Title
Detection and Characterization of O-GlcNAc in Embryonic Stem Cells

Author
Myers, Samuel Anthony

Publication Date
2014

Peer reviewed|Thesis/dissertation

Detection and Characterization of *O*-GlcNAc in Embryonic Stem Cells

by

Samuel Anthony Myers

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Dedication and Acknowledgments

I would like to thank my parents Jill and Joe, and my sister Emily, for instilling the values needed for this work; John Pistorino for incessantly influencing me, gratefully down this path; Brad Dawson, Chuck Leavell, Lisa Lindert and Chris Petzold for providing invaluable initial direction; Michelle Bowlin for putting me here; Megan Riel-Mehan, Dan Le, John Haliburton and Kate Lovero for keeping me here; and Al Burlingame and Barbara Panning for continuously providing the support and the freedom to explore scientifically. It was with the motivation and support from you all that made this possible.

Detection and Characterization of

O-GlcNAc in Embryonic Stem Cells

by

Samuel A. Myers

Abstract

The post-translational modification of serines and threonines of intracellular proteins by *N*-acetylglucosamine (*O*-GlcNAc) is becoming increasing implicated in processes which control embryonic stem cell self-renewal and transcriptional regulation of pluripotency. While *O*-GlcNAc and its requirement for embryonic development were described nearly 30 and 15 years ago, respectively, our molecular understanding of this dependence is severely lacking. Historically, this is due to the lack of strategies to directly detect *O*-GlcNAcylated proteins and unambiguously identify their sites of modification. The work reported here utilizes recent advancements in lectin-based *O*-GlcNAcylated peptide enrichment strategies, coupled with electron transfer dissociation mass spectrometry, to identify and characterize *O*-GlcNAc modified proteins in embryonic stem cells. This work has established that many transcriptional regulatory proteins responsible for stem cell maintenance are *O*-GlcNAc modified. One such example, SOX2, a transcription factor essential for establishing and maintaining pluripotency, possesses a complex modification occupancy pattern where the major form is *O*-GlcNAcylated. This thesis reveals that SOX2 *O*-GlcNAcylation is inhibitory for pluripotency transcriptional networks. Our results suggest a potential crosstalk between *O*-GlcNAc and phosphorylation in regulating SOX2 function, where the *O*-GlcNAc may prevent a nearby phosphorylation. Although *O*-GlcNAc signaling is essential for stem cell self-renewal we show that the connection between *O*-GlcNAc and transcriptional

regulation, especially with respect to pluripotency, is more complex than previously appreciated.

Table of Contents

Chapter 1 - On the analysis of *O*-GlcNAc and its role in Embryonic Stem Cells 1

Chapter 2 – 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 26

Chapter 3 -Polycomb Repressive Complex 2 is Necessary for the Normal Site Specific *O*-GlcNAc Distribution in Mouse Embryonic Stem Cells 58

Chapter 4 – *O*-GlcNAcylation inhibits the epigenetic priming activity of SOX2 during reprogramming 100

Chapter 5 - *O*-GlcNAc regulates SOX2 activity in Embryonic Stem Cells by altering protein-SOX2 interactions 155

Chapter 6- Characterization of *O*-GlcNAcylation of the C-terminal domain of RPB1 204

Chapter 7- Characterization of TET2 post-translational modifications 220

Appendix I - Annotated ETD mass spectra for *O*-GlcNAc modified peptides from Mouse Embryonic Stem Cells 236

Appendix II - Alkaline phosphatase staining and colony morphology of fSOX2-Tg and fS248A-Tg cells 302

List of Tables

Chapter 2

Table 2.1 Post-translationally modified Host Cell Factor C1 peptides 55
Table 2.2 *O*-GlcNAc peptides previously reported though not identified in this study 57

Chapter 3

Table 3.1 *O*-GlcNAc modified peptides identified from mESCs 94
Table 3.2 *O*-GlcNAc modified peptides and SILAC ratios between PRC2-null and WT mESCs 97
Table 3.3 Overlap of *O*-GlcNAc modified peptides identified from mESCs and other studies 99

Chapter 4

Table 4.1 SOX2 TAD peptides identified from MEFs six days after OSKM infection 153
Table 4.2 Primers used in the reprogramming study 154

Chapter 6

Table 6.1 Modified CTD peptides identified from RPB1 purified from cells 219

Chapter 7

Table 7.1 Post-translationally modified TET2 tryptic peptides 231

List of Figures and Illustrations

Chapter 1 1

Figure 1.1 Diagram of *O*-GlcNAc/phosphorylation relationships 4

Figure 1.2 MS/MS spectra of an *O*-GlcNAc modified peptide 9

Figure 1.3 Kinase pathways implicated in mouse pluripotency and self-renewal and their interaction with *O*-GlcNAc/OGT 13

Chapter 2 26

Figure 2.1 Schematic of two widely used strategies for enrichment of *O*-GlcNAc modified peptides 31

Figure 2.2 Purification and characterization of HCF-1 35

Figure 2.3 Comparison of charge state dependence on ETD efficiency for an *O*-GlcNAc modified HCF-1 peptide 37

Figure 2.4 Analysis of GluC-derived *O*-GlcNAc modified peptides 39

Chapter 3 58

Figure 3.1 ETD MS/MS spectra of two positional isomers of the *O*-GlcNAcylated SOX2 peptide 65

Figure 3.2 Western blot analysis of the relationship between PRC2 and OGT 67

Figure 3.3 SILAC LWAC LC-MS/MS workflow 68

Figure 3.4 Diagram for landmarks of HCF-1 71

Figure 3.5 Analysis of whole cell and nuclear lysates for the *O*-GlcNAc hydrolase 76

Chapter 4 100

Figure 4.1 OCT4 is not appreciably *O*-GlcNAc modified in ESCs 107

Figure 4.2 *O*-GlcNAcylation of OCT4 at T228 is undetectable 108

Figure 4.3 Characterization of SOX2 *O*-GlcNAcylation during somatic cell reprogramming 113
Figure 4.4 SOX2^{S248A} increases somatic cell reprogramming efficiency 116
Figure 4.5 Characterization of iPSCs 117
Figure 4.6 Characterization of 3xF-SOX2^{S248A} steady state levels and subcellular localization 119
Figure 4.7 Characterization of 3xF-SOX2^{S248A} activity during somatic cell reprogramming 121
Figure 4.8 Gene expression analysis of MEFs infected with *eGfp*, OS$^{FLAG-WT}$KM or OS$^{FLAG-S248A}$KM for six days 125
Figure 4.9 SOX2^{S248D} also increases somatic cell reprogramming efficiency 128

Chapter 5 155

Figure 5.1 SOX2 possesses a diverse set of PTMs in ESCs 161
Figure 5.2 Stoichiometry determination of SOX2 GlcNAc-S248 in ESCs 163
Figure 5.3 High *O*-GlcNAc stoichiometry is consistent across independently derived ESC lines 164
Figure 5.4 Removal of self-renewing cytokines alters SOX2 TAD PTM profile 167
Figure 5.5 SOX2 TAD PTM patterns differ between ESCs and breast cancer cells 168
Figure 5.6 SOX2^{S248A} mutant ESCs show altered gene expression 170
Figure 5.7 *O*-GlcNAcylation does not affect subcellular localization or steady state protein levels of SOX2 171
Figure 5.8 fS248A-Tg ESCs show minor changes in SOX2 distribution but large increases in DNA occupancy at distal regulatory regions 173
Figure 5.9 SOX2^{S248A} interacts with a different subset of proteins than SOX2^{WT} 176
Figure 5.10 fS248A-Tg cells are more refractory to self-renewal loss while simultaneously activating differentation genes 180
Figure 5.11 The S248A mutation alters the “energy landscape” of cellular identity and alters interactions with transcriptional regulatory machinery 186

Chapter 6 204

Figure 6.1 Heptad peptide repeats of the human pol II CTD 208
Figure 6.2 Identification of CTD *O*-GlcNAcylation 209
Figure 6.3 Detection of pol IIγ in cells 210
Figure 6.4 HCD spectrum of a doubly phosphorylated CTD peptide 212

Chapter 7 220

Figure 7.1 Identification of TET2 O-GlcNAcylation 224

Chapter 1

On the analysis of O-GlcNAc and its role in Embryonic Stem Cells

Introduction

Post-translational modifications (PTMs) expand the capabilities of the proteome far beyond the number of proteins encoded by the genome (1). PTMs, such as phosphorylation, glycosylation, acetylation, etc., can affect a protein's function, localization, stability and biomolecular interactions, depending on the context of the modifcation's addition or neighboring PTMs. Moreover, the crosstalk between several PTMs occurring at the same or nearby residue of a given protein is emerging as an important aspect of controlling chromatin structure and transcription regulation (2). In this work, we focus on the intracelluar glycosylation, *O*-GlcNAc, and its role in regulating pluripotency transcriptional networks.

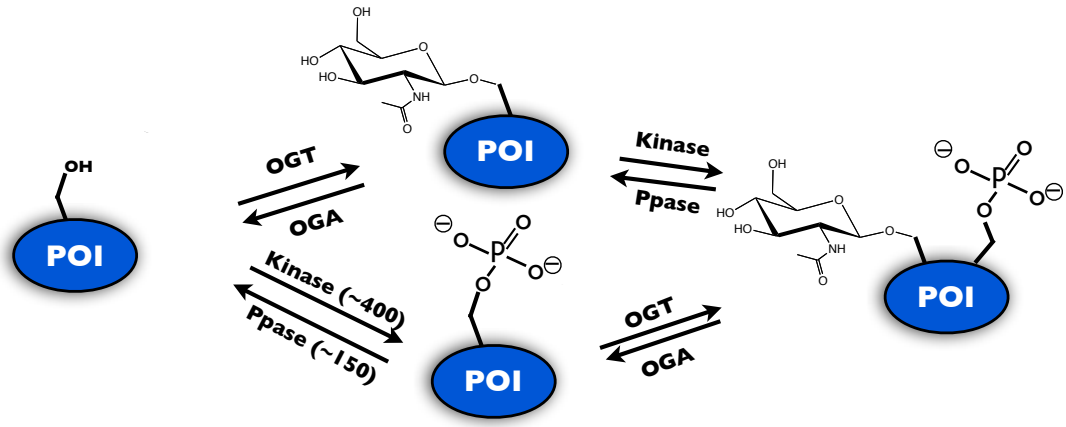
O-linked N-acetylglucosamine

O-linked N-acetylglucosamine (*O*-GlcNAc), discovered in 1984 (3), is a single β-GlcNAc moiety enzymatically coupled to serine and threonine residues of nuclear and cytosolic proteins. Unlike glycosylation of proteins destined for the membrane or for secretion, *O*-GlcNAc is a monosaccharide and is not extended to complex oligosaccharide glycans. Also, unlike "classical" glyc2osylation, *O*-GlcNAc is a dynamic and regulatory modification involved in nearly every cellular process, from cell cycle (4, 5) and transcription regulation (6-10) to synaptic plasticity (11). Because of its ubiquitous role many pathologies have been linked to aberrant *O*-GlcNAcylation such as tauopathies (12), insulin resistance and diabetes (13), cancer progression (14) and cardiovascular disease (15, 16).

O-GlcNAc signaling is analogous to S/T phosphorylation in that it is rapidly added and hydrolyzed, site-specifically, in response to various cues to affect the activity, stability, localization, etc., of substrate proteins. There is evidence of crosstalk between *O*-GlcNAcylation and phosphorylation, since most *O*-GlcNAc modified proteins are also phosphoproteins. There are also examples of the same S/T residues that can be modified by either *O*-GlcNAc or phosphate (17-19). The crosstalk is less likely due to a GlcNAc/phosphate switch at certain residues but more to cross-regulation of kinases by *O*-GlcNAc and *vice versa* (19-22). Kinases are enriched as a class of enzymes modified by *O*-GlcNAc, (19, 23) whereas the same site being a GlcNAc/phospho site occurs less frequently than expected by chance (19).

Unlike phosphorylation, which is catalyzed by over 500 kinases and hydrolyzed by about 150 phopshatases, *O*-GlcNAc is controlled by a single transferase and a single hydrolase (Figure 1). The sole enzyme responsible for catalyzing the addition of GlcNAc from uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) to intracellular proteins is *O*-GlcNAc transferase (OGT). *O*-GlcNAc is considered to be a nutrient sensing modification because the activated sugar donor for OGT is UDP-GlcNAc, which receives metabolic inputs from glutamine, acetyl-CoA, uridine nucleoside, glucose and phosphate metabolism (24). UDP-GlcNAc has the second highest concentration in cells after ATP and OGT-mediated glycosylation is the most affected by depletion of UDP-GlcNAc levels (25, 26). However, the role of OGT itself as a nutrient sensor is poorly

Figure 1 Diagram of *O*-GlcNAcylation/phosphorylation relationships. A protein of interest (POI) can exist unmodified (left most oval), *O*-GlcNAcylated (top, middle oval), phosphorylated (bottom, middle oval), or co-modified (right most oval). Chemical structures of *O*-GlcNAc and *O*-phosphate are shown. OGT, *O*-GlcNAc transferase; OGA, *O*-GlcNAc hydrolase; Ppase, phosphatase.



understood as cellular *O*-GlcNAcylation can increase with glucose deprivation and the biochemical activity of OGT is not affected by UDP-GlcNAc concentrations (27, 28).

Ogt is expressed in nearly every tissue in the body, where the highest expression is in neural tissues and embryonic stem cells (29). OGT is spliced into two distinct forms, a nucleocytosolic and mitochondrial version (30). The two splice variants are distinguished by the number of tetratricopeptide repeats (TPRs) each contains, with the nucleocytosolic variant having 11.5 and the mitochondrial version having nine. TPR domains are characterized as protein-protein interaction domains (31). Evidence suggests the TPR domains define specific sets of targets that allow OGT to bind and catalyze GlcNAc transfer (32-36). However, the mechanism for protein substrate recognition for OGT is not well understood. Unlike kinases, which can recognize protein primary sequence, OGT has little to no discernible motif or consensus sequence (19, 23). OGT also possesses a phosphoinositide-binding domain allowing the nucleocytoplasmic variant to localize to the membrane during phosphoinositide-dependent signaling (37).

