important for the neural induction of ESCs, and that the changes in O-GlcNAc levels are not being modulated by the protein expression of OGT and OGA.

UDP-GlcNAc levels modulate global O-GlcNAcylation state[2]

OGT is thought to function as a metabolic sensor for glucose given the fact that levels of O-GlcNAcylation have been found to be influenced by glucose flux via the hexosamine biosynthetic pathway (HBP), which produces UDP-GlcNAc, the nucleotide sugar donor for OGT [17]. Using this as the possibility for the fluctuations seen in O-GlcNAc levels, we measured the levels of UDP-GlcNAc at every day of differentiation. Analysis of extracted UDP-GlcNAc using high performance anion exchange chromatography (HPAEC) from four biological replicates showed a similar oscillatory pattern during neural induction of hESCs (Fig 3.4). This result suggests that protein O-GlcNAcylation is being modulated by the cellular concentration of UDP-GlcNAc. Given that cell differentiation is generally associated with proliferation and increased metabolic requirement [18, 19], it is possible that the changes in UDP-GlcNAc concentration reflect a difference in glucose flux through HBP to instead produce nucleotides, amino acids and lipids. Another possibility is that the difference in glucose flux through the HBP regulates neural differentiation by influencing growth factor receptor surface expression due to changes in UDP-GlcNAc, and consequently, N-linked cell surface glycosylation [19].

[2] The result described in this section was obtained in collaboration with Nam Pham in the Kohler Group at UT Southwestern.

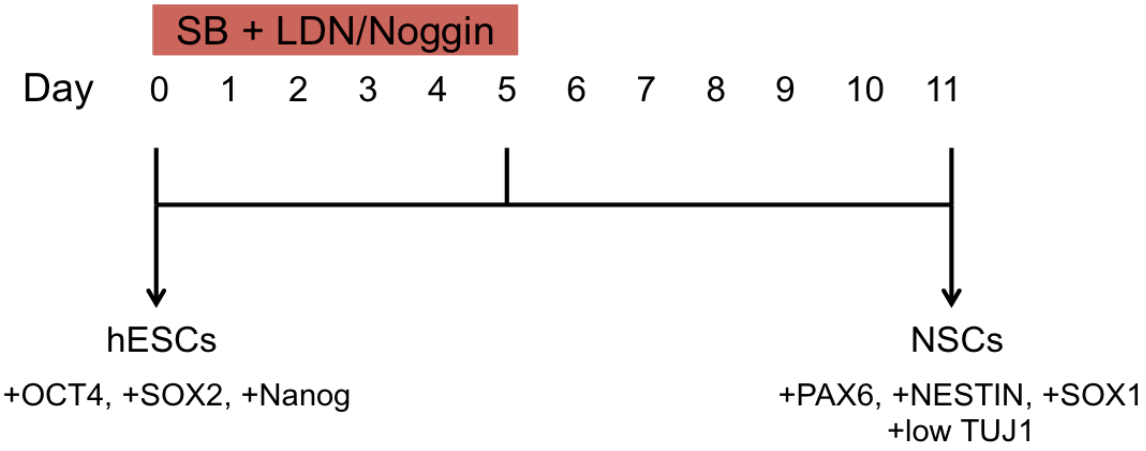


Figure 3.1. Experimental design used to promote the differentiation of hESCs to NSCs, based on dual-SMAD inhibition. hESCs were cultured for 5 days in media containing dual-SMAD inhibitors, SB431542 (SB, 10 μM) and LDN193189 (LDN, 100 nM) or Noggin (200 ng/ml).

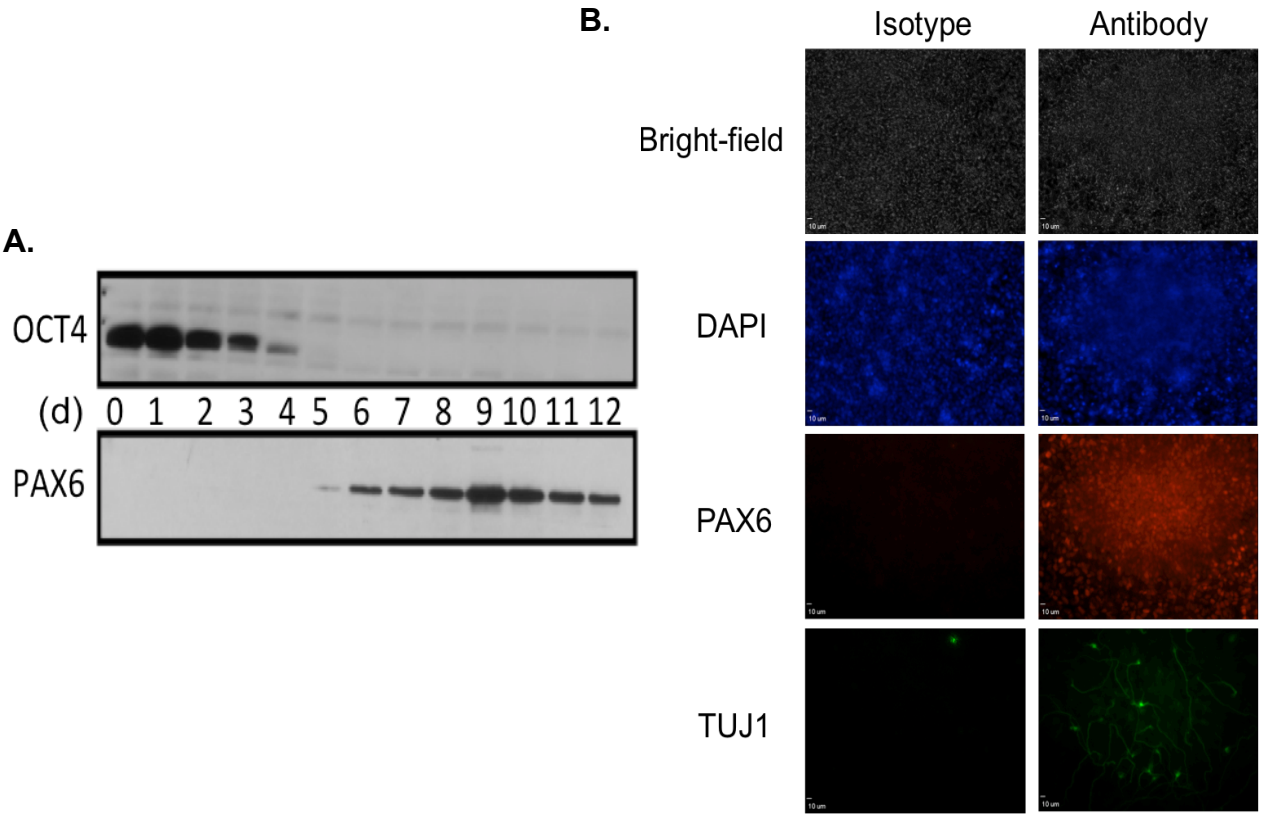


Figure 3.2. hESCs are converted into NSCs. (A) hESCs lose the expression of pluripotency marker OCT4, and start expressing PAX6, a neuroectoderm marker, by day 5, as shown by western blot. (B) Representative immunofluorescent image for PAX6 (red) and TUJ1 (green) expression at day 11 following dual-SMAD inhibition. Control cells were stained with respective isotype controls. DAPI (blue) is a nuclear stain. Scale bar, 10 μm.

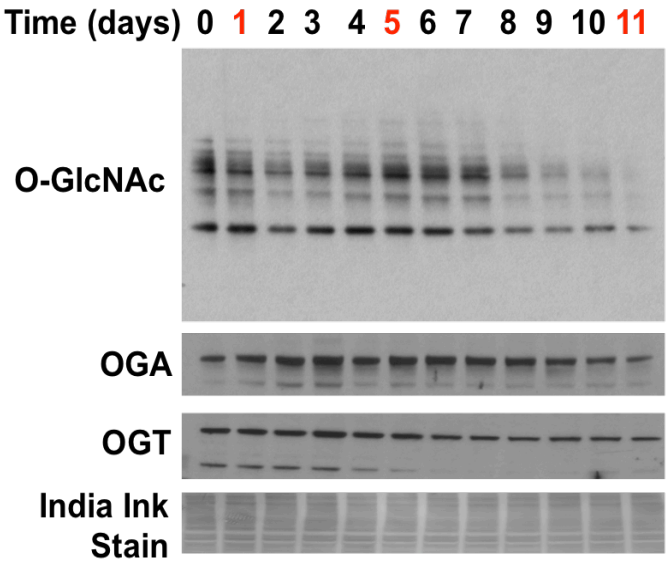


Figure 3.3. Global O-GlcNAc levels fluctuate as hESCs differentiate to NSCs. Cells were harvested at intervals of 24 hours and whole cell lysates where immunoblotted for O-GlcNAc, OGT and OGA. Days of differentiation in red were selected for additional MS based proteomic analysis. India ink staining showed equal loading.

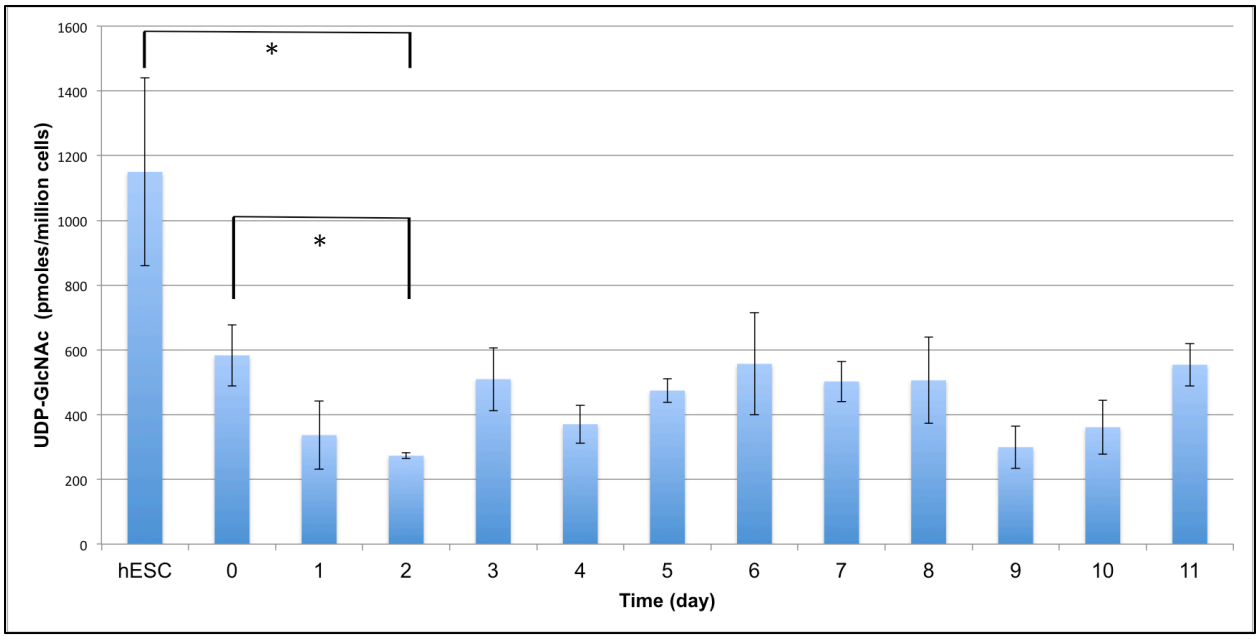


Figure 3.4. Relative concentration of UDP-GlcNAc in hESCs differentiated for 11 days. The amount of UDP-GlcNAc changes during differentiation, markedly dropping during the first days of neural induction. The oscillatory nature of the concentration of UDP-GlcNAc during differentiation resembles that of global O-GlcNAcylation (data are the mean ± S.E.M; *$p < 0.05$ using student t-test; $n = 4$).

Fractionated hESCs show proper cellular and differentiation markers

Based on the results shown previously in Figure 3.3, four time points were chosen to profile the O-GlcNAc-modified proteins during neural differentiation. These four time points represent four different stages of neural differentiation: 1) H1, representing the undifferentiated hESCs before day 0 of differentiation; 2) day 1, representing the start of the neural induction; 3) day 5, representing the end of dual-SMAD inhibition; 4) day 11, representing the final day of neural induction. Isolation of nuclear fractions from each of these time points was important given the inherent difficulties in detecting O-GlcNAc-modified proteins.

The O-GlcNAc modification is dynamic, present at substoichiometric levels, and prevalent on low-abundance regulatory proteins [20], thus making it necessary to perform cellular fractionations and to enrich specifically for the GlcNAc sugar. Here, we used a fractionation protocol developed by the Lamond lab [21] to isolate nuclear extracts. Due to its dynamic nature, and the fact that O-GlcNAc is quickly removed by hydrolases during cell lysis, we pre-treated cells with Thiamet-G, an O-GlcNAcase inhibitor, 1 hour before harvesting and also during the fractionation and lysis steps. Once we had prepared the nuclear fractions, we checked for the presence of nuclear (nuclear pore complex proteins; NPC) and cytosolic (tubulin) protein markers by performing western blot analysis (Fig 3.5). Similarly, we evaluated the differentiation stage of the samples using antibodies against OCT4, a pluripotency marker, and PAX6, a neuroectoderm marker (Fig 3.5). As expected, all the nuclear fractions were expressing NPC, although much less at day 11. This may be attributable to lower expression of certain nuclear pore complex proteins at this stage, since Ponceau staining of the western blot membrane shows similar loading amongst the four samples. The antibody used for NPC detection was raised using a nuclear pore complex mixture, primarily detecting p62, p152, p90 and other proteins. Previous experiments have shown that a specific change in NPC composition is required for neuronal differentiation [22], which suggests that at day 11 the NPC composition has changed and is not being detected by the antibody. Although contamination of cytosolic proteins was present in the nuclear extracts, especially in H1 and day 1 samples, it was minimal in comparison with the enriched nuclear fraction. Also, any possible contaminating N-glycan present from the cytosolic proteins was deglycosylated by treating all samples with Peptide N-glycosidase F (PNGase F) before LWAC enrichment. We also observed, as in Figure 3.2, the expression of OCT4 in H1 and day 1 samples, but not day 5 and 11; conversely, PAX6 was expressed by samples from day 5 and 11, demonstrating proper neural differentiation of samples fractionated.

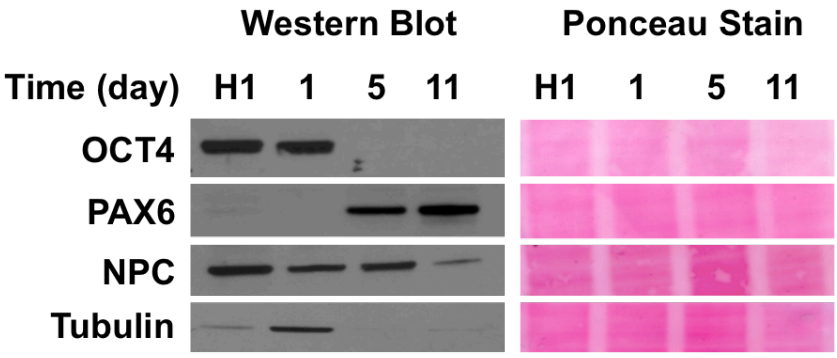


Figure 3.5. Western blot analysis of nuclear extracts corresponding to the 4 stages of differentiation used for LWAC-enrichment and MS analysis. Subcellular fractionation allowed us to probe for the presence of pluripotency (OCT4), neuroectoderm (PAX6), nuclear (NPC; nuclear pore complex), and cytosolic markers (tubulin). Ponceau staining of the western blot membranes was used to show equal loading. (H1 - human embryonic stem cell line used to start neural induction)

LWAC enrichment and ETD MS/MS analysis of O-GlcNAc-modified proteins [3]

To identify O-GlcNAc-modified proteins, we performed a large-scale enrichment of the four differentiation time points described above using a combination of lectin weak affinity chromatography (LWAC) and ETD MS/MS [23, 24] (Fig 3.6). hESCs were differentiated using the dual-SMAD inhibition protocol, and approximately 10 mg of protein were harvested and digested with trypsin. The resulting peptides were separated by lectin chromatography on a wheat germ agglutinin (WGA)-conjugated agarose column. Peptides containing GlcNAc monosaccharides migrated slower through the WGA column and eluted later than the unmodified peptides [23, 24]. The O-GlcNAc fractions were pooled and analyzed on an LTQ-Orbitrap Velos mass spectrometer using electron transfer dissociation (ETD) and higher-energy collisional dissociation (HCD). Of the two fragmentation methods, only ETD provided information regarding the location of the O-GlcNAc modification site. We identified 227 unambiguous O-GlcNAc modification sites on 104 unique proteins and 179 distinct O-GlcNAcylated peptides across all four time points analyzed (Table 3.1). Several of these O-GlcNAc-modified sites have been reported in homologous residues in rodents [6, 13, 14, 20, 23, 25, 26] suggesting evidence for site-specific conservation. Additionally, we identified 11 sites as potentially O-GalNAcylated based upon sub-cellular localization of the modification and annotation in the PhosphoSitePlus online resource [27].

We identified 30 novel O-GlcNAcylated sites in 20 previously unreported O-GlcNAc-modified proteins. Most of these proteins are localized in the nucleus, thus establishing the importance of cellular fractionation for the controlled enrichment of O-GlcNAc-modified proteins. Many of the newly identified O-GlcNAcylated proteins are known to be transcription factors, such as GRHL2, TCF12, RACK7 (ZMYND8) and SCML2. Interestingly, SCML2 is also part of the polycomb group (PcG) complex, along

[3] The results described in this section were obtained in collaboration with Samuel Myers in the Burlingame Group at UC San Francisco.

with another novel O-GlcNAcylated protein identified in this study, PHC1. PcG proteins function as epigenetic regulators by repressing transcription of genes necessary for controlling the self-renewal and differentiation of hESCs [28]. O-GlcNAcylation had been previously found to be important for polycomb repression [29, 30]. Similarly, polycomb repressive complex 2 was found to be necessary for proper O-GlcNAcylation in mouse ESCs [6]. Thus, the identification of two of the members of the PcG complex will help in further elucidating the molecular mechanisms involved in the regulation of O-GlcNAc in polycomb repression. The identification of two O-GlcNAc sites in TCF12 is important given the role of this transcription factor in initiating neuronal differentiation [31]. In addition to transcription factors, we also identified ALMS1, a protein involved in cell cycle regulation, UBN2, a protein involved in chromatin regulation, and SNRPA, a protein important in mRNA processing. We also observed O-GlcNAc modification of proteins previously shown to be involved in phosphorylation and ubiquitination, such as CDK8 and CBL, respectively. Interestingly, CDK8 is known to phosphorylate the CDK7/Cyclin H subunits of the general transcription initiation factor IIH (TFIIH), the subunit involved in phosphorylating RNA Pol II [32, 33]. Phosphorylation of TFIIH by CDK8 acts as a negative regulator of transcription since it is thought to halt the formation of the transcription-initiation complex [33]. O-GlcNAc modification of RNA Pol II occurs at the same site of TFIIH phosphorylation, thereby regulating transcription [34, 35]. Thus, it would be interesting to know if the O-GlcNAc modification of CDK8 has an influence on the O-GlcNAcylation/phosphorylation state of RNA Pol II.

We also localized 75 novel O-GlcNAc sites in proteins that were previously reported to be O-GlcNAcylated. Many of these O-GlcNAcylated proteins are known to be important in the maintenance and differentiation of hESCs. These include HCFC1, POU2F1, SALL2, SOX3, TAF4 and TET1. All of these proteins were enriched in stage 1, except for SOX3, which was enriched only in stage 4. SOX3 has been found to be expressed in the central nervous system and is thought to play a role in specifying neuronal cell fate [36, 37]. Given that SOX3's O-GlcNAcylation was only detected in stage 4, it is possible that the O-GlcNAcylation of SOX3 may be important for its role in mediating neuronal specification. Also, four sites of O-GlcNAcylation were identified in TET1, a protein involved in the hydrolysis of 5-methylcytosine (5mC) to 5-hydroxylmethylcytosine (5hmC) [38]. Previously, one site of O-GlcNAc modification had been reported for TET1, and was found to be important for regulating TET1's protein concentration and stability [6, 39]. Many of the novel O-GlcNAc-modified peptides identified correspond to proteins previously known to interact with pluripotency factor OCT4. These include GATAD2A, PHC1, EP400, ARID3B, HCFC1, RBM14, EMSY, NFRKB and UBN2 [40, 41]. Analysis of the novel O-GlcNAc sites identified here will be necessary to determine the significance of O-GlcNAcylation of these proteins and of their interaction with OCT4.

Sites of O-GlcNAc modification identified here occurred in proximity to proline and valine residues, as has been reported previously [14, 23, 26]. Many of the O-GlcNAcylated sites were within the P-X-T(GlcNAc)-X-A and P-V-S(GlcNAc) sequences. Interestingly, proline and serine residues are also enriched around phosphorylation sites [14, 42], suggesting an evolutionary preference for these residues to promote cross-talk between the two modifications. Of the 227 O-GlcNAc-modified sites identified in this study, 39 of them are also known to be phosphorylated, which again suggests the

potential cross-talk between both modifications. One of these potential cross-talk cases was observed in BPTF (Thr-2056), a component of the nucleosome-remodeling factor (NURF) complex involved in the regulation of genes and signaling pathways essential for embryonic development [43]. It has been demonstrated that phosphorylation of BPTF increases its DNA binding activity and regulates its subcellular localization [44]. Another example is the cell cycle regulator host cell factor 1 (HCFC1), with three O-GlcNAc-modified sites (Thr-1486, Ser-1150 and Thr-515) also known as phosphorylated. Interestingly, Ser-515 is within the region of HCFC1 found to be required for its interaction with OGT [5, 45]. Although it is known that OGT and O-GlcNAcylation are required for HCFC1 cleavage, little is known about the role of phosphorylation in HCFC1 function and stability, other than it is cell cycle-regulated [45, 46]. Other O-GlcNAcylated proteins with reciprocal phosphorylation sites include nuclear pore proteins, NUP153 (Thr-515) and NUP35 (Ser-53 and Ser-298). Phosphorylation of Nup98 has been shown to be important for NPC disassembly [47], and O-GlcNAcylation of Nup62 and Nup88 was shown to modulate their protein levels and interaction [48]. Thus, these results suggest that competition between O-GlcNAcylation and phosphorylation may play a role in maintaining the structure and functional integrity of NPCs.

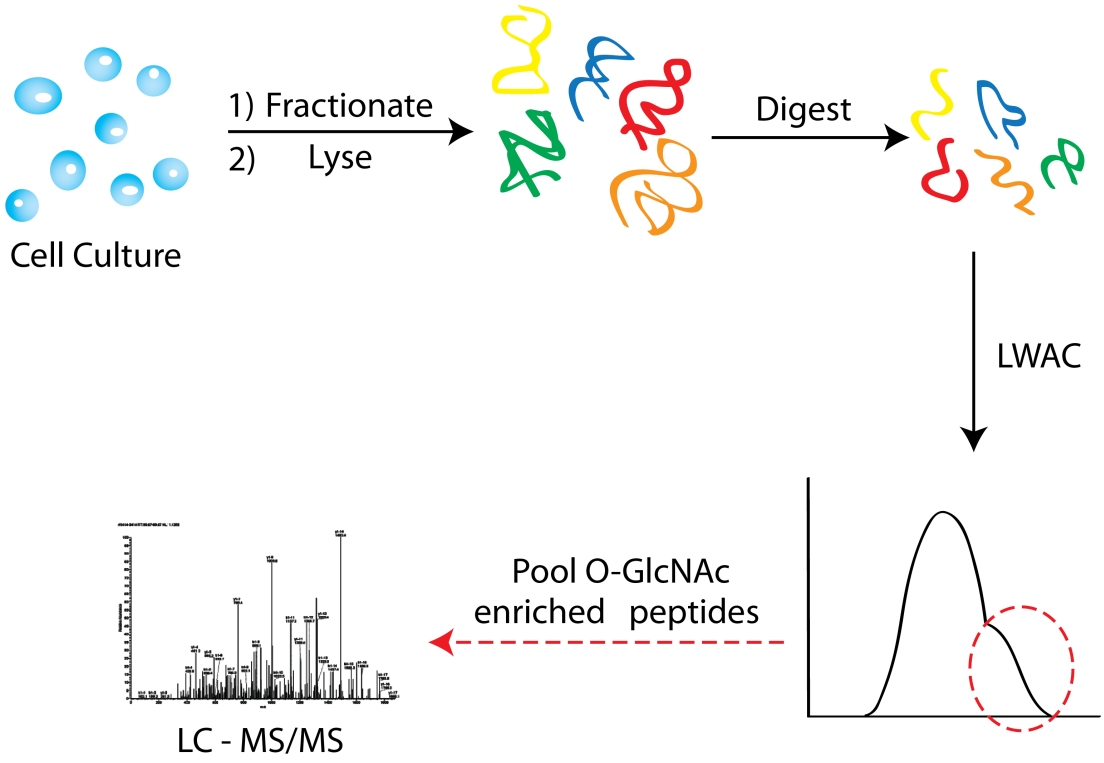
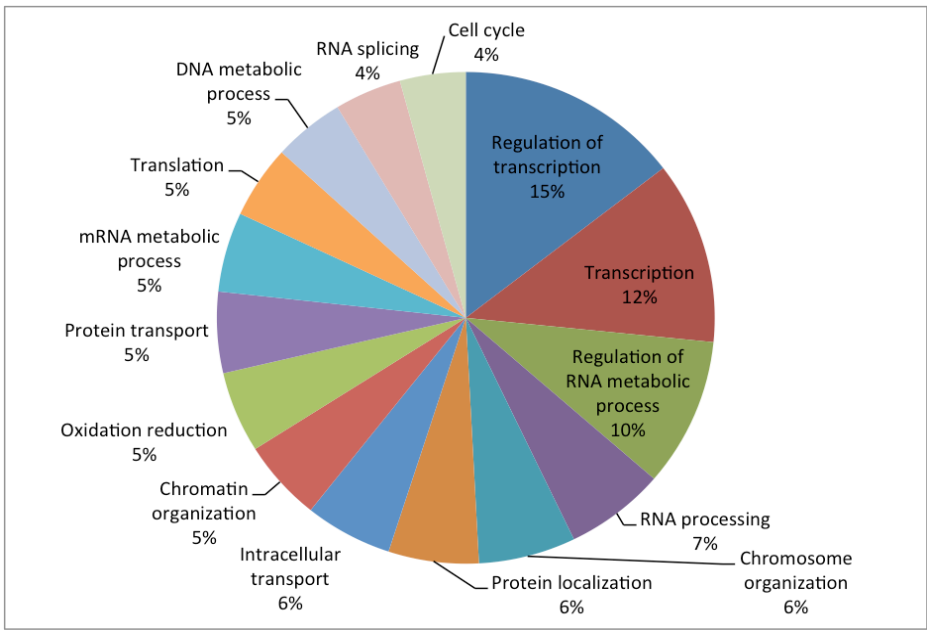


Figure 3.6. Workflow implemented to identify O-GlcNAc-modified peptides in differentiating hESCs. In this strategy, the lectin wheat germ agglutinin (WGA) is attached to agarose to bind terminal GlcNAc residues (LWAC). After the nuclear O-GlcNAcylated peptides are enriched, they are sequenced by mass spectrometry. The red doted circle represents the GlcNAc-peptides, which migrate slower through the WGA column and elute later than the unmodified peptides. LC: liquid chromatography.

Ontology analysis of O-GlcNAc-modified proteins identified in stage 1 and 4

We performed functional annotation and gene ontology analysis using DAVID v.6.7 [49, 50]. Analysis of the O-GlcNAcylated proteins identified across stages 1 (undifferentiated hESCs) and 4 (day 11 of neural differentiation), revealed prevalence for proteins involved in the regulation of transcription and RNA processing, respectively (Fig 3.7). This is consistent with previous studies indicating the importance of O-GlcNAc in regulating transcriptional events [51]. The high percentage of RNA processing proteins in stage 4 is interesting since previous studies have shown that a large fraction of the alternative splicing events occur exclusively in the nervous system (NS), thus suggesting a key role in supporting the complex functions of the NS [52, 53]. Functional annotation analysis of more than 600 unique proteins identified as O-GlcNAcylated in stage 1 indicated that 24% are classified as phosphoproteins and 12% are localized in the nucleus. Meanwhile, out of the more than 1000 unique proteins identified as O-GlcNAcylated in stage 4, 17% are classified as phosphoproteins, 15% as having polymorphisms and 10% as being localized to the nucleus (Fig 3.8). It is possible that the high percentage of polymorphisms and alternative splicing present at stage 4 is due to the necessity of genetic variability as an organism evolves and matures. The lower than expected percentage of nuclear-localized proteins confirms the presence of contaminating cytosolic and membrane proteins as shown previously in Figure 3.5, and possibly of proteins without a known cellular localization.

A.



B.

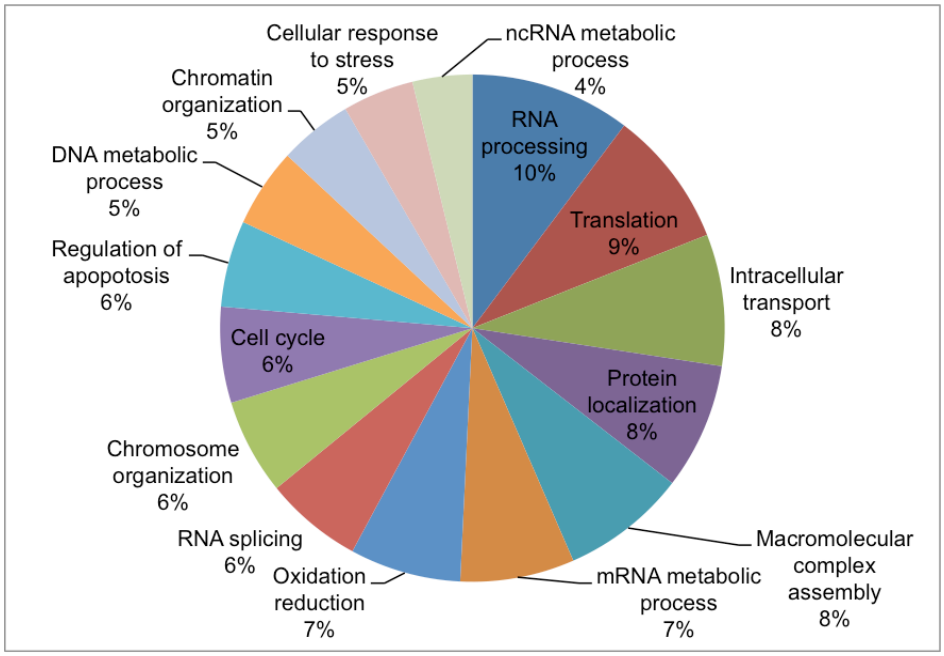
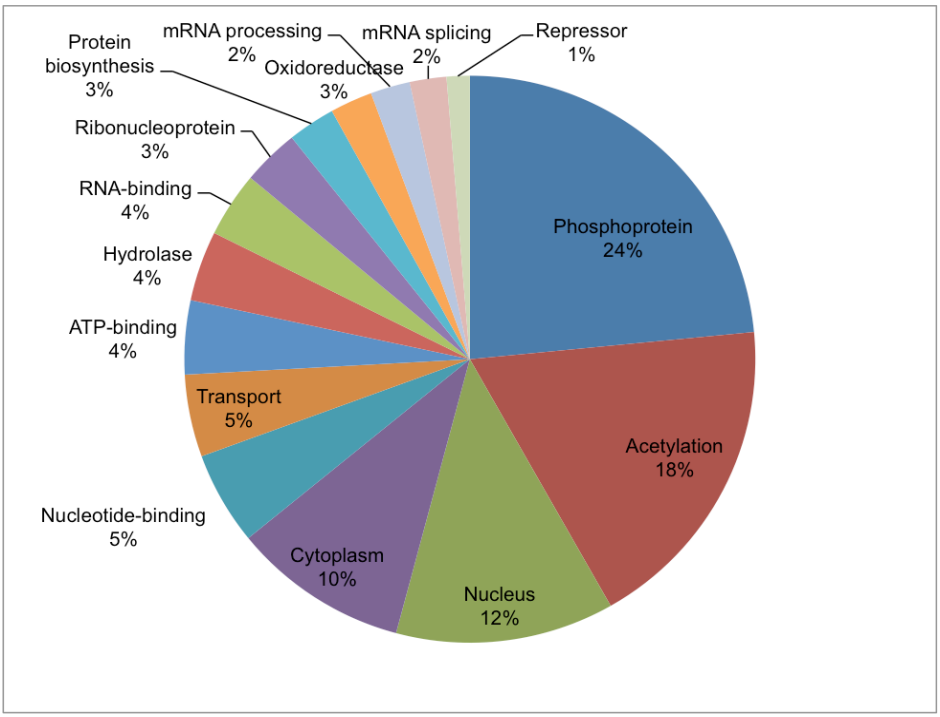


Figure 3.7. Biological processes most-represented from analysis of (A) 615 genes from stage 1 and (B) 1141 genes from stage 4. Charts were generated using DAVID Bioinformatics Resources v.6.7.

A.



B.