The reverse reaction, or the removal of *O*-GlcNAc, is catalyzed by OGA (a.k.a. MGEA5/NCOAT/*O*-GlcNAcase). This glycoside hydrolase has two splice variants and has been shown to have histone acetyltransferase activity (38). The validity of the HAT activity has yet to be rigorously established, as the human

OGA lacks key residues for transferase function (39). While the only atomic resolution data is from bacterial homologs, several inhibitors have been developed for metazoan OGAs, each with distinct cellular phenotypes (40).

Analysis of *O*-GlcNAc

Although *O*-GlcNAcylation was discovered nearly 30 years ago, the functions of this modification remain poorly understood. *O*-GlcNAc site mapping is crucial for allowing the biochemical studies that address the functional role of protein *O*-GlcNAcylation. The lack of knowledge of the exact sites where this ubiquitous modification occurs is due mainly to the lack of techniques to readily identify sites of *O*-GlcNAc modification unambiguously.

Electrospray tandem mass spectrometry (MS/MS) has proven to be an invaluable resource for the direct detection and site-determination of many PTMs. For example, metal-affinity and antibody-based enrichment strategies, paired with liquid chromatography MS/MS (LC-MS/MS) using collision-induced/ activated dissociation (CID/CAD) have enabled the assignment of thousands of phosphorylation and ubiquitinylation sites (19, 41). These studies have contributed greatly to our understanding of signaling pathways in a wide range of organisms and tissues. In contrast, *O*-GlcNAc enrichment has proven challenging and conventional MS/MS is not amenable to *O*-GlcNAc peptide analysis.

Several antibodies against *O*-GlcNAc have been developed though their specificity remains questionable. Western blotting the same sample with more than one of the anti-*O*-GlcNAc antibodies often shows different banding patterns (42, 43). While several of these antibodies have been used to identify potential *O*-GlcNAc modified proteins, there have been no reports that unambiguously identify a site of O-GlcNAcylation when coupled with MS/MS. The lack of high-quality antibodies is thought to be due to the immune system being pre-disposed against making strong antibodies against glycoproteins, which are very abundant in blood (44, 45).

O-GlcNAc modified proteins were originally discovered via galactosyl-transferase (GalT) and tritium-labeled GlcNAc (3). GalT specifically transfers galactose (Gal) to terminal GlcNAc moieties. This method has had several adaptations over the years. In 2003, Khidekel and colleagues developed a mutant GalT (mGalT), which could transfer a ketone-containing Gal, allowing new chemistries to be exploited for detection (46), and eventually enrichment (47). This method has been particularly useful for determination of *O*-GlcNAc stoichiometry by coupling high molecular weight polyethylene glycol adducts (2-10 kDa) to the keto-Gal prior to Western blotting for the protein of interest (48, 49). Subsequently, mGalT chemoenzymatic labeling was combined with azide/ alkyne 1-3 cycloaddition (a.k.a. “click”) chemistry as an additional method to couple a biotinyl groups to GlcNAc-containing molecules (50). When employed with a photo-cleavable, biotinylated linker, *O*-GlcNAc modified peptides can be

enriched and identified by MS/MS. While this method has allowed the identification of over 100 *O*-GlcNAc sites, several drawbacks, including technical challenges and the need to remove biologically-informative protein phosphorylation, prevent its widespread utilization.

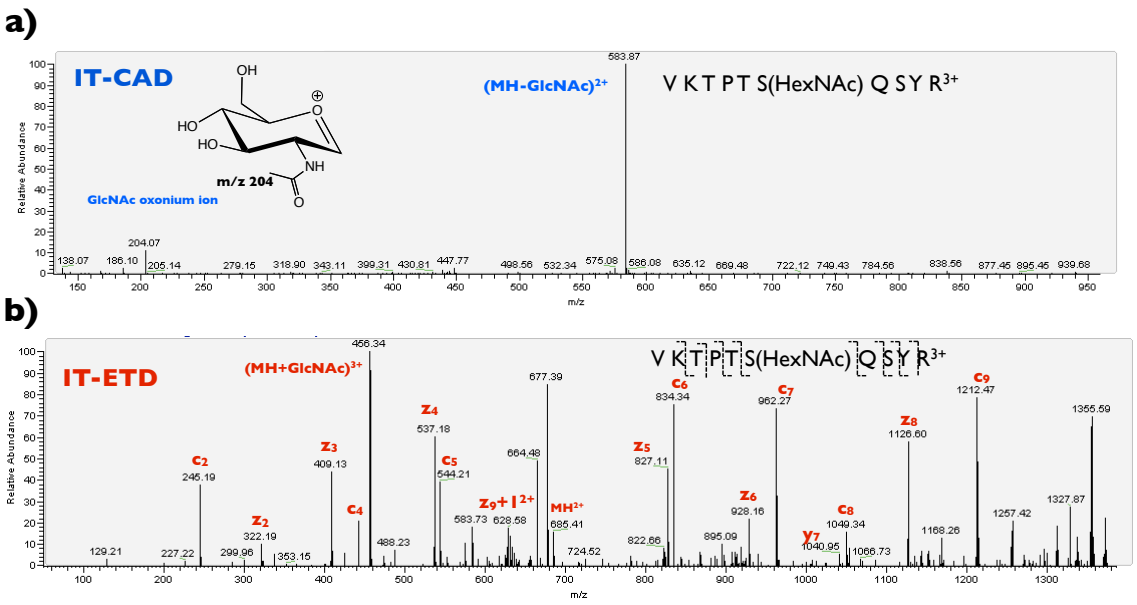
Wheat germ agglutanin (WGA) is a lectin able to bind a variety of carbohydrates including β-GlcNAc. WGA functions as a homodimer with four sugar binding sites and achieves high affinity by binding glycoproteins with multiple GlcNAc moieties. WGA has a much lower affinity for free GlcNAc or single GlcNAcylated species. While the affinity for free GlcNAc is less than desired (~10 mM) for traditional bind and elute enrichment strategies, its use in lectin affinity chromatography has proven to be useful for enriching *O*-GlcNAc modified peptides (51). Termed lectin weak affinity chromatography (LWAC), this method has allowed for the enrichment and identification of thousands of native *O*-GlcNAcylation sites from a variety of tissues and cell lines (19, 23, 52, 53).

While advancements in enrichment have made identifying *O*-GlcNAc modified proteins possible, a major hurdle in *O*-GlcNAc site assignment by mass spectrometry has been conventional peptide fragmentation modes. During ergodic MS-based peptide sequencing, as in CID/CAD in an ion trap, vibrational energy deposited into an unmodified peptide equilibrates to the weakest bond and causes unimolecular dissociation of the amide bond. In *O*-linked glycopeptides, the glycosidic bond is significantly more labile than the peptide

backbone, leading almost exclusively to the elimination of the sugar moiety and very little peptide fragmentation (Figure 2a) (54-59). Beam-type collisional activation, like that performed by a time-of-flight (TOF) mass spectrometers or in higher-energy collisional dissociation (HCD), provides secondary collisions which will allow peptide backbone fragmentation but retain very little evidence of *O*-GlcNAc site occupancy (60-63).

Electron capture and electron transfer dissociation (ECD and ETD, respectively) are radical-based, unimolecular dissociation methods that are independent of bond energies. That is, capture of a thermal electron (ECD) or an electron from a radical, anionic molecule (ETD) by the cationic peptide analyte, undergoes a chemical rearrangement that dissociates the N-Cα bond of the peptide backbone, irrespective of bond energies. This allows peptide fragmentation while retaining the labile *O*-glycosidic bond, thus allowing simultaneous identification and site-assignment of *O*-GlcNAcylated peptides (Figure 2b). The recent development of ETD for use on commercial mass spectrometers (64) have made high-throughput and unambiguous *O*-GlcNAc site assignment possible. As a result, identification of *O*-GlcNAc sites has jumped from less than 80 in 2008 to over 2000 sites in 2012 (19, 66). This information will allow researchers to elucidate specific *O*-GlcNAc function.

Figure 2 MS/MS spectra of an *O*-GlcNAc modified peptide. a) Ion trap CAD of an *O*-GlcNAc modified peptide from the protein ZFP281. The GlcNAc oxonium ion and the peptide minus the sugar loss are the predominant peaks. b) ETD mass spectrum of the same precursor as in a). Both c- and z-ion series allow peptide sequencing as well as site-assignment.



Embryonic Stem Cells

Embryonic stem cells (ESCs) derive from the inner cell mass of a developing embryo and eventually give rise to every cell of an adult organism. Understanding the regulation of this capacity to alter major changes in cellular identity will shed light on both the very early stages of development and put us closer to understanding how these undifferentiated cells might be used for therapeutic purposes.

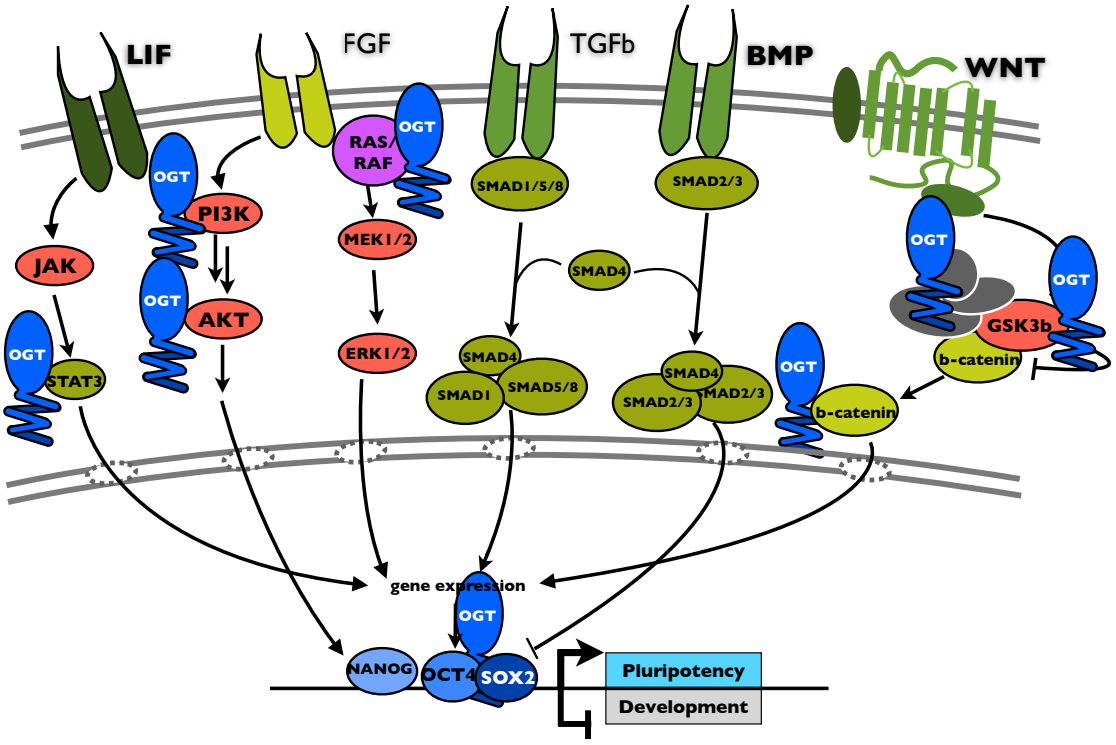
ESCs are defined by the ability to divide in an undifferentiated state in culture indefinitely. ESCs maintain a balance between remaining undifferentiated and having the ability to assume the specific transcriptional repertoire of a specialized cell. This pluripotent and self-renewing state is maintained most notably by three transcription factors: OCT4, SOX2 and NANOG. These transcription factors interact with themselves and one another, creating a core transcription factor network, allowing ESCs to maintain their stem cell characteristics, or stemness (66). These core transcription factors have also been shown to be involved in modulating chromatin structure to maintain these stem cell qualities (67-71).

O-GlcNAc signaling is essential for mouse embryo and embryonic stem cell viability (29, 72-74). OGT is involved in chromatin structure regulation, as many histone-modifying proteins either interact with OGT or are *O*-GlcNAc modified (7, 43, 52, 75, 76). Several of these histone modifiers are also known to

interact through the ESC transcription factor network to maintain stemness (68, 77, 78). *O*-GlcNAc signaling is also potentially important for promoting pluripotency and self-renewal as several studies in other cell types have shown that nodes in relevant kinase pathways are either modified by *O*-GlcNAc or interact with OGT (Figure 3) (19, 37, 52, 73, 79-81).

While the necessity of *O*-GlcNAc in ESCs has been well established, the mechanisms by which *O*-GlcNAc regulates self-renewal and pluripotency remain poorly understood. Therefore, studies which unambiguously identify sites of *O*-GlcNAc on important ESC proteins are required to understand the mechanisms by which *O*-GlcNAc regulates the pluripotency transcriptional network.

Figure 3 Kinase pathways implicated in mouse pluripotency and self-renewal and their interaction with *O*-GlcNAc/OGT. Leukemia inhibitory factor (LIF) and Bone morphogenic protein (BMP) are the main kinase pathways responsible for embryonic stem cell maintenance. WNT signalling plays an important but poorly understood role. Fibroblast growth factor (FGF) and Transforming growth factor beta (TGFb) play roles in differentiation. OGT resides in the diagram where reports of physical association or/and *O*-GlcNAc modification takes place.



References

1. Walsh C. Posttranslational Modification of Proteins: Expanding Nature's Inventory. Englewood, CO: Roberts and Company Publishers, 2006.

2. Hunter T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol Cell. 2007;28(5):730-8.

3. Torres CR, Hart GW. 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):3308-17.

4. Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. Sci Signal.3(104):ra2. PMCID: 2866299.

5. Sakabe K, Hart GW. O-GlcNAc transferase regulates mitotic chromatin dynamics. J Biol Chem. 2010;285(45):34460-8. PMCID: 2966060.

6. Goldberg HJ, Whiteside CI, Hart GW, Fantus IG. Posttranslational, reversible O-glycosylation is stimulated by high glucose and mediates plasminogen activator inhibitor-1 gene expression and Sp1 transcriptional activity in glomerular mesangial cells. Endocrinology. 2006;147(1):222-31.

7. Capotosti F, Guernier S, Lammers F, Waridel P, Cai Y, Jin J, Conaway JW, Conaway RC, Herr W. A PGC-1alpha-O-GlcNAc transferase complex regulates FoxO transcription factor activity in response to glucose. J Biol Chem. 2009;284(8):5148-57. PMCID: 2643526.

8. Yang X, Zhang F, Kudlow JE. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. Cell. 2002;110(1):69-80.

9. Capotosti F, Guernier S, Lammers F, Waridel P, Cai Y, Jin J, Conaway JW, Conaway RC, Herr W. O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. Cell. 2011;144(3):376-88.

10. Daou S, Mashtalir N, Hammond-Martel I, Pak H, Yu H, Sui G, Vogel JL, Kristie TM, Affar EB. 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):2747-52. PMCID: 3041071.

11. Tallent MK, Varghis N, Skorobogatko Y, Hernandez-Cuebas L, Whelan K, Vocadlo DJ, Vosseller K. In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. J Biol Chem. 2009;284(1):174-81.

12. Dias WB, Hart GW. O-GlcNAc modification in diabetes and Alzheimer's disease. Mol Biosyst. 2007;3(11):766-72.

13. Ma J, Hart GW. Protein O-GlcNAcylation in diabetes and diabetic complications. Expert Rev Proteomics. 2013;10(4):365-80.

14. Fardini Y, Dehennaut V, Lefebvre T, Issad T. O-GlcNAcylation: A New Cancer Hallmark? Front Endocrinol (Lausanne). 2013;4:99. PMCID: 3740238.

15. Dassanayaka S, Jones SP. O-GlcNAc and the cardiovascular system. Pharmacol Ther. 2013.

16. Lima VV, Spitler K, Choi H, Webb RC, Tostes RC. O-GlcNAcylation and oxidation of proteins: is signalling in the cardiovascular system becoming sweeter? Clin Sci (Lond). 2012;123(8):473-86. PMCID: 3389386.

17. Cheng X, Hart GW. Alternative O-glycosylation/O-phosphorylation of serine-16 in murine estrogen receptor beta: post-translational regulation of turnover and transactivation activity. J Biol Chem. 2001;276(13):10570-5.

18. Chou TY, Hart GW, Dang CV. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. J Biol Chem. 1995;270(32):18961-5.

19. Trinidad JC, Barkan DT, Gulledge BF, Thalhammer A, Sali A, Schoepfer R, Burlingame AL. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. Mol Cell Proteomics. 2012.

20. Kaasik K, Kivimae S, Allen JJ, Chalkley RJ, Huang Y, Baer K, Kissel H, Burlingame AL, Shokat KM, Ptacek LJ, Fu YH. Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock. Cell Metab. 2013;17(2):291-302. PMCID: 3597447.

21. Tarrant MK, Rho HS, Xie Z, Jiang YL, Gross C, Culhane JC, Yan G, Qian J, Ichikawa Y, Matsuoka T, Zachara N, Etzkron FA, Hart GW, Jeong JS, Blackshaw S, Zhu H, Cole PA. Regulation of CK2 by phosphorylation and O-GlcNAcylation revealed by semisynthesis. Nat Chem Biol. 2012;8(3):262-9. PMCID: 3288285.

22. Dias WB, Cheung WD, Hart GW. O-GlcNAcylation of kinases. Biochem Biophys Res Commun. 2012;422(2):224-8. PMCID: 3387735.

23. Chalkley RJ, Thalhammer A, Schoepfer R, Burlingame 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): 8894-9. PMCID: 2690010.

24. Rossetti L. Perspective: Hexosamines and nutrient sensing. Endocrinology. 2000;141(6):1922-5.

25. Marshall S, Nadeau O, Yamasaki K. Dynamic actions of glucose and glucosamine on hexosamine biosynthesis in isolated adipocytes: differential effects on glucosamine 6-phosphate, UDP-N-acetylglucosamine, and ATP levels. J Biol Chem. 2004;279(34):35313-9.

26. Boehmelt G, Wakeham A, Elia A, Sasaki T, Plyte S, Potter J, Yang Y, Tsang E, Ruland J, Iscove NN, Dennis JW, Mak TW. Decreased UDP-GlcNAc levels abrogate proliferation control in EMeg32-deficient cells. EMBO J. 2000;19(19):5092-104. PMCID: 302091.

27. Cheung WD, Hart GW. AMP-activated protein kinase and p38 MAPK activate O-GlcNAcylation of neuronal proteins during glucose deprivation. J Biol Chem. 2008;283(19):13009-20. PMCID: 2435304.

28. Taylor RP, Parker GJ, Hazel MW, Soesanto Y, Fuller W, Yazzie MJ, McClain DA. Glucose deprivation stimulates O-GlcNAc modification of proteins through up-regulation of O-linked N-acetylglucosaminyltransferase. J Biol Chem. 2008;283(10):6050-7.

29. Shafi R, Iyer SP, Ellies LG, O'Donnell N, Marek KW, Chui D, Hart GW, Marth JD. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc Natl Acad Sci U S A. 2000;97(11):5735-9. PMCID: 18502.

30. Hanover JA, Yu S, Lubas WB, Shin SH, Ragano-Caracciola M, Kochran J, Love DC. Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. Arch Biochem Biophys. 2003;409(2):287-97.

31. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci. 2003;28(12):655-62.

32. Housley MP, Rodgers JT, Udeshi ND, Kelly TJ, Shabanowitz J, Hunt DF, Puigerver P, Hart GW. O-GlcNAc regulates FoxO activation in response to glucose. J Biol Chem. 2008;283(24):16283-92. PMCID: 2423255.

33. Iyer SP, Hart GW. Roles of the tetratricopeptide repeat domain in O-GlcNAc transferase targeting and protein substrate specificity. J Biol Chem. 2003;278(27):24608-16.

34. Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E. The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. Nat Struct Mol Biol. 2004;11(10):1001-7.

35. Lazarus MB, Jiang J, Gloster TM, Zandberg WF, Whitworth GE, Vocadlo DJ, Walker S. Structural snapshots of the reaction coordinate for O-GlcNAc transferase. Nat Chem Biol. 2012;8(12):966-8. PMCID: 3508357.

36. Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. Nature. 2011;469(7331):564-7. PMCID: 3064491.

37. Yang X, Ongusaha PP, Miles PD, Havstad JC, Zhang F, So WV, Kudlow JE, Michell RH, Olefsky JM, Field SJ, Evans RM. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. Nature. 2008;451(7181):964-9.

38. Toleman C, Paterson AJ, Whisenhunt TR, Kudlow JE. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. J Biol Chem. 2004;279(51):53665-73.

39. Rao, F.V., Schuttelkopf, A.W., Dorfmueller, H.C., Ferenbach, A.T., Navratilova, I., and van Aalten, D.M. Structure of a bacterial putative acetyltransferase defines the fold of the human O-GlcNAcase C-terminal domain. Open Biol. 2013;3:130021.

40. Macauley MS, Shan X, Yuzwa SA, Gloster TM, Vocadlo DJ. Elevation of Global O-GlcNAc in rodents using a selective O-GlcNAcase inhibitor does not cause insulin resistance or perturb glucohomeostasis. Chem Biol.17(9):949-58.

41. Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods. 2013;10(7):634-7.

42. Teo CF, Ingale S, Wolfert MA, Elsayed GA, Not LG, Chatham JC, Wells L, Boon GJ. Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. Nat Chem Biol. 2010;6(5):338-43. PMCID: 2857662.

43. Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013;493(7433):561-4. PMCID: 3684361.

44. Nores GA, Lardone RD, Comin R, Alaniz ME, Moyano AL, Irazoqui FJ. Anti-GM1 antibodies as a model of the immune response to self-glycans. Biochim Biophys Acta. 2008;1780(3):538-45.

45. Doores KJ, Fulton Z, Hong V, Patel MK, Scanlan CN, Wormald MR, Finn MG, Burton DR, Wilson IA, Davis BG. A nonself sugar mimic of the HIV glycan shield shows enhanced antigenicity. Proc Natl Acad Sci U S A. 2010;107(40): 17107-12. PMCID: 2951454.

46. Khidekel N, Arndt S, Lamarre-Vincent N, Lippert A, Poulin-Kerstien KG, Ramakrishnan B, Qasba PK, Hsieh-Wilson LC. A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. J Am Chem Soc. 2003;125(52):16162-3.

47. Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. Proc Natl Acad Sci U S A. 2004;101(36):13132-7. PMCID: 516536.

48. Rexach JE, Clark PM, Mason DE, Neve RL, Peters EC, Hsieh-Wilson LC. Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. Nat Chem Biol. 2012;8(3):253-61. PMCID: 3288555.

49. Rexach JE, Rogers CJ, Yu SH, Tao J, Sun YE, Hsieh-Wilson LC. Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. Nat Chem Biol. 2010;6(9):645-51. PMCID: 2924450.

50. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW. 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):153-60. PMCID: 2808261.

51. Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA, Schoepfer R, Burlingame AL. O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. Mol Cell Proteomics. 2006;5(5):923-34.

52. Myers SA, Panning B, Burlingame AL. 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):9490-5. PMCID: 3111310.

53. Nagel AK, Schilling M, Comte-Walters S, Berkaw MN, Ball LE. 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):945-55. PMCID: 3617341.

54. Medzihradszky KF, Gillece-Castro BL, Settineri CA, Townsend RR, Masiarz FR, Burlingame AL. Structure determination of O-linked glycopeptides by tandem mass spectrometry. Biomed Environ Mass Spectrom. 1990;19(12): 777-81.

55. Darula Z, Chalkley RJ, Baker P, Burlingame AL, Medzihradszky KF. Mass spectrometric analysis, automated identification and complete annotation of O-linked glycopeptides. Eur J Mass Spectrom (Chichester, Eng). 2010;16(3):421-8. PMCID: 2963623.

56. Seipert RR, Dodds ED, Lebrilla CB. Exploiting differential dissociation chemistries of O-linked glycopeptide ions for the localization of mucin-type protein glycosylation. J Proteome Res. 2009;8(2):493-501. PMCID: 2680678.

57. Medzihradszkyaff KF, Gillece-Castroaff BL, Townsendaff RR, Burlingameaff AL, Hardyaff MR. Structural elucidation of O-linked glycopeptides by high energy collision-induced dissociation. J Am Soc Mass Spectrom. 1996;7(4):319-28.

58. Carr SA, Huddleston MJ, Bean MF. Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. Protein Sci. 1993;2(2):183-96. PMCID: 2142339.

59. Jebanathirajah J, Steen H, Roepstorff P. Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. J Am Soc Mass Spectrom. 2003;14(7):777-84.

60. Chalkley RJ, Burlingame AL. Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. J Am Soc Mass Spectrom. 2001;12(10):1106-13.

61. Chalkley RJ, Burlingame AL. Identification of novel sites of O-N-acetylglucosamine modification of serum response factor using quadrupole time-of-flight mass spectrometry. Mol Cell Proteomics. 2003;2(3):182-90.

62. Myers SA, Daou S, Affar el B, Burlingame A. 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):982-91.

63. Zhao P, Viner R, Teo CF, Boons GJ, Horn D, Wells L. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. J Proteome Res. 2011;10(9):4088-104. PMCID: 3172619.

64. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V, Coon JJ. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. J Proteome Res. 2008;7(8):3127-36. PMCID: 2601597.

65. Copeland RJ, Bullen JW, Hart GW. Cross-talk between GlcNAcylation and phosphorylation: roles in insulin resistance and glucose toxicity. Am J Physiol Endocrinol Metab. 2008;295(1):E17-28. PMCID: 3751035.

66. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122(6):947-56.

67. van den Boom V, Kooistra SM, Boesjes M, Geverts B, Houtsmuller AB, Monzen K, Komuro I, Essers J, Drenth-Diephuis LJ, Eggen BJ. UTF1 is a chromatin-associated protein involved in ES cell differentiation. J Cell Biol. 2007;178(6):913-24. PMCID: 2064617.

68. Baltus GA, Kowalski MP, Tutter AV, Kadam S. A positive regulatory role for the mSin3A-HDAC complex in pluripotency through Nanog and Sox2. J Biol Chem. 2009;284(11):6998-7006. PMCID: 2652339.

69. Ho L, Jothi R, Ronan JL, Cui K, Zhao K, Crabtree GR. An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A. 2009;106(13): 5187-91. PMCID: 2654397.

70. Engelen E, Akinci U, Bryne JC, Hou J, Gontan C, Moen M, Szumska D, Kockx C, van Ijcken W, Dekkers DH, Demmers J, Rijkers EJ, Bhattacharya S, Philipsen S, Pevny LH, Grosveld FG, Rottier RJ, Lenhard B, Poot RA. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat Genet. 2011;43(6):607-11.

71. Liber D, Domaschenz R, Holmqvist PH, Mazzarella L, Georgiou A, Leleu M, Fisher AG, Labosky PA, Dillon N. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. Cell Stem Cell. 2010;7(1):114-26.

72. O'Donnell N, Zachara NE, Hart GW, Marth JD. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic

cell function and embryo viability. Mol Cell Biol. 2004;24(4):1680-90. PMCID: 344186.

73. Jang H, Kim TW, Yoon S, Choi SY, Kang TW, Kim SY, Kwon YW, Cho EJ, Youn HD. O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell. 2012;11(1): 62-74.

74. Yang YR, Song M, Lee H, Jeon Y, Choi EJ, Jang HJ, Moon HY, Byun HY, Kim EK, Kim DH, Lee MN, Koh A, Ghim J, Choi JH, Lee-Kwon W, Kim KT, Ryu SH, Suh PG. O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. Aging Cell. 2012;11(3):439-48.