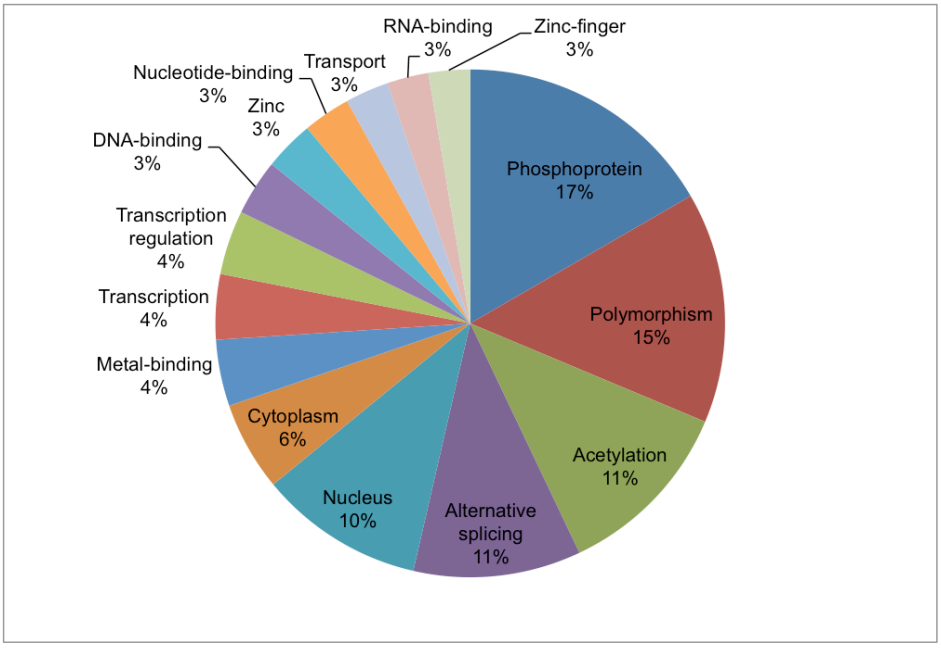


Figure 3.8. Functional categories most-represented from analysis of (A) 740 genes from stage 1 and (B) 1344 genes from stage 4. Charts were generated using DAVID Bioinformatics Resources v.6.7.

Relative quantitation of O-GlcNAc-modified peptides identified in stage 1 and 4

To get an approximation of the abundance of the O-GlcNAc-modified peptides identified in our MS analysis we used Skyline version 2.5.0.6157 [54]. Skyline is an open-source software that supports quantitation of label-free mass spectrometry data [54-56]. The software measures quantitative differences in peptide expression using the MS1 scans from the mass spectrometry data acquired for each stage of differentiation analyzed. First, we imported raw output files from the LTQ-Orbitrap Velos mass spectrometer to extract precursor ion chromatograms from MS1 scans. Next, spectral libraries were generated by importing all of the HCD data from the four stages of differentiation analyzed. This was done to ensure a comprehensive library containing the maximum number of identifications possible from our acquired data. Skyline then created a custom report containing the retention time and peak areas of different isotopic peaks (M, M+1, M+2) for a given peptide. An example of an extracted ion chromatogram of the M (blue), M+1 (purple), and M+2 (red) peaks for the peptide YPSPFFVFGEK from the protein DHX9, generated by Skyline, is displayed in Figure 3.9. This figure shows the chromatographic profiles for the given peptide in stage 1 (top panel) and stage 4 (bottom panel).

For a qualitative assessment of the varying peptide abundances between stage one and stage four, total area under the curves of the M, the M+1, and the M+2 liquid chromatography trace curves were summed for the peptide that was identified and divided by the mean of the peak areas for all peptides analyzed in a given stage. Peptides with a total peak area higher than 1 correspond to peptides present in higher abundance in stage 1 of differentiation (undifferentiated hESCs) in comparison to stage 4 (NSCs). Whereas, peptides with a total peak area lower than 1 correspond to peptides present in higher abundance in stage 4 than in stage 1 of differentiation. As shown in figure 3.10, the majority of peptides have a ratio of 1, with similar abundance in stage 1 and 4. Table 3.2 lists a selection of peptides with the most-significant difference in abundance between stages 1 and 4. This corresponds to peptides with a ratio higher than 100 and lower than 0.001. Many of the most abundant peptides in stage 1 correspond to proteins involved in transcription regulation, as expected. These include SOX2, BPTF, CHD4, GTF2H1, and SUPT16H. Furthermore, the most abundant peptide of SOX2 is one that is also O-GlcNAc-modified and phosphorylated, suggesting a regulatory role of both modifications in the control of pluripotency through SOX2, as has been described before [6, 7, 57]. Additionally, we also see high abundance of peptides for chromosomal proteins such as HIST1H2BK and HIST1H3A. Many of the abundant peptides in stage 4 correspond to proteins involved in mRNA splicing, such as CDC5L, PRPF19, CRNKL1, HNRNPK, and HNRNPM. Since the data is a representation of only one biological replicate for each stage of differentiation, we cannot make a definitive quantitative comparison between samples. Thus, an orthogonal method needs to be used to validate the difference in peptide abundance observed between samples.

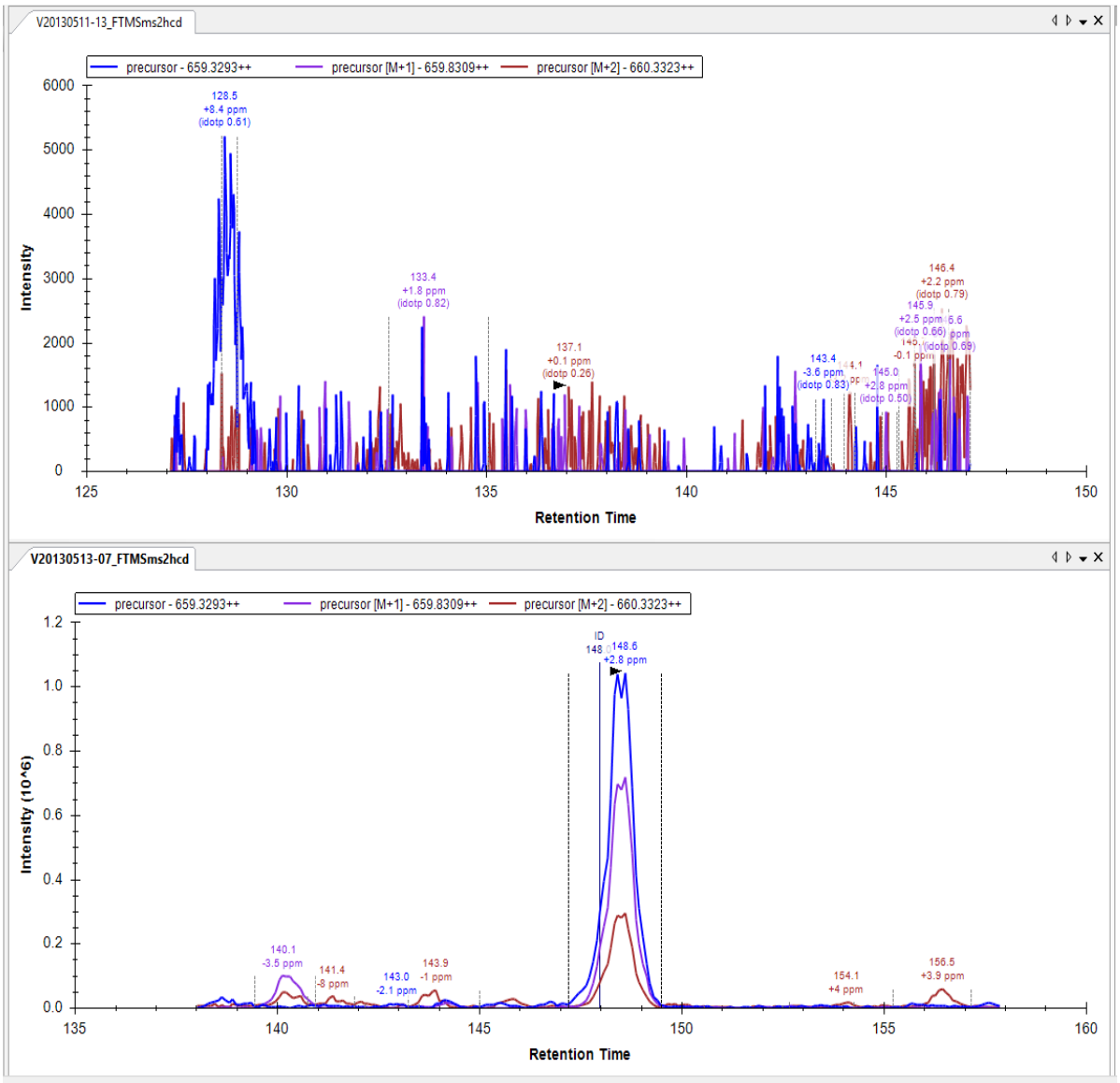


Figure 3.9. Relative quantitation of peptide abundance using Skyline. Skyline display from peak areas and imported targeted full MS and MS/MS data from stage 1 (top panel) and 4 (bottom panel) of the peptide YPSPFFVFGEK. Three precursor ion isotope peaks, M (blue), M+1 (purple), and M+2 (red), are shown. At low abundance, there is no apparent signal observed above the level of noise, indicating that the peptide is likely absent from the sample in stage 1; however, the filtered precursor ions from the stage 4 are abundant, producing a clear chromatographic peak.

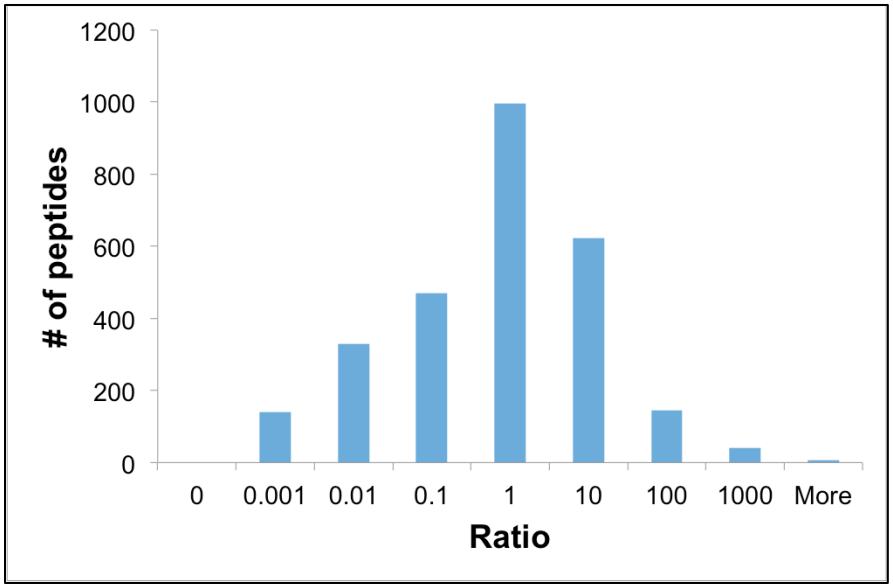


Figure 3.10. Distribution of peptide abundance in stage 1 and 4. Histogram representing the distribution of peptides with an abundance ratio higher or lower than 1, as analyzed using Skyline. Ratio corresponds to (Total peak area in stage 1/mean of all peak areas in stage 1) / (Total peak area in stage 4/mean of all peak areas in stage 4).

3.3. Conclusion

The precise mechanism by which hESCs self-renew while maintaining the ability to differentiate into virtually all cell types is not well understood. In order to differentiate, cells need to undergo multiple developmental steps regulated by epigenetic factors, transcription factors, and signaling pathways. Analysis of the molecular mechanisms involved in this process is essential for broadening our understanding of hESC fate determination. In addition to regulating protein activity, O-GlcNAcylation also influences cell identity by modifying proteins involved in epigenetics, transcription activation and repression, and signaling. We used MS-based proteomics to study the O-GlcNAc state of hESCs undergoing neural differentiation. The results described in this chapter provide a valuable resource for understanding the role of O-GlcNAc during neural differentiation, and they complement genomic, transcriptomic and epigenetic analysis to reveal insights into the mechanisms underlying neural differentiation of hESCs.

Table 3.1. Unambiguous O-GlcNAc-modified sites identified in this study

Gene	Peptide sequence	O-GlcNAc site(s)	Novel
AHDC1	KASGTYAGPPT**S**ALPAQR	S877	site
ALMS1	TETPSV**S**SSLYSYR	S1707	protein and site
	TGVSTVT**S**TSYSHR	S1802	protein and site
ANKHD1	SIHANFSSGVGT**T**AASSK	T182	site
ANKRD17	IGSSAPT**TT**AANTSLMGIK	T1825, T1826	site (in human)
ARID2	VDSVPDVSPAPSPAGIPHGSQ**T**IGNHFQR	T645	site
ARID3B	LAVPV**T**LASQQAGTR	T408	site
ARNT	HSNPTQGATPTWTPT**T**R	T643	site
ATF7IP	NPV**S**LPSLPNP**T**KPNNVPSVPSPSIQR	S834, T842	site
BPTF	VMVAPI**S**GSVTTGTK	S2081	no
	STVT**T**TTTTV**T**K	S1749, T1753, T1759	no
	VGSPATV**T**FQQNK	T2103	no
	LEQQKPTVIA**T**STTSPTSSTTSTISPAQK	T2056	T2056
	GQPVSTAVSAPN**T**VSSTPGQK	T2237	T2237
	QTVV**SST**ENCAK	S1741, S1742, T1743	site (in human)
	TVITEVTTMTSTVATESK	T1709	site (in human)
	FLFTPLA**TT**ATTASTT**T**TTVSTTAAGTGEQR	T2342, T2343, T2351	no
C11orf30 (EMSY)	ITF**T**KPS**T**QT**T**NTTTQK	T265, T269, T272	T265 and T269
	LVTTPTG**T**QATYTRPTVSPSIGR	T507	no
	IIS**S**NIV**S**GTTTK	S558, S562	S562
	TITVPV**S**GSPK	S237	no
	T**T**SGSIITVVPK	T536	T536
CANX	V**T**YKAPVP**T**GEVYFADSFDR	T59, T66	no
CBL	VPV**S**APSSSDPWTGR	S601	protein and site
CDK8	VVPPT**T**TSGGLIMTSDYQR	T411	protein and site
CDKN2AIP	**SS**SSTNTSLLTSK	S349, S350	S350(in human)
Unknown	TPPS**TT**VGSHSPPETPVLTR	T687, T688	protein and site
Unknown	QHGVNV**S**VNASATPFQQPSGYGSHGYNTGR	S860	protein and site
CHD8	APGYPSSPVT**T**ASGTTLR	T2523	site
CNOT1	APLAGQVSTMV**TTS**TTTTVAK	T1041, T1042, S1043	s (in human)
	TVTVTRPTGVSFK	T1051	s (in human)
CNPY3	ASPL**T**HSPPDEL	T271	O-GalNAc
CPVL	QAIHVGNQ**T**FNDGTIVEK	T167	site
CRB2	EHFASWPG**T**PAPILGCR	T1047	O-GalNAc
DKFZp451J085	IGGDLTAAV**T**K	T15	protein and site
DKFZp666C163	AFLYEPT**T**QASGR	T738	protein and site
DKFZp686A01127	AT**S**GAATPVIASTK	S472	protein and site

DKFZp686L2367	**S**APSPTSNSTYLTMNAASR	S292	protein and site
	EIPNT**T**VSNFR	T316	protein and site
EGR1	VLVET**S**YPSQTTR	S120	no
EIF4ENIF1	SVLHPPGSG**S**HAAAVSVQTTPQNVPSR	S914	site
ELF2	SPT**TT**ASVSATAAPR	T375, T376	site (in human)
EP400	AAAAPFQT**S**QASASAPR	S1527	site
	TAAPTTASAAPQGPLR	T1488	site
ETAA1	**S**LSSQVDTPIMTK	S342	protein and site
FAM208B	VASY**S**GTV**T**QATFTR	S969, T973	protein and site
FIP1L1	ETALPS**T**KAEFTSPPSLFK	T253	no
GATAD2A	GTQNIPAGKPSLQT**S**SAR	S210	site
	EATAQKPTG**S**VGSTVTTPPPLVR	S182	site (in human)
GATAD2B	**S**ISQSISGQK	S584	no
GIGYF2	ISDQNIIPSV**T**R	T680	site (in human)
GLTSCR1	HGGAGGSPSV**T**WAR	T1260	protein and site
	QVPV**S**GYLASAAGPSEPVTLASAGVSPQGAGLVIQK	S260	protein and site
GRHL2	EQYSISFPESSAIIPV**S**GITVVK	S153	protein and site
HCFC1	AVT**T**VTQSTPVPGPSVPK	T1486	site
	TIPMSAII**T**QAGATGVTSSPGIK	T779	no
	TMAVTPGTTTLPATVK	T579	no
	VM**S**VVQTKPVQTSAVTGQASTGPVTQIIQTK	S685	no
	VASSPVMV**S**NPATR	S603	site
	TAAAQVGT**S**V**SS**A**T**NTSTRPIITVHK	S620, S622, S623, T625	no
	IATGHGQQGV**T**QVVLK	T831	site (in human)
	I**S**VATGALEAAQG**S**K	S1150, S1162	S1162
	SGTVTVAQQAQVV**TT**VVGGVTK	S651, S652	no
	ALT**T**EVPIPAK	T1514	site
	V**T**GPQATTGTPLVTMRPASQAGK	T490	no
	HSHAV**ST**AAMTR	S1238, T1239	no
	VMT**S**GTGAPAK	S806	no
	APV**T**VT**S**LPAGVR	T515, S518	S518
	SPITIIT**T**K	T801	no
	TQGVPAVLK	T480	no
	LVTPV**T**VSAVKPAV**T**TLVVK	T861, T870	no
HIPK1	**T**VVGAAATTTTVTTK	T151	site (in human)
IRF2BPL	SRFEYPPPPV**S**LGSSSHTAR	S179	site (in human)
JMJD1C	SPTHLTV**S**STNTLR	S1192	site
	IPEHLPHQIA**S**HSVTTFR	S1173	site
	HSVPQ**S**LPQSNYFTTLSNSVVNEPPR	S1087	site
	V**S**TTAPV**T**LASSK	S1451, T1457	S1451
KRT18	SLGSVQAPSYGARPVS**S**AASVYAGAGGSGSR	S31	no

KRT8	SYKV**S**TSGPR	S13	no
LMAN1	LVSGMQHPG**S**AGGVYETTQHFIDIK	S425	O-GalNAc
LMNA	SVGGSGGGSFGDNLV**T**R	T643	no
LRP2	**T**CHPEYFQCTSGHCVHSELK	T3797	O-GalNAc
	AYVCDHDNDCQDGSDEHACNYP**T**CGGYQFTC PSGR	T221	O-GalNAc
	R**T**CSENEFTCGYGLCIPK	T2992	O-GalNAc
LRPAP1	QV**T**SNSLSGTQEDGLDDPRLEK	T134	O-GalNAc
MAP4	RASPSKPASAPA**S**R	S797	no
MINPP1	DPVAS**S**LSPYFGTK	S42	O-GalNAc
NCOR1	RTPV**S**YQNTMSR	S1487	no
NCOR2	YPPH**S**LSYPVQIAR	S23	site
	VV**T**LAQHISEVITQDYTR	T2142	no
	GSEPRPLVPPV**S**GHATIAR	S1997	site (in human)
	GSPVT**T**REPTPR	T1570	no
NFRKB	IQTVPASHLQQG**T**ASGSSK	T1273	no
	VV**S**HSGSAGLSQVR	S795	site
	TVAVA**S**GAASTPISISTGAPTVR	T1141, S1146	site
NUP153	VQM**T**SPSSTGSPMFK	T515	site
	CQPVF**S**FGNSEQTKDENSSK	S1134	site
	QQEPV**TS**TSLVFGK	T1112, S1113	T1112
	STFSF**S**M**T**KPSEKESEQPAK	S1154, T1156	S1154
	FGV**S**SSSSGPSQTLTSTGNFK	S908	no
	NVF**S**SSGTSFSGR	S1456	site
	TAELSGSSSTLEPIISSSAHHVTTVNSTNCK	T556	site
NUP214	SHLVHG**S**SPGVMGTSVATSASK	S1044	site
	AAPGPGPSTFSFVPP**S**K	S526	site
	LGELLFPS**S**LAGETLGSFSGLR	S1325	site
NUP35	**S**ISGPSVGVMEMR	S53	site (in human)
	AST**S**DYQVISDR	S298	site (in human)
PAMR1	CACLAGY**T**GQR	T267	O-GalNAc
PHACTR4	FIIS**T**SI**T**TAPAATTAATSLAK	T202, T205	site
PHC1	KAEADGSGQQNVGMNL**T**R	T354	protein and site
PHF21A	FTPTTLPT**S**QNSIHPVR	S277	site (in human)
	TV**TT**ASMITTK	T124, T125	site (in human)
PODXL	**T**PSP**T**VAHESNWAK	T330, T334	O-GalNAc
POGZ	V**T**SSIPVFDLQDGGR	T361	site (in human)
POM121	APPTLQAETATKPQA**TS**APSPAPK	T693, S694	no
POU2F1	**T**IAATPIQTLPQSQSTPK	T255	site
PPP1R12A	RQDDLISSSVPS**T**TSTPTVTSAAGLQK	T570	site (in human)
PRRC2B	GGLPV**S**QSQEIFSSLQPFR	S1982	site
QSER1	TAQAAA**S**GTTLLPQFR	S17	site

	T**S**QGTVPTALAFER	S105	no
	QSSLSCSPIGDSTQV**S**NGGLQQK	S245	site (in human)
	N**S**TNLIQTPQIR	S660	site
	FLPAVQSS**S**FASSTHCQTLQNNITSPDPK	S523	site
RAD54L2	VVTT**T**DIVIPGLNSSTDVQAR	T1075	site
RANBP2	EGFSIPV**S**ADGFK	S1894	no
RBM12	VNLPTTVSNFNNPSPSVVTA**TT**SVHESNK	T137, T138	site
RBM27	MMSKPQT**S**GAYVLNKVPVK	S738	site
RGPD8	**S**TSGEGFQFGK	S977	protein and site
RBM14 (SIP)	AQP**S**A**S**LGVGYR	S188, S190	S190 (in human); S188
	ASYVAPL**T**AQPATYR	T165	site (in human)
	AQPSV**S**LGAAYR	S178	no
	AQPSV**S**LGAPYR	S214	site (in human)
RPRD2	SAVS**T**SVP**T**KPTENISK	T400, T404	T404
	TPAPAT**T**TSHNPLANILSK	T501	site
SALL2	**T**LASSSSSSSSSSGAETPK	T253	site
SAP130	I**T**LPSHPALGTPK	T319	site (in human)
	SSLIPI**S**GHR	S433	site
SAP30BP	KG**TTT**NATSTT**T**TTASTAVADAQK	T231, T232, T233, T240	T231,T232,T233 (in human); T240
SCML2	NPMYIHT**S**VSQDFSR	S83	protein and site
SEC16A	**T**LENPVNVYNPSHSDSLASQQSVASHPR	T823	site
SEC24B	TPPTANHPVEPV**T**SVTQPSELLQQK	T341	site
SH3RF1	IGVFPGNYVAPV**T**R	T505	site (in human)
SNRPA	KAVQGGGATPVVGAVQGPVPGMPPM**T**QAPR	T148	protein and site
SOX2	**S**EASSSPPVVTSSSHSR	S246	site (in human)
SOX3	**S**EPSSPPPAIASHSQR	S376	site
SP2	**S**STTTTPVQSGANVVK	S187	site (in human)
SRCAP	AETQGANHTPVI**S**AHQTR	S2416	no
TAB1	VYPVSVPY**S**SAQSTSK	S395	no
TAF4	QV**S**QAQTTVQPSATLQR	S528	no
	TVPGATTTSSAATETMENVKK	T575	site
	GAAGAV**T**QSLSR	T409	site
TCF12	LSYPPH**S**VSPTDINTSLPPMSSFHR	S281	protein and site
	GSTSSSPYVAA**S**HTPPINGSDSILGTR	S311	protein and site
TCF4	LSYPSH**S**SADINSSLPPMSTFHR	S269	site (in human)
TET1	NVNVVC**S**GGITVVSTK	S1361	site
	SIAQGIITLDNCSNDLHQLPPR	S1020	site
	QVHISFLPAN**T**QGFPLAPER	T512	site
	GLFHA**S**LGIAQLSQAGPSK	S527	site
TLE4	**T**DAPTPGSNSTPGLRPVPGKPPGVDPLASSLR	T330	site (in human)
TNK2	TMPT**T**QSFASDPK	T818	site (in human)

TNRC6C	FLAQGQALPP**T**SSWQSSSASSQPR	T1602	site
UBAP2	LLQLPST**T**IENISVSVHQPQPK	T487	site
UBN2	VHQHSAVQQNYVSPLQATI**S**K	S948	protein and site
	TVPS**TT**TSSNYLAK	T1003, T1007, T1008	protein and site
VEZF1	KTPT**T**VVPLI**ST**IAGDSSR	T111, S117, T118	site (in human)
VIM	SYVT**T**STR	T33	no
	TYSLGSALRPST**S**R	S49	site (in human)
WNK1	EGPVLAT**S**SGAGVFK	S1849	site (in human)
WNK3	QIMAPV**T**NSSSYSTTSVR	T974	site
YEATS2	QEPGEAPHVPA**T**GAASQSPLPQYVTVK	T604	site (in human)
	GFPAS**T**EAER	T389	site
	TSGQQQVCV**S**QATVGTCK	S998	site
	AIV**S**GGGGTIVAQPVQTLTK	S703	site
	QAVAI**S**GGQILVAK	S670	site (in human)
ZFR	AGY**S**QGATQY**T**QAQQTR	S195, T202	no
	QQEAPPPPPPA**T**TQNYQDSYSYVR	T135	site (in human)
ZMYND8	TPPS**TT**VGSHSPPETPVLTR	T750, T751	protein and site
ZNF281	VKTPT**S**QSYR	S891	no
ZNF318	TVVAHTSPWMPVV**TT**STQTK	T1341, T1342	T1342
ZYX	V**S**SGYVPPPVATPFSSK	S169	site

Residues in bold represent site of O-GlcNAc modification within the given peptide sequence. Residues in red represent site of phosphorylation, as identified in this study. O-GlcNAcylated sites and/or proteins are defined as: (protein and site) if both are novel; (site) if only the site is novel but not the protein; (site in human) if the site has been identified for the first time in human; (T/S) if only the given site is novel; (no) if neither the site nor the protein are novel; (O-GalNAc) if the site is possibly O-GalNAc modified, not O-GlcNAc modified. (T = Threonine, S = Serine)

Table 3.2. Relative quantitation of selected peptides using Skyline

Gene	Peptide sequence	Ratio (Stage 1 / Stage 4)[4]
GTF2H1	FFQSHYFHR	1526.234569
HIST1H3A	YRPGTVALR	1204.422933
SUPT16H	RLYSNWR	1008.529353
SF3B1	HFWQHR	553.5188548
HIST1H2BK	EIQTAVR	314.6471494
PPP2R5C	ILPIMFPSLYR	290.4840611
ZMYM3	HFCNQQCLLR	275.1923504
KPNA2	FVSFLGR	245.239902
CHD4	QSYWNHR	242.2322785
SOX2	SEASSSPPVVTSSSHSR	189.9649049
BPTF	KYWFLNR	182.5992191
HCFC1	LYIWSGR	147.9966748
HNRNPL	VFNVFCLYGNVEK	103.0150003
PRPF19	ALQDEWDAVMLHSFTLR	0.000525697
BRD3	HQFAWPFYQPVDAIK	0.000485843
TRIP12	LFLQFVTGSPR	0.000409523
CRNKL1	ANPHNYDAWFDYLR	0.00037645
NUP155	YMDLLWR	0.000354956
TUBB	ISEQFTAMFR	0.000349809
GARS	TFFSFPAVVAPFK	0.000336301
CHD1L	GIPTYIYYFPR	0.000312092
NUP133	YMTQFADQNFSDFLFR	0.00026959
ROCK2	AFVGNQLPFIGFTYYR	0.000241153
MYO1B	NFHVFYQLLSGASEELLNK	0.000193739
CDC5L	WYEWLDPSIK	0.000173985
HNRNPM	NLPFDFTWK	0.000171788
PSIP1	LPIFFFGTHETAFLGPK	0.000157014
SMC5	QYGFFSYLR	0.000151489
HNRNPUL1	WMGIAFR	0.000151005
CARM1	ANFWYQPSFHGVDLSALR	0.000146006
DHX9	YPSPFFVFGEK	0.000130569
HSP90AB3P	HSQFIGYPITLYLEK	0.000114303
PPM1G	LPLPYGFSAMQGWR	0.000102955
ARIH2	YTLQYTYPYAYYMESGPR	9.32769E-05
STIP1	FMNPFNMPNLYQK	2.99274E-05
TUBA1B	AVFVDLEPTVIDEVR	1.38956E-05
HNRNPK	IITITGTQDQIQNAQYLLQNSVK	1.10243E-05

[4] Ratio = (Total area in stage 1/mean of all areas in stage 1) / (Total area in stage 4/mean of all areas in stage 4)

3.4. Experimental Methods

Human embryonic stem cell (hESC) culture

The NIH approved human embryonic stem cell line H1 (WiCell Research Institute) was cultured on 10 cm polystyrene dishes coated with hESC-qualified basement membrane matrix Matrigel (BD Biosciences). Matrigel plates were made according to manufacturer specifications. Cells were grown under standard conditions (37 °C, 5% CO_2) in TeSR2 media (STEMCELL Technologies Inc). Media was exchanged daily after the first 48 h in culture, and cells were passaged every 3 to 5 days using 1 mg/ml of dispase (Gibco) in Knockout DMEM media (Invitrogen) and mechanically removed. Karyotype analysis was routinely performed and indicated that all samples were diploid and had no chromosomal abnormalities.

Neural differentiation

Human embryonic stem cell line H1 was differentiated into neural progenitors that have the capacity to become neurons and glia. The differentiation protocol was performed as reported by Chambers et al [15]. Briefly, cells were dissociated into single cells using Accutase (STEMCELL Technologies Inc.) and plated on Matrigel-treated dishes at a density of 40,000–50,000 cells/cm^2 in the presence of MEF-conditioned hESC medium containing 10 ng/ml FGF-2 (R&D Systems) and 10 μM ROCK inhibitor (Y-27632; Tocris) for single cell survival. MEF-conditioned hESC medium was obtained by culturing mytomycin-C-treated mouse embryonic fibroblasts (MEF; Millipore PMEF-CF) in DMEM/F12, 20% (v/v) Knockout Serum Replacement (Invitrogen), 1 mM GlutaMAX, 100 μM MEM nonessential amino acids, 0.1 mM β-mercaptoethanol (Invitrogen), and 6 ng/ml FGF-2. After 24 hours of cell growth, media was harvested and sterile filtered. The ROCK inhibitor was withdrawn the day after plating, and hESCs were allowed to expand for 3-4 days or until they were 80-90% confluent. Neural differentiation was initiated using KSR medium, which contained Knockout DMEM (Invitrogen), 15% (v/v) Knockout Serum Replacement, 1 mM GlutaMAX, 100 μM MEM nonessential amino acids and 0.1 mM β-mercaptoethanol. To inhibit SMAD signaling, 100 nM LDN-193189 (Stemgent) or 200 ng/ml Noggin (3344-NG-050, R&D Systems) and 10 μM SB431542 (Tocris) were added on days 0–5 of induction. Cells were fed daily, and N2 medium was added in increasing 25% increments every other day starting on day 4 and leading to 100% N2 on day 10. N2 media consisted of DMEM/ F12 powder, 1:1 (Gibco/Invitrogen) resuspended in distilled water, glucose, sodium bicarbonate, putrescine, progesterone, sodium selenite, insulin (Sigma), and apo-transferrin (Kamada Ltd).

Sample preparation and fractionation

Biological samples were collected at four different time points, corresponding to undifferentiated hESCs and days 1, 5 and 11 of neural differentiation (as described above). hESCs at different stages of neural induction were pre-treated for 1 hour with 100 μM of O-GlcNAcase inhibitor Thiamet-G (Cayman Chemical), before harvesting. Cells were then scraped in the presence of ice-cold 1X PBS (Thermo Scientific), pooled

into 50 ml conical tubes and centrifuged for 5 minutes at 1200 rpm at 4 °C. Cell pellets were washed twice with ice-cold 1x PBS and centrifuged for 5 minutes each time. After the second PBS wash, nuclear extracts were made as follows (Lamond's protocol: http://www.lamondlab.com/f7nucleolarprotocol.htm), with protease and phosphatase inhibitors (Roche) and 2 μM Thiamet-G included throughout: Pellets were placed on ice and resuspended with a homogenization buffer consisting of 10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, and 0.5 mM DTT. Cells were swelled for 5 minutes on ice, and broken by Dounce homogenization using a tight pestle (Kimble Chase). Crude nuclei were pelleted by centrifugation (1200 rpm, 10 min at 4 °C). Nuclear pellet was resuspended in buffer S1 consisting of 250 mM sucrose and 10 mM $MgCl_2$, and layered over an equal volume of buffer S3 consisting of 880 mM sucrose and 0.5 mM $MgCl_2$ and pelleted by centrifugation (3500 rpm, 15 min at 4 °C). Finally, the nuclear pellets were washed twice with ice-cold 1X PBS and centrifuged at 3650 rpm for 10 minutes at 4 °C. After the last PBS wash, cell pellets were frozen in dry ice and kept at -80 °C until used for LWAC enrichment. A fraction of the nuclear extract was kept for western blotting analysis and the rest was used for LWAC enrichment of O-GlcNAc-modified peptides.