75. Ruan HB, Han X, Li MD, Singh JP, Qian K, Azarhoush S, Zhao L, Bennett AM, Samuel VT, Wu J, Yates JR 3rd, Yang X. O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1alpha stability. Cell Metab. 2012;16(2):226-37. PMCID: 3480732.

76. Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K, Pasini D. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013;49(4):645-56.

77. Fazzio TG, Huff JT, Panning B. An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. Cell. 2008;134(1):162-74.

78. Landry J, Sharov AA, Piao Y, Sharova LV, Xiao H, Southon E, Matta J, Tessarollo L, Zhang YE, Ko MS, Kuehn MR, Yamaguchi TP, Wu C. Essential role

of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. PLoS Genet. 2008;4(10):e1000241. PMCID: 2570622.

79. Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell. 2010;6(4):382-95. PMCID: 2860244.

80. Sayat R, Leber B, Grubac V, Wiltshire L, Persad S. O-GlcNAc-glycosylation of beta-catenin regulates its nuclear localization and transcriptional activity. Exp Cell Res. 2008;314(15):2774-87.

81. van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell. 2010;6(4):369-81. PMCID: 2860243.

Chapter 2

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

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

Samuel A. Myers[1], Salima Daou[2], El Bachir Affar[2] and AL Burlingame[1]

[1]Department of Pharmaceutical Chemistry, UCSF, San Francisco, CA 94158-2517

[2]Maisonneuve-Rosemont Hospital Research Center, Department of Medicine, University of Montréal, Montréal PQ H1T 2M4, Canada

Abstract The development of electron-based, unimolecular dissociation mass spectrometric methods, i.e. electron capture and electron transfer dissociation (ECD and ETD, respectively), has greatly increased the speed and reliability of labile post-translational modification (PTM) site assignment. The field of intracellular *O*-GlcNAc signaling has especially advanced with the advent of ETD mass spectrometry. Only within the last several years have proteomic-scale experiments utilizing ETD allowed the assignment of hundreds of *O*-GlcNAc sites within cells and subcellular structures. Our ability to identify and unambiguously assign the site of *O*-GlcNAc modifications using ETD is rapidly increasing our understanding of this regulatory glycosylation and its potential interaction with other PTMs. Here, we discuss the advantages of using ETD, complimented with collisional-activation mass spectrometry (CID/CAD/HCD), in a study of *O*-GlcNAc modified peptides of the extensively *O*-GlcNAcylated protein Host Cell Factor C1 (HCF-1). HCF-1 is a transcriptional co-regulator, forms a stable complex with *O*-GlcNAc transferase and is involved in control of cell cycle progression. ETD, along with higher energy collisional dissociation (HCD) mass spectrometry, was employed to assign the PTMs of the HCF-1 protein isolated from HEK293T cells. These include nineteen sites of *O*-GlcNAcylation, two sites of phosphorylation and two sites bearing dimethylarginine, and showcase the residue-specific, PTM complexity of this regulator of cell proliferation.

Introduction

While the post-translational modification (PTM) of nuclear and cytosolic proteins by *O*-linked *N*-acetylglucosamine of serine and threonine residues (*O*-GlcNAc) was discovered almost three decades ago, our knowledge of its biological functions has lagged far behind that of serine and threonine phosphorylation [1-4]. Development of robust phosphopeptide enrichment strategies and electrospray tandem mass spectrometry using collision-induced dissociation (CID/CAD) have enabled the accelerated assignment of thousands of phosphorylation sites [5]. In contrast, *O*-GlcNAc enrichment combined with conventional mass spectrometric methods have not proven to be as effective for high throughput site identification. While methods for *O*-GlcNAc peptide enrichment have improved, the major hurdle to identifying residues modified by *O*-GlcNAc with mass spectrometry has been the lability of the *O*-glycosidic bond upon collisional activation [6].

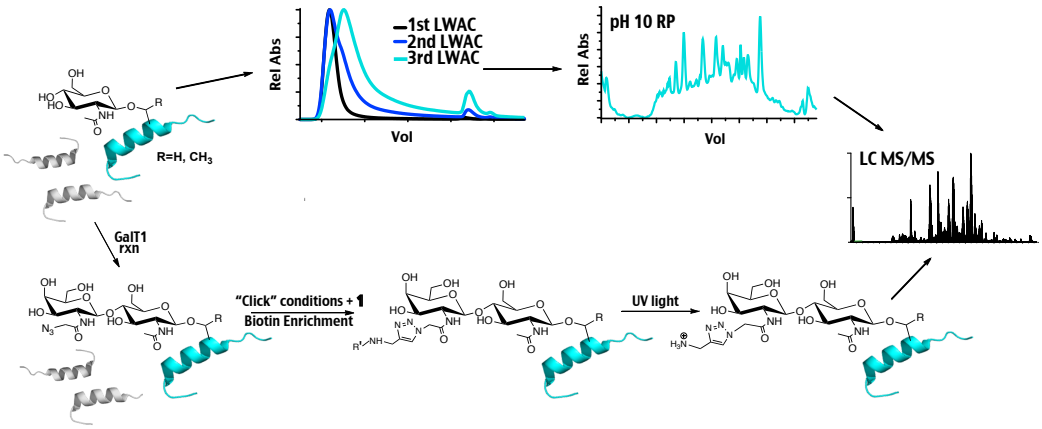
However, with the discovery of electron capture dissociation (ECD) by gas-phase polyprotonated peptide species and recognition that this internal energy deposition process was non-ergodic, it became clear that peptide backbone bond cleavage(s) occurred rapidly without vibronic energy randomization to side chain residues or to any covalently linked modifications. It was quickly shown by McLafferty and collegues that γ-carboxyglutamic acid could be observed intact, shifted appropriately in peptide sequence ion series [7].

Zubarev and colleagues then established that *O*-linked hexosamine modified peptides from mucin derived synthetic glycopeptides were stable in the sequence ion series as well [8]. ECD was soon followed by the development of the electron transfer process using energetically suitable anion-radical electron donor species [9, 10]. This new technology was then implemented on the LTQ-Orbitrap platform by Coon and co-workers, setting the stage for highly sensitive, rapid electron-transfer dissociation (ETD) analysis of complex mixtures [11]. Studies from our laboratory then showed that 58 *O*-GlcNAc modified peptides could be assigned by interpretation of ETD spectra recorded during in a single 60-minute capillary UPLC analysis coupled with an LTQ Orbitrap XL [12].

With a reliable mass spectrometry strategy for high sensitivity *O*-GlcNAc peptide site assignment in hand, there is the additional need for methodologies that will enable the enrichment of cytosolic and nuclear components bearing this sub-stoichiometric PTM. Two primary strategies are presently being developed and utilized to enrich for *O*-GlcNAc at the peptide level: one based on enzymatic and chemical derivitization of *O*-GlcNAc peptides, where the most recent iteration utilizes a cleavable, biotinylated reagent [13, 14]. The other enriches native *O*-GlcNAc modified peptides using lectin weak affinity chromatography (LWAC), specifically wheat germ agglutinin (Figure 1, [12, 15, 16]).

Overall, these recent studies have established that a wide variety of proteins, involved in regulating most cellular processes, are modified by

Figure 1 Schematic of two widely used strategies for enrichment of *O*-GlcNAc modified peptides. The unmodified (grey peptides) and *O*-GlcNAc modified peptides (blue) can be enriched using single or multiple rounds of LWAC, fractionated by high pH reverse phase and analyzed using LC MS/MS (upper arm of diagram). Alternatively, *O*-GlcNAc modified peptides can be labeled with an azide containing galactosamine, conjugated to a photocleavable-biotin alkynyl linker (Reagent 1), enriched via biotin and cleaved off solid support with UV light (lower arm of diagram). The aminotriazole-GalNAc-*O*-GlcNAc peptides are then analyzed by LC MS/MS.



O-GlcNAc. Thus far the most extensive studies have focused on assignment of *O*-GlcNAc sites in cerebrocortical brain tissue and the murine synapse. Alfaro and co-workers reported the identification of sites on 195 proteins from cerebrocortical tissue using the chemical/enzymatic photocleavable *O*-GlcNAc enrichment strategy [13]. This study also described evidence of *O*-GlcNAc on extracellular domains of membrane proteins. Recent studies from our laboratory have extended earlier work on the postsynaptic density pseudo-organelle [12] to synaptosomes and have reported over 1700 *O*-GlcNAc sites together with 16,500 phosphorylation sites from the *same* biological sample [17]. This study revealed 439 peptides containing multiple sites of *O*-GlcNAcylation. Both *O*-GlcNAc and phosphorylation sites were found to cluster preferentially on unstructured or disordered regions, but were found not to co-cluster on the same residues. Kinases were the class of proteins modified most extensively, suggesting cross talk at the level of regulation of enzyme activity [18, 19].

In addition to the assignment of *O*-GlcNAc sites at the proteomic scale now possible with ETD mass spectrometry, ETD has also greatly facilitated protein-specific studies of single *O*-GlcNAcylation events. Housley *et al.* employed ETD to identify sites of *O*-GlcNAcylation on the nutrient-responsive transcription factor FOXO1, and provided

insight to how the sole, mammalian *O*-GlcNAc transferase (OGT) targets its substrates [20]. More recently, Rexach and colleagues were able to identify the *O*-GlcNAc sites on the transcription factor CREB from analysis of a chymotryptic peptide and were able to show a single *O*-GlcNAcylation site can influence

neurite outgrowth and development of long-term memory in mice [21]. These types of site-specific *O*-GlcNAc studies will continue to contribute to our knowledge of the function(s) of protein *O*-GlcNAcylation.

One of the most remarkable and heavily *O*-GlcNAc modified proteins is Host Cell Factor C1 (HCF-1). HCF-1 is a large, chromatin associated scaffolding protein that plays an important role in cell proliferation [22-24]. HCF-1 has been repeatedly described as *O*-GlcNAcylated [12, 15, 16, 25, 26] and to form a stable complex with OGT [12, 15, 25-27, 28]. Interestingly, it was recently shown that OGT promotes the proteolytic maturation of HCF-1, which is critical for normal HCF-1 function [29, 30]. However, the complex relationship between OGT and HCF-1 is not well understood. *O*-GlcNAc modification is necessary for HCF-1 maturation, but might also be involved in regulating HCF-1 function after its proteolytic cleavage. Therefore, to understand the role of the *O*-GlcNAc modification of HCF-1, one must be able to identify the precise sites of modification so that biochemical and cell biological studies can take place. Here, we describe mass spectrometric methods used to identify the sites of *O*-GlcNAcylation and other PTM's on HCF-1 and illustrate ways to increase confidence in both glycopeptide identification and site assignment.

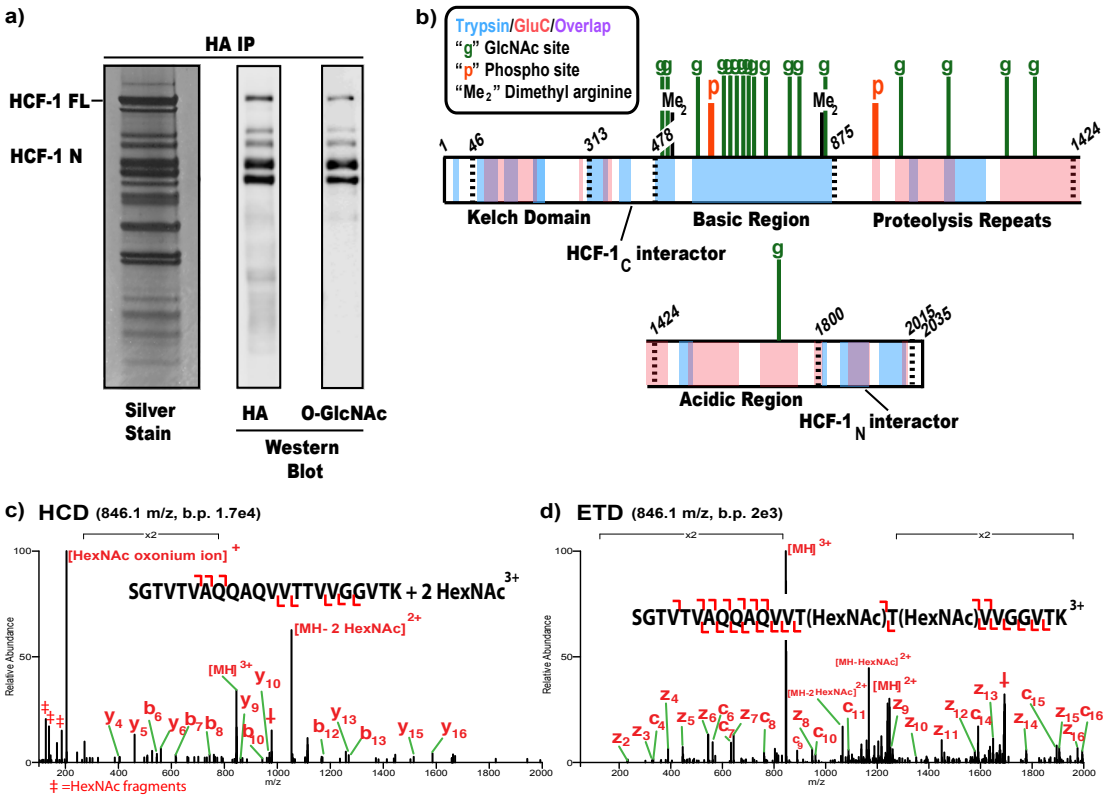
Results

HCF-1 purification yielded several polypeptides that were readily visible with silver stain and Coomassie blue. Immunodetection of *O*-GlcNAc using the RL2 antibody confirmed that HCF-1 is strongly *O*-GlcNAcylated (Figure 2a).

HCF-1 was then subjected to proteolytic digestion using both GluC and trypsin. Both peptide digests were analyzed using capillary UPLC coupled to an ESI LTQ Velos Orbitrap mass spectrometer using both HCD and ETD. The results from this combined mass spectral characterization of the trypsin and GluC digests provided nearly complete coverage of the HCF-1 protein sequence. GluC was important in providing peptides that covered the latter C-terminal portion of the protein (Figure 2b).

The same doubly *O*-GlcNAcylated tryptic peptide was analyzed using HCD and ETD energy deposition (Figure 2c and 2d). Collisional activation of this *O*-linked glycopeptide caused a prominent loss of GlcNAc due to the lability of the *O*-glycosidic bond(s) (Figure 2c). The ETD spectrum provided both c- and $z^{\cdot}$-ions series that permitted both sequence determination of the peptide and the site assignments for both *O*-GlcNAc moieties through appropriate mass shifts (203.08 Da) of product sequence ions (Figure 2d). Note that the supplemental activation required to promote efficient product ion dissociation can cause sugar

Figure 2 Purification and characterization of HCF-1. a) HA tagged HCF-1 was purified form HEK293T cells. Eluates were silver stained or western blotted against HA and *O*-GlcNAc. b) Functional domains of HCF-1 are labeled (below diagram) where the landmark residue for the domains are numbered (above). Blue regions are trypsin-derived peptides, red are GluC derived and purple was covered by both enzymes. Green "g" signifies an *O*-GlcNAc site, while orange "p" phosphorylation and black, "Me_2", arginine dimethylation. b) HCD and c) ETD MS/MS spectra for a doubly *O*-GlcNAc modified HCF-1 tryptic peptide. "b.p" designates base peak intensity.

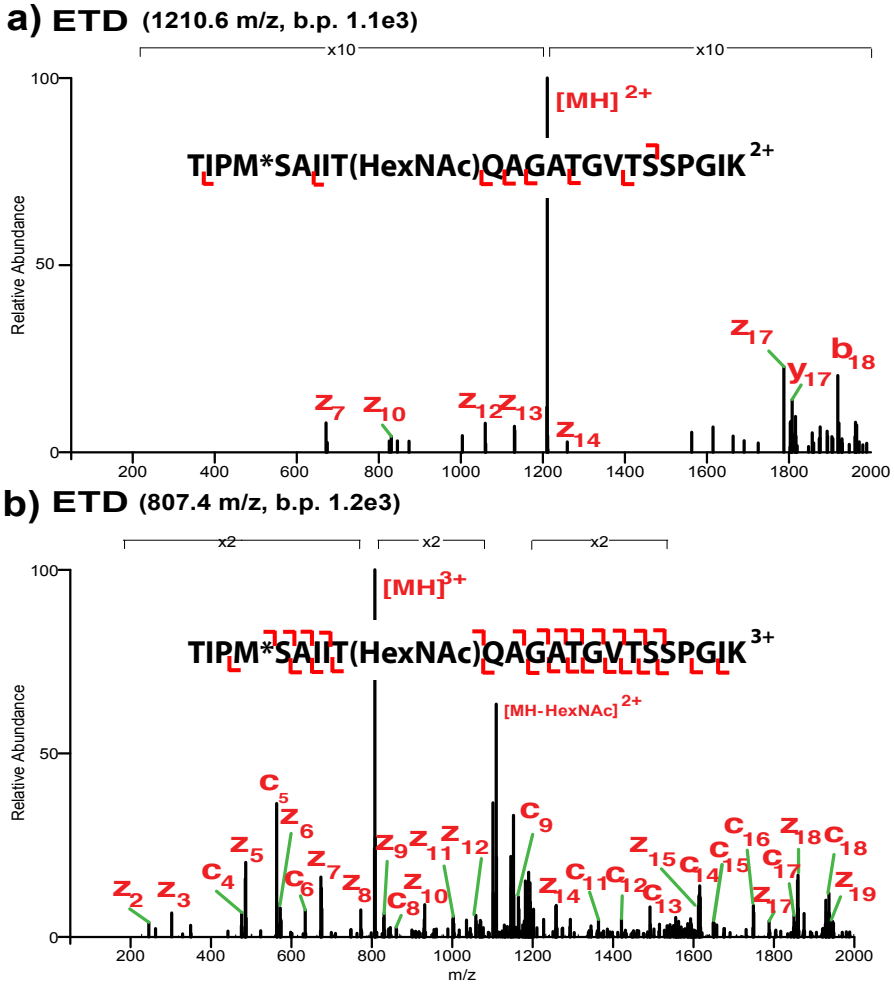


losses [34]. This example illustrates how HCD, while not providing site assigment, can be used to complement ETD in *O*-GlcNAc peptide analysis.

The efficiency of the electron transfer process for glycopeptides follows similar trends as has been described for unmodified peptides, where the greater the charge density, i.e. lower m/z, the more extensive the product ion series is obtained [35]. This phenomenon is especially relevant for glycopeptides, since the GlcNAc moiety itself adds a large, uncharged mass that acts to reduce the overall peptide charge density. This point is especially important since several commercially available bioinformatic, proteomic-searching software strategies do not adjust scoring algorithms accordingly. Without appropriate product ion scoring, some peptides remain unidentified leading to lower protein coverage than is actually present in the dataset. Protein Prospector and pFind are examples of software that adjust scoring based on charge state for ETD spectra [36-38]. The difference in ETD efficiency and resulting spectrum quality from the same doubly and triply charged ion is presented in Figure 3.

While the loss of the GlcNAc prevents reliable site assignment, there is utility in collisionally activating *O*-glycopeptides in beam-type collision instruments, especially HCD [26]. The sugar loss is detected as a singly charged GlcNAc oxonium ion at 204.087 m/z and often as GlcNAc fragments at 138.056, 144.066, 168.066 and 186.077 m/z. When measured with high mass accuracy, as done by the Orbitrap measuring HCD product ions, no amino acid composition can be mistaken for an elemental composition of 204.087m/z. The detection of

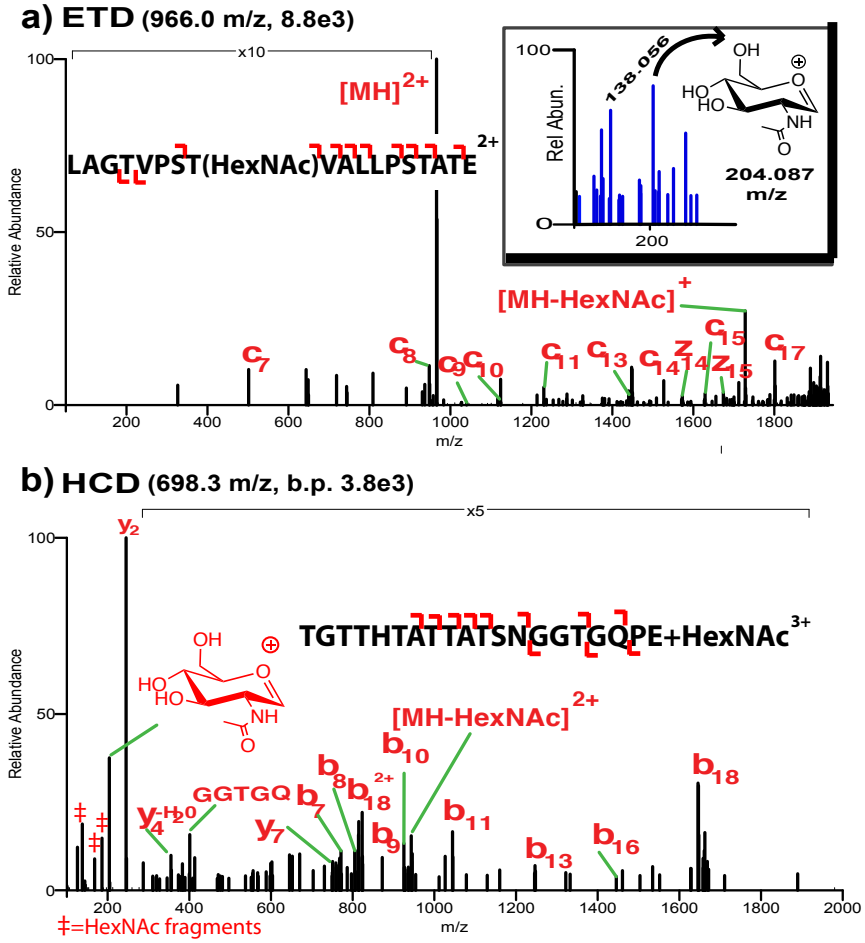
Figure 3 ETD MS/MS spectra for the HCF-1 glycopeptide containing modification site T779 comparing ETD efficacies for the a) doubly charged and the b) triply charged precursors.



GlcNAc diagnostic ions becomes useful when poor quality ETD spectra suggest an *O*-GlcNAc modified peptide. Identifying the GlcNAc ion and its fragments in HCD spectra from the same precursor of a suspect ETD *O*-GlcNAc modified peptide assignment provides certainty for the identification of an *O*-linked glycopeptide (Figure 4a). Though there is a possibility of co-eluting/co-fragmented *O*-glycopeptides that can contribute to the HexNAc oxonium ion presence, an appropriate precursor isolation window can minimize such "contaminating ions." This, again, illustrates the utility of ETD product ions retaining the HexNAc mass addition. While there are exceptions to every rule, the absence of a GlcNAc oxonium ion in collisional activation suggests a false, algorithm based, *O*-linked glycopeptide identification. Finally, another advantage of using HCD activation, that is, over ion trap collisional
activation, is beam-type collisions in the HCD cell allow for secondary collisions, which can provide peptide sequence information (Figure 4b).

Nineteen sites of *O*-GlcNAcylation on HCF-1 were identified in this study, several of which were located in the *N*-terminal region (Table 1). Four of these *O*-GlcNAc sites reside in the proteolytic processing domain (PPD) of HCF-1, two of which were found in the conserved cleavage sequence [29, 30]. Two S/T phosphorylation sites, as well as two dimethylarginine sites (one of which was also identified at singly methylated) were identified in the PPD or basic region of HCF-1 [29, 30].

Figure 4 a) GluC derived *O*-GlcNAcylated HCF-1 peptide lacking extensive z·-ions but containing c-ions allowing assignment of the modified residue. The insert, a zoom of the HCD spectrum from the same precursor, shows a prominent GlcNAc oxonium ion, providing further evidence of an O-linked hexosamine modified peptide. b) HCD spectrum of a GluC-derived glycopeptide. Sufficient b- and y-ions allow for peptide identification though no site-occupancy data is available. Prominent GlcNAc oxonium ion, and its fragments, provide evidence of O-GlcNAc modification.



Discussion

HCF-1 is one of the most highly *O*-GlcNAcylated proteins found in cells. In this study we have identified 19 sites of *O*-GlcNAcylation on HCF-1 purified from HEK293T cells, six of which are novel sites (Table 1). Nearly 30 additional HCF-1 *O*-GlcNAc sites have previously been reported from different cell or tissue types in studies enriching for *O*-GlcNAcylated peptides that were not detected in this study (Table 2). Combining the *O*-GlcNAc sites from this and other studies, including coverage of the C-terminal portion of HCF-1 using GluC digestion described here, we show that nearly all of the *O*-GlcNAc sites are located in the *N*-terminal half of the mature HCF-1 protein (Figure 2b). These *O*-GlcNAc region-specific assignments are consistent with previously published literature showing that OGT forms a stable complex with the *N*-terminal portion of HCF-1, including the PPD [28-30]. It has been proposed that *O*-GlcNAcylation of HCF-1 provides a signal required for its proteolytic maturation. Of note, OGT remains associated with HCF-1 polypeptides suggesting that these complexes may have other functions distinct from proteolytic maturation of HCF-1, such as modulation of epigenetic enzymes. HCF-1 is well known to associate with transcription regulatory proteins such as the Sin3a-HDAC complex (several components of which are *O*-GlcNAc modified) and Set/1Ash2 methyltransferase complex [6, 15, 28, 39, 40]. Thus, OGT may be recruited to HCF-1 to promote maturation, and allow the OGT-HCF-1 complex to modify transcriptional regulators.

Evidence that the OGT-HCF-1 complex recruits other transcriptional regulatory proteins is further implied by the methylation and dimethylation of arginine residues on HCF-1 (Table 1). CARM1, also known as protein arginine methyltransferase 4 (PRMT4), is known to mono- and asymmetrically dimethylate arginine residues on proteins involved in transcriptional regulation and mRNA stability [41-43]. Moreover, CARM1 interacts with, and is modified by, OGT [44]. While this study does not establish that CARM1 is responsible to HCF-1 arginine methylation, is it not unreasonable to suggest CARM1 may associate with the OGT-HCF-1 complex, and influence transcription.

The identification of phosphorylation within the heavily *O*-GlcNAcylated area of the basic region of HCF-1 lends evidence to a possible interaction between *O*-GlcNAc and phosphorylation. Both of the phosphorylated residues of HCF-1 are proline-directed, where a proline residue directly follows a phosphorylated serine or threonine. In fact, one phosphopeptide identified in this study contains the GSK3β phosphorylation motif, SXXXS, where the latter serine is phosphorylated. Inhibition of GSK3β with LiCl causes a decrease in *O*-GlcNAcylation of HCF-1 in COS7 cells [25]. Perhaps HCF-1 phosphorylation by GSK3β promotes OGT recruitment and *O*-GlcNAcylation, forming the mature OGT-HCF-1 complex as a mechanism to fine tune transcription through a wide reaching regulatory complex. It remains to be determined how these *O*-GlcNAcylation states of HCF-1, as well as the other post-translationally modified states, affect transcription and cell proliferation.

Conclusions

Electron-based unimolecular dissociation mass spectrometric methods, i.e. ECD and ETD, have revolutionized *O*-GlcNAc analytical biochemistry. Glycopeptides previously unidentifiable are now readily identifiable on a chromatographic time-scale from purified proteins, or enriched peptides mixtures. ETD mass spectrometry provided this study the means to characterize the extent of *O*-GlcNAc site occupancy of HCF-1, a heavily modified protein involved in many aspects of transcriptional regulation and cell proliferation. This emerging technology will continue to allow us to further our fledgling understanding of *O*-GlcNAc biology and its interaction with other PTM systems, namely phosphorylation and arginine methylation.