Immunofluorescence

hESCs were fixed with 4% (v/v) paraformaldehyde (Thermo Scientific) for 40 minutes, rinsed three times with 1X PBS for 5 minutes, and blocked and permeabilized for at least one hour with 5% (v/v) FBS plus 0.3% (v/v) Triton X-100 in PBS. Cells were stained overnight at 4 °C with the following primary antibodies: PAX6 (Covance, PRB-278P, 1:500) and TUJ1 (Millipore, MAB1637, 1:1,000). Cells were rinsed three times with 1X PBS, and then were incubated with Alexa Fluor 488-conjugated goat-anti-mouse secondary antibody and Alexa Fluor 594-conjugated goat-anti-rabbit secondary antibody (1:400, Invitrogen) for 2 hours at room temperature. After, cells were rinsed three times with 1X PBS, and DAPI (Invitrogen) was added for 5 minutes before analysis using an Olympus IX71 fluorescent microscope. Control cells were stained with mouse and rabbit IgG isotype (Millipore, PP54, PP64). All image acquisition and processing was performed under identical conditions for test and control samples.

Western blot analysis

Nuclear extracts were prepared as described before. Pellets were lysed by sonication in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% (v/v) Triton X-100, 1 μM Thiamet-G, EDTA-free protease inhibitors and phosphatase inhibitors (Roche). Insoluble material was removed by centrifugation at 12,000 x g for 10 minutes. Protein concentration was determined by bicinchoninic acid assay (Pierce). Samples were separated by SDS-PAGE (4-12% Bio-Rad), transferred to nitrocellulose membranes, and blocked with 5% milk (LabScientific, inc.) in PBS with 0.1% Tween 20 (Sigma). Membranes were incubated with primary antibodies overnight at 4 °C. The following primary antibodies were used to detect the indicated proteins: anti-OGT (1:1000, DM-17, Sigma, sc-32921, Santa Cruz Biotech), anti-OCT4 (1:2500, MAB4401, Millipore), anti-nuclear pore complex (1:2000, MMS-120R, Covance), anti-α-tubulin (1:5000, T5168, Sigma), anti-O-

GlcNAc (1:1000, RL2, Thermo Scientific), anti-OGA (1:1000, SAB4200267, Sigma) and anti-PAX6 (1:2000, PRB-278P, Covance). The following horseradish peroxidase-linked secondary antibodies were used for detection at a dilution ratio of 1:5000: anti-rabbit HRP-conjugated secondary antibody (4010-05, Southern Biotech) and anti-mouse-HRP-conjugated secondary antibody (1010-05, Southern Biotech).

High Performance Anion Exchange Chromatography (HPAEC)

Cells were induced to neural progenitors as indicated before. During harvest, cells were treated with Accutase to dissociate colonies into single cells and counted with a hemocytometer. Approximately 3 x 10^6 cells were harvested per sample. Cells were spun, and washed with cold 1X PBS twice, immediately placed in dry ice and kept at -80 °C until lysis. Cells were then lysed with 80% "super-cold" methanol (on dry ice) [58]. Lysate was spun at 2,000 x g for 15 min. The supernatant was dried by speed vacuum for 4-5 hours. The intracellular metabolite pellet was either used immediately, or it was stored at -80 °C for future use.

Metabolite pellet was resuspended in 40 mM sodium phosphate buffer (pH 7.4 40 µL per million cells), and filtered through an Amicon® Ultra centrifugal filter unit (Millipore, 10,000 MWCO). Filtrates were analyzed by HPAEC (ICS-3000 system, Dionex) with CarboPac™PA1 (Dionex) with a pulsed amperometry detector (PAD) and UV-detector in-line [59, 60]. Typically, 20 µL of metabolite was injected into the sample loading loop before the sample enters a guard column (Dionex, 4 × 50 mm) and then an analytical column (Dionex, 4 × 250 mm). The eluents used were 1.0 mM NaOH (C) and 1.0 M NaOAc and 1.0 mM NaOH (D). Low-carbonate NaOH (50% in water) was obtained from Fisher Scientific (SS254-1) and NaOAc was from Sigma (71183). HPAEC was run with a flow rate = 1 mL/min and the following gradient elution was performed: T_0 min = 95% C, T_5 = 85% C, T_{15} = 70% C, T_{20} = 60% C, T_{45} = 60% C, T_{50} = 0% C, T_{60} = 0% E2, T_{65} = 95%, T_{75} = 95%. UDP-GlcNAc standards (50 µM, 25 µM, 10 µM, and 2.5 µM) were injected at the same time as cellular samples. UDP-GlcNAc peak areas were input into excel and raw data were converted to pmoles of UDP-GlcNAc by comparing to a standard curve generated by analyzing the peak areas of the UDP-GlcNAc standards. Data was normalized to cell number.

Enrichment of O-GlcNAc-modified peptides

Nuclear extracts were resuspended in 6 M guanidine hydrochloride, 25 mM ammonium bicarbonate, 2 µM Thiamet-G (Cayman Chemical) and 2 µM PUGNAc (Tocris), phosphatase inhibitors (Sigma) and 2 mM tris(2-carboxyethyl)phosphine (TCEP, Thermo Scientific). The resuspended nuclei were sonicated and cleared by centrifugation. Soluble proteins were reduced for 30 minutes at 55 °C and alkylated using 10 mM iodoacetamide (IAM, Sigma) for 15 minutes at room temperature. Proteins were digested with TPCK trypsin (Sigma) at a 1:50 protease:substrate ratio for 18 hours. After digestion, the sample was acidified with formic acid and desalted using a C18 Sep-Pak (Waters), before drying down by vacuum centrifugation.

O-GlcNAc modified peptides were enriched by lectin weak affinity chromatography (LWAC) as previously described [14]. Briefly, wheat germ agglutinin (WGA) coupled in-house to aldehyde-coated POROS resin was packed into a 2 mm x 250 mm column. An isocratic flow of LWAC buffer (25 mM Tris, pH 7.8, 300 mM NaCl, 5 mM $CaCl_2$, 1 mM $MgCl_2$, 5% acetonitrile) was used at 100 µL/min for the entire run except for a 200 mM GlcNAc plug to elute any glycopeptdes bound to WGA. Three mg of protein per LWAC run was used to prevent WGA-column overloading. The O-GlcNAc enriched fractions were pooled, acidified with formic acid and desalted with C18 before drying by vacuum centifugation.

Separation and MS/MS analysis of O-GlcNAc-modified peptides

Chromatography was performed on a Nanoacquity HPLC (Waters) at 400 nl/min with an EASY-spray C18, 3 µm particle size, 75 µm ID x 15 cm column and source (Thermo). A 180-minute gradient from 2% solvent A (0.1% formic acid) to 35% solvent B (0.1% formic acid in acetonitrile) was used. Mass spectrometry was performed on an LTQ-Orbitrap Velos equipped with ETD (Thermo). Data dependent analysis selected the six most highly abundant, multiply charged ions within a 3 Da isolation window for subsequent HCD. ETD was automatically triggered and performed on any precursor in which the HCD scan produced a product ion of 204.09 m/z and/or 138.05 m/z in the top 20 peaks. Precursor scans and HCD product ions were measured in the Orbitrap at a resolution of 30,000 and 7,500, respectively. ETD product ions were measured in the ion trap with one microscan, allowing 100 ms for ion injection time. Normalized activation energy and activation time for HCD was set to 32 and 30 msec, respectively. ETD activation time was charge state dependent, where doubly charged precursors reacted for 100 ms, triply charged for 66.6 msec, and so on. Automatic gain control for precursor ions was set at 1e6, 5e4 MS/MS scans and 1e6 for the ETD reagent, fluoranthene. Supplemental activation was enabled for ETD. Dynamic exclusion was set for 45 seconds.

Data Analysis

Raw data was converted to peaklists using in-house software called PAVA. HCD and ETD peaklists were searched separately using Protein Prospector v 5.10.0. against the Swissprot database with a concatenated, decoy database (21 March, 2012) where 36,775 entries were searched. The false discovery rate was set to be less than 0.1 percent for HCD. Only human and mouse genomes were searched. Precursor mass tolerance was set to 10 ppm, where fragment ion error was allowed at 20 ppm and 0.6 Da for HCD and ETD, respectively. Cysteine residues were assumed to be carbamidomethylated, variable modifications considered were *N*-terminal acetylation, *N*-terminal pyroglutamine conversion, methionine oxidation, O-linked HexNAc, N-linked HexNAc, and S/T/Y phosphorylation. SLIP scoring [61] was reported for all modification site localization unless the modification site was ambiguous (a false localization rate of less than 5%), then manual interpretation was employed [62].

Relative quantitation was obtained using Skyline version 2.5.0.6079 [54]. Protein Prospector and MS Convert software were used to create .pep and .mzML files. Spectral libraries were generated using the combined HCD data from all four stages of differentiation analyzed. For a qualitative assessment of the varying peptide abundances between stage one and stage four, total area under the curves of the M, the M+1, and the M+2 liquid chromatography trace curves were summed for the peptide that was identified. This ensured that, even without identification from Protein Prospector, the peptide could be compared between the two stages of differentiation. The data were then exported, and those peptides with the most-significant differences between stages one and four were selected for additional validation studies.

Gene ontology analysis of the O-GlcNAc-modified proteins identified by Protein Prospector was performed using Database for Annotation, Visualization and Integrated Discovery (DAVID) v.6.7 software [49, 50].

3.5. References

1. Ma, J. and G.W. Hart, *O-GlcNAc profiling: from proteins to proteomes.* Clin Proteomics, 2014. **11**(1): p. 8.

2. Shafi, R., et al., *The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny.* Proceedings of the National Academy of Sciences, 2000. **97**(11): p. 5735-5739.

3. O'Donnell, N., et al., *Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability.* Mol Cell Biol, 2004. **24**(4): p. 1680-90.

4. Jackson, S.P. and R. Tjian, *O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation.* Cell, 1988. **55**(1): p. 125-33.

5. Wysocka, J., et al., *Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation factor HCF-1.* Genes Dev, 2003. **17**(7): p. 896-911.

6. Myers, S.A., B. Panning, and A.L. Burlingame, *Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9490-5.

7. Jang, H., et al., *O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network.* Cell Stem Cell, 2012. **11**(1): p. 62-74.

8. Vella, P., et al., *Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.* Mol Cell, 2013. **49**(4): p. 645-56.

9. Akimoto, Y., et al., *Localization of the O-GlcNAc transferase and O-GlcNAc-modified proteins in rat cerebellar cortex.* Brain Res, 2003. **966**(2): p. 194-205.

10. Liu, K., et al., *Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism.* Journal of Neurochemistry, 2004. **89**(4): p. 1044-1055.

11. Liu, Y., et al., *Developmental regulation of protein O-GlcNAcylation, O-GlcNAc transferase, and O-GlcNAcase in mammalian brain.* PLoS One, 2012. **7**(8): p. e43724.

12. Rexach, J.E., P.M. Clark, and L.C. Hsieh-Wilson, *Chemical approaches to understanding O-GlcNAc glycosylation in the brain.* Nat Chem Biol, 2008. **4**(2): p. 97-106.

13. Chalkley, R.J., et al., *Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides.* Proc Natl Acad Sci U S A, 2009. **106**(22): p. 8894-9.

14. Trinidad, J.C., et al., *Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse.* Mol Cell Proteomics, 2012. **11**(8): p. 215-29.

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

16. Yanagisawa, M. and R.K. Yu, *O-linked beta-N-acetylglucosaminylation in mouse embryonic neural precursor cells.* J Neurosci Res, 2009. **87**(16): p. 3535-45.

17. Wells, L., K. Vosseller, and G.W. Hart, *Glycosylation of Nucleocytoplasmic Proteins: Signal Transduction and O-GlcNAc.* Science, 2001. **291**(5512): p. 2376-2378.

18. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.

19. Wellen, K.E., et al., *The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism.* Genes Dev, 2010. **24**(24): p. 2784-99.

20. Khidekel, N., et al., *Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(36): p. 13132-13137.

21. Boisvert, F.M., et al., *A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage.* Mol Cell Proteomics, 2010. **9**(3): p. 457-70.

22. D'Angelo, Maximiliano A., et al., *A Change in Nuclear Pore Complex Composition Regulates Cell Differentiation.* Developmental Cell, 2012. **22**(2): p. 446-458.

23. Vosseller, K., et al., *O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry.* Mol Cell Proteomics, 2006. **5**(5): p. 923-34.

24. Ma, Z., Y. Skorobogatko, and K. Vosseller, *Tandem Lectin Weak Affinity Chromatography for Glycoprotein Enrichment*, in *Mass Spectrometry of Glycoproteins*, J.J. Kohler and S.M. Patrie, Editors. 2013, Humana Press. p. 21-31.

25. Teo, C.F., et al., *Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc.* Nat Chem Biol, 2010. **6**(5): p. 338-43.

26. Alfaro, J.F., et al., *Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets.* Proc Natl Acad Sci U S A, 2012. **109**(19): p. 7280-5.

27. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Research, 2012. **40**(D1): p. D261-D270.

28. Di Croce, L. and K. Helin, *Transcriptional regulation by Polycomb group proteins.* Nat Struct Mol Biol, 2013. **20**(10): p. 1147-55.

29. Gambetta, M.C., K. Oktaba, and J. Muller, *Essential role of the glycosyltransferase sxc/Ogt in polycomb repression.* Science, 2009. **325**(5936): p. 93-6.

30. Sinclair, D.A., et al., *Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc).* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13427-32.

31. Uittenbogaard, M. and A. Chiaramello, *Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis.* Brain Res Gene Expr Patterns, 2002. **1**(2): p. 115-21.

32. Hengartner, C.J., et al., *Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases.* Mol Cell, 1998. **2**(1): p. 43-53.

33. Akoulitchev, S., S. Chuikov, and D. Reinberg, *TFIIH is negatively regulated by cdk8-containing mediator complexes.* Nature, 2000. **407**(6800): p. 102-6.

34. Kelly, W.G., M.E. Dahmus, and G.W. Hart, *RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc.* J Biol Chem, 1993. **268**(14): p. 10416-24.

35. Ranuncolo, S.M., et al., *Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo.* J Biol Chem, 2012. **287**(28): p. 23549-61.

36. Collignon, J., et al., *A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2.* Development, 1996. **122**(2): p. 509-20.

37. Mojsin, M., et al., *Mapping of the RXRalpha binding elements involved in retinoic acid induced transcriptional activation of the human SOX3 gene.* Neurosci Res, 2006. **56**(4): p. 409-18.

38. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.* Nature, 2010. **466**(7310): p. 1129-1133.

39. Shi, F.T., et al., *Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells.* J Biol Chem, 2013. **288**(29): p. 20776-84.

40. van den Berg, D.L., et al., *An Oct4-centered protein interaction network in embryonic stem cells.* Cell Stem Cell, 2010. **6**(4): p. 369-81.

41. Pardo, M., et al., *An Expanded Oct4 Interaction Network: Implications for Stem Cell Biology, Development, and Disease.* Cell Stem Cell, 2010. **6**(4): p. 382-395.

42. Swaney, D.L., et al., *Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry.* Proc Natl Acad Sci U S A, 2009. **106**(4): p. 995-1000.

43. Landry, J., et al., *Essential Role of Chromatin Remodeling Protein Bptf in Early Mouse Embryos and Embryonic Stem Cells.* PLoS Genet, 2008. **4**(10): p. e1000241.

44. Jordan-Sciutto, K.L., J.M. Dragich, and R. Bowser, *DNA binding activity of the fetal Alz-50 clone 1 (FAC1) protein is enhanced by phosphorylation.* Biochem Biophys Res Commun, 1999. **260**(3): p. 785-9.

45. Daou, S., et al., *Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2747-52.

46. Wysocka, J., et al., *Developmental and cell-cycle regulation of Caenorhabditis elegans HCF phosphorylation.* Biochemistry, 2001. **40**(19): p. 5786-94.

47. Laurell, E., et al., *Phosphorylation of Nup98 by Multiple Kinases Is Crucial for NPC Disassembly during Mitotic Entry.* Cell, 2011. **144**(4): p. 539-550.

48. Kodiha, M., et al., *Dissecting the signaling events that impact classical nuclear import and target nuclear transport factors.* PLoS One, 2009. **4**(12): p. e8420.

49. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.

50. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

51. Bond, M.R. and J.A. Hanover, *O-GlcNAc cycling: a link between metabolism and chronic disease.* Annu Rev Nutr, 2013. **33**: p. 205-29.

52. Zhu, H., et al., *A nuclear function of Hu proteins as neuron-specific alternative RNA processing regulators.* Mol Biol Cell, 2006. **17**(12): p. 5105-14.

53. Makeyev, E.V., et al., *The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing.* Mol Cell, 2007. **27**(3): p. 435-48.

54. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-8.

55. Sherrod, S.D., et al., *Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides.* J Proteome Res, 2012. **11**(6): p. 3467-79.

56. Liu, S., et al., *Exploring skyline for both MS(E) -based label-free proteomics and HRMS quantitation of small molecules.* Proteomics, 2014. **14**(2-3): p. 169-80.

57. Hanover, J.A., M.W. Krause, and D.C. Love, *Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation.* Nat Rev Mol Cell Biol, 2012. **13**(5): p. 312-21.

58. Dettmer, K., et al., *Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols.* Anal Bioanal Chem, 2011. **399**(3): p. 1127-39.

59. Tomiya, N., et al., *Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: sugar nucleotide contents in cultured insect cells and mammalian cells.* Anal Biochem, 2001. **293**(1): p. 129-37.

60. Yu, S.H., et al., *Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4834-9.

61. Baker, P.R., J.C. Trinidad, and R.J. Chalkley, *Modification site localization scoring integrated into a search engine.* Mol Cell Proteomics, 2011. **10**(7): p. M111.008078.

62. Medzihradszky, K.F., *Peptide sequence analysis.* Methods Enzymol, 2005. **402**: p. 209-44.

Chapter 4

Perturbing O-GlcNAc cycling during neural differentiation of hESCs using chemical inhibitors

Chapter 4. Perturbing O-GlcNAc cycling during neural differentiation of hESCs using chemical inhibitors[1]

4.1. Introduction

Stem cells maintain a balance of self-renewal and differentiation programs through the activation and repression of specific developmental genes. The transcriptional and epigenetic mechanisms involved in regulating the expression of these genes have been the focus of research of many labs around the world. Still, very little is known about the role that posttranslational modifications play in maintaining hESC pluripotency and cell-fate determination. However, recent studies have suggested a potential role of the O-GlcNAc modification in regulating these processes [1-4]. O-GlcNAc is a non-canonical form of protein glycosylation that occurs within the nucleus and cytoplasm [5, 6]. The enzymes involved in the addition and removal of O-GlcNAc, OGT and OGA, respectively, have been shown to target key transcriptional and epigenetic regulators. For example, OGT is required for the proper processing of the cell cycle regulator host cell factor 1 (HCF-1) [7-9]. O-GlcNAcylation was found to modulate the function of the Polycomb group (PcG) proteins, which are chromatin regulators implicated in the repression of lineage-specific regulatory genes [10, 11]. Moreover, OGT modifies and interacts with several core regulators of pluripotency, such as SOX2 and OCT4 [1, 12-14]. Very recently, members of the ten-eleven translocation protein family, or TET proteins, were shown to act as binding partners of OGT during gene transcription, directly affecting the O-GlcNAcylation of histone tails [15-21].

O-GlcNAc modification is the final step in the hexosamine biosynthetic pathway (HBP), which ends with the production of UDP-GlcNAc, the nucleotide sugar donor for OGT. Levels of UDP-GlcNAc are highly responsive to intracellular concentrations of glucose, amino acids, and lipids [22]. Because of this, O-GlcNAcylation is sensitive to insulin, to nutrients, and to cellular stress. It therefore functions to link cellular signaling with nutrient availability [23]. Moreover, maintaining a proper balance of O-GlcNAc in the cell is important, since abnormal amounts of O-GlcNAcylation underlie the cause of developmental defects, and diseases such as cancer and Alzheimer's [24]. For example, in *C. elegans* dysregulation of O-GlcNAc cycling leads to changes in transcription and metabolic function [22, 25, 26]. In zebrafish, altering O-GlcNAcylation by overexpressing OGT or OGA, or by reducing OGT expression, resulted in embryos with shortened body axes and reduced brains [13]. Additionally, elevated O-GlcNAcylation has been described in various cancers [27, 28]. Finally, glucose metabolism was found to be impaired in Alzheimer's disease neurons, which is associated with reduced O-GlcNAcylation of proteins such as tau [29, 30].

O-GlcNAc has also been shown to play a crucial role in maintaining proper neuronal function. Both OGT and OGA are highly expressed in the brain and are specifically enriched at neuronal synapses, where they localize to synaptic vesicles [31-36]. The O-GlcNAc modification is present on many proteins important for cell signaling, synaptic plasticity, learning, and memory. Examples include cAMP-responsive element binding protein (CREB) [37, 38], β-amyloid precursor protein (APP) [39], tau [40], and bassoon [41]. Additionally, perturbations to normal O-GlcNAcylation

[1] Ian Blong contributed to the work described in this chapter.

can display profound developmental defects. For example, neuron-specific OGT deletion contributed to neuronal dysfunction and resulted in neonatal death [42]. Accumulation of protein O-GlcNAcylation in the mouse hippocampus and in mouse neural precursor cells coincided with neuronal apoptosis [43, 44]. Furthermore, overexpression of OGA was found to increase the percentage of neurons exhibiting axon branching [45]. Finally, excess levels of O-GlcNAc, induced by inhibition of OGA, in differentiating mouse and human ESCs impaired the ability of cells to differentiate into the ectodermal lineage [4, 46]. Altogether, these results suggest that the O-GlcNAc modification and the enzymes controlling its addition and removal to proteins are involved in maintaining proper neuronal development.

In this chapter, we investigated the effects of perturbing O-GlcNAcylation, during differentiation of hESCs to neural stem cells (NSCs) via inhibition of OGT. We found that chemical inhibition of OGT caused a decrease in both global O-GlcNAcylation and UDP-GlcNAc levels, while maintaining the oscillations in O-GlcNAc and UDP-GlcNAc levels previously described in chapter 3. Additionally, upon inhibition of OGT, we observed that NSCs gained morphology reminiscent of premature neurons, acquiring the neuronal markers β-III tubulin (TUJ1) and microtubule-associated protein 2 (MAP2). Gene expression analysis revealed that genes involved in neuronal differentiation and axonal guidance signaling were significantly up-regulated in OGT-inhibited cells. Interestingly, among the most down-regulated genes were PPP1R3C, a protein phosphatase, and FGF23, which is involved in regulating phosphate concentrations. This might allude to the well-established crosstalk between phosphorylation and O-GlcNAcylation.

The mechanism by which OGT inhibition induces neuronal differentiation was analyzed using the Ingenuity Pathway Analysis (IPA) software. Transforming growth factor beta 1 (TGFβ1) was identified as the most represented upstream regulator in our data set, in addition to being up-regulated in OGT-inhibited cells. These findings illustrate a novel mechanism by which modulation of O-GlcNAcylation influences cellular differentiation.

4.2. Results and Discussion

Inhibition of OGT lowers global O-GlcNAc levels in hESCs

Chemical inhibitors of OGT and OGA are extremely important tools for understanding the biological functions of O-GlcNAc in the cell. Until recently, the chemical approaches available for reducing the levels of O-GlcNAc in cells involved the use of non-specific and toxic inhibitors [47]. Here we utilized an OGT-specific inhibitor developed by Gloster et al. that relies on ‘hijacking’ the biosynthetic pathway of UDP-GlcNAc [48]. The sugar analogue, per-O-acetylated 2-acetamido-2-deoxy-5-thio-D-glucopyranose (Ac-5SGlcNAc), is converted through HBP into UDP-5SGlcNAc, which binds to OGT and inhibits its activity, resulting in the global decrease of O-GlcNAc levels in cells. To evaluate the effect of treating hESCs with Ac-5SGlcNAc, we started by probing different concentrations of the inhibitor (OGTi) for 24 h. As a control, we treated a separate plate of cells with vehicle (DMSO). As shown in Figure 4.1A, treatment with Ac-5SGlcNAc lowered O-GlcNAc levels in a dose-dependent manner.

We also observed a time-dependent decrease in O-GlcNAc for up to 3 days following treatment with Ac-5SGlcNAc. After 3 days, O-GlcNAc levels start to increase. This is possibly due to the decrease in O-GlcNAcase (OGA) protein levels and the slight increase in OGT protein levels (Fig. 4.1B). OGT inhibitor did not affect hESC pluripotency, as measured by OCT4 expression. Similarly, small interfering RNA (siRNA) silencing of OGT for 72 hours did not affect OCT4 expression (not shown). Most importantly, treatment of cells with the inhibitor did not cause apparent changes in cell morphology or cell proliferation (Figure 4.1C).

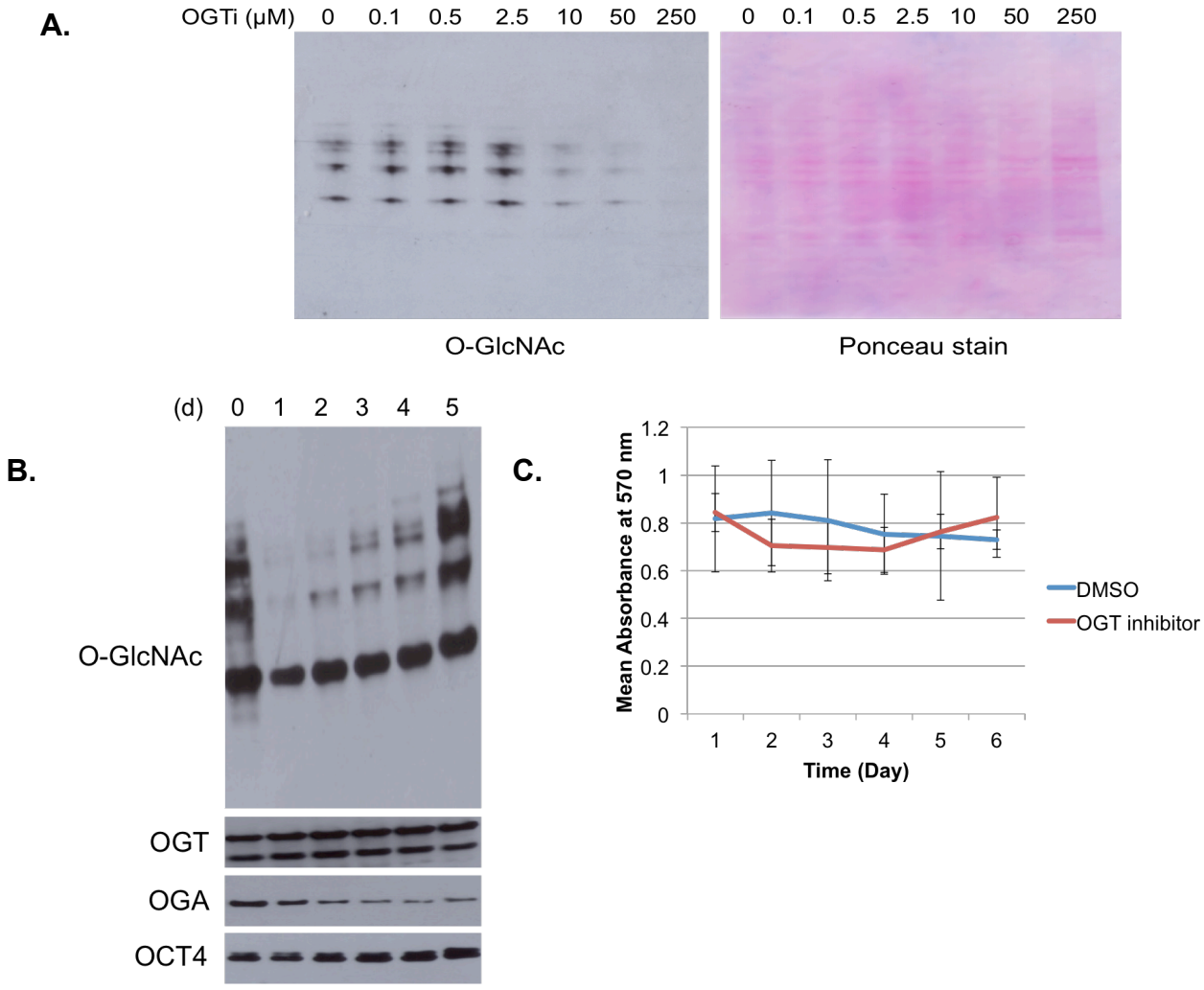


Figure 4.1. Ac-5SGlcNAc lowers cellular O-GlcNAc. (A) Dose-dependent decrease in protein O-GlcNAcylation after treating hESCs with OGT inhibitor (OGTi) for 24 h, as measured by western blot. Ponceau stain of western blot membrane showed equal loading. (B) Treatment of hESCs with 50 μM OGTi for 5 days caused an increase in O-GlcNAc levels, coinciding with a decrease in OGA expression (d = day). (C) OGT inhibition for 6 days with 50 μM Ac-5SGlcNAc did not affect hESC proliferation. DMSO (vehicle) treated cells were used as control (mean ± stdev; n = 2).

OGT inhibition accelerates neural differentiation of hESCs

We next evaluated the effect of treating neural induced hESCs with Ac-5SGlcNAc. Because 50 μM Ac-5SGlcNAc notably decreased O-GlcNAcylation while not being overly toxic to the cells, we treated hESCs with 50 μM OGT inhibitor during neural induction. As described in chapter 3, we used the dual-SMAD inhibition neural induction protocol developed by Chambers et al. [49] to differentiate hESCs into NSCs in 11 days (Fig. 4.2A). This protocol relies on the synergistic action of SB431542 (SB), which inhibits the Lefty/Activin/TGFβ pathways, and LDN-193189 (LDN), an inhibitor of bone morphogenic protein (BMP). Previously we measured the levels of O-GlcNAc during hESC neural induction by western blot and observed a fluctuation in cellular O-GlcNAc (Figure 3.3). Inhibition of OGT during neural induction of hESCs decreased O-GlcNAcylation, as seen in Figure 4.1A, while maintaining the oscillatory expression of O-GlcNAc during differentiation (Fig. 4.2B). Consistent with previous reports [48, 50, 51], treatment with OGT inhibitor induced a compensatory decline in OGA and an increase in OGT. Furthermore, we also observed a decrease in the levels of UDP-GlcNAc after differentiating cells in the presence of the OGT inhibitor, which was previously reported by Gloster et al [48], while maintaining the same oscillatory pattern (Fig. 4.2C). Addition of Ac-5SGlcNAc to cells results in the formation UDP-5SGlcNAc, a nucleotide sugar analogue of UDP-GlcNAc. Thus, these results confirm that the OGT inhibitor can proficiently be metabolized by the cells and inhibit OGT, as the decrease in O-GlcNAc levels correlates with a decrease in UDP-GlcNAc. It is possible that this is a mechanism for the cell to compensate for the accumulation of UDP-5SGlcNAc, given that it is a poor substrate for OGT [48]. Also, the presence of the oscillatory pattern in the levels of both O-GlcNAc and UDP-GlcNAc during neural induction, even after OGT inhibition, is an indication of a possible regulatory mechanism utilized by hESCs to properly differentiate into NSCs.

We next asked whether treatment of hESCs with the OGT inhibitor during neural induction caused changes in the expression of NSC and neuronal markers. Several lines of evidence suggested the possibility that inhibition of OGT would have a positive regulatory effect during early neuronal differentiation. These lines of evidence included: 1) neuronal apoptosis occurred after neuron-specific deletion of OGT, suggesting that O-GlcNAcylation may regulate neural development [42]; 2) O-GlcNAc levels declined as cells further differentiated to NSCs, as shown by Figure 4.2A, and as reported by Liu et al [52]; 3) Liu et al also reported high neuronal staining of O-GlcNAc, OGT, and OGA over non-neuronal cells in rat brain, also suggesting direct involvement of O-GlcNAc in regulating neuronal differentiation [52]; 4) maintaining high levels of O-GlcNAcylation by inhibiting OGA prevented ectodermal differentiation of human pluripotent stem cells and mouse ESCs [4, 46] and impaired axonal branching [45]; 5) low glucose, and possibly low O-GlcNAcylation, during differentiation facilitated NSC differentiation into neurons [53]. Inhibition of OGT during neural conversion of hESCs induced a higher expression of neuronal markers TUJ1, MAP2, and NeuN in comparison to cells induced in the presence of DMSO (control) (Fig. 4.3A). In contrast, when hESCs were differentiated in the presence of the OGA inhibitor Thiamet-G, which causes an increase in cellular O-GlcNAcylation [54], expression of neuronal markers TUJ1 and MAP2 was much lower than in control cells. This result corresponds to previous reports described above in item

4. Enhanced neural induction of OGT-inhibited cells is specific to neuronal formation, as immunofluorescence staining with astrocyte (GFAP) and oligodendrocyte (Olig1) markers was negative (Fig. 4.3A). We also tested the effects of the OGT inhibitor during neural induction of two different human embryonic stem cell lines, H7 and H9, and also saw formation of a premature neuronal cell type with higher expression of TUJ1 and MAP2 than in non-inhibited cells (Fig. 4.3B,I-K). A decrease in proliferation was also observed in OGT-inhibited NSCs compared to cells treated with DMSO (control) (Fig. 4.3C), indicating that OGT-inhibited cells are losing their "stem cell" proliferative characteristics and acquiring more mature cell fates.

We then optimized NSC induction by varying the duration and starting point of OGT inhibition. Cells were monitored for the loss of neural progenitor marker NESTIN and acquisition of neuronal marker β-III tubulin (TUJ1). We found that inhibition of OGT was optimal 48 hours before initiating neural induction with SMAD inhibitors LDN and SB, which corresponds to day -2 (Fig. 4.2A, 4.3D). We next wanted to identify the shortest duration of OGT inhibition necessary for premature expression of neuronal marker TUJ1 and loss of neural progenitor marker NESTIN, but found that inhibition during the entire differentiation protocol resulted in the best balance of expression of both markers (Fig. 4.3E). We also quantified the co-expression of neuronal markers TUJ1 and MAP2 to compare the efficiency of neuronal differentiation of hESCs under different conditions (Fig. 4.3F). These included treating cells either from day -2 to day 11 or from day 6 to day 11 with one of the following: OGT inhibitor, OGA inhibitor (Thiamet-G), transforming growth factor β (TGFβ), OGT inhibitor with no SB, and DMSO (vehicle). LDN and SB were added on day 0 for all conditions, except where noted. Although the results confirmed that neural induction was most efficient in the presence of OGT inhibitor from day -2 to day 11, surprisingly OGT inhibitor and LDN, without SB, where sufficient to convert hESCs to NSCs and induce expression of TUJ1 and MAP2. This result correlates with improved neural induction when cells were treated with TGFβ, as shown in Figure 4.3F, suggesting that inhibition of TGFβ with SB is perhaps not as vital for the successful neural induction of hESCs as previously thought. As observed previously, inhibition of OGA during neural induction did not enhance expression of neuronal markers. Thus, all further neural inductions were performed in the presence of OGT inhibitor from day -2 to day 11, and LDN and SB from day 0 to 5.

We next sought to characterize lineage progression of hESCs and gain insight into the potential mechanism that could contribute to the neuronal enhancement of OGT-inhibited cells. Expression analysis of pluripotent markers OCT4 and SOX2 revealed similar expression pattern for both markers in control and OGT-inhibited cells (Fig. 4.3G,H). Loss of OCT4 expression was observed by day 5, whereas SOX2 expression was maintained at later days of differentiation. These results are not surprising since SOX2 is also necessary for the maintenance of neural progenitor cells [55]. However, this data suggests that the improved neuronal induction produced by OGT-inhibition is not due to a premature exit from the pluripotency and neural progenitor state controlled by OCT4 and SOX2 expression, respectively. Finally, temporal analysis of TUJ1 and MAP2 co-expression in three hESC lines (H1, H7 and H9) revealed that OGT inhibition induced a rapid increase in expression of these markers as early as day 6 of neural differentiation, and peaking at day 8 with an approximate 4-fold increase above DMSO-treated cells (Fig. 4.3I-K). There is also a 2-

fold increase in the expression of these markers from day 6 to day 8 of induction in OGT-inhibited cells. Overall, these results demonstrate the reproducibility of OGT inhibitor treatment during neural differentiation of hESCs.

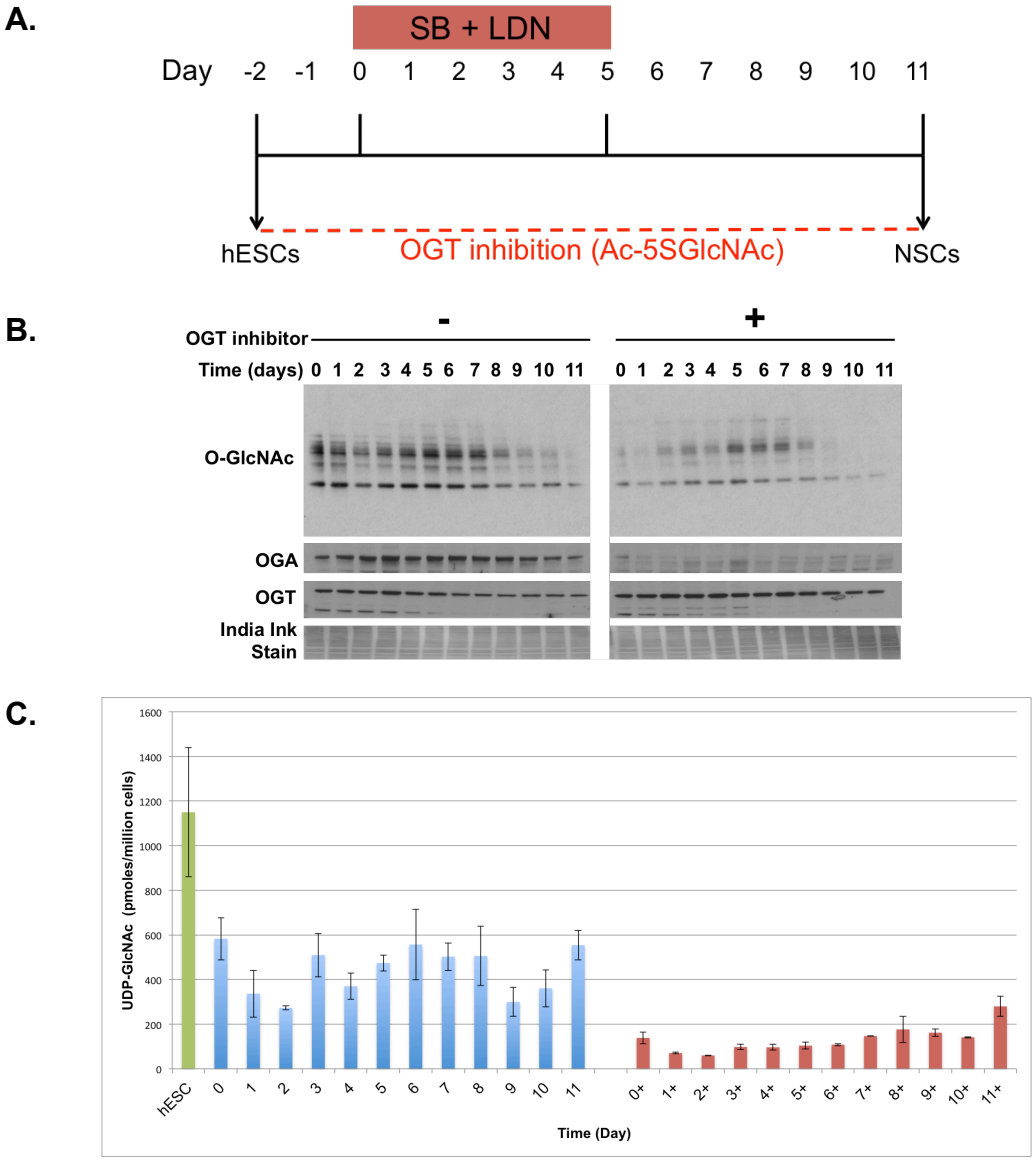
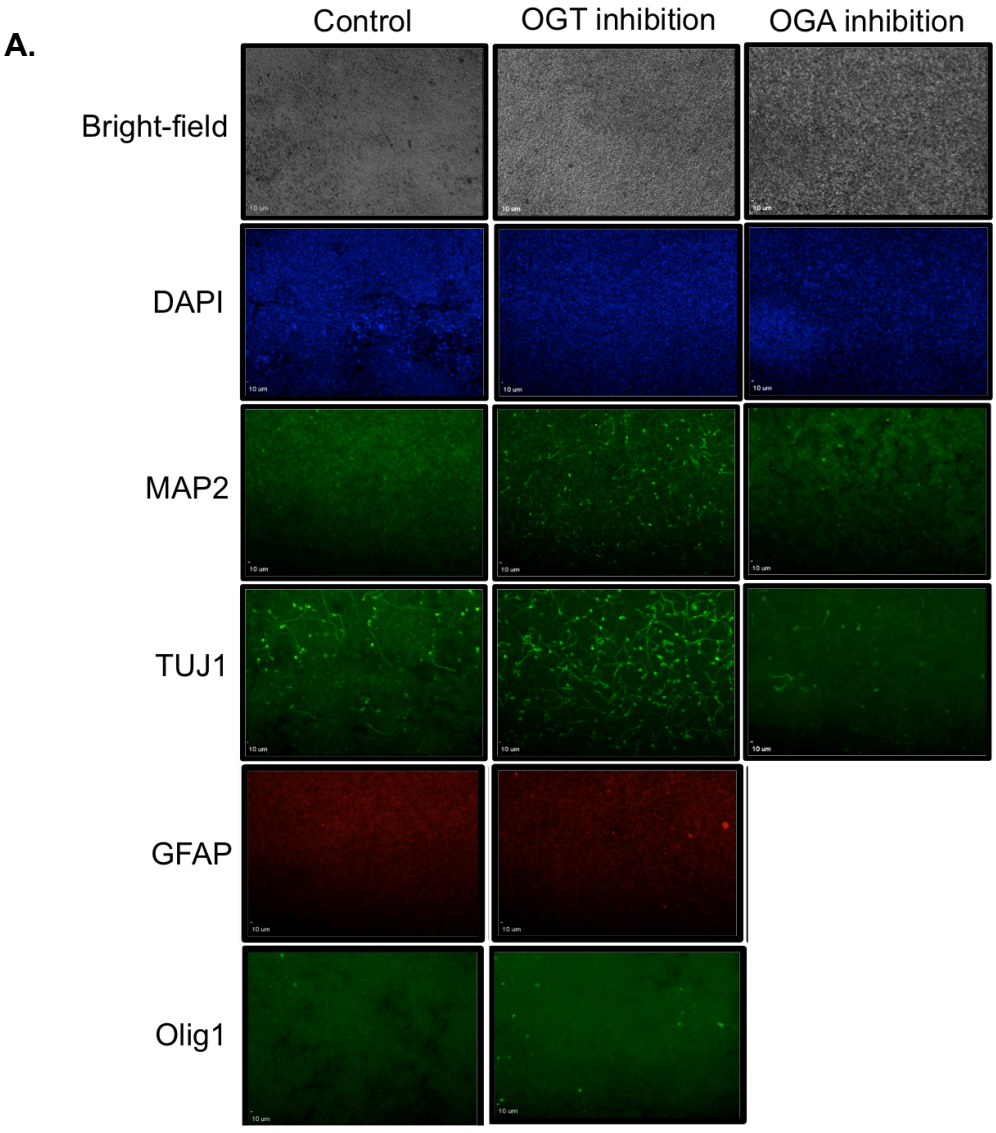
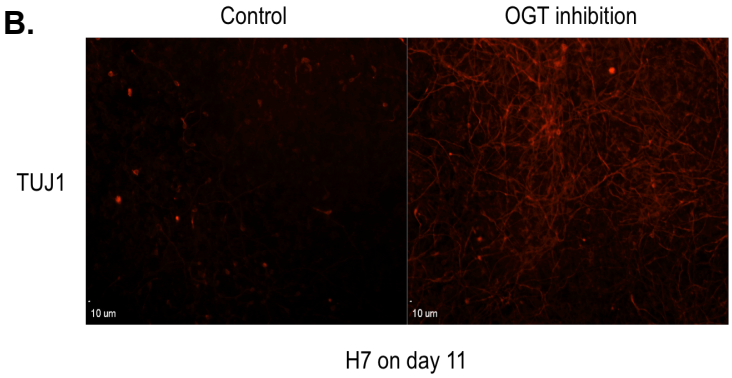
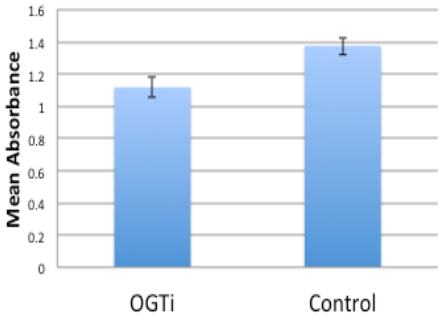


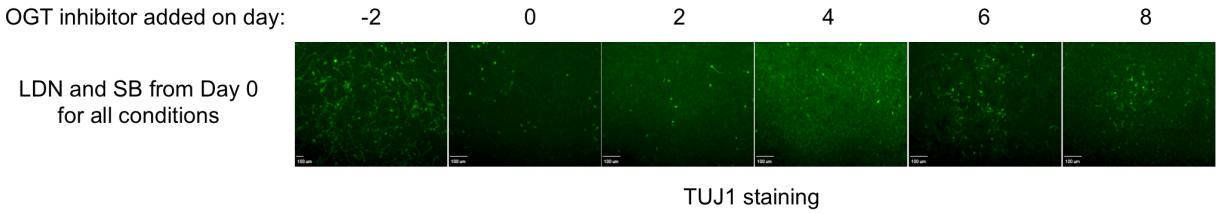
Figure 4.2. OGT inhibitor perturbs O-GlcNAc and UDP-GlcNAc levels in hESCs undergoing neural differentiation. (A) Schematic of the neural induction protocol and duration of OGT inhibition. Cells were grown to 90% confluency on days -2 and -1 before starting induction on day 0. (B) Western blot analysis of hESCs differentiated in the presence of OGT inhibitor (50 μM) showed a decrease in O-GlcNAc levels and in OGA expression while enhancing OGT; India ink staining showed equal loading. (C) Analysis of UDP-GlcNAc levels of hESCs treated with vehicle (blue bars, as in Figure 3.4) or 50 μM OGT inhibitor (+; red bars) during neural differentiation. hESC (green bar) was used as a reference. (Data: mean ± S.E.M; blue n = 4; red n = 2; green n = 3)


A.
Control
OGT inhibition
OGA inhibition
Bright-field
DAPI
MAP2
TUJ1
GFAP
Olig1
10 um

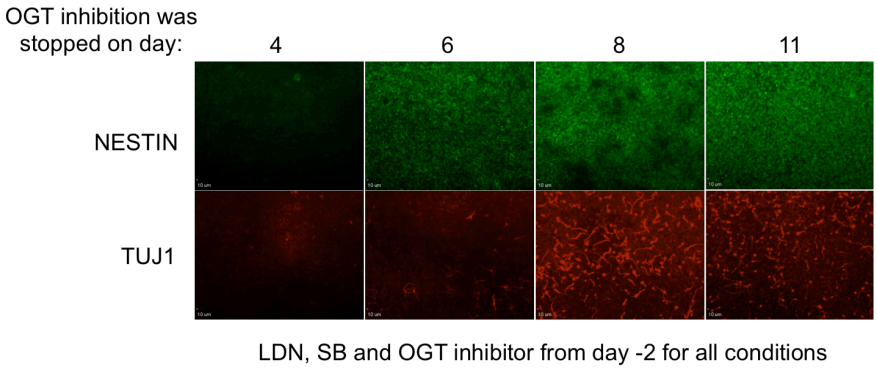

B.
Control
OGT inhibition
TUJ1
10 um
H7 on day 11


C.
Mean Absorbance
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
OGTi
Control

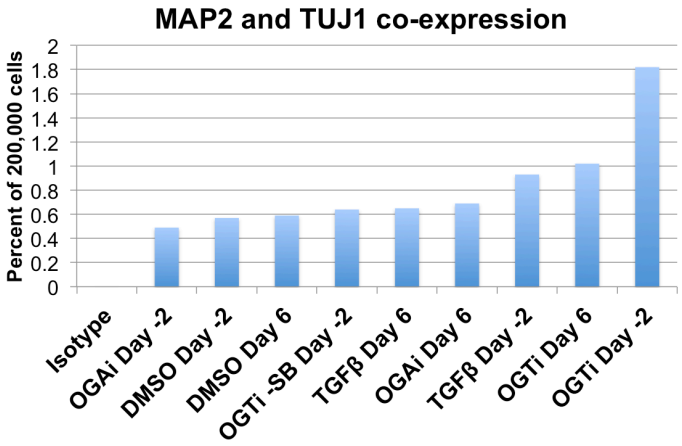
D.


OGT inhibitor added on day:
-2
0
2
4
6
8
LDN and SB from Day 0 for all conditions
TUJ1 staining

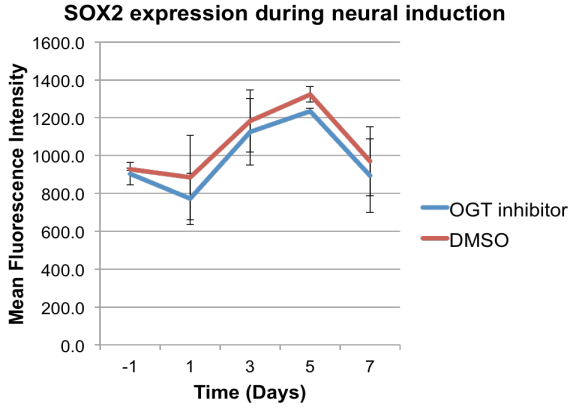
E.


OGT inhibition was stopped on day:
4
6
8
11
NESTIN
TUJ1
LDN, SB and OGT inhibitor from day -2 for all conditions

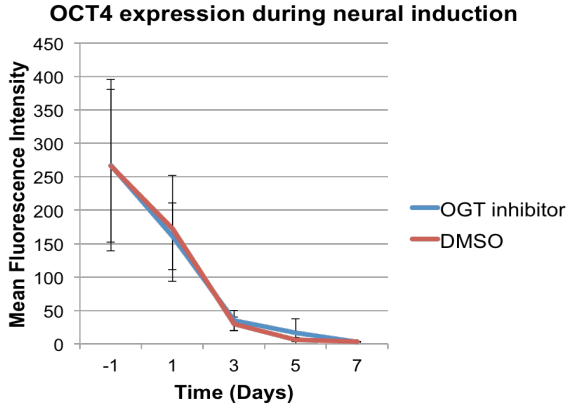
F.


MAP2 and TUJ1 co-expression
Percent of 200,000 cells
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Isotype
OGAi Day -2
DMSO Day -2
DMSO Day 6
OGTi -SB Day -2
TGFβ Day 6
OGAi Day 6
TGFβ Day -2
OGTi Day 6
OGTi Day -2

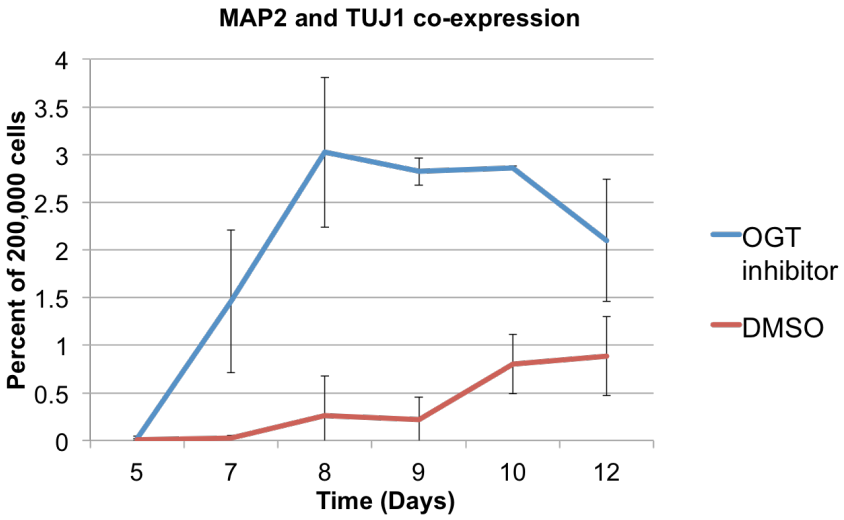
G.


SOX2 expression during neural induction
Mean Fluorescence Intensity
1600.0
1400.0
1200.0
1000.0
800.0
600.0
400.0
200.0
0.0
-1
1
3
5
7
Time (Days)
OGT inhibitor
DMSO

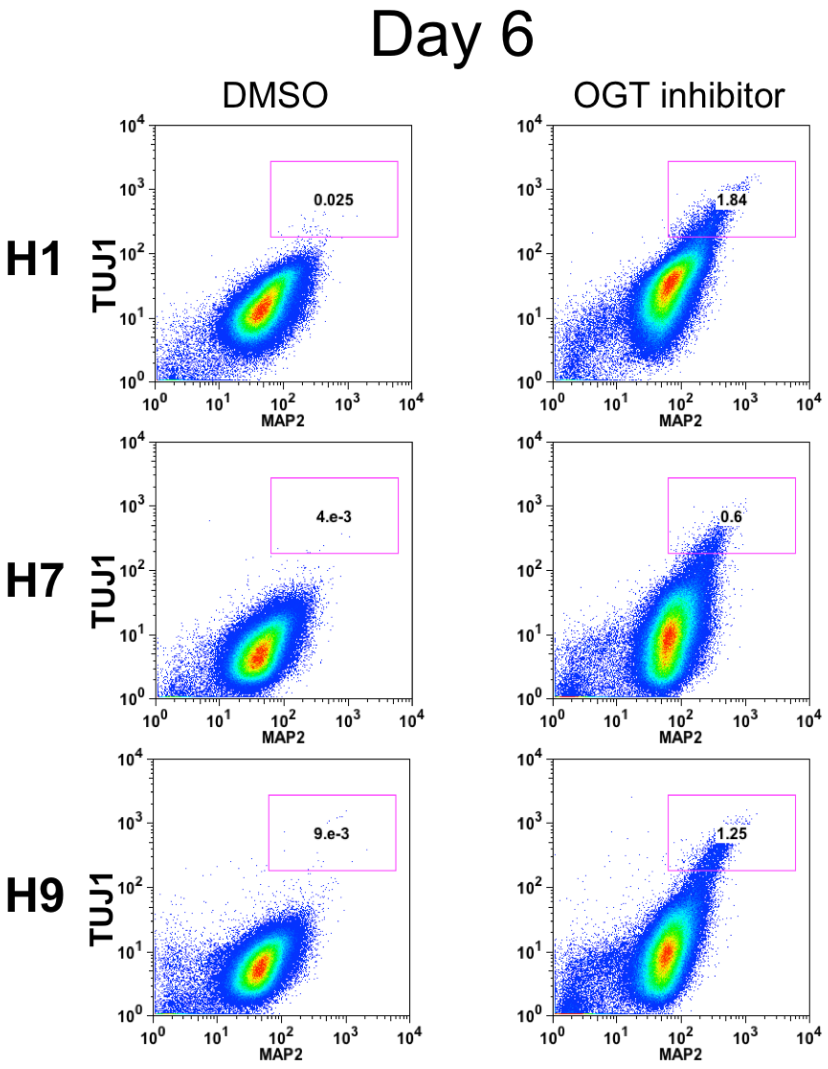
H.


OCT4 expression during neural induction
Mean Fluorescence Intensity
450
400
350
300
250
200
150
100
50
0
-1
1
3
5
7
Time (Days)
OGT inhibitor
DMSO

I.


MAP2 and TUJ1 co-expression
Percent of 200,000 cells
4
3.5
3
2.5
2
1.5
1
0.5
0
5
7
8
9
10
12
Time (Days)
OGT inhibitor
DMSO

J.


Day 6
DMSO
OGT inhibitor
H1
H7
H9
TUJ1
MAP2
0.025
1.84
4.e-3
0.6
9.e-3
1.25

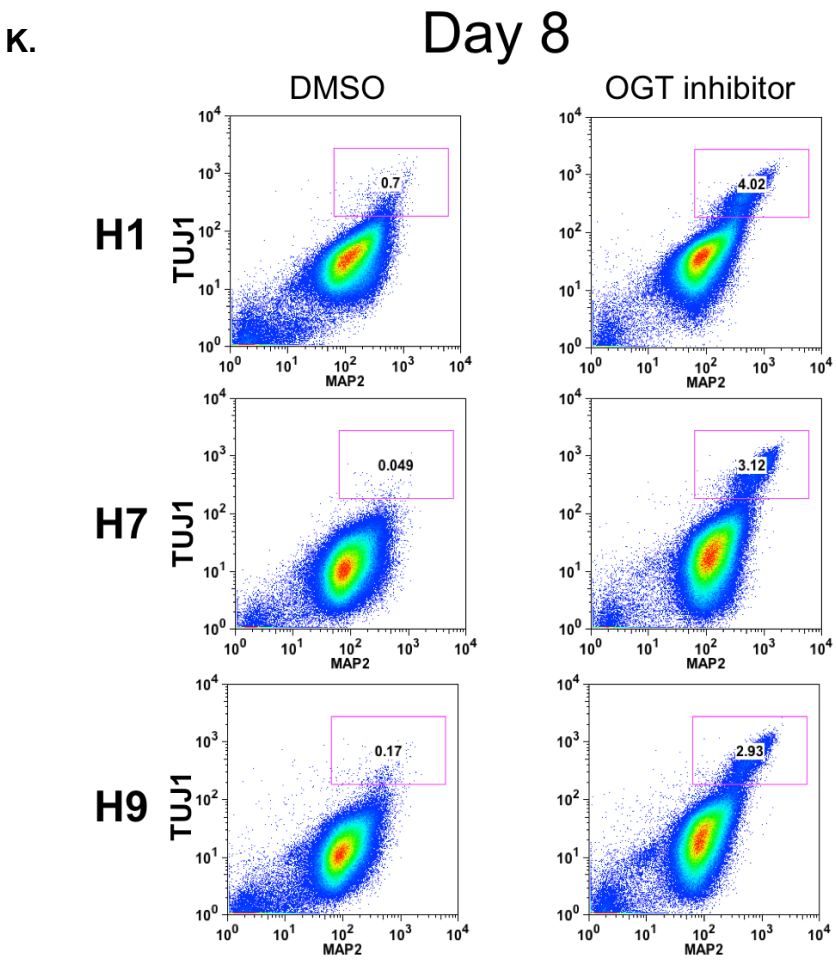


Figure 4.3. Ac-5SGlcNAc-treated hESCs prematurely acquire a neuronal phenotype within 11 days. (A) Staining for TUJ1 and MAP2 showed many more positive cells when hESCs were induced in the presence of OGT inhibitor (50 μM) for 11 days, compared with non-inhibited cells. OGA inhibition with Thiamet-G (100 nM) for 11 days caused a reduction in TUJ1 expression. (B) Similar TUJ1 staining was observed when H7 hESCs were differentiated in the presence of OGT inhibitor for 11 days. (C) Proliferation assay performed on day 8 of NSC induction revealed that OGT-inhibited cells have a decreased ability to proliferate, which for NSC is indicative of an early fate determination (n = 3). (D) Day -2 until (E) day 11 were determined to be the optimal days for addition of OGT inhibitor (Ac-5SGlcNAc; 50 μM) during neural induction. Cells were examined for the loss of NESTIN and acquisition of TUJ1 by immunofluorescence. Scale bars: 100 μm (D); 10 μm (E). (F) Higher TUJ1 and MAP2 expression was detected in cells treated with OGT inhibitor from day -2 to day 11 by flow cytometry (OGAi - OGA inhibitor; OGTi - OGT inhibitor). Isotype control was used to subtract background fluorescence. (G) SOX2 and (H) OCT4 expression was similar between OGT-inhibited and DMSO-treated cells (mean ± stdev; n = 3). (I-K) Flow cytometry analysis revealed that co-expression of TUJ1 and MAP2 was accelerated, and maximal expression (3% by day 8) occurred earlier in OGT-inhibited cells than in DMSO-treated cells for all cell lines tested (H1, H7 and H9).

Gene expression profile suggests regulation of O-GlcNAc through the activation of TGFβ signaling[2]

As we observed the greatest fold difference in TUJ1 expression on day 8 of neural induction, we performed global gene expression analysis of OGT- and DMSO-treated hESCs to characterize changes in transcript expression on day 8 of differentiation. The expression of 670 genes was significantly altered due to OGT inhibition of hESCs on day 8 of differentiation (p-value < 0.0023) (Fig. 4.4A) (for the complete list of genes that showed significant changes in expression see Appendix III). Of these, 168 genes decreased in expression, while 502 genes increased. Consistent with the induction of neural development in OGT-inhibited cells, markers for neuroectoderm and neurons showed an increase in expression (Fig. 4.4B). For instance, TUBB3, NCAM1, SOX3, SALL3, NTRK2, L1CAM, HES1, NEDD9, POU3F4, POU3F2, SOX1, and NEUROD4 were all up-regulated in OGT-inhibited cells. The up-regulation of the TUBB3 gene, which encodes for β-III tubulin, in OGT-inhibited cells confirmed our previous results (Fig. 4.3). Conversely, we observed down-regulation of FOXG1, ID1, ID2, FGF8, and IGF2. FGF8 is involved in the expression of neural crest cells [56], and FOXG1, ID1, ID2, and IGF2 are involved in repressing neuronal differentiation [57-59]. The dual-SMAD inhibition neural induction protocol can also promote the formation of neural crest cells, which are the cells that form most of the peripheral nervous system (PNS) and other non-neural cell types. Initial cell density on day 0 determines the ratio of neuronal (central nervous system; CNS) and neural crest (PNS) progeny obtained by day 11. We influenced the formation of neuronal progeny (CNS) in our protocol by initiating the induction with a high-density of hESCs. Thus, FGF8 down-regulation confirmed the presence of a primarily neuronal population in our day 8 NSCs. The down-regulation of FOXG1, ID1, ID2, and IGF2, and the observation that only 3% of the cells on day 8 were expressing TUJ1, suggests that the OGT-inhibited NSCs are present in a mixed population of cells that are either in a progenitor proliferative state or that have already differentiated to neurons. We also observed both the up-regulation (HES1) and down-regulation (PTEN) of genes involved in the proliferation of neural progenitor cells, validating this hypothesis. It would be necessary to examine the gene expression profile of NSCs at a later stage of neural induction to examine the developmental progression of this cell population. Finally, the gene that encodes for the O-GlcNAc enzyme OGA, MGEA5, was also down-regulated in OGT-inhibited cells, which correlates with the result obtained by western blot analysis, shown in Fig. 4.2B.

We next analyzed the Gene Ontology (GO) of significantly expressed genes, and observed that genes involved in neuron differentiation, neuron development, neuron projection morphogenesis, and the regulation of neurological system processes were significantly overrepresented among up-regulated genes in OGT inhibitor-treated cells compared to DMSO-treated cells (Fig. 4.4C). We also performed pathway analysis of all the differentially expressed genes using the Ingenuity Pathway Analysis (IPA) software. Analysis revealed that the transforming growth factor β1 (TGFβ1) was the top upstream regulator (Fig. 4.4D). Ingenuity's Upstream Regulator Analysis in IPA is a tool that

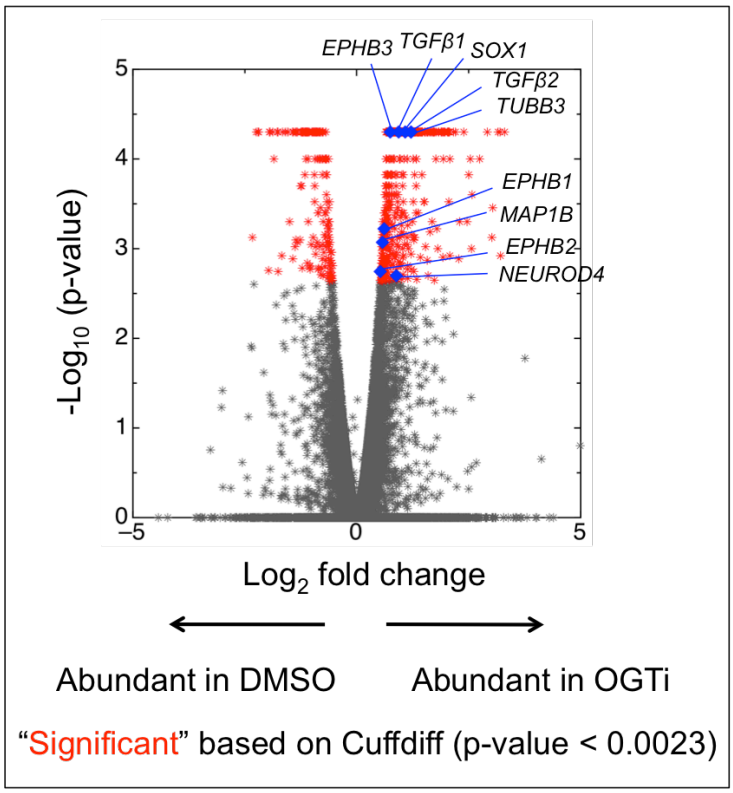
[2] The results described in this section were obtained in collaboration with Teppei Yamaguchi from the Tjian Group at UC Berkeley.

predicts upstream regulators from gene expression data based on literature findings compiled in the Ingenuity® Knowledge Base. A Fisher's Exact Test p-value is calculated to assess the significance of enrichment of the gene expression data for the genes downstream of an upstream regulator [60]. Thus, this indicates that TGFβ1, also up-regulated in OGT-inhibited cells, was the top effector of the genes down-regulated and up-regulated due to OGT-inhibition. TGFβ2 was also up-regulated in the OGT-inhibited cells.