Acknowledgements

We are grateful to Winship Herr for his generous gift of HCF-1 construct and antibodies and Kati Medzihradszky and Barbara Panning for reading the manuscript. This work was supported by the Biomedical Technology Research Centers program of the NIH National Institute of General Medical Sciences, NIH NIGMS 8P41GM103481 and Howard Hughes Medical Institute. Work in E.B.A's laboratory was supported by grants from the Canadian Institutes for Health Research (CIHR) (MOP115132). E.B.A. is a scholar of the Canadian Institutes for Health Research (CIHR) and Le Fonds de la Recherche en Santé du Québec (FRSQ).

References

1. Torres CR, Hart GW: 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):3308-3317.
2. Hanover JA, Krause MW, Love DC: The hexosamine signaling pathway: O-GlcNAc cycling in feast or famine. *Biochim Biophys Acta* 2005, 1800(2): 80-95.
3. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O: Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annu Rev Biochem* 2011, 80:825-858.
4. Love DC, Krause MW, Hanover JA: O-GlcNAc cycling: emerging roles in development and epigenetics. *Semin Cell Dev Biol* 2010, 21(6):646-654.
5. Beltrao P, Albanese V, Kenner LR, Swaney DL, Burlingame A, Villen J, Lim WA, Fraser JS, Frydman J, Krogan NJ: Systematic functional prioritization of protein posttranslational modifications. *Cell* 2012, 150(2):413-425.
6. Chalkley RJ, Burlingame AL: Identification of novel sites of O-N-acetylglucosamine modification of serum response factor using quadrupole time-of-flight mass spectrometry. *Mol Cell Proteomics* 2003, 2(3):182-190.
7. Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT: Localization of labile posttranslational modifications by electron

capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* 1999, 71(19):4250-4253.

8. Mirgorodskaya E, Roepstorff P, Zubarev RA: Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem* 1999, 71(20):4431-4436.

9. Huang TY, Emory JF, O'Hair RA, McLuckey SA: Electron-transfer reagent anion formation via electrospray ionization and collision-induced dissociation. *Anal Chem* 2006, 78(21):7387-7391.

10. Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF: Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci U S A* 2005, 102(27): 9463-9468.

11. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, Stafford G, Swaney DL, Syka JE, Zabrouskov V *et al*: A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. *J Proteome Res* 2008, 7(8):3127-3136.

12. Chalkley RJ, Thalhammer A, Schoepfer R, Burlingame 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): 8894-8899.

13. Alfaro JF, Gong CX, Monroe ME, Aldrich JT, Clauss TR, Purvine SO, Wang Z, Camp DG, 2nd, Shabanowitz J, Stanley P *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):7280-7285.

14. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW: 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):153-160.

15. Myers SA, Panning B, Burlingame AL: 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):9490-9495.

16. Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA *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):923-934.

17. Trinidad JC, Barkan DT, Gulledge BF, Thalhammer A, Sali A, Schoepfer R, Burlingame AL: Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. *Mol Cell Proteomics* 2012.

18. Butkinaree C, Park K, Hart GW: O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* 2009, 1800(2):96-106.

19. Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW: Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal*, 3(104):ra2.

20. Housley MP, Udeshi ND, Rodgers JT, Shabanowitz J, Puigserver P, Hunt DF, Hart GW: A PGC-1alpha-O-GlcNAc transferase complex regulates FoxO transcription factor activity in response to glucose. *J Biol Chem* 2009, 284(8):5148-5157.

21. Rexach JE, Clark PM, Mason DE, Neve RL, Peters EC, Hsieh-Wilson LC: Dynamic O-GlcNAc modification regulates CREB-mediated gene expression and memory formation. *Nat Chem Biol* 2012, 8(3):253-261.

22. Mangone M, Myers MP, Herr W: Role of the HCF-1 basic region in sustaining cell proliferation. *PLoS One* 2010, 5(2):e9020.

23. Wysocka J, Liu Y, Kobayashi R, Herr W: Developmental and cell-cycle regulation of Caenorhabditis elegans HCF phosphorylation. *Biochemistry* 2001, 40(19):5786-5794.

24. Wysocka J, Reilly PT, Herr W: Loss of HCF-1-chromatin association precedes temperature-induced growth arrest of tsBN67 cells. *Mol Cell Biol* 2001, 21(11):3820-3829.

25. Wang Z, Pandey A, Hart GW: Dynamic interplay between O-linked N-acetylglucosaminylation and glycogen synthase kinase-3-dependent phosphorylation. *Mol Cell Proteomics* 2007, 6(8):1365-1379.

26. Zhao P, Viner R, Teo CF, Boons GJ, Horn D, Wells L: Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. *J Proteome Res* 2011, 10(9): 4088-4104.

27. Mazars R, Gonzalez-de-Peredo A, Cayrol C, Lavigne AC, Vogel JL, Ortega N, Lacroix C, Gautier V, Huet G, Ray A *et al*: The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. *J Biol Chem*, 285(18):13364-13371.

28. Wysocka J, Myers MP, Laherty CD, Eisenman RN, Herr W: 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):896-911.

29. Daou S, Mashtalir N, Hammond-Martel I, Pak H, Yu H, Sui G, Vogel JL, Kristie TM, Affar el B: 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):2747-2752.

30. Capotosti F, Guernier S, Lammers F, Waridel P, Cai Y, Jin J, Conaway JW, Conaway RC, Herr W: O-GlcNAc transferase catalyzes site-specific proteolysis of HCF-1. *Cell* 2011, 144(3):376-388.

31. Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL: A data processing pipeline for mammalian proteome dynamics studies using

stable isotope metabolic labeling. *Mol Cell Proteomics* 2011, 10(12):M111 010728.

32. Baker PR, Trinidad JC, Chalkley RJ: Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 2011, 10(7):M111 008078.

33. Medzihradszky KF: Peptide sequence analysis. *Methods Enzymol* 2005, 402:209-244.

34. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JE, Coon JJ: Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 2007, 79(2):477-485.

35. Swaney DL, McAlister GC, Coon JJ: Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 2008, 5(11):959-964.

36. Baker PR, Medzihradszky KF, Chalkley RJ: Improving software performance for peptide electron transfer dissociation data analysis by implementation of charge state- and sequence-dependent scoring. *Mol Cell Proteomics* 2010, 9(9):1795-1803.

37. Darula Z, Chalkley RJ, Lynn A, Baker PR, Medzihradszky KF: Improved identification of O-linked glycopeptides from ETD data with optimized scoring for different charge states and cleavage specificities. *Amino Acids* 2011, 41(2):321-328.

38. Sun RX, Dong MQ, Song CQ, Chi H, Yang B, Xiu LY, Tao L, Jing ZY, Liu C, Wang LH *et al*: Improved peptide identification for proteomic analysis

based on comprehensive characterization of electron transfer dissociation spectra. *J Proteome Res* 2010, 9(12):6354-6367.

39. Narayanan A, Ruyechan WT, Kristie TM: The coactivator host cell factor-1 mediates Set1 and MLL1 H3K4 trimethylation at herpesvirus immediate early promoters for initiation of infection. *Proc Natl Acad Sci U S A* 2007, 104(26):10835-10840.

40. Yokoyama A, Wang Z, Wysocka J, Sanyal M, Aufiero DJ, Kitabayashi I, Herr W, Cleary ML: Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol Cell Biol* 2004, 24(13):5639-5649.

41. Bedford MT, Clarke SG: Protein arginine methylation in mammals: who, what, and why. *Mol Cell* 2009, 33(1):1-13.

42. Hassa PO, Covic M, Bedford MT, Hottiger MO: Protein arginine methyltransferase 1 coactivates NF-kappaB-dependent gene expression synergistically with CARM1 and PARP1. *J Mol Biol* 2008, 377(3):668-678.

43. Schurter BT, Koh SS, Chen D, Bunick GJ, Harp JM, Hanson BL, Henschen-Edman A, Mackay DR, Stallcup MR, Aswad DW: Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry* 2001, 40(19):5747-5756.

44. Cheung WD, Sakabe K, Housley MP, Dias WB, Hart GW: O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J Biol Chem* 2008, 283(49):33935-33941.

Materials and Methods

Purification and preparation of HCF-1

HEK293T (293T) cells were maintained in DMEM with 5 % FBS. The cells were plated at 70 % confluency (10 X 10 cm dishes) and transfected with 7 μg of pGCN-HA-HCF-1 FL using Polyethylenimine (2 mg/ml) in serum free media. Next day, the transfected cells were plated into 15X15cm dishes. Three days post-transfection, cells were harvested for HCF-1 purification. Total cell extracts were prepared using the lysis buffer (50 mM Tris-HCl, pH 7.3; 5 mM EDTA; 300 mM Nacl; 10 mM NAF; 1% NP-40; 1 mM phenylmethylsulfonyl fluoride (PMSF); 1 mM dithiothreitol; protease inhibitors cocktail (Sigma), and 5 μM PUGNAC (TRC)). Following centrifugation at 15000 rpm/30 min, the supernatant was incubated overnight at 4 ºC with 200 μl of anti-HA-agarose beads (Sigma) (100 μl packed beads for 10 ml extract). The beads were washed several times with the lysis buffer and then transferred into a 800 μl chromatography columns (Biorad) for elution. Bound proteins were eluted with 200 μg/ml of HA peptide (sigma). The purified proteins were precipitated with cold TCA (30%), washed with cold acetone and resuspended in SDS-PAGE sample buffer for western blotting and Coomassie staining. Anti-HCF-1 antibodies, a generous gift from Winship Herr's laboratory, and the anti-*O*-GlcNAc antibody RL-2 (Santa Cruz Biotechnology) were used as previously described [29]. Following gel destaining, HCF-1 bands were cuts from the gel, reduced with 2 mM TCEP (Thermo) for 60 minutes at 56°C, alkylated with iodoacetamide (Sigma) for 30 minutes at room temperature

in the dark and digested in-gel with 1:100 trypsin or GluC (Roche). Peptides were extracted with 5% formic acid and 50% acetonitrile, concentrated using C18 Ziptips (Millipore) and vacuum centrifugation followed by LC-MS/MS analysis.

LC MS/MS analysis

Chromatography was performed on a Nanoacquity HPLC (Waters) at 600 nl/min with a BEH130 C18 75 μM ID x150mm column (Waters). A 120-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 three most highly abundant, multiply charged ions within a 3 Da isolation window for subsequent HCD and ETD. Precursor scans and HCD product ions were measured in the Orbitrap at a resolution 30,000 and 7,500, respectively. For ions measured in the Orbitrap, one microscan of 250 ms injection time was used. 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 30 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 [31]. 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. 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 and methionine oxidation. Trypsin was allowed one missed cleavage, while GluC was allowed up to two. Proteins identified from this analysis were searched again allowing for HexNAc (neutral loss for HCD data) and phosphorylation of serine and threonines, acetylation, methylation and carbamidomethylation of lysines though only the former two for arginine residues. SLIP scoring was reported for all modification site localization [32] unless the modification site was ambiguous (a false localization rate of less than 5%), then manual interpretation was employed [33].

Acknowledgements

We are grateful to Winship Herr for his generous gift of HCF-1 construct and antibodies and Kati Medzihradszky and Barbara Panning for reading the manuscript. This work was supported by the Biomedical Technology Research

Centers program of the NIH National Institute of General Medical Sciences, NIH NIGMS 8P41GM103481 and Howard Hughes Medical Institute. Work in E.B.A's laboratory was supported by grants from the Canadian Institutes for Health Research (CIHR) (MOP115132). E.B.A. is a scholar of the Canadian Institutes for Health Research (CIHR) and Le Fonds de la Recherche en Santé du Québec (FRSQ).

Table 1 Modified HCF-1 peptides identified in this study. Residue numbering refers to Uniprot accession number P51610. Within the peptide sequence, bold black residues indicate unambiguous site assignment of HexNAc moiety, unless another modification is specified. Red residues indicate the site is ambiguous between one of the colored sites. The blue and green indicate a mixture of the two, singly HexNAc modified positional isomers is seen. * indicates oxidation, and ‡ indicates lysine carbamidomethylation. Modified sites previously identified from studies reporting mass spectrometric statistics or annotated spectra are referenced, and correspond with the text. Otherwise, the identification is labeled “novel”. Peptides identified by ETD and/or HCD are designated as “ETD ID” and “HCD ID,” respectively. “Site info” refers to whether or not the dissociation method provided information on the site localization of the modification. Protein Prospector outputs for modified peptides are available for ETD (http://prospector.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=gjvopep7rv&search_name=msviewer) and HCD (http://prospector.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=vkwayutqbu&search_name=msviewer) at their respective links.