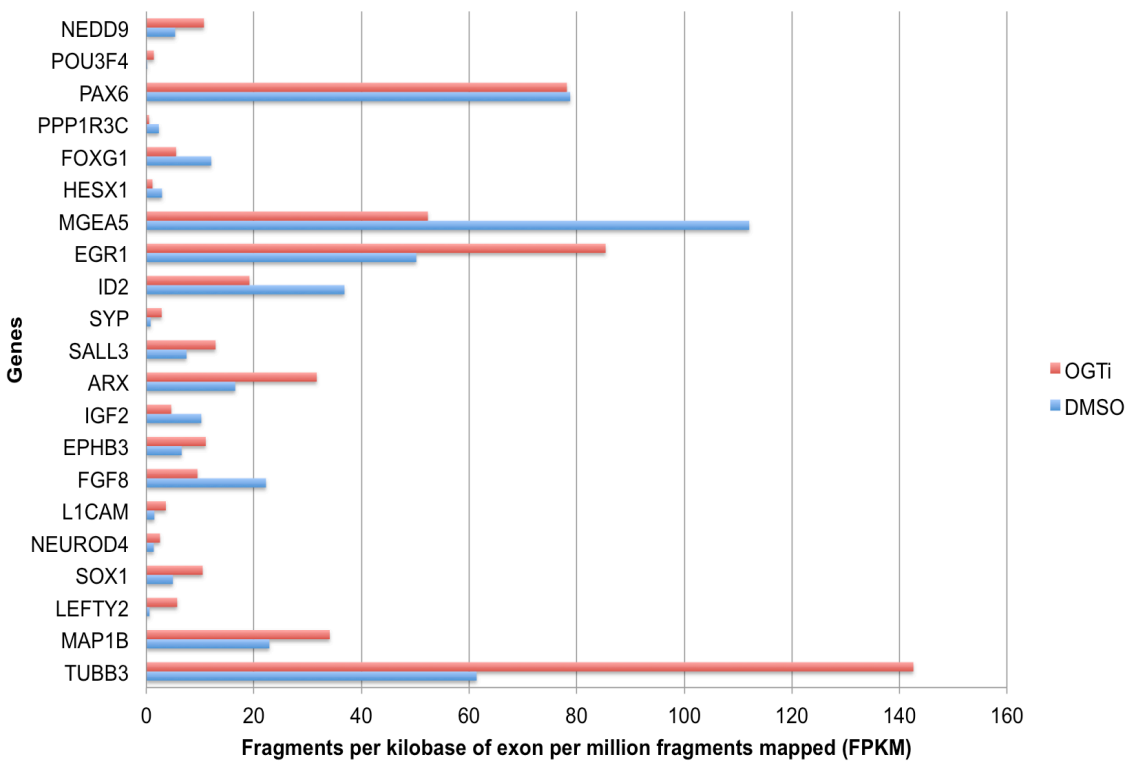
Both TGFβ1 and TGFβ2 are expressed in the brain [61]. Previous gene expression profiling studies have suggested a potential role for the TGFβ signaling pathway in the priming and differentiation of human NSCs [62]. In their gene expression analysis, Cai et al. observed a decrease in the expression of TGFβ signaling pathway suppressors in NSCs that were primed to become neurons. The authors also mention the importance of priming NSCs, by the addition of a cocktail of heparin, laminin, and basic fibroblast growth factor (bFGF) to preferentially differentiate NSCs to neurons. It is possible that the addition of the OGT inhibitor during neural induction is inducing a priming effect on the NSCs, perhaps through the activation of the TGFβ signaling pathway, and consequently, the expression of neuronal genes. NEDD9, up-regulated in the OGT-inhibited cells, was recently identified as an essential downstream component of TGFβ-mediated neuronal differentiation [63]. This study suggested that NEDD9 might promote a progenitor status that primes NSCs to neuronal differentiation.

NEDD9 is heavily phosphorylated and it is involved in coordinating tyrosine-kinase-based signaling with cell adhesion. In the gene expression analysis of OGT-inhibited cells, we observed a high number of genes involved in protein kinase activity (29 genes), and an even higher number of genes that were denominated as encoding for a phosphorylated protein (235 genes). Of these, FGFR2, FGFR3, EPHA3, EPHB1, EPHB2, EPHB3, TEK, EGFR, NTRK2, and ERBB4 were classified as tyrosine-protein kinases. In addition, many of the up-regulated genes were classified as part of the transmembrane receptor protein tyrosine kinase signaling pathway. In a recent study, Angiopoietin-1 induced neurite outgrowth and enhanced neuronal differentiation through the activation of Akt and of the protein encoded by TEK, angiopoietin-1 receptor [64]. Akt is a serine/threonine protein kinase, and a known substrate of OGT [65]. O-GlcNAcylation at Thr-305 and Thr-312 inhibited Akt phosphorylation at Thr-308 by disrupting the interaction between Akt and PDK1, impairing Akt activation [66]. It is possible that OGT inhibition activated Akt signaling, and consequently induced neuronal differentiation, via cross-talk between O-GlcNAcylation and phosphorylation of Akt. Furthermore, TGFβ was shown to induce expression of β-III tubulin (TUJ1) in cultured mouse retinal pigment epithelial cells via tyrosine phosphorylation and activation of the extracellular-signal-regulated kinase (ERK) signaling pathway [67]. ERKs are a subfamily of mitogen-activated protein (MAP) kinases, and are known to play important roles in brain development [68]. Several MAP kinases were up-regulated in OGT-inhibited cells, including MAPK12 and MAPK15. Overall, these results suggest that, by inhibiting OGT and maintaining low levels of O-GlcNAc in cells, we are possibly promoting the phosphorylation and activation of molecules downstream of TGFβ, such as NEDD9 and Akt, and the transcription of neuronal genes. The molecular mechanisms involved in this process still need to be elucidated.

A.


EPHB3
TGFβ1
SOX1
TGFβ2
TUBB3
EPHB1
MAP1B
EPHB2
NEUROD4
-Log10 (p-value)
Log2 fold change
Abundant in DMSO
Abundant in OGTi
"Significant" based on Cuffdiff (p-value < 0.0023)

B.


NEDD9
POU3F4
PAX6
PPP1R3C
FOXG1
HESX1
MGEA5
EGR1
ID2
SYP
SALL3
ARX
IGF2
EPHB3
FGF8
L1CAM
NEUROD4
SOX1
LEFTY2
MAP1B
TUBB3
Genes
0
20
40
60
80
100
120
140
160
Fragments per kilobase of exon per million fragments mapped (FPKM)
OGTi
DMSO

C.

OGT inhibitor

GO Description	P-value
Cell motion	5.49E-13
Neuron differentiation	5.49E-13
Neuron development	5.64E-13
Cell morphogenesis	9.89E-11
Regulation of system process	2.84E-09
Cell adhesion	2.84E-09
Biological adhesion	7.13E-09
Neuron projection morphogenesis	7.29E-09
Regulation of neurological system process	5.18E-08
Neuron projection development	5.18E-08

DMSO

GO Description	P-value
Regulation of DNA metabolic process	5.49E-05
Axon guidance	3.12E-04
Extracellular region part	3.51E-04
Cell projection morphogenesis	0.001220848
Neuron projection development	0.001613313
Regulation of DNA replication	0.001952319
Neuron projection morphogenesis	0.002380412

D.

Top Upstream Regulators

Upstream Regulator	p-value of overlap	Predicted Activation State
TGFB1	2.71E-26	Activated
PDGF BB	1.80E-17	Activated
beta-estradiol	3.48E-17	Activated

Figure 4.4. Gene expression analysis of day 8 of neural differentiation. Gene expression analysis was done on day 8 for both OGT inhibitor- and DMSO-treated cells. (A) Volcano plot showing the distribution of differentially expressed genes on day 8. (B) Abundance of select developmental genes identified by RNA-sequencing from hESCs differentiated for 8 days in the presence of OGT inhibitor (OGTi; 50 uM) or DMSO (vehicle). (C) The enriched Gene Ontology (GO) categories of genes up-regulated (left) and down-regulated (right) on day 8 of induction. Categories for genes up-regulated are consistent with enhanced neural induction of OGT-inhibited cells. (D) Activated upstream regulator genes identified by analyzing the RNA-sequencing data presented above using the Ingenuity Pathway Analysis software (IPA).

4.3. Conclusion

Our understanding of how O-GlcNAcylation regulates cellular functions is steadily increasing with the development of new tools to study O-GlcNAc cycling. Through the use of chemical inhibitors, we were able to show that the inhibition of OGT, in combination with dual SMAD-inhibition, results in an enhanced neural differentiation of hESCs in less than 11 days. The increase in expression of neuronal markers, TUJ1 and MAP2, due to OGT inhibition suggests a regulatory role of O-GlcNAc in modulating neuronal differentiation. It is possible that O-GlcNAc exerts this role through the activation of TGFβ signaling. Studying the molecular mechanisms involved in this activation, and consequently, in the induction of neuronal differentiation, should provide insights into the basic process of CNS development as well as a better understanding of the role of O-GlcNAc in neurodegenerative disorders.

4.4. Experimental Methods

Human embryonic stem cell (hESC) culture

Human embryonic stem cell lines H1, H7 and H9 (WA01, WA07, WA09; WiCell) were cultured on 10 cm plates coated with hESC-qualified basement membrane matrix Matrigel (BD Biosciences). Cells were grown under standard conditions (37 °C, 5% CO2) in TeSR2 media (STEMCELL Technologies Inc). Media was exchanged daily after the first 48 hours in culture, and cells were passaged every 3 to 5 days using 1 mg/ml of dispase (Gibco) in Knockout DMEM media (Invitrogen) and mechanically removed. Karyotype analysis was routinely performed and indicated that all samples were diploid and had no chromosomal abnormalities.

Neural Induction

Human embryonic stem cell lines H1, H7 and H9 were differentiated into neural progenitors that have the capacity to become neurons and glia. The differentiation protocol was performed as reported by [49]. Briefly, cells were dissociated into single cells using Accutase (STEMCELL Technologies Inc.) and plated on Matrigel (BD Biosciences)-treated dishes at a density of 40,000–50,000 cells/cm^2 in the presence of MEF-conditioned hESC medium containing 10 ng/ml FGF-2 (R&D Systems) and 10 μM ROCK inhibitor (Y-27632; Tocris) for single cell survival. MEF-conditioned hESC medium was obtained by culturing mytomycin-C-treated mouse embryonic fibroblasts (MEF; Millipore PMEF-CF) in DMEM/F12, 20% (v/v) Knockout Serum Replacement, 1 mM GlutaMAX, 100 μM MEM nonessential amino acids, 0.1 mM β-mercaptoethanol (Invitrogen), and 6 ng/ml FGF-2. After 24 hours of cell growth, media was harvested and sterile filtered. The ROCK inhibitor was withdrawn the day after plating, and hESCs were allowed to expand for 3-4 days or until they were 80-90% confluent. Neural differentiation was initiated using KSR medium, which contained Knockout DMEM (Invitrogen), 15% (v/v) Knockout Serum Replacement, 1 mM GlutaMAX, 100 μM MEM nonessential amino acids and 0.1 mM β-mercaptoethanol. To inhibit SMAD signaling, 100 nM LDN-193189 (Stemgent) or 200 ng/ml Noggin (3344-NG-050, R&D Systems) and 10 μM SB431542 (Tocris) were added on days 0–5 of induction. Cells were fed daily, and N2 medium was added in increasing 25% increments every other day starting on day 4 and leading to 100% N2 on day 10. N2 media consisted of DMEM/ F12 powder, 1:1 (Gibco/Invitrogen) resuspended in distilled water, glucose, sodium bicarbonate, putrescine, progesterone, sodium selenite, insulin (Sigma), and apo-transferrin (Kamada Ltd).

Chemical inhibitors

Synthesis of the OGT inhibitor (Ac-5SGlcNAc) was carried out as described previously [48]. Thiamet-G, an OGA inhibitor, was purchased from Cayman Chemical. hESCs undergoing neural differentiation were treated with OGT inhibitor starting two days before the beginning of the neural induction. The inhibitor was added to the media every day, until the end of the neural induction protocol. The best times for OGT inhibition

were determined by observing the expression of neural markers, β-III tubulin (TUJ1) and MAP2, and of neuroectoderm marker PAX6, by immunofluorescence, between OGT-inhibited and DMSO (control)-treated cells. To determine the effects of OGA inhibition during neural induction, cells were treated at different times during induction with 100 nM Thiamet-G. DMSO (Sigma) was used as a control for all of the inhibitions.

Western Blot analysis

Whole cell protein extracts were lysed with a buffer consisting of 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% Triton X-100, and 2 mM EDTA (including protease and phosphatase inhibitors (Roche) and 1 μM Thiamet-G). Samples were sonicated and cleared by centrifugation (12000 x g, 20 min at 4 °C). Protein concentration was determined by bicinchoninic acid assay (Pierce). Samples were separated by SDS-PAGE (4-12% Bio-Rad), transferred to nitrocellulose membranes, and blocked with 5% milk (LabScientific, inc.) or BSA (Sigma) in PBS with 0.1% Tween 20 (Sigma). Membranes were incubated with primary antibodies overnight at 4 °C. The following primary antibodies were used to detect the indicated proteins: anti-OGT (1:1000, DM-17, Sigma and sc-32921, Santa Cruz Biotech), anti-OCT4 (1:2500, MAB4401, Millipore), anti-O-GlcNAc (1:1000, RL2, Thermo Scientific) and anti-OGA (1:1000, SAB4200267, Sigma). The following horseradish peroxidase-linked secondary antibodies were used for detection at a dilution ratio of 1:5000: goat anti-rabbit-HRP-conjugated secondary antibody (4010-05, Southern Biotech) and goat anti-mouse-HRP-conjugated secondary antibody (1010-05, Southern Biotech).

Immunofluorescence

Cells were fixed with 4% (v/v) paraformaldehyde (Thermo Scientific) for 40 minutes, rinsed 3 times with 1X PBS for 5 minutes, and blocked and permeabilized for at least one hour with 5% (v/v) FBS plus 0.3% (v/v) Triton X-100 in PBS. Cells were stained overnight at 4 °C with the following primary antibodies: TUJ1 (Millipore, MAB1637, 1:1,000), MAP2 (Millipore, MAB3418, 1:500), Nestin (Millipore, MAB353, 1:500), GFAP (Millipore, MAB360, 1:250) and Olig1 (Millipore, MAB344, 1:500). Cells were rinsed three times with 1X PBS, and then were incubated with Alexa Fluor 488-conjugated goat-anti-mouse secondary antibody and Alexa Fluor 594-conjugated goat-anti-rabbit secondary antibody (1:400, Invitrogen) for 2 hours at room temperature. After, cells were rinsed three times with 1X PBS, and DAPI (Invitrogen) was added for 5 minutes before analysis using an Olympus IX71 fluorescent microscope. Control cells were stained with mouse and rabbit IgG isotypes (Millipore, PP54, PP64). All image acquisition and processing was performed under identical conditions for test and control samples.

Flow cytometry

Colonies were dissociated into single cell suspensions with Accutase and rinsed in 1X PBS with 0.5% (w/v) BSA twice. Cells were counted using a hemocytometer and roughly 1 x 10^6 cells per sample were used for analysis. All cells were fixed with 4%

(v/v) paraformaldehyde for 40 minutes at 4 °C, washed twice with PBS, permeabilized using 0.1% (v/v) Triton X-100 in PBS, and blocked using 0.5% (w/v) BSA in PBS. Cells were stained with conjugated antibodies for 20 minutes in the dark on ice. Primary conjugated antibodies used for flow cytometry were the following: MAP2-PE (Millipore, FCMAB318PE, 1:20), TUJ1 Alexa Fluor 488 (BD Pharmingen, 560381, 1:5), MAP2B Alexa Fluor 647 (BD Pharmingen, 560382, 1:5), OCT3/4 Alexa Fluor 647 (BD Pharmingen, 560329, 1:5) and SOX2 Alexa Fluor 488 (BD Pharmingen, 561593, 1:20). Flow cytometry (BD FACs Calibur from BD Biosciences) was performed and data was analyzed using FlowJo software (TreeStar Inc). At least two independent assays were carried out. Unstained cells from each sample type and cells stained with isotype control were used as controls.

High Performance Anion Exchange Chromatography (HPAEC)

Cells were induced to neural progenitors as indicated before. During harvest, cells were treated with Accutase to dissociate colonies into single cells and counted with a hemocytometer. Approximately 3 x 10^6 cells were harvested per sample. Cells were spun and washed twice with cold 1X PBS, immediately placed in dry ice, and kept at -80 °C until lysis. Cells were lysed with 80% “super-cold” methanol (on dry ice) [69]. Lysate was spun at 2,000 x g for 15 min. The supernatant was dried by speed vacuum for 4-5 hours. The intracellular metabolite pellet was either used immediately, or it was stored at -80 °C for future use.

Metabolite pellet was resuspended in 40 mM sodium phosphate buffer (pH 7.4 40 µL per million cells), and filtered through an Amicon® Ultra centrifugal filter unit (Millipore, 10,000 MWCO). Filtrates were analyzed by HPAEC (ICS-3000 system, Dionex) with CarboPac™PA1 (Dionex) with a pulsed amperometry detector (PAD) and UV-detector in-line [70, 71]. Typically, 20 µL of metabolite was injected into the sample loading loop before the sample enters a guard column (Dionex, 4 × 50 mm) and then an analytical column (Dionex, 4 × 250 mm). The eluents used were 1.0 mM NaOH (C) and 1.0 M NaOAc and 1.0 mM NaOH (D). Low-carbonate NaOH (50% in water) was obtained from Fisher Scientific (SS254-1) and NaOAc was from Sigma (71183). HPAEC was run with a flow rate = 1 mL/min and the following gradient elution was performed: T_0 min = 95% C, T_5 = 85% C, T_{15} = 70% C, T_{20} = 60% C, T_{45} = 60% C, T_{50} = 0% C, T_{60} = 0% E2, T_{65} = 95%, T_{75} = 95%. UDP-GlcNAc standards (50 µM, 25 µM, 10 µM, and 2.5 µM) were injected at the same time as cellular samples. UDP-GlcNAc peak areas were input into excel and raw data were converted to pmoles of UDP-GlcNAc by comparing to a standard curve that was generated by analyzing the peak areas of the UDP-GlcNAc standards. Data were normalized to cell number.

RNA-sequencing and data analysis

Total RNA was isolated from cells using TRIzol Reagent (Invitrogen) and RNeasy Mini kit (Qiagen). RNA quality was assessed using the Agilent 2100 Bioanalyzer. One microgram of each sample was used to isolate mRNA and prepare the library following the protocol of the Illumina TruSeq RNA sample prep kit v2 (RS-122-2001). Prepared

libraries were sequenced on an Illumina HiSeq 2500 sequencer at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. Sequenced reads (27 million) were mapped to the human RefSeq (hg19) using TopHat [72] with default parameters. Differentially expressed genes were identified using Cuffdiff [73]. Biological duplicates for each condition were subjected to the analysis and the q-value cutoff 0.05 was applied to select significant differences in gene expression for scatter plots and for GO term analysis using DAVID Bioinformatics Resources 6.7 [74, 75]. Pathway analysis was performed using Ingenuity Pathway Analysis (Qiagen).

Cell proliferation

Human embryonic stem cells were cultured and differentiated as described above. Cells were dissociated with Accutase, and grown in media plus 50 μM OGT inhibitor or DMSO as a control. Media was changed every day. The effect of the OGT inhibitor on cell proliferation was measured using a calorimetric (MTT) kit for cell survival and proliferation (CT02, Millipore) following the manufacturers' instructions.

4.5. References

1. Jang, H., et al., *O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network.* Cell Stem Cell, 2012. **11**(1): p. 62-74.

2. Yang, Y.R., et al., *O-GlcNAcase is essential for embryonic development and maintenance of genomic stability.* Aging Cell, 2012. **11**(3): p. 439-48.

3. Kim, H.S., et al., *Excessive O-GlcNAcylation of proteins suppresses spontaneous cardiogenesis in ES cells.* FEBS Lett, 2009. **583**(15): p. 2474-8.

4. Maury, J.J., et al., *Excess of O-linked N-acetylglucosamine modifies human pluripotent stem cell differentiation.* Stem Cell Res, 2013. **11**(2): p. 926-37.

5. Hart, G.W., M.P. Housley, and C. Slawson, *Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins.* Nature, 2007. **446**(7139): p. 1017-22.

6. Yuzwa, S.A. and D.J. Vocadlo, *O-GlcNAc and neurodegeneration: biochemical mechanisms and potential roles in Alzheimer's disease and beyond.* Chem Soc Rev, 2014.

7. Capotosti, F., et al., *O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1.* Cell, 2011. **144**(3): p. 376-88.

8. Lazarus, M.B., et al., *HCF-1 is cleaved in the active site of O-GlcNAc transferase.* Science, 2013. **342**(6163): p. 1235-9.

9. Daou, S., et al., *Crosstalk between O-GlcNAcylation and proteolytic cleavage regulates the host cell factor-1 maturation pathway.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 2747-52.

10. Gambetta, M.C., K. Oktaba, and J. Muller, *Essential role of the glycosyltransferase sxc/Ogt in polycomb repression.* Science, 2009. **325**(5936): p. 93-6.

11. Sinclair, D.A., et al., *Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc).* Proc Natl Acad Sci U S A, 2009. **106**(32): p. 13427-32.

12. Myers, S.A., B. Panning, and A.L. Burlingame, *Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells.* Proc Natl Acad Sci U S A, 2011. **108**(23): p. 9490-5.

13. Webster, D.M., et al., *O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development.* BMC Dev Biol, 2009. **9**: p. 28.

14. Khidekel, N., et al., *Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(36): p. 13132-13137.

15. Shi, F.T., et al., *Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells.* J Biol Chem, 2013. **288**(29): p. 20776-84.

16. Balasubramani, A. and A. Rao, *O-GlcNAcylation and 5-methylcytosine oxidation: an unexpected association between OGT and TETs.* Mol Cell, 2013. **49**(4): p. 618-9.

17. Deplus, R., et al., *TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS.* Embo j, 2013. **32**(5): p. 645-55.

18. Vella, P., et al., *Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells.* Mol Cell, 2013. **49**(4): p. 645-56.

19. Chen, Q., et al., *TET2 promotes histone O-GlcNAcylation during gene transcription.* Nature, 2013. **493**(7433): p. 561-4.

20. Zhang, Q., et al., *Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT).* J Biol Chem, 2014. **289**(9): p. 5986-96.

21. Ito, R., et al., *TET3-OGT interaction increases the stability and the presence of OGT in chromatin.* Genes Cells, 2014. **19**(1): p. 52-65.

22. Love, D.C., et al., *Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7413-8.

23. Hart, G.W., et al., *Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease.* Annu Rev Biochem, 2011. **80**: p. 825-58.

24. Bond, M.R. and J.A. Hanover, *O-GlcNAc cycling: a link between metabolism and chronic disease.* Annu Rev Nutr, 2013. **33**: p. 205-29.

25. Forsythe, M.E., et al., *Caenorhabditis elegans ortholog of a diabetes susceptibility locus: oga-1 (O-GlcNAcase) knockout impacts O-GlcNAc cycling, metabolism, and dauer.* Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11952-7.

26. Hanover, J.A., et al., *A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11266-71.

27. Ferrer, C.M., et al., *O-GlcNAcylation Regulates Cancer Metabolism and Survival Stress Signaling via Regulation of the HIF-1 Pathway.* Mol Cell, 2014. **54**(5): p. 820-31.

28. Lynch, T.P. and M.J. Reginato, *O-GlcNAc transferase: a sweet new cancer target.* Cell Cycle, 2011. **10**(11): p. 1712-3.

29. Alexander, G.E., et al., *Longitudinal PET Evaluation of Cerebral Metabolic Decline in Dementia: A Potential Outcome Measure in Alzheimer's Disease Treatment Studies.* Am J Psychiatry, 2002. **159**(5): p. 738-45.

30. Liu, F., et al., *Reduced O-GlcNAcylation links lower brain glucose metabolism and tau pathology in Alzheimer's disease.* Brain, 2009. **132**(Pt 7): p. 1820-32.

31. Kreppel, L.K., M.A. Blomberg, and G.W. Hart, *Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats.* J Biol Chem, 1997. **272**(14): p. 9308-15.

32. Gao, Y., et al., *Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain.* J Biol Chem, 2001. **276**(13): p. 9838-45.

33. Akimoto, Y., et al., *Localization of the O-GlcNAc transferase and O-GlcNAc-modified proteins in rat cerebellar cortex.* Brain Res, 2003. **966**(2): p. 194-205.

34. Okuyama, R. and S. Marshall, *UDP-N-acetylglucosaminyl transferase (OGT) in brain tissue: temperature sensitivity and subcellular distribution of cytosolic and nuclear enzyme.* Journal of Neurochemistry, 2003. **86**(5): p. 1271-1280.

35. Cole, R.N. and G.W. Hart, *Cytosolic O-glycosylation is abundant in nerve terminals.* J Neurochem, 2001. **79**(5): p. 1080-9.

36. Rexach, J.E., P.M. Clark, and L.C. Hsieh-Wilson, *Chemical approaches to understanding O-GlcNAc glycosylation in the brain.* Nat Chem Biol, 2008. **4**(2): p. 97-106.

37. Lamarre-Vincent, N. and L.C. Hsieh-Wilson, *Dynamic Glycosylation of the Transcription Factor CREB: A Potential Role in Gene Regulation.* Journal of the American Chemical Society, 2003. **125**(22): p. 6612-6613.

38. Rexach, J.E., et al., *Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation.* Nat Chem Biol, 2012. **8**(3): p. 253-61.

39. Griffith, L.S., M. Mathes, and B. Schmitz, *β-Amyloid precursor protein is modified with O-linked N-acetylglucosamine.* Journal of Neuroscience Research, 1995. **41**(2): p. 270-278.

40. Liu, F., et al., *O-GlcNAcylation regulates phosphorylation of tau: A mechanism involved in Alzheimer's disease.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(29): p. 10804-10809.

41. Vosseller, K., et al., *O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry.* Mol Cell Proteomics, 2006. **5**(5): p. 923-34.

42. O'Donnell, N., et al., *Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability.* Mol Cell Biol, 2004. **24**(4): p. 1680-90.

43. Liu, K., et al., *Accumulation of protein O-GlcNAc modification inhibits proteasomes in the brain and coincides with neuronal apoptosis in brain areas with high O-GlcNAc metabolism.* Journal of Neurochemistry, 2004. **89**(4): p. 1044-1055.

44. Yanagisawa, M. and R.K. Yu, *O-linked beta-N-acetylglucosaminylation in mouse embryonic neural precursor cells.* J Neurosci Res, 2009. **87**(16): p. 3535-45.

45. Francisco, H., et al., *O-GLcNAc post-translational modifications regulate the entry of neurons into an axon branching program.* Dev Neurobiol, 2009. **69**(2-3): p. 162-73.

46. Speakman, C.M., et al., *Elevated O-GlcNAc levels activate epigenetically repressed genes and delay mouse ES cell differentiation without affecting naive to primed cell transition.* Stem Cells, 2014.

47. Ostrowski, A. and D.M. van Aalten, *Chemical tools to probe cellular O-GlcNAc signalling.* Biochem J, 2013. **456**(1): p. 1-12.

48. Gloster, T.M., et al., *Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells.* Nat Chem Biol, 2011. **7**(3): p. 174-81.

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

50. Slawson, C., et al., *Perturbations in O-linked beta-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis.* J Biol Chem, 2005. **280**(38): p. 32944-56.

51. Palaniappan, K.K., et al., *A chemical glycoproteomics platform reveals O-GlcNAcylation of mitochondrial voltage-dependent anion channel 2.* Cell Rep, 2013. **5**(2): p. 546-52.

52. Liu, Y., et al., *Developmental regulation of protein O-GlcNAcylation, O-GlcNAc transferase, and O-GlcNAcase in mammalian brain.* PLoS One, 2012. **7**(8): p. e43724.

53. Horie, N., et al., *Lowered glucose suppressed the proliferation and increased the differentiation of murine neural stem cells in vitro.* FEBS Lett, 2004. **571**(1-3): p. 237-42.

54. Yuzwa, S.A., et al., *A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo.* Nat Chem Biol, 2008. **4**(8): p. 483-90.

55. Graham, V., et al., *SOX2 Functions to Maintain Neural Progenitor Identity.* Neuron, 2003. **39**(5): p. 749-765.

56. Creuzet, S., et al., *Reciprocal relationships between Fgf8 and neural crest cells in facial and forebrain development.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(14): p. 4843-4847.

57. Regad, T., et al., *The neural progenitor-specifying activity of FoxG1 is antagonistically regulated by CKI and FGF.* Nat Cell Biol, 2007. **9**(5): p. 531-540.

58. Bracko, O., et al., *Gene Expression Profiling of Neural Stem Cells and Their Neuronal Progeny Reveals IGF2 as a Regulator of Adult Hippocampal Neurogenesis.* The Journal of Neuroscience, 2012. **32**(10): p. 3376-3387.

59. Sun, X.H., et al., *Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins.* Mol Cell Biol, 1991. **11**(11): p. 5603-11.

60. Kramer, A., et al., *Causal analysis approaches in Ingenuity Pathway Analysis.* Bioinformatics, 2014. **30**(4): p. 523-30.

61. Gomes, F.C., O. Sousa Vde, and L. Romao, *Emerging roles for TGF-beta1 in nervous system development.* Int J Dev Neurosci, 2005. **23**(5): p. 413-24.

62. Cai, Y., et al., *Gene expression profiling and analysis of signaling pathways involved in priming and differentiation of human neural stem cells.* Neuroscience, 2006. **138**(1): p. 133-148.

63. Vogel, T., et al., *Transforming Growth Factor β Promotes Neuronal Cell Fate of Mouse Cortical and Hippocampal Progenitors In Vitro and In Vivo: Identification of Nedd9 as an Essential Signaling Component.* Cerebral Cortex, 2010. **20**(3): p. 661-671.

64. Bai, Y., et al., *Ectopic expression of angiopoietin-1 promotes neuronal differentiation in neural progenitor cells through the Akt pathway.* Biochemical and Biophysical Research Communications, 2009. **378**(2): p. 296-301.

65. Vosseller, K., et al., *Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes.* Proceedings of the National Academy of Sciences, 2002. **99**(8): p. 5313-5318.

66. Wang, S., et al., *Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling.* PLoS One, 2012. **7**(5): p. e37427.

67. Chung, E.J., et al., *TGF-beta-stimulated aberrant expression of class III beta-tubulin via the ERK signaling pathway in cultured retinal pigment epithelial cells.* Biochem Biophys Res Commun, 2011. **415**(2): p. 367-72.

68. Samuels, I.S., S.C. Saitta, and G.E. Landreth, *MAP'ing CNS Development and Cognition: An ERKsome Process.* Neuron, 2009. **61**(2): p. 160-167.

69. Dettmer, K., et al., *Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols.* Anal Bioanal Chem, 2011. **399**(3): p. 1127-39.

70. Tomiya, N., et al., *Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: sugar nucleotide contents in cultured insect cells and mammalian cells.* Anal Biochem, 2001. **293**(1): p. 129-37.

71. Yu, S.H., et al., *Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners.* Proc Natl Acad Sci U S A, 2012. **109**(13): p. 4834-9.

72. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-1111.

73. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nat Biotech, 2010. **28**(5): p. 511-515.

74. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Res, 2009. **37**(1): p. 1-13.

75. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.