m/z	z	ppm	Peptide sequence	PTM site of HCF-1 (P51610)	Prev. Id. or Novel	ETD ID/site info	HCD ID/site info
830.096	3	1.3	V**T**GPQA**T**TGTPLVTM*RPASQAGK	T490 & T495 mix	15, 17	yes/yes	yes/no
846.211	4	5.3	TQGVPAVLKV**T**GPQATTGTPLVTM*RPASQAGK	T490	17, 19	yes/yes	no/N.A.
931.767	4	3.3	IPPSSAPTVL**S**VPAGTTIVKTM*AVTPGTTTLPATVK	S569	Novel	yes/yes	yes/no
701.124	4	3.5	TAAAQVGT**S**VSSATNTSTRPIITVHK	S620	15, 17, 25	yes/yes	yes/no
701.124	4	4.1	TAAAQVGTSV**S**SATNTSTRPIITVHK	S622	13, 15, 17, 25	yes/yes	yes/no
701.124	4	3.5	TAAAQVGTSVS**S**ATNTSTRPIITVHK	S623	13, 15, 17, 25	yes/yes	yes/no
701.124	4	2.9	TAAAQVGTSVSSA**T**NTSTRPIITVHK	T625	25	yes/yes	yes/no
751.895	4	4.3	TAAAQVGTSV**S**S**T**N**TS**TRPIITVHK	S620, ambiguous 2nd site.	15, 17	yes/yes	no/N.A.
778.428	3	5.8	SGTVTVAQQAQVV**T**TVVGGVTK	T651	13, 15, 17, 25	yes/yes	no/N.A.
846.120	3	4.2	2/13/14	T651 & T652	13, 15, 17, 25	yes/yes	no/N.A.
1135.277	3	4.8	VM*SVVQTKPVQTSAVTGQASTGPVTQIIQTK	S685	13, 17	yes/yes	yes/no
984.136	5	-1.0	LV**T**SADGKPTTIITTTQASGAGTK‡PTILGISSVSPSTTK‡PGTTTIIK	T726	19	yes/yes	no/N.A.
807.428	3	4.2	TIPM*SAII**T**QAGATGVTSSPGIK	T779	13, 15, 17, 25	yes/yes	yes/no
731.730	3	2.9	SPITIIT**T**KVM*TSGTGAPAK	T801	13, 15, 16, 17, 25	yes/yes	no/N.A.
742.452	3	1.5	LVTPV**T**VSAVKPAVTTLVVK	T861	15, 17	yes/yes	yes/no
1211.590	3	0.3	TGTTN**T**A**T**TTVVANLGGHPQPTQVQFVCDRQE	T1030 or T1031	Novel (either)	yes/yes	no/N.A.
482.590	3	1.7	RACAAG**T**PAVIR	T1143	Novel	no/N.A.	no/N.A.
925.970	2	0.01	SLQGGSPSTTVTV**T**ALE	T1273	Novel	yes/yes	no/N.A.
698.319	3	###	TGTTHTAT**T**ATSNGGTGQPE	T1335	Novel	yes/yes	yes/no
966.018	2	3.5	LAGTVPS**T**VALLPSTATE	T1743	Novel	yes/yes	no/N.A.
767.453	4	1.5	TVPM*GGV**R(Me$_2$)**LVTPV**T**VSAVKPAVTTLVVK	Dimethyl@R855;HexNAc@T861	Novel	yes/yes	no/N.A.
767.077	3	3.4	VTGPQATTGTPLVTM***R(Me)**PASQAGK	Methyl@504	Novel	yes/yes	yes/yes
771.748	3	2.1	VTGPQATTGTPLVTM***R(Me$_2$)**PASQAGK	Dimethyl@504	Novel	yes/yes	yes/yes
716.685	4	5.1	TVPM*GGV**R(Me$_2$)**LVTPVTVSAVKPAVTTLVVK	Dimethyl@855	Novel	yes/yes	no/N.A.
756.352	2	3.4	VAS**S**PVM*VSNPATR	Phospho-S598	19	yes/yes	yes/yes
977.983	2	###	AQPVHDLPVSILA**S**PTTE	Phospho-S984	Novel	yes/yes	no/N.A.

Table 2 *O*-GlcNAcylation sites identified on HCF-1 not identified in this study. Unambiguous sites of *O*-GlcNAc modification are reported from studies that report mass spectrometric statistics or annotated spectra, along with appropriate references. A) Alfaro *et al.,* 2012, B) Myers *et al.,* 2011, C) Trinidad *et al.,* 2012, D) Wang *et al.,* 2010 (Ref. 14 in text), E) Wang *et al.,* 2010 (Ref. 19 in text) and F) Zhao *et al.* 2011

O-GlcNAc HCF-1 Site	Ref.
T405	F
T480	B
T496	C
T515	B, C
S562	E
S563	C
S563 & T579	B
T579	B
T587 & T588	C
T588	C
S628	F
T640	F
T658	F
S685	E
T694	E
T739	E
T784	C
S806	C, E
T808	B, C, E
T831	C
T870	B
T1139	B
T1148	B, D
S1150	E
T1239	E
T1241	A
T1246	B

Chapter 3

Polycomb Repressive Complex 2 is Necessary for the Normal Site Specific O-GlcNAc Distribution in Mouse Embryonic Stem Cells

Polycomb Repressive Complex 2 is Necessary for the Normal Site Specific *O*-GlcNAc Distribution in Mouse Embryonic Stem Cells

Samuel A. Myers[1], Barbara Panning[2] and A. L. Burlingame[1]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th Street, Genentech Hall Suite N472B, San Francisco, CA 94158

[2]Department of Biochemistry and Biophysics, University of California San Francisco, 600 16th Street, Genentech Hall Suite S372B, San Francisco, CA 94158

Abstract

The monosaccharide addition of an *N*-acetylglucosamine to serine and threonine residues (*O*-GlcNAc) of nuclear and cytosolic proteins is emerging as a general regulator of many cellular processes including signal transduction, cell division and transcription. The sole mouse *O*-GlcNAc transferase (OGT) is essential for embryonic development. To better understand the role of OGT in mouse development we mapped sites of *O*-GlcNAcylation of nuclear proteins in mouse embryonic stem cells. Here, we unambiguously identify over 60 nuclear proteins as *O*-GlcNAcylated, several of which are crucial for mouse embryonic stem cell maintenance. Furthermore, we extend the connection between OGT and Polycomb Group (PcG) genes from flies to mammals, showing Polycomb Repressive Complex 2 is necessary to maintain normal levels of OGT and for the correct cellular distribution of *O*-GlcNAc. Together these results provide insight into how OGT may regulate transcription in early development, possibly by modifying proteins important to maintain the embryonic stem cell transcriptional repertoire.

Introduction

O-GlcNAc is an enzyme-mediated modification to serine and threonine residues on cytosolic and nuclear proteins by a single *N*-acetylglucosamine (1). Unlike membrane-bound and secreted protein glycosylation, such as *N*-linked glycans, mucin-type *O*-glycosylation or GPI anchors, *O*-GlcNAcylation is a reversible and dynamic post-translational modification (PTM). Although *O*-GlcNAcylation has been implicated in nearly every cellular process, from nutrient sensing and insulin signaling (2) to synaptic plasticity (3), our knowledge of the exact functions of this widely distributed monosaccharide PTM is still in its infancy.

O-GlcNAcylation is akin to phosphorylation - it is coupled to and hydrolyzed from specific sites on proteins to modulate their function. Hence, cycling of *O*-GlcNAc may affect the activity, stability, subcellular localization, and biomolecular interactions of modified proteins. Some of the activities of *O*-GlcNAcylation may occur as a result of crosstalk between *O*-GlcNAcylation and other PTMs – a phenomenon documented for protein phosphorylation (4, 5) and ubiquitination (6). The crosstalk with phosphorylation is particularly interesting since there is only one conserved, mammalian *O*-GlcNAc transferase (OGT), with its extensive tetratricopeptide repeat (TPR) protein-protein interaction domain, that must cooperate and/or compete with hundreds of protein kinases that recognize their substrates individually (7-9).

OGT is essential for mouse development, as *Ogt* mutant embryos die shortly after implantation (10). OGT interacts with and may modify OCT4, a transcription factor that is essential for post-implantation development (11-14). Studies in *Drosophila melanogaster* provide additional mechanistic insight into the developmental role for OGT. Mutants of the fly *Ogt* homologue, *super sex combs* (*sxc*), exhibit phenotypes similar to the loss of Polycomb group (PcG) proteins (15-17). PcG proteins assemble into two Polycomb Repressive Complexes (PRC1 and PRC2) that are necessary for developmentally regulated transcriptional silencing of many genes crucial for early embryonic patterning (18). Two lines of evidence link fly OGT to PcG proteins. First, there is significant overlap in genomic occupancy between *O*-GlcNAc and subunits of PRC2 and PRC1. It is likely that *O*-GlcNAcylation regulates a function other than PRC recruitment to target genes since PRC binding to the majority of its targets is unaffected in *sxc* mutants (16). Second, a *Drosophila* PRC1 protein, Polyhomeotic (Ph), interacts with wheat germ agglutinin (WGA), a feature of *O*-GlcNAcylated proteins (16). *O*-GlcNAcylation of the mammalian Ph homolog, PHL3 and a second mammalian PcG protein homologue, YY1, extends the relationship between OGT and PcG to mammals (19-21).

Mouse embryonic stem cells (mESCs) provide a mammalian model system for deciphering the role of OGT in early development. ESCs are pluripotent and can differentiate into each cell type in the adult and developing embryo. In addition, they can be cultured in the pluripotent state indefinitely (22).

While OGT mutant ESC lines cannot be derived (10, 23), PRC2 mutant ESCs are viable and exhibit inappropriate expression of genes that are important in later developmental stages (24). Thus PRC2 mutant ESC lines provide a system to study the interaction between OGT and PcG function.

To explore the role for *O*-GlcNAc modification in ESCs, we employed an unbiased strategy for enrichment of native *O*-GlcNAcylated nuclear peptides and used high-resolution electron transfer dissociation tandem mass spectrometry (ETD MS/MS) to map their sites of modification. Here, we report the unambiguous identification of *O*-GlcNAc modification sites on over 60 proteins, many of which have documented roles in ESC pluripotency or self-renewal. In addition, we discover that PRC2 is necessary for maintaining normal levels of OGT in ESCs. The decrease in OGT abundance in PRC2 mutant ESCs is accompanied by an alteration in the amounts and distribution of *O*-GlcNAcylation in the ESC nuclear *O*-GlcNAcylated proteome. Together these results suggest that normal OGT activity in ESCs is influenced by PRC2, providing a functional connection between the two pathways.

Results

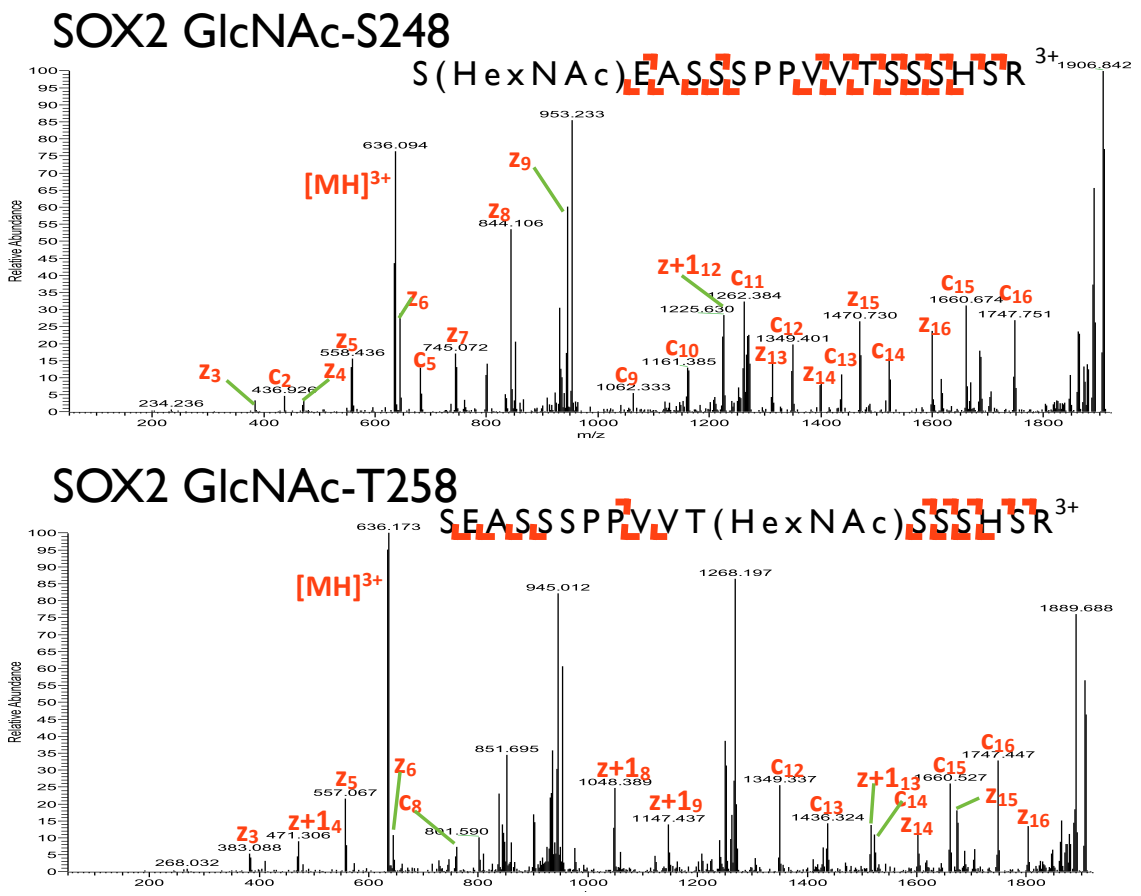
Identification of *O*-GlcNAc Modified Peptides in Mouse ESC Nuclei

A combination of lectin weak affinity chromatography (LWAC) and ETD MS/MS (19, 25, 26) was used to identify native *O*-GlcNAc modified peptides from mouse ESC nuclear proteins. One hundred forty-two *O*-GlcNAc sites were detected on 62 proteins (Appendix I and Table 1). While proteins implicated in virtually every nuclear function were found to be *O*-GlcNAc modified, the majority are involved in transcriptional regulation. For example, two transcription factors essential for ESC self-renewal, SOX2 (Figure 1) and ZFP281 (Appendix I), were modified. BPTF, which is a subunit of the histone remodeling complex NURF and is necessary for pluripotency (27, 28), was also *O*-GlcNAc modified. Sin3a, an essential transcriptional regulator in ESCs (29), and several of its interacting proteins were *O*-GlcNAcylated. For some of these targets, such as SOX2 and the Sin3a interaction protein Host Cell Factor C1 (HCF-1), *O*-GlcNAcylation was previously reported, although no specific modification sites were determined (30, 31). ETD MS/MS enabled the unambiguous site assignments for these proteins.

PRC2 is Required for Normal OGT Stability and *O*-GlcNAc Distribution

The link between OGT and PcG proteins in *D. melanogaster* prompted us to investigate whether PRC2-null ESCs exhibited any alterations in *O*-

Figure 1 ETD MS/MS spectra of the two positional isomers of the *O*-GlcNAcylated SOX2 peptide. Adequate c- and z-series ions allow the unambiguous site assignment of the *N*-acetylhexosamine (HexNAc) mass modification to serine 248 (upper panel) and threonine 258 (lower panel). Charged reduced species and common neutral losses (55) are not labeled.