Appendices

Appendix I. Complete list of proteins identified in MS analysis of hESCs treated with or without $Ac_4GalNAz$ (chapter 2)

Accession	Gene	$Ac_4GalNAz$ sequence coverage	Control sequence coverage
IPI00009104.7	RUVBL2	0%	5.00%
IPI00945749.1	POR	0%	2.90%
IPI00024580.4	MCCC1	0%	3.70%
IPI00942257.1	KRT6A	0%	11.30%
IPI00100160.3	CAND1	0%	2.00%
IPI00789750.1	KRT14	0%	21.90%
IPI00790342.1	RPL6	0%	5.90%
IPI00027834.3	HNRNPL	0%	6.60%
IPI00215719.6	RPL18	0%	10.60%
IPI00217466.3	HIST1H1D	0%	10.40%
IPI00922181.1	MCM2	0%	2.50%
IPI00054042.1	GTF2I	0%	2.30%
IPI00022774.3	VCP	0%	5.20%
IPI00022434.4	ALB	0%	4.90%
IPI00300567.1	DCI	0%	9.30%
IPI00304692.1	RBMX	0%	9.00%
IPI00023086.3	MRPL15	0%	11.10%
IPI00473136.3	CTNNA1	0%	2.50%
IPI00855856.2	API5	0%	4.60%
IPI00297084.7	DDOST	5.70%	0%
IPI00216691.5	PFN1	20.00%	0%
IPI00009904.1	PDIA4	5.30%	3.90%
IPI00917964.1	MPP6	3.40%	0%
IPI00794461.1	HIST1H2BN	14.50%	0%
IPI00219729.3	SLC25A11	9.60%	0%
IPI00011274.3	HNRPDL	5.70%	7.60%
IPI00021405.3	LMNA	2.70%	0%
IPI00018349.5	MCM4	2.50%	0%
IPI00302674.3	CCDC134	13.50%	0%
IPI00144003.3	SALL2	3.00%	0%
IPI00009771.6	LMNB2	2.90%	3.40%
IPI00395887.4	TMX1	7.50%	8.20%
IPI00025815.2	TARDBP	5.00%	0%
IPI00164018.5	CYP2S1	4.30%	0%
IPI00007032.1	SALL4	2.30%	0%
IPI00550906.6	CSTF2T	4.40%	0%
IPI00006663.1	ALDH2	6.00%	0%
IPI00171903.2	HNRNPM	3.30%	0%

IPI00013214.2	MCM3	2.50%	4.30%
IPI00384280.5	PCYOX1	5.70%	0%
IPI00329650.1	NUP35	9.50%	0%
IPI00419373.1	HNRNPA3	6.90%	2.60%
IPI00015973.1	EPB41L2	2.40%	0%
IPI00013122.1	MIR1181	10.10%	0%
IPI00409717.1	MIR1248	5.60%	0%
IPI00418313.3	ILF3	3.50%	5.50%
IPI00022793.5	HADHB	3.80%	4.60%
IPI00306516.1	TIMM44	5.30%	0%
IPI00003527.5	SLC9A3R1	7.00%	0%
IPI00171438.2	TXNDC5	4.20%	0%
IPI00011107.2	IDH2	4.20%	0%
IPI00302927.6	CCT4	5.40%	0%
IPI00031517.1	MCM6	3.20%	4.40%
IPI00005158.1	LONP1	2.40%	0%
IPI00784224.1	ZFR	3.10%	0%
IPI00215998.5	CD63	4.20%	0%
IPI00011253.3	RPS3	10.70%	23.90%
IPI00102864.3	HK2	2.90%	0%
IPI00027713.1	CREB1	6.70%	0%
IPI00017292.1	CTNNB1	5.50%	0%
IPI00399307.2	PRCP	5.00%	0%
IPI00014236.6	SLC39A14	6.50%	0%
IPI00032563.3	ZNF281	4.40%	0%
IPI00027146.1	GLUD2	4.10%	4.10%
IPI00216230.3	TMPO	4.80%	0%
IPI00017451.1	SF3A1	4.40%	0%
IPI00944951.1	TAF6	3.70%	0%
IPI00003802.2	MAN2A1	3.40%	0%
IPI00604664.5	NDUFS1	5.90%	0%
IPI00218319.3	TPM3	9.70%	0%
IPI00004962.2	GOLIM4	3.20%	0%
IPI00940373.2	REPS1	5.00%	0%
IPI00003704.4	RBM14	9.30%	0%
IPI00217468.3	HIST1H1B	10.20%	10.20%
IPI00844578.1	DHX9	2.70%	3.60%
IPI00871851.2	SEPT2	6.60%	0%
IPI00384541.4	RPRD2	2.70%	0%
IPI00215918.3	ARF4	11.70%	0%
IPI00018350.3	MCM5	4.90%	3.40%

IPI00294879.1	RANGAP1	4.40%	0%
IPI00017726.1	HSD17B10	13.40%	0%
IPI00140420.4	SND1	4.40%	0%
IPI00028954.1	MCM3AP	1.20%	0%
IPI00479786.5	KHSRP	6.80%	0%
IPI00883857.2	HNRNPU	3.30%	3.30%
IPI00023542.6	TMED9	16.20%	0%
IPI00419916.4	ALPL	8.40%	0%
IPI00848226.1	GNB2L1	11.00%	6.90%
IPI00900293.1	FLNB	1.10%	0%
IPI00031030.1	APLP2	3.40%	0%
IPI00002220.4	SAP130	3.10%	0%
IPI00018146.1	YWHAQ	9.80%	0%
IPI00008530.1	RPLP0	16.70%	15.80%
IPI00797038.1	PCK2	4.10%	0%
IPI00514424.4	PPT1	9.90%	0%
IPI00334587.1	HNRNPAB	9.90%	0%
IPI00926491.1	PTCD1	3.20%	0%
IPI00903213.1	IGF2R	1.00%	0%
IPI00025239.2	NDUFS2	6.70%	0%
IPI00010346.1	NLN	6.80%	3.30%
IPI00011229.1	CTSD	11.90%	0%
IPI00418497.1	TIMM50	5.90%	0%
IPI00010697.2	ITGA6	3.00%	0%
IPI00017303.1	MSH2	3.50%	4.60%
IPI00013698.3	ASAH1	4.20%	0%
IPI00027252.6	PHB2	10.40%	9.70%
IPI00012011.6	CFL1	27.10%	0%
IPI00921118.1	ACTN1	4.70%	0%
IPI00022608.2	SORL1	2.30%	0%
IPI00008982.1	ALDH18A1	6.50%	4.00%
IPI00186290.6	EEF2	2.30%	2.30%
IPI00219217.3	LDHB	12.00%	7.50%
IPI00305383.1	UQCRC2	9.90%	0%
IPI00011416.2	ECH1	9.50%	0%
IPI00293845.5	RIF1	1.30%	0%
IPI00028635.5	RPN2	5.50%	0%
IPI00549248.4	NPM1	11.90%	0%
IPI00217563.4	ITGB1	3.80%	0%
IPI00640598.1	ZMYND8	4.70%	0%
IPI00216308.5	VDAC1	10.60%	8.50%

IPI00940000.1	DDX17	3.30%	3.30%
IPI00644712.4	XRCC6	4.80%	3.40%
IPI00031169.1	RAB2A	20.30%	0%
IPI00304596.3	NONO	10.00%	6.60%
IPI00910194.1	NUP153	3.70%	0%
IPI00418169.3	ANXA2	14.60%	7.80%
IPI00449049.5	PARP1	3.70%	4.80%
IPI00009703.1	SOX2	6.30%	0%
IPI00011084.2	CLDN6	17.30%	10.00%
IPI00005614.6	SPTBN1	2.90%	1.30%
IPI00966900.1	MFGE8	5.70%	0%
IPI00296337.2	PRKDC	1.40%	1.00%
IPI00018500.1	ARID3A	6.40%	0%
IPI00012540.1	PROM1	8.60%	0%
IPI00297492.2	STT3A	4.50%	3.00%
IPI00444945.2	BAX	11.90%	0%
IPI00220687.1	HM13	5.20%	0%
IPI00789551.1	MATR3	6.60%	0%
IPI00922751.1	SLC2A14	5.80%	3.00%
IPI00396131.4	YTHDF3	7.10%	0%
IPI00903226.1	HK1	6.90%	2.30%
IPI00000816.1	YWHAE	23.90%	0%
IPI00019912.3	HSD17B4	3.70%	3.70%
IPI00021263.3	YWHAZ	11.40%	0%
IPI00410034.2	SLC38A2	8.10%	0%
IPI00000874.1	PRDX1	10.10%	0%
IPI00028888.1	HNRNPD	11.50%	6.20%
IPI00418991.5	QSER1	2.70%	0%
IPI00292975.4	RBM27	4.00%	0%
IPI00022334.1	OAT	13.00%	0%
IPI00397904.6	NUP93	5.10%	0%
IPI00946792.1	PTK7	3.90%	0%
IPI00744889.2	CDH1	3.80%	0%
IPI00031522.2	HADHA	10.00%	3.10%
IPI00926935.1	GNAI2	12.60%	0%
IPI00216047.3	SMARCC2	4.50%	0%
IPI00013808.1	ACTN4	4.70%	4.20%
IPI00386119.4	SF1	3.80%	0%
IPI00873829.1	ACAA2	8.30%	0%
IPI00014053.3	TOMM40	14.70%	0%
IPI00001639.2	KPNB1	3.40%	0%

IPI00018140.3	SYNCRIP	9.80%	8.00%
IPI00171411.4	GOLM1	17.10%	0%
IPI00303300.3	FKBP10	9.60%	0%
IPI00930710.1	SLC25A13	8.30%	0%
IPI00217357.2	CCAR1	2.10%	0%
IPI00220834.8	XRCC5	3.00%	0%
IPI00396485.3	EEF1A1	5.60%	7.60%
IPI00025796.3	NDUFS3	16.30%	14.00%
IPI00478128.2	GATAD2A	4.60%	0%
IPI00893541.1	PDIA3	19.50%	19.50%
IPI00646304.4	PPIB	12.00%	12.00%
IPI00438229.2	TRIM28	5.90%	4.10%
IPI00009960.6	IMMT	9.90%	3.70%
IPI00939304.1	IPO5	4.60%	0%
IPI00382470.3	HSP90AA1	6.60%	0%
IPI00941649.1	HNRNPR	9.60%	8.50%
IPI00003881.5	HNRNPF	8.20%	0%
IPI00000877.1	HYOU1	10.70%	0%
IPI00219078.5	ATP2A2	2.70%	4.20%
IPI00010740.1	SFPQ	6.20%	7.10%
IPI00334190.4	STOML2	28.40%	0%
IPI00879810.1	SPTAN1	3.40%	2.10%
IPI00103467.4	ALDH1B1	13.00%	0%
IPI00301936.4	ELAVL1	9.90%	0%
IPI00604773.2	PODXL	12.00%	0%
IPI00400826.1	CLU	8.00%	0%
IPI00299571.5	PDIA6	8.50%	0%
IPI00216457.7	HIST2H2AA4	21.50%	0%
IPI00026942.5	ERLIN2	11.50%	0%
IPI00748411.2	SHMT2	11.10%	11.10%
IPI00219171.3	POU2F1	9.70%	0%
IPI00005719.1	RAB1A	11.70%	0%
IPI00005198.2	ILF2	11.30%	14.10%
IPI00025366.4	CS	8.20%	8.20%
IPI00003925.6	PDHB	16.70%	0%
IPI00220739.3	PGRMC1	11.30%	0%
IPI00166555.3	RBM26	5.30%	0%
IPI00328243.2	PLD3	8.40%	0%
IPI00300371.5	SF3B3	3.70%	3.90%
IPI00646361.3	NUP214	2.10%	0%
IPI00794221.1	DBN1	9.40%	0%

IPI00017510.3	MT-CO2	14.50%	7.50%
IPI00171127.1	UBAP2	5.10%	0%
IPI00965550.1	PRRC1	7.80%	0%
IPI00025874.2	RPN1	7.90%	0%
IPI00465248.5	ENO1	14.50%	8.50%
IPI00796337.1	PCBP2	14.20%	6.60%
IPI00019472.4	SLC1A5	8.10%	0%
IPI00871890.1	PICALM	13.20%	0%
IPI00796333.1	ALDOA	16.70%	15.80%
IPI00295698.6	SLC7A3	8.90%	0%
IPI00218019.1	BSG	8.60%	0%
IPI00014230.1	C1QBP	20.90%	0%
IPI00294779.1	VDAC3	14.10%	0%
IPI00025252.1	PDIA3	14.70%	7.10%
IPI00028931.2	DSG2	10.50%	0%
IPI00414676.6	HSP90AB1	9.70%	8.00%
IPI00022462.2	TFRC	8.80%	0%
IPI00016610.2	PCBP1	14.60%	0%
IPI00337397.1	NUP98	3.10%	0%
IPI00016342.1	RAB7A	28.00%	0%
IPI00396378.3	HNRNPA2B1	8.20%	10.50%
IPI00909140.7	TUBB	13.80%	11.60%
IPI00604620.3	NCL	11.30%	11.50%
IPI00453473.6	HIST1H4F	17.50%	29.10%
IPI00219365.3	MSN	8.00%	0%
IPI00030275.5	TRAP1	14.80%	4.30%
IPI00170596.1	SIN3A	6.00%	0%
IPI00020557.1	LRP1	2.20%	0%
IPI00479191.2	HNRNPH1	7.20%	0%
IPI00221325.3	RANBP2	3.20%	0%
IPI00401264.5	ERP44	8.10%	0%
IPI00387144.5	TUBA1C	10.20%	6.40%
IPI00216746.1	MIR7-1	21.60%	13.60%
IPI00414717.1	GLG1	4.20%	0%
IPI00872684.1	EZR	11.80%	5.60%
IPI00017334.1	PHB	27.20%	15.80%
IPI00328266.4	ARID3B	17.50%	0%
IPI00024919.3	PRDX3	17.20%	14.10%
IPI00027107.5	TUFM	11.90%	9.90%
IPI00783271.1	LRPPRC	12.40%	4.40%
IPI00018206.3	GOT2	15.80%	11.60%

IPI00253050.2	L1TD1	12.80%	6.10%
IPI00022891.3	SLC25A4	12.10%	0%
IPI00789134.4	GAPDH	12.80%	15.20%
IPI00644431.2	DDX39	5.10%	5.10%
IPI00872375.3	SLC2A1	6.40%	0%
IPI00217975.4	LMNB1	18.80%	16.00%
IPI00215965.3	HNRNPA1	7.00%	2.70%
IPI00232571.1	GPC4	5.80%	0%
IPI00216592.2	HNRNPC	22.90%	22.90%
IPI00003362.2	HSPA5	16.20%	20.20%
IPI00514856.4	UBAP2L	10.30%	0%
IPI00902560.1	VDAC2	16.50%	6.50%
IPI00003865.1	HSPA8	18.90%	16.40%
IPI00646182.5	ATP1A1	13.00%	7.90%
IPI00299402.1	PC	10.40%	11.30%
IPI00291006.2	MDH2	27.20%	17.80%
IPI00291467.7	SLC25A6	15.10%	8.10%
IPI00032140.4	SERPINH1	12.00%	6.50%
IPI00554481.1	SLC3A2	13.90%	3.20%
IPI00333541.6	FLNA	11.70%	7.70%
IPI00024242.3	EBF3	1.20%	1.20%
IPI00744115.1	PCCA	19.20%	24.20%
IPI00007188.5	SLC25A5	15.10%	8.10%
IPI00641743.2	HCFC1	7.00%	0%
IPI00107122.1	NUPL1	9.00%	0%
IPI00303476.1	ATP5B	26.50%	21.60%
IPI00440493.2	ATP5A1	25.00%	22.20%
IPI00293533.4	NUP62	9.00%	0%
IPI00969566.1	UBC	3.60%	3.60%
IPI00027230.3	HSP90B1	22.70%	14.10%
IPI00007765.5	HSPA9	23.40%	22.70%
IPI00172580.4	NUP54	27.60%	0%
IPI00784154.1	HSPD1	31.10%	21.80%
IPI00003269.1	ACTBL2	12.00%	0%
IPI00021428.1	ACTA1	15.10%	0%
IPI00021440.1	ACTG1	24.80%	20.00%

Appendix II. Complete list of peptides analyzed for relative quantitation using Skyline (chapter 3)

Gene	Peptide	Ratio (Stage 1/Stage 4)[1]
Dhcr7	LLVSGFWGVAR	3773.422782
GTF2H1	FFQSHYFHR	1526.234569
COX6B1	NCWQNYLDFHR	1288.433538
Mcm6	NLYHNLCTSLFPTIHGNDEVKR	1280.996936
HIST1H3A	YRPGTVALR	1204.422933
Mll	VFPWFSVK	1045.332155
SUPT16H	RLYSNWR	1008.529353
SF3B1	HFWQHR	553.5188548
RPL15	YIQELWR	516.9517682
CPOX	WCDDYFFIAHR	516.9344087
Mobkl3	IFSHAYFHHR	490.2383431
CHD4	TYEIWHR	479.2639061
RANGAP1	HSLLQTLYKV	455.8217581
MCM3AP	DWYDFVWNR	377.987694
PITPNB	FKWWGLQSK	326.4899294
HIST1H2BK	EIQTAVR	314.6471494
MT-CO1	WLFSTNHK	303.3822228
PPP2R5C	ILPIMFPSLYR	290.4840611
ZMYM3	HFCNQQCLLR	275.1923504
KPNA2	FVSFLGR	245.239902
CHD4	QSYWNHR	242.2322785
RPL7A	VAPAPAVVK	227.7665223
GNPDA1	GLMLVHNK	218.6818896
STT3A	EAYYWLR	207.1711264
VANGL1	WLSTQWR	199.291109
RCOR2	YYYSWK	195.0782902
GTF2H1	FFQSHYFHR	193.923597
SOX2	SEASSSPPVVTSSSHSR	189.9649049
IDH1	IWYEHR	186.7184491
BPTF	KYWFLNR	182.5992191
LETM1	HYYHGFR	176.0656537
NDUFA8	FYFWTK	164.7303273
MYO18A	GVKDWPWWK	161.2128831
RANGAP1	CHWSDMFTGR	154.2864631
C19orf66	QFACSSCDHVWWR	151.1318612
HCFC1	LYIWSGR	147.9966748
SUPT16H	LYSNWR	147.7026776
ACO1	LFFWNSK	117.8649561
ACTG1	GYSFTTTAER	113.0533803

[1] Ratio = (Total area in stage 1/mean of all areas in stage 1) / (Total area in stage 4/mean of all areas in stage 4)

PNP	FHMYEGYPLWK	109.4709743
SURF4	MWFQWSEQR	107.2162012
EDIL3	WYPYYAR	106.449722
HNRNPL	VFNVFCLYGNVEK	103.0150003
ACAA2	HNFTPLAR	102.1875956
CENPV	HFIVPASR	100.4373394
NCL	GLSEDTTEETLKESFDGSVR	0.009907388
ADNP	YLPTDTLLNHMLIHGLSCPYCR	0.009831289
C2orf43	AFLLFPTIER	0.009816332
XPO5	YLESFLAFTTHPSQFLR	0.009737487
GEMIN5	SLEAFFLYGR	0.009723086
RBPJ	FFCPPPCVYLMGSGWK	0.009696396
LAMP1	FFLQGIQLNTILPDAR	0.009641063
FOXP1	MFAYFR	0.009598027
PRPF8	GTYFPTWEGLFWEK	0.009450427
RPL6	QLASGLLLVTGPLVLNR	0.009321942
MYOF	WVTFLLK	0.009316704
SFPQ	AVVIVDDR	0.009301622
Cbx3	GFTDADNTWEPEENLDCPELIEAFLNSQK	0.009213794
NPM1	MSVQPTVSLGGFEITPPVVLR	0.009130426
UBB	TLSDYNIQK	0.008943268
ABCF1	LQGQLEQGDDTAAERLEK	0.00893046
EEF2	AYLPVNESFGFTADLR	0.008921156
NSDHL	AFHITNDEPIPFWTFLSR	0.008911145
RTL1	NFIEFVFPYR	0.00890282
CTBP1	AFGFNVLFYDPYLSDGVER	0.008879921
GCN1L1	HAYLQCMLASYR	0.008865926
CRSP3	FLPVFDIVIHR	0.008668006
PDIA4	DLGLSESGEDVNAAILDESGKK	0.008629768
VIM	NLQEAEEWYK	0.008621773
TF	SAGWNIPIGLLYCDLPEPR	0.008546808
SRSF1	EAGDVCYADVYR	0.008514144
IARS	LMAPYTPFLTELMYQNLK	0.008474969
SFPQ	FGQGGAGPVGGQGPR	0.008427019
MAVS	DTLWHLFNTLQR	0.008362043
HNRNPL	SSSGLLEWESK	0.008242741
ACTR3	TLTGTVIDSGDGVTHVIPVAEGYVIGSCIK	0.008208308
DICER1	FLLFTDTFLR	0.008187483
TPM2	LEEAEKAADESER	0.008180153
SFPQ	DKLESEMEDAYHEHQANLLR	0.00813656
TCERG1	TYYYNNR	0.008098767
ERP29	LDKESYPVFYLFR	0.008084177
ATP5A1	AVDSLVPIGR	0.007993007
RPL8	ASGNYATVISHNPETK	0.007910723

VIM	FANYIDKVR	0.007838837
ENO1	YISPDQLADLYK	0.00767608
RPLP2	ILDSVGIEADDDRLNK	0.00764162
PDIA6	LAAVDATVNQVLASR	0.007620854
DDX21	GVTFLFPIQAK	0.00761566
ACTN4	VGWEQLLTTIAR	0.0076036
NOLC1	ATGATQQDANASSLLDIYSFWLK	0.007460853
PRKDC	QLFSSLFSGILK	0.007344706
KIAA1797	YLLISAPLWIK	0.007343401
DSC2	YTYSEWHSFTQPR	0.007328239
PHB2	IYLTADNLVLNLQDESFTR	0.007280251
ALDH2	GYFIQPTVFGDVQDGMTIAK	0.007217433
RPS20	LIDLHSPSEIVK	0.007208107
CANX	APVPTGEVYFADSFDR	0.007185411
VKORC1L1	GFGLLGSIFGK	0.007102521
EEF2	KIWCFGPDGTGPNILTDITK	0.007069405
RPLP2	ILDSVGIEADDDRLNK	0.007054167
SYNE2	ALALWDKLFNLK	0.00703268
PHB	QVSDDLTER	0.007013167
COQ9	MLIPYIEHWPR	0.006969451
RPS24	TTPDVIFVFGFR	0.006845271
LONP1	IVSGEAESVEVTPENLQDFVGKPVFTVER	0.006613621
P4HB	YQLDKDGVVLFK	0.006555989
PHB	AAELIANSLATAGDGLIELR	0.006452953
NOP16	RFYPAEWQDFLDSLQK	0.006412315
RPN2	SIVEEIEDLVAR	0.006392088
ACTR2	HLWDYTFGPEK	0.006383947
HNRNPK	LLIHQSLAGGIIGVK	0.006344126
FLAD1	LGLGSYPDWGSNYYQVK	0.006288914
PDIA6	ALDLFSDNAPPPELLEIINEDIAKR	0.006121837
HUWE1	EMFNPMYALFR	0.006039379
CCT2	VQDDEVGDGTTSVTVLAAELLR	0.00602554
LECT1	VIMPCSWWVAR	0.006018695
MRPL13	IWYLLDGK	0.006017386
ETFA	LLYDLADQLHAAVGASR	0.005990683
RPL37A	YTCSFCGK	0.005975802
HNRNPU	DLPEHAVLK	0.005962828
DLD	NLGLEELGIELDPR	0.005938677
SCFD2	VTPGQLMSYIQLFK	0.005927014
MYH9	IIGLDQVAGMSETALPGAFK	0.005866895
FLNA	ASGPGLNTTGVPASLPVEFTIDAK	0.005845191
RPS3A	VFEVSLADLQNDEVAFR	0.005746917
NES	SLGAWNLENLR	0.005744278
CSNK2A1	FYMYEILK	0.005666803

MDH2	FVFSLVDAMNGK	0.005627541
MDH2	TIIPLISQCTPK	0.005550424
VIM	LQDEIQNMKEEMAR	0.005426995
STOML2	ILEPGLNILIPVLDR	0.00535415
DHX8	FSNPWCYENFIQAR	0.005307596
PRPF38B	LEWFSTLFPR	0.005301449
RPS18	FQHILR	0.005299752
KPNA2	NLTWTLSNLCR	0.005259851
NOP58	FQDTAEALAAFTALMEGK	0.005225369
TRIM28	LSPPYSSPQEFAQDVGR	0.005103485
RPS3	IMLPWDPTGK	0.005101629
MGAT1	AFWDDWMR	0.005090703
SF3A1	EPPPEFEFIADPPSISAFDLDVVK	0.005038419
MANBA	IEAYNICHLNYFTFSPIYDK	0.004956908
VDAC1	EHINLGCDMDFDIAGPSIR	0.004952735
ILF2	ILPTLEAVAALGNK	0.004912055
SERPINH1	LQIVEMPLAHK	0.004905131
SCYL1	LGCLIWEVFNGPLPR	0.004872093
RALA	VFFDLMR	0.004848146
PRPSAP2	ASPFLLQYIQEEIPDYR	0.004814696
ATIC	ANYWWLR	0.0048096
G6PD	LFYLALPPTVYEAVTK	0.004807863
ATP5O	FSPLTTNLINLLAENGR	0.004794564
NCAPG	HLLQGWLR	0.004776513
NT5C	LLSWSDNWR	0.004762966
VIM	FLEQQNK	0.00473811
MATR3	YQLLQLVEPFGVISNHLILNK	0.004685585
LIN28B	MGFGFISMINR	0.004685215
TRAP1	FFEDYGLFMR	0.004680734
XRCC5	HNYECLVYVQLPFMEDLR	0.004661443
CAMK2D	HPWICQR	0.004652337
CFL1	NIILEEGKEILVGDVGQTVDDPYATFVK	0.004573646
DYNC1H1	QNLFTTWSHHLQQANIQFR	0.004503769
P4HB	HNQLPLVIEFTEQTAPK	0.004501641
PTD004	IPAFLNVVDIAGLVK	0.004498675
TARS	GAYIYNALIEFIR	0.004483942
RPL37A	TVAGGAWTYNTTSAVTVK	0.0044597
TIA1	GYGFVSFFNK	0.00443066
FLNA	TFSVWYVPEVTGTHK	0.004377703
SPATA20	SPYLLQHAYNPVDWYPWGQEAFDK	0.004346328
NUP107	YLFTLIR	0.004340017
RPS5	TIAECLADELINAAK	0.004329242
ATP5B	IPSAVGYQPTLATDMGTMQER	0.004283681
HIST1H4A	TVTAMDVVYALKR	0.00426037

WDR77	VWDLAQQVVLSSYR	0.004249951
TPI1	VVLAYEPVWAIGTGK	0.004230147
RIF1	LWPLFVK	0.004224684
ATP5H	KYPYWPHQPIENL	0.004194345
EEF2	YVEPIEDVPCGNIVGLVGVDQFLVK	0.00418757
TPI1	ELASQPDVDGFLVGGASLKPEFVDIINAK	0.004166804
PDIA3	FISDKDASIVGFFDDSFSEAHSEFLK	0.004112169
ILF3	LFPDTPLALDANK	0.004111806
NUP93	EALQYFYFLR	0.004091241
DDX50	GAVDALAAALAHISGASSFEPR	0.00404009
FUS	AAIDWFDGKEFSGNPIK	0.004026228
KRT8	SNMDNMFESYINNLRR	0.004011727
CPOX	FGLFTPGSR	0.004005244
EEF1A1	DGNASGTTLLEALDCILPPTRPTDKPLR	0.003910333
RTN4	GSSGSVVVDLLYWR	0.00386902
LACTB	FGNAMLYGYQVGLFK	0.003862435
HSPA9	LLGQFTLIGIPPAPR	0.003798603
VIM	LGDLYEEEMR	0.003763989
VIM	LQEEMLQREEAENTLQSFR	0.003763094
RUVBL1	QAASGLVGQENAR	0.003747802
HNRNPA2B1	GFGFVTFDDHDPVDKIVLQK	0.003679248
GALNT2	GGFDWNLVFK	0.003668603
ARF4	HYFQNTQGLIFVVDSNDRER	0.003663198
UQCRC1	HLGGIPWTYAEDAVPTLTPCR	0.003646388
ESYT1	WFTLSSGQGQVLLR	0.003631249
MDH2	VNVPVIGGHAGK	0.003615408
TUBB	AILVDLEPGTMDSVR	0.003597115
PLOD1	FLGSGGFIGYAPNLSK	0.003589167
HSPD1	VGEVIVTK	0.003582948
TBC1D15	FLLGYFPWDSTKEER	0.003545318
KDM1A	VFWDPSVNLFGHVGSTTASR	0.003531195
EEF2	GHVFEESQVAGTPMFVVK	0.00346416
QRICH1	LYDFYLFK	0.003461634
UFL1	TQSTWVDSFFR	0.003398796
PSMD3	SLMPYFLLTQAVR	0.003376276
CKB	LAVEALSSLDGDLAGR	0.00328248
Actc1	YPIEHGIITNWDDMEKIWHHTFYNELR	0.003271396
CCT4	DALSDLALHFLNK	0.003248267
ZCCHC11	LFGTPFYPLIGR	0.003243364
PPP1CA	EIFLSQPILLELEAPLK	0.003226175
PDIA6	GSTAPVGGGAFPTIVEREPWDGR	0.003224474
HSPA8	QTQTFTTYSDNQPGVLIQVYEGER	0.003204405
ACTB	LCYVALDFEQEMATAASSSSLEK	0.003198855
FLNA	FGGEHVPNSPFQVTALAGDQPSVQPPLR	0.00316898

OGDH	HWLDSPWPGFFTLDGQPR	0.003131874
HNRNPH1	STGEAFVQFASQEIAEK	0.003080042
RPL34	IVYLYTK	0.003059629
XRCC5	EPLPPIQQHIWNMLNPPAEVTTK	0.003027947
SNRNP200	GLFYFDNSFRPVPLEQTYVGITEK	0.003024402
TRA2B	GYDDRDYYSR	0.003008705
ILF2	NQDLAPNSAEQASILSLVTK	0.002992089
RPSA	FTPGTFTNQIQAAFR	0.002986143
RPL14	LVAIVDVIDQNR	0.002932295
HYOU1	LSAASTWLEDEGVGATTVMLK	0.002894972
CRAT	KTENWLSEWWLK	0.002889805
RBM4	TPYTMSYGDSLYYNNAYGALDAYYKR	0.002886485
GANAB	VSQGSKDPAEGDGAQPEETPR	0.002879022
YWHAZ	TAFDEAIAELDTLSEESYK	0.002875104
FUT8	VHGDPAVWWVSQFVK	0.002704402
ANXA5	FITIFGTR	0.002647548
HIST1H4A	DAVTYTEHAKR	0.002558909
SNRNP200	QVLDLEDLVFTQGSHFMANKR	0.002557693
RPL5	RFPGYDSESK	0.002487471
SLC25A1	FGMFEFLSNHMR	0.002486956
HNRNPK	ILSISADIETIGEILKK	0.002473548
UBN2	YGGFYINTGTLQFR	0.002459246
EHD1	VYIGSFWSHPLLIPDNR	0.002443485
NUP85	FADAASLLLSLMTSR	0.002438036
PPP1CA	IKYPENFFLLR	0.002432583
DYNC1H1	LYQEMFAWK	0.002413236
MTMR1	WLSNVDGTHWLEYIR	0.002405666
YWHAE	AAFDDAIAELDTLSEESYK	0.002398491
CCT3	NVLLDPQLVPGGGASEMAVAHALTEK	0.002395274
GNB2L1	HLYTLDGGDIINALCFSPNR	0.002393918
PDIA3	GFPTIYFSPANK	0.002376261
RPL15	FFEVILIDPFHK	0.002363185
KIAA1432	VWLPLFPR	0.002350754
MTFMT	SLTATDFYNGYLHPWYQK	0.002345144
CHID1	HFAGDVLGYVTPWNSHGYDVTK	0.002344604
SLC30A7	WILSQTHNIFTQAGVR	0.002315947
KHDRBS1	DSLDPSFTHAMQLLTAEIEK	0.002288915
HSPD1	IMQSSSEVGYDAMAGDFVNMVEK	0.002286067
AP2A1	IVSSASTDLQDYTYYFVPAPWLSVK	0.002256091
DDB1	FLYGCQAPTICFVYQDPQGR	0.002243724
MDH2	VAVLGASGGIGQPLSLLLK	0.002229161
GBAS	SYQLRPGTMIEWGNYWAR	0.002223989
KPNA2	NKNPAPPIDAVEQILPTLVR	0.002223248
40422	STLINSLFLTDLYPER	0.00221944

PDIA3	FLQDYFDGNLK	0.002212618
CLU	LFDSDPITVTVPVEVSR	0.002188643
HSP90B1	KYSQFINFPIYVWSSK	0.002165538
CSNK2A1	LIDWGLAEFYHPGQEYNVR	0.002156441
HNRNPA2B1	GFGFVTFDDHDPVDKIVLQK	0.002127519
MDH2	GCDVVVIPAGVPR	0.00210301
NCL	VEGTEPTTAFNLFVGNLNFNK	0.002067108
ALDH4A1	LYVPHSLWPQIK	0.002037425
RPL3	ERLEQQVPVNQVFGQDEMIDVIGVTK	0.002032302
EIF3E	LFIFETFCR	0.002028663
CTSD	ISVNNVLPVFDNLMQQK	0.002025209
PRPF19	VTSVVFHPSQDLVFSASPDATIR	0.002024654
PRPF8	TEDPDLPAFYFDPLINPISHR	0.001971463
UNQ2523	FIYITPEELAAVANFIR	0.001968701
RPS13	GLAPDLPEDLYHLIK	0.001968066
TSR1	WTYDPYVPEPVPWLK	0.001946305
TP53BP2	LLPFLSNPYR	0.001919737
ENO1	LAMQEFMILPVGAANFR	0.001910286
TRAP1	HLAEHSPYYEAMK	0.001906616
TUFM	LLDAVDTYIPVPAR	0.001902669
RPS28	EGDVLTLLESER	0.00189616
NDUFB8	LNWGEPMHWHLDMYNR	0.001877058
PTBP1	IAIPGLAGAGNSVLLVSNLNPER	0.001864723
CALR	KPEDWDEEMDGEWEPPVIQNPEYK	0.001848864
GOT2	IAAAILNTPDLR	0.001847683
PTBP1	KLPIDVTEGEVISLGLPFGK	0.001826867
NUP62	ILNAHMDSLQWIDQNSALLQR	0.001826119
Nbla00271	TSVVDLLYWR	0.001814834
TCEAL4	LGAFLWMQR	0.001775494
YWHAB	AVTEQGHELSNEER	0.001731697
XRCC5	KYAPTEAQLNAVDALIDSMSLAK	0.001687038
ZSCAN18	LLLWGYQLSQPDAASR	0.001679295
ENO1	SFIKDYPVVSIEDPFDQDDWGAWQK	0.00167759
U2AF2	RPHDYQPLPGMSENPSVYVPGVVSTVVPDSAHK	0.001668503
AKR7A2	FYAYNPLAGGLLTGK	0.001664865
HDGFRP2	YPIFFFGTHETAFLGPK	0.001653871
PLOD1	LQLNYLGNYIPR	0.001639685
RP11-34E5.1	YFSLPFCVGSK	0.001636227
BANF1	DFVAEPMGEKPVGSLAGIGEVLGK	0.001616908
HSPA9	STNGDTFLGGEDFDQALLR	0.001595298
FAM82B	LAAFWLMK	0.001583723
ACTR1A	IWQYVYSK	0.001568235
PPP2CB	YSFLQFDPAPR	0.001565338
NUP160	FVSSPQTIVELFFQEVAR	0.00153827

ACAP2	WFSIQNNQLVYQK	0.001517518
ZNF195	VFMWFSDITK	0.001499108
NUP93	LESLSAATTFEPLEPVKDTDIQGFLK	0.001488632
FLNA	VGSAADIPINISETDLSLLTATVVPPSGR	0.001486636
TTLL12	MPVWYIMDEFGSR	0.001472927
LAMA5	TYQPWQFFASSK	0.001469421
CLTC	ADDPSSYMEVVQAANTSGNWEELVK	0.001452194
HSP90B1	EFEPLLNWMK	0.001442174
COPB2	NNVAFMSYFLQGK	0.00143843
SRSF1	SHEGETAYIR	0.001428776
CCT2	MLPTIIADNAGYDSADLVAQLR	0.001419284
TBC1D7	LPYDLWFK	0.001364156
HSPA9	QAASSLQQASLK	0.001344565
ILF3	VLECLASGIVMPDGSGIYDPCEKEATDAIGHLDR	0.001336907
ARIH2	YTLQYTYPYAYYMESGPR	0.001319329
ILVBL	LGLFYQLLHK	0.001302796
HNRNPK	IITITGTQDQIQNAQYLLQNSVK	0.001273675
PDIA6	TGEAIVDAALSALR	0.001266218
TTC38	IYPFWTPDIPLSSYVK	0.001257834
MYL6	VLDFEHFLPMLQTVAK	0.001256393
PKM2	IYVDDGLISLQVK	0.001255863
KIDINS220	LASWINLTEQWPYR	0.001251707
ARSB	GFDTYFGYLLGSEDYYSHER	0.001248724
HELLS	WYQVEGMEWLR	0.001242867
GANAB	FSFSGNTLVSSSADPEGHFETPIWIER	0.00122995
FERMT2	YYSFFDLNPK	0.001203595
METTL1	MFFLFPDPHFK	0.001196795
RPLP2	NIEDVIAQGIGK	0.001168061
PTBP1	NFQNIFPPSATLHLSNIPPSVSEEDLK	0.001165649
HNRNPA2B1	GFGFVTFDDHDPVDKIVLQK	0.001160877
POLDIP2	VTVIPFYMGMR	0.001158212
MYH9	TQLEELEDELQATEDAK	0.001155986
NUP160	TAPLLLSYYLIK	0.00114783
HSPD1	ISSIQSIVPALEIANAHR	0.001144201
CECR2	HGSQGPGQGTWWLLCQTEEEWR	0.001137865
MED17	FQPSLWPWDSVR	0.001127616
SLC25A10	WYFGGLASCGAACCTHPLDLLK	0.00111595
SUZ12	IFYQFLYNNNTR	0.001089174
PTGR1	LGFDVVFNYK	0.001071408
ATP2A2	VSFYQLSHFLQCK	0.00105328
LARS	EVWDYVFFK	0.001030396
VIM	EMEENFAVEAANYQDTIGR	0.001029556
NCL	FGYVDFESAEDLEK	0.001028757
PFKL	EQWWLSLR	0.001008654

FUCA1	FFHPEEWADLFQAAGAK	0.0010021
MATR3	IGPYQPNVPVGIDYVIPK	0.000995305
TNPO2	FLQFFK	0.000943387
ZFC3H1	YPIPFSSADYWSNYEFHNR	0.000939238
PTBP1	KLPIDVTEGEVISLGLPFGK	0.000914649
PRSS23	DFLLNYPFSTSVK	0.000912797
TUBB	ALTVPELTQQVFDAK	0.000908944
SSR4	NNEDISIIPPLFTVSVDHR	0.00090625
40422	STLINSLFLTDLYPER	0.000860043
WDHD1	FMVWNSIGIIR	0.000858674
CCT4	DALSDLALHFLNK	0.000851102
AGRN	IFFVNPAPPYLWPAHK	0.000847426
RCC1	VFLWGSFR	0.000822108
TUBB	MAVTFIGNSTAIQELFK	0.000816706
TMLHE	FDYVWLR	0.0008135
HSP90B1	FQSSHHPTDITSLDQYVER	0.000811909
ITGA2	GNWLLVGSPWSGFPENR	0.000795527
SRSF7	NPPGFAFVEFEDPR	0.000782685
COPB1	LLHEMILVCDAYRK	0.000766677
CKMT1A	GWEFMWNER	0.000765429
KHSRP	TSMTEEYRVPDGMVGLIIGR	0.000748636
OAT	IVFAAGNFWGR	0.000741839
FLNA	GLVEPVDVVDNADGTQTVNYVPSR	0.000740903
P4HB	ILEFFGLK	0.000740266
MDH2	LTLYDIAHTPGVAADLSHIETK	0.000728705
CTSC	ILHLPTSWDWR	0.000719103
MRPS18B	DLDFSTSHGAVSATPPAPTLVSGDPWYPWYNWK	0.000715249
MYH9	VISGVLQLGNIVFK	0.000706094
TRA2B	YGPIADVSIVYDQQSR	0.000705259
NID1	GNLYWTDWNR	0.000694165
ACTR3	DITYFIQQLLR	0.000669521
IARS	FVDILTNWYVR	0.000665536
C2orf56	FNFFALLPHQR	0.000633845
PRPF8	YWIDIQLR	0.000611307
ITGA6	LIATFPDTLTYSAYR	0.000610818
UNQ2523	FIYITPEELAAVANFIR	0.000606422
COX5A	WVTYFNKPDIDAWELRK	0.000603854
STIP1	FMNPFNMPNLYQK	0.000601306
SRSF1	GGPPFAFVEFEDPRDAEDAVYGR	0.000594142
ACTN4	LASDLLEWIR	0.000574389
RPLP2	LASVPAGGAVAVSAAPGSAAPAAGSAPAAAEEK	0.00057006
VIM	ILLAELEQLKGQGK	0.000567665
HIST1H4A	TVTAMDVVYALK	0.000553524
UBR4	WFDFPFTR	0.000548013

BYSL	FYNLVLLPR	0.00053243
ZMYND8	FGVFNYSPFR	0.000526746
PRPF19	ALQDEWDAVMLHSFTLR	0.000525697
NID1	IYTYQWR	0.000525657
ATP5B	FLSQPFQVAEVFTGHMGK	0.00051789
HNRNPA2B1	LFIGGLSFETTEESLR	0.000488607
BRD3	HQFAWPFYQPVDAIK	0.000485843
ESYT1	IVAQVWPFLGQYMEK	0.000434213
ARFGAP3	STELDSNWSWFQLR	0.000432808
RCN2	VIDFDENTALDDAEEESFR	0.000423522
HSP90B1	KYSQFINFPIYVWSSK	0.000422213
NAGLU	FLLGSWLEQAR	0.000418596
PKM2	TATESFASDPILYRPVAVALDTK	0.000414007
H2afv	ATIAGGGVIPHIHK	0.000411345
TRIP12	LFLQFVTGSPR	0.000409523
CRNKL1	ANPHNYDAWFDYLR	0.00037645
SHMT2	ISATSIFFESMPYK	0.000369793
Tubb3	LHFFMPGFAPLTAR	0.000355531
NUP155	YMDLLWR	0.000354956
TUBB	ISEQFTAMFR	0.000349809
GOT2	TCGFDFTGAVEDISK	0.000342665
TF	SAGWNIPIGLLYCDLPEPR	0.00034254
C4orf27	LSLPEDFYHFWK	0.000342462
GARS	TFFSFPAVVAPFK	0.000336301
MYL12B	ELLTTMGDRFTDEEVDELYR	0.000325632
NBEA	YYYWVINPADSSGITPK	0.000321141
OGDH	SWDIFFR	0.000318054
RRP12	SWLLPVIR	0.000312797
CHD1L	GIPTYIYYFPR	0.000312092
GANAB	QYASLTGTQALPPLFSLGYHQSR	0.000305743
RPS7	AIIIFVPVPQLK	0.00030008
XPO1	IYLDMLNVYK	0.000292716
PTDSS2	WQGLWNIPTYK	0.000276148
NUP133	YMTQFADQNFSDFLFR	0.00026959
HNRNPA3	GFAFVTFDDHDTVDKIVVQK	0.000267651
MOGS	HLWSPFGLR	0.000253576
CCDC51	TWWDRYEEFVGLNEVR	0.000247986
HSP90AB3P	HSQFIGYPITLY	0.00024797
ROCK2	AFVGNQLPFIGFTYYR	0.000241153
ATP5A1	FENAFLSHVVSQHQALLGTIR	0.000233602
RAN	SNYNFEKPFLWLAR	0.000224953
RAB1B	EFADSLGIPFLETSAK	0.00022404
HNRNPM	MGPAMGPALGAGIER	0.000222952
EIF3B	VDNAYWLWTFQGR	0.000218327

PMEL	TWGQYWQVLGGPVSGLSIGTGR	0.000216012
PDIA3	ELSDFISYLQR	0.000204407
GSTP1	FQDGDLTLYQSNTILR	0.000200771
SERPINH1	AVLSAEQLRDEEVHAGLGELLR	0.000200636
MYO1B	NFHVFYQLLSGASEELLNK	0.000193739
RPS24	TTGFGMIYDSLDYAK	0.000191568
TUBA1B	EIIDLVLDR	0.000184248
DKFZp686D20108	AFFEVLAHPQNYFK	0.000184109
GTPBP1	MFLNLLSPR	0.000183019
SERPINH1	LFYADHPFIFLVR	0.000182201
HNRNPL	YGPQYGHPPPPPPPPEYGPHADSPVLMVYGLDQSK	0.000174205
CDC5L	WYEWLDPSIK	0.000173985
HNRNPM	NLPFDFTWK	0.000171788
IDH2	IIWQFIK	0.000159613
RPL31	LYTLVTYVPVTTFK	0.000159146
RBPMS	ELYLLFRPFK	0.000158885
EEF1G	WFLTCINQPQFR	0.000158386
PSIP1	LPIFFFGTHETAFLGPK	0.000157014
SMC5	QYGFFSYLR	0.000151489
HNRNPUL1	WMGIAFR	0.000151005
HSP90AA1	HSQFIGYPITLFVEK	0.000146504
CARM1	ANFWYQPSFHGVDLSALR	0.000146006
TNFRSF10B	YGQDYSTHWNDLLFCLR	0.000139674
DHX9	YPSPFFVFGEK	0.000130569
SART3	YANMWLEYYNLER	0.000130371
FAM40A	FFNQNIMSYITAK	0.00011721
HSP90AB3P	HSQFIGYPITLYLEK	0.000114303
ACTB	TTGIVMDSGDGVTHTVPIYEGYALPHAILR	0.00011405
EPB41L5	IGLFFWPK	0.000107773
ACTB	DLYANTVLSGGTTMYPGIADR	0.000107324
RPLP0	TSFFQALGITTK	0.000105048
PPM1G	LPLPYGFSAMQGWR	0.000102955
TUBA1B	TIGGGDDSFNTFFSETGAGK	9.71533E-05
ARIH2	YTLQYTYPYAYYMESGPR	9.32769E-05
NARS	NLMFLVLR	9.01431E-05
ATP5A1	TGAIVDVPVGEELLGR	8.53179E-05
RPS2	TYSYLTPDLWK	8.11597E-05
MRPS5	YGFLWPGLNVPLMK	6.93752E-05
CTSC	ILHLPTSWDWR	5.75017E-05
STRBP	FPTYPVPHYSFF	5.38866E-05
STIP1	FMNPFNMPNLYQK	2.99274E-05
RPS10	IAIYELLFK	1.95667E-05
ATP5B	VALTGLTVAEYFR	1.61719E-05
TUBA1B	AVFVDLEPTVIDEVR	1.38956E-05

HNRNPK	IITITGTQDQIQNAQYLLQNSVK	1.10243E-05

Appendix III. Complete list of genes significantly up-regulated and down-regulated on day 8 of neuronal differentiation of OGT-inhibited hESCs (Chapter 4)

Gene	DMSO (FPKM)[2]	OGT inhibitor (FPKM)[2]	Log_2 (fold change)	p value	q value
FGF23	0.626834	0.124623	-2.33052	0.00075	0.0222552
SERPINA5	2.29459	0.485075	-2.24196	5.00E-05	0.00247621
PPP1R3C	2.32028	0.504181	-2.20229	5.00E-05	0.00247621
ADAMTS19	1.85785	0.407806	-2.18768	5.00E-05	0.00247621
ATP8B5P	0.652912	0.16699	-1.96713	0.00175	0.0406034
SLC13A4	1.35714	0.351561	-1.94872	5.00E-05	0.00247621
RHOH	2.01416	0.523958	-1.94266	5.00E-05	0.00247621
EYA1	1.07283	0.280191	-1.93694	5.00E-05	0.00247621
TC2N	1.18916	0.324345	-1.87434	5.00E-05	0.00247621
YBX2	2.11252	0.586855	-1.84789	0.0001	0.00439205
GRB7	2.55018	0.746392	-1.77259	5.00E-05	0.00247621
SLC38A4	7.986	2.36231	-1.75727	5.00E-05	0.00247621
PPM1N	1.09022	0.323156	-1.75432	0.0018	0.0412965
APCDD1	0.777456	0.243918	-1.67237	0.0012	0.0315648
KLF11	2.26428	0.718865	-1.65526	5.00E-05	0.00247621
C4orf33	3.6655	1.21564	-1.59229	5.00E-05	0.00247621
EFEMP1	16.9458	5.8315	-1.53899	5.00E-05	0.00247621
DIRAS2	13.3779	4.60983	-1.53706	5.00E-05	0.00247621
KCND2	0.582295	0.205976	-1.49927	0.0005	0.0162466
FZD1	10.2624	3.80263	-1.4323	5.00E-05	0.00247621
TMEM139	1.47827	0.549925	-1.4266	0.0013	0.0330453
KITLG	2.44672	0.91668	-1.41636	5.00E-05	0.00247621
SFN	2.52187	0.970865	-1.37715	0.00095	0.0263396
FAM160A1	1.95437	0.755105	-1.37195	5.00E-05	0.00247621
HESX1	2.9113	1.13525	-1.35865	0.0008	0.0234482
LPAR1	5.45616	2.14337	-1.348	5.00E-05	0.00247621
FAM222A	12.3218	4.91666	-1.32547	5.00E-05	0.00247621
PYCR1	31.8455	12.7072	-1.32544	5.00E-05	0.00247621
ZNF57	1.64769	0.659544	-1.3209	0.00085	0.0245628
RAB25	3.71102	1.54593	-1.26335	0.0008	0.0234482
UGT3A1	2.12948	0.89338	-1.25316	0.0002	0.00789121
ARHGEF19	5.10816	2.15377	-1.24594	5.00E-05	0.00247621
ADCYAP1R1	0.883894	0.376715	-1.2304	0.0002	0.00789121
FGF8	22.237	9.50047	-1.22689	5.00E-05	0.00247621
SLC7A8	46.0829	19.7156	-1.22489	5.00E-05	0.00247621
HIST3H2A	10.858	4.67164	-1.21675	0.00165	0.038848
IP6K3	4.23009	1.83072	-1.20828	5.00E-05	0.00247621
HAS2	3.3595	1.45681	-1.20544	5.00E-05	0.00247621
ENPP5	3.17862	1.3888	-1.19456	5.00E-05	0.00247621
C1orf151-NBL1	39.8146	17.4648	-1.18885	5.00E-05	0.00247621
MAP2K6	33.9233	15.1418	-1.16374	5.00E-05	0.00247621

[2] Fragments per kilobase of exon per million fragments mapped (FPKM)

FZD5	31.3365	14.023	-1.16004	5.00E-05	0.00247621
ARID3C	7.60367	3.41051	-1.15671	0.00015	0.00613761
IGF2	10.2179	4.62515	-1.14352	5.00E-05	0.00247621
BCHE	4.26905	1.94802	-1.13191	0.0001	0.00439205
ADCK3	8.35471	3.82092	-1.12867	5.00E-05	0.00247621
FOXG1	12.0623	5.5314	-1.12479	5.00E-05	0.00247621
ACBD4	9.97344	4.60628	-1.11449	5.00E-05	0.00247621
ICAM5	1.43047	0.666058	-1.10277	0.00145	0.0355981
C1orf51	7.47194	3.48467	-1.10046	5.00E-05	0.00247621
MGEA5	112.067	52.3333	-1.09857	5.00E-05	0.00247621
ELAVL2	3.03722	1.43141	-1.08532	0.0001	0.00439205
LOC400680	1.54274	0.728906	-1.08169	0.0005	0.0162466
LGR4	45.9525	21.7372	-1.07998	5.00E-05	0.00247621
FBLN1	17.3707	8.23931	-1.07606	5.00E-05	0.00247621
BBS9	56.7568	27.3519	-1.05316	5.00E-05	0.00247621
IL1RAPL1	1.78092	0.860234	-1.04982	0.0016	0.0379821
LINC00648	1.86668	0.905813	-1.04319	0.00155	0.0372883
AURKC	4.54496	2.2098	-1.04035	0.0017	0.0396998
COL9A3	10.034	4.962	-1.0159	0.0009	0.0254947
BST2	178.103	88.7462	-1.00495	5.00E-05	0.00247621
PRSS8	10.0238	4.998	-1.00401	5.00E-05	0.00247621
ETV4	12.4431	6.29002	-0.98421	5.00E-05	0.00247621
CD40	5.09282	2.59485	-0.972812	0.00075	0.0222552
CCNG2	38.2052	19.5025	-0.970111	5.00E-05	0.00247621
MICB	8.50296	4.34299	-0.969275	5.00E-05	0.00247621
CRELD1	17.5133	8.96758	-0.965657	5.00E-05	0.00247621
SLC25A33	11.3063	5.87344	-0.944848	0.00015	0.00613761
TNNI3	26.7278	13.8939	-0.943886	5.00E-05	0.00247621
ID2	36.8261	19.1644	-0.9423	5.00E-05	0.00247621
EYA4	2.12977	1.11018	-0.9399	0.0003	0.0110477
GCA	23.5091	12.2547	-0.939881	5.00E-05	0.00247621
MUM1L1	3.24476	1.69429	-0.937435	0.0002	0.00789121
SGK110	15.7318	8.24105	-0.932779	0.0001	0.00439205
NRP2	8.11324	4.28625	-0.920563	5.00E-05	0.00247621
SEMA5A	8.24827	4.36161	-0.919232	5.00E-05	0.00247621
TMEM108	4.89174	2.59003	-0.917383	5.00E-05	0.00247621
PXMP4	1.61655	0.855955	-0.917312	0.00115	0.030814
IRX2	3.36135	1.78183	-0.915686	0.00095	0.0263396
TNNT1	76.0005	40.6527	-0.902657	5.00E-05	0.00247621
LTBP4	36.8024	19.8224	-0.892667	5.00E-05	0.00247621
DDIT4L	11.1286	6.04026	-0.881587	5.00E-05	0.00247621
PDGFD	3.21166	1.74323	-0.881559	0.00085	0.0245628
ID3	34.1671	18.6182	-0.875893	0.0001	0.00439205
RCN3	5.26121	2.8904	-0.864127	0.0017	0.0396998
LOC100216545	4.94221	2.71536	-0.86401	0.0004	0.0138096
SHISA2	19.6427	10.8033	-0.862525	5.00E-05	0.00247621
TMEM45A	12.8298	7.07971	-0.857737	0.0004	0.0138096
ART5	6.89167	3.8213	-0.85079	0.00165	0.038848

CDON	15.428	8.61506	-0.840619	5.00E-05	0.00247621
LOC100505806	23.6331	13.2414	-0.835756	0.0001	0.00439205
ID1	59.3004	33.231	-0.835511	5.00E-05	0.00247621
VWDE	3.41315	1.91917	-0.830625	0.0001	0.00439205
VSNL1	6.72702	3.78839	-0.828382	0.00095	0.0263396
ATF7IP	67.7986	38.4567	-0.818021	5.00E-05	0.00247621
CREG1	20.9442	11.886	-0.817283	0.0001	0.00439205
JRKL	13.6712	7.76707	-0.815696	5.00E-05	0.00247621
MICA	18.1524	10.3803	-0.806319	0.00035	0.0125042
GJA1	645.356	369.081	-0.806159	5.00E-05	0.00247621
DDX25	6.1008	3.50753	-0.798543	0.00195	0.0439653
ARHGAP15	7.22108	4.18831	-0.785847	0.00165	0.038848
DANCR	85.0956	49.5494	-0.780218	0.0001	0.00439205
VGF	19.9457	11.6731	-0.772893	5.00E-05	0.00247621
PER3	2.43151	1.42856	-0.76729	0.0013	0.0330453
LOC650368	38.9827	22.957	-0.763898	0.00015	0.00613761
LOC100272216	7.89032	4.65495	-0.761319	0.0009	0.0254947
IGFBP5	162.801	96.0627	-0.761062	5.00E-05	0.00247621
MPP7	2.94573	1.7456	-0.754901	0.002	0.0446719
AK2	31.2701	18.5541	-0.753041	0.0001	0.00439205
GUCY1B3	14.4887	8.62894	-0.747674	5.00E-05	0.00247621
DIO3	15.9195	9.48766	-0.746668	0.0005	0.0162466
ST6GALNAC3	6.77654	4.0753	-0.733645	0.0011	0.0296959
C5orf54	10.452	6.37462	-0.713375	0.0007	0.0214358
MTHFD2	18.8887	11.5341	-0.711618	0.0004	0.0138096
IGFBP3	49.1236	30.0654	-0.708312	0.0001	0.00439205
IFI16	6.7138	4.11041	-0.707849	0.0016	0.0379821
CREBRF	6.9476	4.25991	-0.705691	0.0001	0.00439205
GPR180	3.82717	2.35604	-0.699916	0.00075	0.0222552
SLCO1A2	4.09466	2.52368	-0.698213	0.00135	0.0339557
PAMR1	74.8301	46.3176	-0.692057	5.00E-05	0.00247621
ARRDC4	33.7553	20.9165	-0.690474	5.00E-05	0.00247621
LRBA	10.7211	6.64441	-0.690233	0.00095	0.0263396
EXO1	9.91589	6.14796	-0.689633	0.0004	0.0138096
EFNA5	30.9921	19.246	-0.687343	0.0001	0.00439205
IL17RD	31.1457	19.4379	-0.680159	0.0001	0.00439205
KIAA0101	55.0569	34.4081	-0.678177	0.00015	0.00613761
NELL2	262.02	163.992	-0.676049	0.00025	0.00957467
ATAD5	19.0954	11.981	-0.67247	0.0001	0.00439205
LAMB1	121.288	76.3709	-0.667344	0.0003	0.0110477
RANBP6	27.8625	17.5803	-0.664368	0.00015	0.00613761
SCML2	12.3013	7.81485	-0.65452	0.00085	0.0245628
TCN2	21.1269	13.4273	-0.653912	0.00055	0.0175926
CCDC28A	17.6222	11.2618	-0.645965	0.0021	0.0466155
PRCP	22.0695	14.1294	-0.643355	0.00085	0.0245628
AMOT	4.13391	2.64703	-0.64313	0.0019	0.0431769
PTN	51.7204	33.1744	-0.640662	0.0006	0.0189806
CDV3	80.2852	51.7592	-0.633319	0.0005	0.0162466

GLCCI1	28.8832	18.671	-0.629436	0.00035	0.0125042
RGS16	29.1373	18.8648	-0.627169	0.0007	0.0214358
CEP152	7.44373	4.82317	-0.626045	0.00095	0.0263396
GAS5	476.432	309.004	-0.624643	0.0001	0.00439205
TXNIP	531.373	345.442	-0.621279	0.0013	0.0330453
EPHA4	43.2845	28.2509	-0.615554	0.0005	0.0162466
PDP1	24.0253	15.709	-0.612964	0.0009	0.0254947
DPYSL3	217.653	142.359	-0.612494	0.00135	0.0339557
ZNF521	70.0134	45.8648	-0.610243	0.00075	0.0222552
SLC3A2	153.595	100.714	-0.608858	0.0005	0.0162466
CTNNBIP1	30.97	20.3205	-0.607937	0.0006	0.0189806
FBXO33	11.8721	7.79928	-0.606158	0.00105	0.0285608
SLC29A1	48.7619	32.1158	-0.602472	0.0003	0.0110477
NBN	12.6476	8.33351	-0.601868	0.00155	0.0372883
CDC6	15.1328	9.98147	-0.600355	0.0023	0.0493024
NRXN1	3.93223	2.60442	-0.594387	0.00215	0.0470706
TIAM1	19.2262	12.7594	-0.591514	0.00085	0.0245628
ATAD2	14.8163	9.87258	-0.585688	0.00165	0.038848
BBS10	22.4121	14.9633	-0.582849	0.0018	0.0412965
ACTN3	23.0023	15.3899	-0.579799	0.00205	0.045576
F11R	30.5513	20.4686	-0.577823	0.0011	0.0296959
RAD51AP1	28.2527	18.9827	-0.573705	0.00135	0.0339557
FKBP5	17.998	12.1874	-0.562451	0.00185	0.0422412
SERPING1	39.0403	26.5316	-0.557253	0.002	0.0446719
MTRNR2L1	122.61	83.4688	-0.554762	0.0011	0.0296959
ZNF738	13.5923	9.2615	-0.553471	0.00145	0.0355981
MCM6	24.8464	16.9802	-0.549183	0.002	0.0446719
P4HA1	98.4894	67.4406	-0.546352	0.00215	0.0470706
PCDH18	35.3794	24.2504	-0.5449	0.002	0.0446719
NCALD	60.0018	41.3031	-0.538754	0.00225	0.048375
SAT1	302.339	209.208	-0.531228	0.00155	0.0372883
APLP1	64.0113	92.0952	0.5248	0.00225	0.048375
EPHB2	21.0299	30.3349	0.528534	0.0018	0.0412965
CAPN2	33.0418	47.8974	0.535656	0.00175	0.0406034
ACTN1	39.4928	57.3381	0.537904	0.00195	0.0439653
CKB	278.63	407.217	0.547446	0.00225	0.048375
TNS3	12.536	18.3897	0.552821	0.00145	0.0355981
KLC2	19.169	28.1363	0.553655	0.00175	0.0406034
RAI14	21.5279	31.63	0.555082	0.0012	0.0315648
SYNE1	1.79399	2.63981	0.55726	0.0016	0.0379821
TBC1D9	16.5092	24.3306	0.5595	0.0012	0.0315648
BCAR1	26.5628	39.1559	0.559822	0.0012	0.0315648
LASP1	36.5619	53.9425	0.561081	0.00095	0.0263396
GPRC5B	24.467	36.1222	0.56205	0.0016	0.0379821
CLN8	6.30422	9.30888	0.56229	0.0019	0.0431769
ATP6V0E2	15.2223	22.4822	0.562595	0.0022	0.0477287
DMPK	16.9032	24.9719	0.563008	0.00145	0.0355981
TMEM132A	23.9705	35.4701	0.565343	0.00135	0.0339557

DPYSL5	45.8254	67.9856	0.569082	0.0012	0.0315648
EHD1	12.5387	18.6067	0.569434	0.00215	0.0470706
MEIS3	39.0577	58.006	0.570596	0.00105	0.0285608
FARP1	20.8702	31.0154	0.571542	0.00185	0.0422412
NRBP2	10.625	15.7945	0.57196	0.00205	0.045576
ZNF503	17.803	26.4861	0.573118	0.0023	0.0493024
PLXNA3	5.48774	8.17887	0.575689	0.00215	0.0470706
C7orf50	49.0453	73.112	0.575992	0.00175	0.0406034
WDR1	43.8033	65.3019	0.576084	0.00075	0.0222552
RGMA	57.9654	86.425	0.576259	0.0007	0.0214358
MAP1B	22.8645	34.0928	0.576358	0.00085	0.0245628
USP44	22.8361	34.1076	0.578777	0.00095	0.0263396
MYRF	4.78931	7.15805	0.579747	0.0022	0.0477287
TCF25	14.7327	22.0363	0.580854	0.0022	0.0477287
ARHGEF40	12.7181	19.0387	0.582056	0.0005	0.0162466
RFK	15.171	22.7157	0.582371	0.0018	0.0412965
LDLR	10.1793	15.2513	0.583296	0.0012	0.0315648
CPNE2	19.7631	29.6107	0.583308	0.0016	0.0379821
HGSNAT	7.61351	11.4106	0.583736	0.0016	0.0379821
TMEM170B	6.47914	9.71809	0.584872	0.00145	0.0355981
PLEC	14.1422	21.2511	0.587531	0.00065	0.0202941
KIAA0930	10.4371	15.6857	0.58773	0.0009	0.0254947
IER5L	28.62	43.0699	0.589653	0.00085	0.0245628
ITPKB	5.19437	7.83042	0.592142	0.0015	0.0365136
MYOF	10.8883	16.436	0.594085	0.0008	0.0234482
WASF1	29.0889	43.9433	0.595174	0.0009	0.0254947
P4HTM	15.1975	22.9627	0.595458	0.0014	0.0349076
MFAP3L	4.61958	6.98131	0.595738	0.00195	0.0439653
PLEKHB1	23.1716	35.0916	0.598765	0.001	0.0273562
MYO9B	9.18916	13.9226	0.599424	0.00075	0.0222552
PDZD4	8.20741	12.457	0.601963	0.00145	0.0355981
POR	32.3182	49.095	0.603231	0.00145	0.0355981
SCD	92.194	140.059	0.603287	0.00125	0.0322887
RING1	19.4204	29.5233	0.604277	0.00125	0.0322887
COL6A2	23.5078	35.7976	0.606724	0.00065	0.0202941
MAPK8IP1	6.88865	10.4907	0.60682	0.0016	0.0379821
PPP1R12C	21.761	33.1913	0.609057	0.00075	0.0222552
PKNOX2	7.38034	11.258	0.609195	0.0014	0.0349076
FADS3	14.5438	22.233	0.6123	0.00215	0.0470706
ARNT2	8.16569	12.4867	0.612745	0.0004	0.0138096
EPHB1	7.86741	12.0337	0.613123	0.0006	0.0189806
KLF12	2.1623	3.31051	0.614488	0.00165	0.038848
ZFP36L2	42.1714	64.6142	0.615584	0.00065	0.0202941
DKK3	16.7444	25.6811	0.617034	0.00075	0.0222552
FOXP4	13.4908	20.7053	0.618026	0.0005	0.0162466
MAPK15	13.8489	21.2725	0.619217	0.0015	0.0365136
TUBB2A	21.8365	33.5762	0.620696	0.0013	0.0330453
ARVCF	15.1534	23.391	0.626308	0.0019	0.0431769

STC2	12.5714	19.411	0.626731	0.0004	0.0138096
MXRA8	12.7671	19.7238	0.627508	0.0013	0.0330453
DHCR24	18.197	28.2073	0.632369	0.00025	0.00957467
RHOC	80.3655	124.659	0.633341	0.0005	0.0162466
LIPG	4.95626	7.69143	0.634	0.0015	0.0365136
MAPK11	22.146	34.4687	0.638243	0.0005	0.0162466
HES1	18.8515	29.3878	0.64054	0.0016	0.0379821
PLAGL1	12.2148	19.0552	0.641554	0.0007	0.0214358
KIF1A	24.6525	38.4638	0.641766	0.00035	0.0125042
SOGA2	3.47989	5.43528	0.643314	0.00135	0.0339557
CELSR1	6.56514	10.26	0.64414	0.0002	0.00789121
SMS	72.8782	114.065	0.646296	0.00015	0.00613761
JPH3	8.32127	13.0323	0.647217	0.00195	0.0439653
ASTN1	4.89319	7.67558	0.6495	0.001	0.0273562
TRANK1	6.5328	10.2651	0.651968	5.00E-05	0.00247621
RASA3	5.96583	9.37814	0.652579	0.0007	0.0214358
SGSM2	18.4173	29.0006	0.655026	0.0001	0.00439205
RSU1	4.19555	6.61111	0.656034	0.00215	0.0470706
MAPK12	22.6044	35.6194	0.656058	0.0004	0.0138096
KLHL21	2.73791	4.31544	0.65643	0.00225	0.048375
H2AFY	47.8541	75.4582	0.657034	0.00045	0.0152068
CCDC50	26.6017	41.9638	0.657624	0.0002	0.00789121
TLE3	11.155	17.6061	0.658381	0.00025	0.00957467
SHC2	21.7291	34.3163	0.659262	5.00E-05	0.00247621
LINC00617	8.06387	12.7379	0.659585	0.00055	0.0175926
C17orf70	4.30956	6.81807	0.661823	0.00125	0.0322887
COL6A1	33.738	53.5506	0.666528	5.00E-05	0.00247621
CREB5	7.34722	11.6774	0.66845	0.0001	0.00439205
FLNA	134.015	213.13	0.669336	0.00095	0.0263396
NACAD	2.9709	4.72701	0.670027	0.0014	0.0349076
CDH4	6.54119	10.4162	0.671203	0.0002	0.00789121
ABLIM1	18.5416	29.5314	0.671482	0.0002	0.00789121
ZYX	31.1391	49.602	0.671669	5.00E-05	0.00247621
AKAP6	2.44637	3.89702	0.671725	0.0005	0.0162466
ARHGEF4	13.9474	22.221	0.671931	5.00E-05	0.00247621
CTTN	39.4145	62.8339	0.672816	0.00015	0.00613761
FAM19A5	9.51572	15.1748	0.673296	0.00035	0.0125042
XYLT1	15.328	24.4549	0.673949	0.0003	0.0110477
ATXN1	1.50074	2.39667	0.675355	0.00125	0.0322887
NKAIN4	33.4642	53.4731	0.676196	0.00035	0.0125042
TUBA1A	598.021	956.309	0.677282	0.0007	0.0214358
AP2A2	12.5476	20.0692	0.677571	0.00025	0.00957467
PLXNB1	31.8479	50.9553	0.678033	5.00E-05	0.00247621
EPHB6	2.38323	3.8164	0.679294	0.00205	0.045576
FNDC4	8.57917	13.7391	0.679381	0.00125	0.0322887
PNMA3	5.22615	8.37129	0.679701	0.0009	0.0254947
GAS2L1	5.34833	8.56754	0.679792	0.0013	0.0330453
PTPRN2	4.53188	7.26174	0.680207	0.00045	0.0152068

SEPT8	26.4409	42.4156	0.681825	5.00E-05	0.00247621
PLAT	3.49114	5.60149	0.682114	0.0018	0.0412965
SPTBN2	7.27234	11.6911	0.684925	0.00015	0.00613761
SULF2	27.8606	44.8095	0.685575	0.00015	0.00613761
AARS2	4.33711	6.99047	0.688656	0.0005	0.0162466
NCAN	7.02289	11.3286	0.689826	0.00015	0.00613761
BAIAP2	27.2089	43.9215	0.690847	0.00045	0.0152068
CARD10	3.77315	6.09254	0.691273	0.00095	0.0263396
PDE1B	3.15804	5.10026	0.691541	0.0018	0.0412965
SOX6	4.53227	7.33231	0.694032	5.00E-05	0.00247621
WSCD1	10.1691	16.4615	0.694906	5.00E-05	0.00247621
NPY2R	3.67951	5.97551	0.699549	0.0013	0.0330453
HBEGF	8.56647	13.9284	0.701252	0.00045	0.0152068
PLEKHA6	2.08792	3.399	0.703043	0.00075	0.0222552
SOCS3	9.11171	14.8337	0.703088	0.0005	0.0162466
HS6ST1	25.3242	41.2444	0.70368	5.00E-05	0.00247621
STON2	2.75055	4.48412	0.705104	0.00145	0.0355981
NAB2	35.7405	58.2804	0.705452	0.00015	0.00613761
QPCT	15.5746	25.4032	0.705819	0.00045	0.0152068
MPPED1	4.49825	7.35142	0.70866	0.0013	0.0330453
ZHX3	2.13174	3.48619	0.709619	0.00055	0.0175926
CTGF	120.208	196.596	0.709697	5.00E-05	0.00247621
DRD4	34.7353	56.8924	0.711836	0.0002	0.00789121
TTYH1	232.259	380.94	0.713828	5.00E-05	0.00247621
MEIS1	7.94317	13.0345	0.714546	0.0003	0.0110477
ELOVL2	6.61193	10.8683	0.716979	0.00035	0.0125042
NPTXR	3.17798	5.23359	0.71969	0.0003	0.0110477
PLEKHG4B	5.95085	9.80858	0.720949	5.00E-05	0.00247621
SEPT3	13.7298	22.6323	0.721077	0.0001	0.00439205
DTX1	3.86846	6.38906	0.723842	0.00035	0.0125042
COTL1	98.5485	162.775	0.72397	5.00E-05	0.00247621
WDR86	22.1192	36.591	0.726192	5.00E-05	0.00247621
KNDC1	4.44587	7.36316	0.727859	0.0002	0.00789121
KCNG1	3.82115	6.33134	0.728506	0.00225	0.048375
FAM219A	8.91406	14.7853	0.730008	0.00025	0.00957467
C14orf132	2.65273	4.40218	0.73074	0.0003	0.0110477
CHST3	5.10191	8.46675	0.730769	5.00E-05	0.00247621
STMN4	10.7934	17.9232	0.731677	0.00065	0.0202941
DENND3	1.80035	2.99165	0.732662	0.0014	0.0349076
CDC42EP1	8.56081	14.2377	0.733902	0.0003	0.0110477
CORO2B	2.98395	4.96709	0.735176	0.00105	0.0285608
KIF26A	6.29237	10.4964	0.738223	5.00E-05	0.00247621
CNN1	25.8812	43.2534	0.740909	5.00E-05	0.00247621
BCAR3	11.6223	19.447	0.742655	5.00E-05	0.00247621
GLT1D1	3.79376	6.3516	0.743492	0.001	0.0273562
SOX3	36.1846	60.6008	0.74396	5.00E-05	0.00247621
PEA15	82.476	138.373	0.746517	0.0001	0.00439205
NAT8L	4.57024	7.67508	0.747913	0.0001	0.00439205

JDP2	3.68022	6.18224	0.748339	0.00035	0.0125042
EPHB3	6.56558	11.0421	0.750024	5.00E-05	0.00247621
MSRB3	1.86323	3.13367	0.750046	0.00215	0.0470706
GNAO1	3.09047	5.20306	0.751535	0.00035	0.0125042
ID4	54.5795	91.9997	0.75327	0.00015	0.00613761
PITPNM1	21.2604	35.8729	0.754727	5.00E-05	0.00247621
VASH1	11.0365	18.6302	0.755367	5.00E-05	0.00247621
OSBPL5	4.52731	7.6599	0.758672	0.00015	0.00613761
PTCH1	15.4382	26.1805	0.761985	5.00E-05	0.00247621
FGFR2	21.3356	36.1927	0.762437	5.00E-05	0.00247621
FLNC	49.2578	83.5783	0.762775	5.00E-05	0.00247621
EGR3	2.78653	4.73031	0.763466	0.00115	0.030814
FAT4	2.78231	4.72349	0.763571	5.00E-05	0.00247621
IRS1	5.20809	8.84317	0.763809	5.00E-05	0.00247621
EGR1	50.1813	85.3675	0.766535	5.00E-05	0.00247621
PFKFB3	5.75311	9.7871	0.766541	5.00E-05	0.00247621
LEPREL1	5.43538	9.25386	0.767675	0.0003	0.0110477
ANXA3	5.24611	8.93378	0.768023	0.00155	0.0372883
CHPF2	4.74569	8.10461	0.772126	5.00E-05	0.00247621
TPPP3	103.524	177.282	0.776084	5.00E-05	0.00247621
ACTB	846.905	1452.5	0.778268	0.0002	0.00789121
ALPK3	3.63995	6.24662	0.779157	5.00E-05	0.00247621
WNT2B	11.7005	20.0857	0.779601	0.00015	0.00613761
ENPP2	26.9902	46.3509	0.780163	5.00E-05	0.00247621
CCDC92	5.45406	9.38506	0.783036	0.00055	0.0175927
SALL3	7.46922	12.8651	0.78443	5.00E-05	0.00247621
KCNQ2	15.8725	27.341	0.784541	0.00025	0.00957467
RNF175	6.29481	10.8441	0.784675	0.0008	0.0234482
DNM1P41	4.52734	7.81473	0.787533	5.00E-05	0.00247621
WLS	166.811	288.069	0.788197	5.00E-05	0.00247621
CXXC5	15.3202	26.4603	0.788399	0.0001	0.00439205
ACTG1	957.936	1655.75	0.789483	0.00025	0.00957467
HS3ST4	4.13779	7.1529	0.789669	0.00015	0.00613761
CROT	26.8502	46.7144	0.798934	5.00E-05	0.00247621
RCAN1	3.72734	6.48635	0.79926	0.0008	0.0234482
HTRA1	4.79408	8.34634	0.799891	0.0009	0.0254947
POU3F2	23.5021	40.9313	0.800412	5.00E-05	0.00247621
RNF152	4.87435	8.49018	0.800584	0.00115	0.030814
MCAM	15.609	27.2554	0.804161	5.00E-05	0.00247621
CCND1	44.2194	77.2623	0.805086	5.00E-05	0.00247621
AIM1L	1.73675	3.03462	0.805123	0.00055	0.0175926
SHC3	2.53595	4.43369	0.805981	5.00E-05	0.00247621
SPATA18	4.00511	7.01014	0.8076	5.00E-05	0.00247621
LMO2	3.64415	6.40707	0.814081	0.00185	0.0422412
USP32P1	1.71116	3.01309	0.816271	0.0015	0.0365136
ME3	46.9218	82.7797	0.819016	5.00E-05	0.00247621
METRNL	11.1209	19.6637	0.822262	5.00E-05	0.00247621
TRIM24	53.9825	95.5151	0.823238	5.00E-05	0.00247621

PIM1	14.584	25.8407	0.82526	5.00E-05	0.00247621
VIM	724.295	1284.43	0.826478	0.0001	0.00439205
SOX13	13.324	23.6543	0.828079	5.00E-05	0.00247621
UPP1	11.0131	19.5881	0.830748	5.00E-05	0.00247621
IGF2R	11.3012	20.1156	0.831835	5.00E-05	0.00247621
OLFML2A	3.52778	6.29136	0.834611	0.0001	0.00439205
BAI2	11.5844	20.675	0.835702	5.00E-05	0.00247621
C21orf2	3.77593	6.7438	0.836729	0.00025	0.00957467
ANKRD1	9.57193	17.1181	0.838639	5.00E-05	0.00247621
GRM3	5.80035	10.3762	0.83907	0.0001	0.00439205
PTRF	23.5052	42.0706	0.83983	5.00E-05	0.00247621
SEPT6	10.2722	18.3858	0.839847	5.00E-05	0.00247621
MYL9	54.6795	98.4135	0.847856	5.00E-05	0.00247621
PDLIM3	4.1447	7.46303	0.848494	5.00E-05	0.00247621
CARS	12.2725	22.0987	0.84853	5.00E-05	0.00247621
EPPK1	2.19438	3.95318	0.849197	5.00E-05	0.00247621
SNCAIP	1.96327	3.54274	0.851608	0.00125	0.0322887
CYR61	95.6345	172.757	0.853144	5.00E-05	0.00247621
SHANK1	0.890123	1.60868	0.853801	0.0006	0.0189806
NR5A1	2.44144	4.41825	0.855745	0.00045	0.0152068
PTPRM	12.1302	21.9995	0.858863	5.00E-05	0.00247621
COL3A1	1.8627	3.37986	0.859571	0.00045	0.0152068
C2CD2	3.6178	6.57021	0.860828	5.00E-05	0.00247621
VAT1L	24.7792	45.0046	0.860943	5.00E-05	0.00247621
ATP1A3	1.39459	2.53325	0.861148	0.00125	0.0322887
COL12A1	0.886366	1.61243	0.86326	0.0004	0.0138096
LOXL4	3.74656	6.82621	0.865519	0.00015	0.00613761
DNAH5	0.586297	1.06826	0.865565	0.00015	0.00613761
BRSK2	5.12308	9.35138	0.868168	5.00E-05	0.00247621
ANKRD33B	7.94309	14.5434	0.872588	5.00E-05	0.00247621
TNFRSF12A	46.5718	85.6839	0.879568	5.00E-05	0.00247621
IQCE	2.07707	3.82507	0.880935	0.0002	0.00789121
DDAH1	11.0226	20.3256	0.882833	5.00E-05	0.00247621
RNF150	1.01411	1.87157	0.884032	0.00215	0.0470706
NHSL1	3.46897	6.40318	0.884281	5.00E-05	0.00247621
PWWP2B	1.88457	3.48098	0.88526	0.0012	0.0315648
NEUROD4	1.37136	2.53442	0.886052	0.002	0.0446719
EGFR	1.50875	2.79073	0.887284	0.00025	0.00957467
F3	1.68568	3.12094	0.888648	0.0022	0.0477287
LGALS3	4.86702	9.01292	0.888956	0.0017	0.0396998
TMEM132C	0.900105	1.66704	0.889119	0.0011	0.0296959
GFRA1	4.52425	8.39211	0.891356	5.00E-05	0.00247621
SFRP2	390.173	723.833	0.891543	5.00E-05	0.00247621
TTBK1	1.28711	2.39614	0.89658	0.00015	0.00613761
ABCB4	1.58677	2.95607	0.897586	0.0004	0.0138096
CSDC2	2.34614	4.37426	0.898748	0.0009	0.0254947
HDAC7	17.486	32.7974	0.907382	5.00E-05	0.00247621
LOC441204	4.12021	7.74589	0.910714	0.00075	0.0222552

CADPS	1.16682	2.19839	0.913868	0.0003	0.0110477
RBP1	15.749	29.704	0.915395	5.00E-05	0.00247621
NCAM1	14.0673	26.5531	0.916534	5.00E-05	0.00247621
PLA2G3	2.26201	4.27479	0.918244	0.00035	0.0125042
FAM110B	1.61601	3.05615	0.91928	0.00035	0.0125042
LYST	1.15357	2.18298	0.920192	5.00E-05	0.00247621
GREM1	2.05343	3.8869	0.920584	0.00025	0.00957467
ARRB1	1.37348	2.60899	0.925659	5.00E-05	0.00247621
WWC3	6.81259	12.9748	0.929435	5.00E-05	0.00247621
ARHGAP22	1.86443	3.55344	0.930481	0.00095	0.0263396
SERPINE1	6.65055	12.6931	0.932504	5.00E-05	0.00247621
TGFB1	14.1503	27.1049	0.937718	5.00E-05	0.00247621
RIN1	1.42372	2.72815	0.938258	0.0014	0.0349076
ARX	16.5212	31.6727	0.938925	5.00E-05	0.00247621
CSRP1	7.52192	14.4564	0.942541	5.00E-05	0.00247621
MKRN3	3.57023	6.88939	0.948358	5.00E-05	0.00247621
TAGLN	69.4369	134.584	0.954727	5.00E-05	0.00247621
RIMBP2	1.39251	2.70224	0.956465	0.0001	0.00439205
SORBS2	7.95063	15.4534	0.958785	5.00E-05	0.00247621
ZNF385D	2.64141	5.14958	0.963145	0.00075	0.0222552
TNFRSF19	1.54008	3.0071	0.965371	5.00E-05	0.00247621
PTX3	61.576	120.681	0.970758	5.00E-05	0.00247621
MXRA5	0.979344	1.92	0.971221	5.00E-05	0.00247621
IQCA1	1.70707	3.34832	0.971916	0.0001	0.00439205
ARMCX2	167.882	329.32	0.972041	5.00E-05	0.00247621
SCN4B	0.731154	1.43488	0.972686	0.0015	0.0365136
PLEKHA2	0.646417	1.27486	0.979804	0.00145	0.0355981
LYPD1	26.2785	51.8296	0.979894	5.00E-05	0.00247621
RAB36	1.2732	2.51149	0.980084	0.00045	0.0152068
TMEM178B	1.82142	3.60158	0.98357	5.00E-05	0.00247621
IPW	9.37312	18.5753	0.986783	5.00E-05	0.00247621
ACTA2	23.8476	47.3149	0.988449	5.00E-05	0.00247621
ITGA2	1.95161	3.87323	0.988876	5.00E-05	0.00247621
CCDC164	1.42253	2.82459	0.989584	0.0007	0.0214358
SPHK1	6.42553	12.7842	0.992475	5.00E-05	0.00247621
CACNG4	2.4366	4.84822	0.992588	5.00E-05	0.00247621
RHOB	30.9268	61.6373	0.994948	5.00E-05	0.00247621
SSUH2	1.63194	3.25754	0.997196	0.00155	0.0372883
LHFPL4	0.60565	1.20924	0.997548	0.00155	0.0372883
PODXL	32.4776	64.8608	0.997902	5.00E-05	0.00247621
TTYH2	2.72607	5.44803	0.998911	5.00E-05	0.00247621
NEDD9	5.35655	10.7123	0.999897	5.00E-05	0.00247621
CDHR1	0.699811	1.40226	1.00271	0.00055	0.0175926
FHAD1	0.368829	0.739875	1.00433	0.0022	0.0477287
CXCR7	5.38374	10.8122	1.00598	5.00E-05	0.00247621
CACNG6	2.3702	4.76099	1.00626	0.00065	0.0202941
IL11	2.99984	6.03055	1.0074	5.00E-05	0.00247621
C12orf68	1.8477	3.71478	1.00754	0.0004	0.0138096

ANKRD34B	6.06804	12.2393	1.01222	5.00E-05	0.00247621
FAM174B	2.16281	4.36247	1.01223	0.00015	0.00613761
CDK6	1.98794	4.0296	1.01936	5.00E-05	0.00247621
EMP1	8.78742	17.8388	1.02151	5.00E-05	0.00247621
AGAP2	0.755254	1.53378	1.02206	0.0004	0.0138096
PSTPIP2	1.83945	3.74579	1.02599	0.0002	0.00789121
CHST15	0.747587	1.52489	1.02839	0.001	0.0273562
ROBO3	1.71204	3.49233	1.02847	5.00E-05	0.00247621
PRDM16	5.21585	10.6618	1.03148	5.00E-05	0.00247621
C1QTNF5	0.823658	1.68411	1.03187	0.00125	0.0322887
EPB41L3	0.470497	0.965595	1.03723	0.00195	0.0439653
ABCA3	4.89505	10.0487	1.03762	5.00E-05	0.00247621
FAIM2	1.19929	2.46736	1.04078	0.00015	0.00613761
FANK1	4.0582	8.35253	1.04137	0.0003	0.0110477
P2RY2	2.56334	5.28509	1.04391	0.00015	0.00613761
CAMKV	5.14408	10.6073	1.04408	5.00E-05	0.00247621
TRPS1	0.342133	0.706574	1.04628	0.0012	0.0315648
CDH6	3.13067	6.48096	1.04974	5.00E-05	0.00247621
CTNND2	6.35545	13.1741	1.05164	5.00E-05	0.00247621
VPS37B	15.0565	31.2204	1.0521	5.00E-05	0.00247621
OLFM1	1.42032	2.94965	1.05433	0.0004	0.0138096
ANK1	1.57276	3.26828	1.05523	5.00E-05	0.00247621
BOK	1.22612	2.55238	1.05775	0.0007	0.0214358
LIFR	0.853682	1.77969	1.05986	0.0001	0.00439205
SPAG6	2.26065	4.72982	1.06505	5.00E-05	0.00247621
PLCE1	1.63666	3.43425	1.06925	5.00E-05	0.00247621
RASSF10	1.17958	2.48039	1.0723	0.0009	0.0254947
TMEM151B	1.99633	4.22046	1.08005	5.00E-05	0.00247621
CELF3	5.3234	11.2918	1.08485	5.00E-05	0.00247621
SOX1	4.93236	10.4684	1.08568	5.00E-05	0.00247621
SFXN5	1.29736	2.75396	1.08593	5.00E-05	0.00247621
PRUNE2	1.51671	3.22148	1.08678	5.00E-05	0.00247621
EDNRB	4.58862	9.83902	1.10046	5.00E-05	0.00247621
DOK6	0.374571	0.805109	1.10394	0.00045	0.0152068
COL1A1	115.07	247.366	1.10414	5.00E-05	0.00247621
DTX4	26.0553	56.0358	1.10477	5.00E-05	0.00247621
ADCY5	1.94833	4.21214	1.11232	5.00E-05	0.00247621
ROBO2	2.47006	5.34909	1.11475	5.00E-05	0.00247621
EYA2	1.48491	3.23089	1.12156	5.00E-05	0.00247621
SLC1A4	0.881163	1.92603	1.12815	5.00E-05	0.00247621
PNMA2	8.74272	19.1328	1.12989	5.00E-05	0.00247621
ATP8A2	0.478293	1.04713	1.13047	0.00075	0.0222552
PTPLA	2.14375	4.70391	1.13372	0.0009	0.0254947
FAM20C	2.23884	4.91615	1.13478	0.0001	0.00439205
SLC7A5	8.2456	18.1219	1.13604	5.00E-05	0.00247621
CAV1	0.778158	1.71312	1.13849	0.0017	0.0396998
SOCS2	1.19078	2.62467	1.14023	0.00055	0.0175926
LOC645166	4.4103	9.7219	1.14036	0.00155	0.0372883

ME1	0.516093	1.14059	1.14408	0.00215	0.0470706
SLC1A3	3.65752	8.08456	1.1443	5.00E-05	0.00247621
WNT7B	12.6323	27.9974	1.14818	5.00E-05	0.00247621
TEK	0.83986	1.86946	1.1544	5.00E-05	0.00247621
TSPAN11	5.17693	11.5247	1.15456	5.00E-05	0.00247621
NDRG2	1.85518	4.13544	1.15648	5.00E-05	0.00247621
GFRA2	0.674901	1.5046	1.15663	0.001	0.0273562
COL22A1	0.714445	1.59892	1.1622	5.00E-05	0.00247621
TMOD2	0.848072	1.89935	1.16325	5.00E-05	0.00247621
NMNAT2	0.716074	1.61776	1.17581	5.00E-05	0.00247621
KCNC1	0.714753	1.63309	1.19208	5.00E-05	0.00247621
RGS20	3.59405	8.22662	1.19469	5.00E-05	0.00247621
NEFL	1.53304	3.51713	1.198	5.00E-05	0.00247621
LGR5	2.7607	6.33822	1.19904	5.00E-05	0.00247621
PTPRN	0.720081	1.65537	1.20092	5.00E-05	0.00247621
ERBB4	0.330946	0.765421	1.20966	5.00E-05	0.00247621
ZSWIM5	0.989661	2.29425	1.21302	5.00E-05	0.00247621
CLEC18A	3.25546	7.55076	1.21376	5.00E-05	0.00247621
WBSCR17	5.75923	13.3622	1.21421	5.00E-05	0.00247621
TGFB2	1.26274	2.93204	1.21535	5.00E-05	0.00247621
TUBB3	61.3901	142.584	1.21574	5.00E-05	0.00247621
KBTBD8	0.444028	1.03234	1.2172	0.00115	0.030814
WNT8B	12.7957	29.9049	1.22472	5.00E-05	0.00247621
SYTL4	0.346589	0.811895	1.22807	0.0017	0.0396998
AK5	0.560229	1.31384	1.2297	0.0007	0.0214358
BHLHE40	0.837827	1.97293	1.23561	5.00E-05	0.00247621
KIAA0226L	0.545139	1.29531	1.2486	0.0005	0.0162466
TDRD7	1.10434	2.62561	1.24946	5.00E-05	0.00247621
ADCY2	1.96736	4.71126	1.25985	5.00E-05	0.00247621
PHLDA2	5.34235	12.8405	1.26515	0.0001	0.00439205
CAV2	0.737685	1.77597	1.26753	0.0003	0.0110477
JAM2	5.68018	13.7012	1.27029	5.00E-05	0.00247621
LIMCH1	1.84283	4.46735	1.2775	5.00E-05	0.00247621
GRIN2B	0.221609	0.537617	1.27856	0.00075	0.0222552
ITIH5	0.283741	0.688672	1.27924	0.0018	0.0412965
BARHL1	0.849325	2.06366	1.28082	0.0008	0.0234482
SCUBE3	17.1653	41.8336	1.28517	5.00E-05	0.00247621
L1CAM	1.48643	3.62327	1.28544	5.00E-05	0.00247621
ADAMTS15	0.701778	1.71507	1.28918	5.00E-05	0.00247621
SLIT3	0.212153	0.521156	1.29661	0.00015	0.00613761
TFPI2	4.05721	9.97406	1.29769	5.00E-05	0.00247621
GPRIN1	0.33012	0.813916	1.30189	0.00075	0.0222552
GDAP1L1	1.16223	2.87568	1.30701	5.00E-05	0.00247621
SLIT2	3.17612	7.86013	1.30729	5.00E-05	0.00247621
MMRN1	5.93017	14.69	1.30869	5.00E-05	0.00247621
TPT1-AS1	16.317	40.6838	1.31808	5.00E-05	0.00247621
SDK1	1.40843	3.51192	1.31817	5.00E-05	0.00247621
ITGB8	1.49223	3.72314	1.31905	5.00E-05	0.00247621

C8orf46	0.726027	1.82516	1.32993	0.0002	0.00789121
RSPO1	0.569771	1.43312	1.3307	0.0004	0.0138096
NPAS4	0.700569	1.76419	1.3324	0.0001	0.00439205
HHIPL1	0.594633	1.50184	1.33666	0.0001	0.00439205
MTUS1	0.971406	2.45852	1.33964	5.00E-05	0.00247621
AKAP3	0.30067	0.762701	1.34294	0.0022	0.0477287
CADM3	5.25872	13.3563	1.34473	5.00E-05	0.00247621
LHFP	9.66415	24.6731	1.35223	5.00E-05	0.00247621
CLSTN2	0.521588	1.34278	1.36424	5.00E-05	0.00247621
DNER	1.03894	2.67513	1.3645	5.00E-05	0.00247621
C3orf70	0.780024	2.0142	1.36862	5.00E-05	0.00247621
MIR100HG	1.75675	4.54159	1.37029	5.00E-05	0.00247621
SCN3B	0.258821	0.67047	1.37322	0.0004	0.0138096
PDE3A	1.09223	2.83841	1.37781	5.00E-05	0.00247621
PRMT8	0.917633	2.38629	1.37878	5.00E-05	0.00247621
CADM2	0.924515	2.41977	1.3881	5.00E-05	0.00247621
DPP6	2.48291	6.54305	1.39793	5.00E-05	0.00247621
MTTP	0.828702	2.18923	1.4015	5.00E-05	0.00247621
FGFR3	6.72275	17.789	1.40386	5.00E-05	0.00247621
GPR50	10.9657	29.4112	1.42336	5.00E-05	0.00247621
GAS7	2.65614	7.12967	1.4245	5.00E-05	0.00247621
CCDC19	0.504846	1.35579	1.42522	0.00125	0.0322887
LGI1	5.74771	15.4494	1.42649	5.00E-05	0.00247621
LRRN2	2.32979	6.27406	1.4292	5.00E-05	0.00247621
CLIC6	3.30952	9.01801	1.44619	5.00E-05	0.00247621
EEF1A2	1.21177	3.31664	1.4526	5.00E-05	0.00247621
EPHA3	2.15092	5.90066	1.45592	5.00E-05	0.00247621
FAM181A	6.35083	17.4549	1.45862	5.00E-05	0.00247621
ZNF536	0.26857	0.738377	1.45906	0.00015	0.00613761
TMEM255A	1.04052	2.87701	1.46727	5.00E-05	0.00247621
RHBDL3	0.551077	1.53393	1.47691	5.00E-05	0.00247621
PCDH10	0.247582	0.702065	1.5037	0.0003	0.0110477
KIAA1683	0.262731	0.751179	1.51557	0.0003	0.0110477
CDH20	1.52036	4.35044	1.51675	5.00E-05	0.00247621
ABCB1	0.55402	1.59986	1.52993	5.00E-05	0.00247621
ONECUT2	0.222127	0.644963	1.53783	5.00E-05	0.00247621
PPP1R14C	0.435354	1.26507	1.53895	0.00095	0.0263396
CCDC144B	2.12715	6.3008	1.56661	5.00E-05	0.00247621
CCDC153	1.06041	3.16643	1.57824	0.00135	0.0339557
KIFC3	1.70079	5.10407	1.58545	5.00E-05	0.00247621
OR7E156P	1.02725	3.09854	1.5928	5.00E-05	0.00247621
PGM5	0.621577	1.87658	1.5941	5.00E-05	0.00247621
PTPRH	0.418023	1.26264	1.59478	5.00E-05	0.00247621
AJAP1	0.294918	0.892154	1.59698	0.002	0.0446719
NR2E1	4.02008	12.4532	1.63123	5.00E-05	0.00247621
RSPO2	0.295894	0.92217	1.63995	0.00065	0.0202941
EMILIN2	2.23412	7.0071	1.64911	5.00E-05	0.00247621
GPM6A	5.44079	17.3987	1.67709	5.00E-05	0.00247621

NEFM	1.37322	4.40624	1.68198	5.00E-05	0.00247621
NPY5R	0.488905	1.57323	1.6861	0.0005	0.0162466
MDGA1	0.509448	1.66419	1.70781	5.00E-05	0.00247621
FBXO32	0.156844	0.513474	1.71096	5.00E-05	0.00247621
FAM106A	0.549538	1.79967	1.71144	5.00E-05	0.00247621
RELN	0.201947	0.667066	1.72385	5.00E-05	0.00247621
C15orf59	1.02322	3.38925	1.72785	5.00E-05	0.00247621
AMER2	0.143746	0.478689	1.73556	0.00225	0.048375
DCC	0.313568	1.05249	1.74696	5.00E-05	0.00247621
ADCYAP1	0.284995	0.972799	1.7712	0.0001	0.00439205
GCNT1	7.42954	25.4323	1.77532	5.00E-05	0.00247621
CDH18	0.203587	0.704981	1.79194	0.0006	0.0189806
FOXL1	0.178219	0.626164	1.81289	0.001	0.0273562
CHL1	0.23294	0.827698	1.82914	5.00E-05	0.00247621
USP32P2	2.64254	9.40473	1.83146	5.00E-05	0.00247621
SYP	0.800344	2.84953	1.83203	5.00E-05	0.00247621
NHLH1	1.37944	5.10235	1.88708	5.00E-05	0.00247621
KLHL14	0.737431	2.73034	1.8885	5.00E-05	0.00247621
TRH	1.08997	4.04204	1.89079	5.00E-05	0.00247621
DOCK8	0.23464	0.88067	1.90815	5.00E-05	0.00247621
HMP19	0.198668	0.755298	1.92668	0.00125	0.0322887
NPPB	3.57627	13.6226	1.92947	5.00E-05	0.00247621
LY6H	0.962976	3.68685	1.93681	0.0001	0.00439205
ACTC1	3.07341	11.8442	1.94627	5.00E-05	0.00247621
P2RX3	1.65462	6.48214	1.96997	5.00E-05	0.00247621
NPTX1	0.338058	1.32601	1.97175	5.00E-05	0.00247621
OLFM3	1.8381	7.23323	1.97642	5.00E-05	0.00247621
STMN2	2.37796	9.6728	2.02421	5.00E-05	0.00247621
OAF	0.875582	3.60572	2.04197	5.00E-05	0.00247621
PDE4B	0.40186	1.66502	2.05078	5.00E-05	0.00247621
INSM1	0.277946	1.15839	2.05924	0.0003	0.0110477
FGD3	0.358611	1.49461	2.05928	5.00E-05	0.00247621
SHROOM4	0.187671	0.791618	2.0766	5.00E-05	0.00247621
DGCR9	0.165019	0.698286	2.08119	0.0012	0.0315648
CPLX2	0.119623	0.50645	2.08193	0.0001	0.00439205
FAM107A	1.77424	7.94644	2.16311	5.00E-05	0.00247621
KCNA1	0.22109	1.02097	2.20723	5.00E-05	0.00247621
HRH2	0.125359	0.60898	2.28033	0.0013	0.0330453
EPB49	0.101915	0.501746	2.29959	0.0005	0.0162466
NTRK2	0.628092	3.29867	2.39284	5.00E-05	0.00247621
RSPO3	0.184896	1.02137	2.46572	0.0005	0.0162466
NIPAL4	0.130212	0.737868	2.5025	0.00015	0.00613761
LINC00461	0.208765	1.21238	2.53789	0.0001	0.00439205
SCRT1	0.0951294	0.562382	2.56359	0.001	0.0273562
EBF2	0.0951412	0.56775	2.57711	0.00025	0.00957467
TFAP2A	0.164425	1.10296	2.74588	0.0001	0.00439205
SYT4	0.130066	0.983029	2.918	5.00E-05	0.00247621
POU3F4	0.171495	1.39072	3.01959	0.00075	0.0222552

TLE2	0.109752	0.901967	3.03883	0.00035	0.0125042
VCAM1	0.834394	7.44607	3.15768	5.00E-05	0.00247621
NOS2	4.52825	41.6564	3.20151	5.00E-05	0.00247621
EOMES	0.0877691	0.812729	3.21099	0.0012	0.0315648
LEFTY2	0.580218	5.73106	3.30414	5.00E-05	0.00247621