GlcNAcylation patterns. Mouse ESC lines with mutations in *eed* or *suz12*, two core components of PRC2, were analyzed for levels of OCT4, a pluripotency marker, as well as OGT and *O*-GlcNAcylated proteins (Figure 2). These PRC2 mutant ESCs exhibited levels of OCT4 similar to those in wild type ESCs, indicating that they are largely undifferentiated. There was a marked drop in nuclear and whole cell OGT (Figure 2a). Consistent with a decrease in OGT levels, there was a large decrease in abundance of *O*-GlcNAcylated nuclear proteins in PRC2 ESCs (Figure 2a) while whole cell *O*-GlcNAc patterns were altered compared to wild type (Figure 2c).

We next asked whether changes in OGT levels affect PRC2 activity in ESCs. Upon knockdown of OGT in wild type ESCs, no disruption of PRC2 activity was seen, as measured by EZH2 and H3K27me3 levels (Figure 2b).

PRC2 Mutants Exhibit Altered *O*-GlcNAcylation Patterns

Western blot analysis of PRC2-null ESCs revealed changes in *O*-GlcNAcylation of nuclear proteins when compared to wild type cells. To identify specifically what *O*-GlcNAcylation changes were occurring in the absence of PRC2, we employed stable isotopic labeling of amino acids in culture (SILAC) (32) along with LWAC ETD MS/MS to measure the relative abundance of *O*-GlcNAcylated peptides between wild type and $eed^{-/-}$ nuclear proteins (Figure 3). We found 56 of the 80 *O*-GlcNAc modified peptides exhibited alterations in

Figure 2 Western blot analysis of the relationship between PRC2 and OGT. a) Disruption of PRC2 by mutations in either two of the three core components, *eed* or *suz12*, abrogates complex formation and methyltransferase activity as seen by EZH2 levels (third core component) and H3K27me3 levels, respectively. However, pluripotency marker, OCT4, is unaffected. OGT and general *O*-GlcNAc levels are altered by PRC2 disruption. b) siRNA knockdown of OGT does not significantly affect levels of OCT4, EZH2 or H3K27me3. c) Whole cell lysate analysis of H3K27me3, OGT, *O*-GlcNAc and H3 in ESC mutants for *eed* or *suz12*.

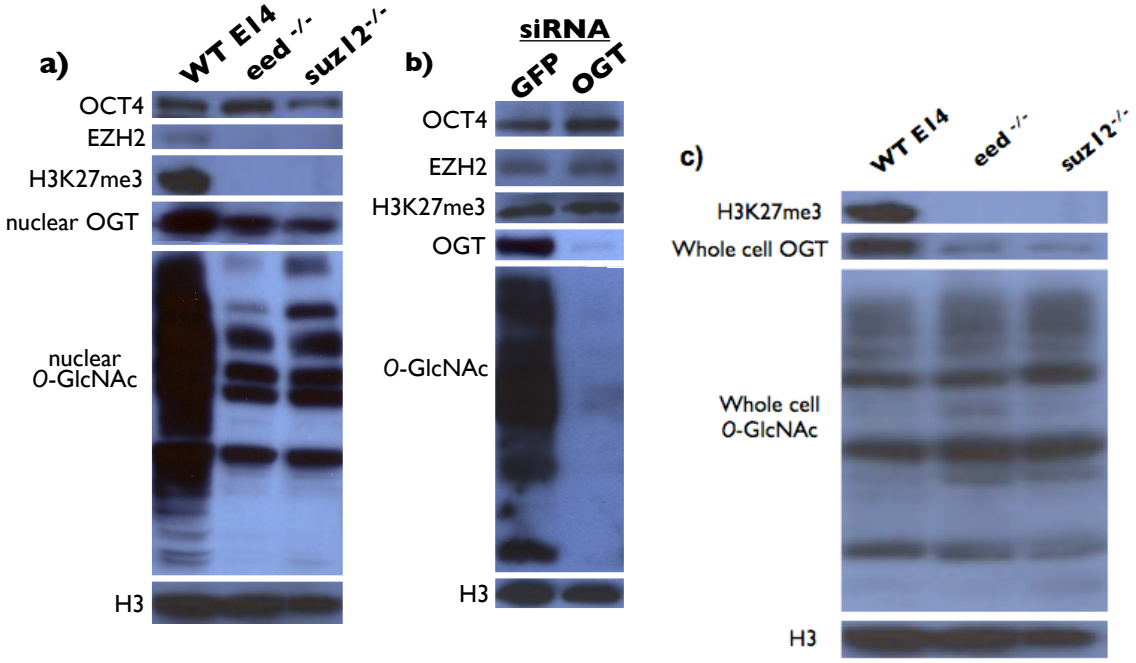
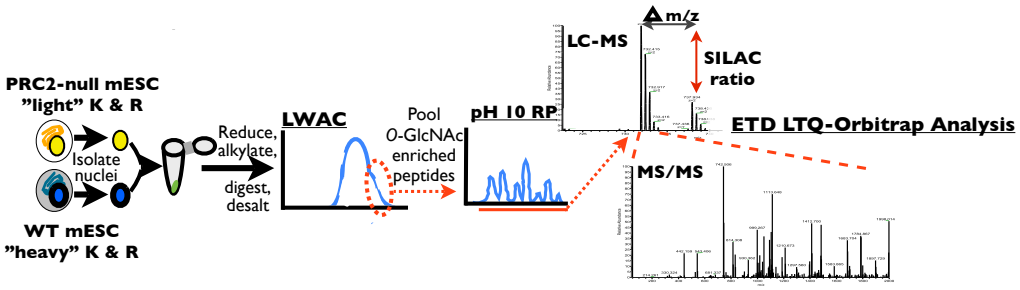


Figure 3 SILAC LWAC LC-MS/MS workflow. Non-quantitative samples were prepared similarly except a single cell type was used.



SILAC ratios greater than 25% between wild type and *eed*$^{-/-}$ mutant ESCs (Table 2). While proteins implicated in a wide set of biological functions were differentially modified in *eed*$^{-/-}$ ESCs compared to wild type ESCs, only *O*-GlcNAc modified peptides from nuclear pore complex (NPC) proteins were consistently less abundant in *eed*$^{-/-}$ ESCs (Table 2).

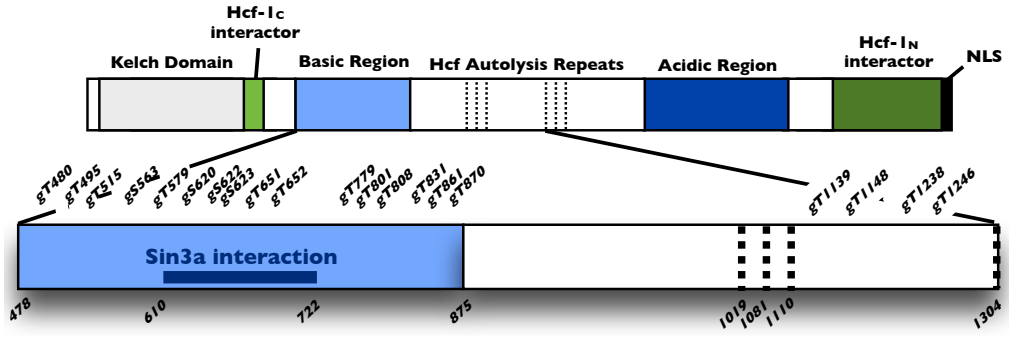
In contrast to NPC proteins, the glycopeptides identified from other proteins exhibited variable differences between wild type and *eed*$^{-/-}$ ESCs (Table 2). For example, the four different peptides derived from EMSY, a protein that interacts with BRCA2 and several chromatin-modifying proteins (33), varied greatly in their SILAC ratios. Two *O*-GlcNAcylated peptides exhibited decreased abundance in *eed*$^{-/-}$ mutant ESCs: the I518-K530 peptide *O*-GlcNAcylated on S520 (gS520) exhibited a decrease of approximately 95% and the gS200-containing T194-K204 peptide showed a decrease of roughly 50%. One glycopeptide, T498-K509 containing either gT499 or gS500, was equivalent between cell types. The fourth glycopeptide, I225-K241, exhibited 30-40% increase in relative abundance in *eed*$^{-/-}$ mutant cells if modified at T228. Thus, the loss of EED affects some sites of modification and not others, suggesting that the *O*-GlcNAcylation of different regions within the same protein can be regulated independently.

In some instances the same peptide exhibited variable SILAC ratios depending on which sites were modified. For example, the gT1710 containing T1710-K1727 BPTF peptide was equally abundant between cell lines, while the

same peptide doubly modified at two other residues, gT1713 and gT1716, was 80% less abundant in the *eed*$^{-/-}$ ESCs. Similarly, a second BPTF glycopeptide, S1750-K1761, showed variable differences depending on which modification sites were used: the gS1750 positional isomer was nearly 50% less abundant in *eed*$^{-/-}$ mutant cells, while the isomer modified at two other sites, gT1755 and gT1760, was approximately 80% less abundant. These data suggest that the *O*-GlcNAc modification of the same peptide can be regulated independently, with different positional isomers having different requirements for EED.

In other instances a decrease in abundance of one *O*-GlcNAcylated positional isomer of a peptide was accompanied by an increase in another, as is illustrated by HCF-1, which contains 20 sites of *O*-GlcNAcylation (Figure 4). In *eed*$^{-/-}$ ESCs, the triply modified T612-K637 glycopeptide containing gS620/ gS622/gS623 was almost 70% lower and the doubly modified peptide ~25% lower, whereas the singly modified gS623 containing peptide was 75% higher. These data indicate that for this HCF-1 peptide the extent of glycosylation is hampered when PRC2 function in disrupted.

Figure 4 Diagram for landmarks of HCF-1. Zoom in of the basic region and region of alternative proteolysis shows the extent and location of *O*-GlcNAc sites identified in this study. Numbers indicate landmark residues (bottom) or *O*-GlcNAc sites (top, "g" for glycosylation)**.** Adapted from ref. 44.



Discussion

Although the existence of *O*-GlcNAcylation of nucleocytoplasmic serine and threonine residues has been known for almost three decades, the technologies to precisely identify these modification sites have only been recently developed (19, 34), opening up many new lines of investigation into the function of this PTM. Our analysis of ESC nuclear *O*-GlcNAcylation revealed modifications of more than 60 proteins, involved in virtually every nuclear process, though predominantly in transcriptional regulation. In addition we find that PRC2 is necessary for normal levels of OGT itself. The decrease in OGT levels in PRC2 mutant ESCs is consistent with an altered distribution of *O*-GlcNAcylation of nuclear proteins relative to wild type ESCs. Together these results show that proper PRC2 function is necessary for normal OGT abundance and *O*-GlcNAc distribution.

SOX2 and ZFP281 May Participate in Phosphorylation/*O*-GlcNAcylation Crosstalk

SOX2, OCT4 and Nanog form the core transcription factor network for pluripotency in ESCs (35). These transcription factors interact with themselves and a variety of other transcriptional regulatory proteins to up-regulate genes involved in pluripotency and down-regulate developmental genes. SOX2 and OCT4 co-occupy and co-regulate many genes and form one arm of the self-

renewal transcription factor network (36). While there is some overlap with known SOX2/OCT4 targets, Nanog regulates a largely unique set of genes and appears to function in a separate pathway (37). ZFP281 interacts with Nanog and is also necessary for self-renewal (38, 39). Here, we have identified two sites of *O*-GlcNAcylation, gS248 and gT258, in the transactivation domain of SOX2, and three *O*-GlcNAcylation sites on ZFP281: gS691, gS889 and gT888. These results suggest that OGT may regulate both the SOX2/OCT4 and Nanog/ZFP281 pathways. Therefore, altered function of these pluripotency transcription factors may contribute to the lack of viability observed for OGT mutant ESCs.

The molecular mechanisms by which *O*-GlcNAcylation regulates transcription factor function are not fully understood. One well-documented mechanism for control of transcription factor activity is phosphorylation (40, 41), and crosstalk between these two PTMs is emerging as a potential means of regulation (5). Our findings suggest the potential for crosstalk between *O*-GlcNAcylation and phosphorylation in the modulation of SOX2 and ZFP281 function. In human ESCs, the residue homologous to S248 of SOX2, which is *O*-GlcNAcylated in mouse ESCs, is phosphorylated (42). Since S248 cannot be simultaneously phosphorylated and *O*-GlcNAcylated, this may represent an *O*-GlcNAcylation-phosphorylation switch involved in the regulation of SOX2 activity. In this study, ZFP281 *O*-GlcNAc modified at gT888 is concurrently phosphorylated at T886, while no phosphorylation is detected when ZFP281 is *O*-GlcNAcylated at S889. This result suggests that at least two populations of

modification states for ZFP281 exist within ESCs, suggesting combinatorial control by these two PTMs. Quantitative assessments of the extent of each combination of modifications and mutational analysis of PTM sites will be necessary to fully understand whether *O*-GlcNAcylation/phosphorylation crosstalk plays a role in the control of pluripotency and self-renewal by SOX2 and ZFP281.

O-GlcNAcylation of Multiple Proteins Associated with Sin3a

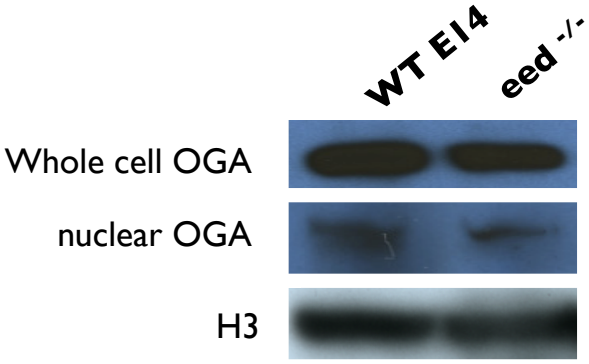
Sin3a is essential for ESC viability (29) and is a core component of several transcriptional co-repressor complexes, including the Sin3-HDAC and NCor/Sin3a complexes (43). We identified several proteins implicated in Sin3a-mediated repression as *O*-GlcNAcylated in ESCs, including RBBP7, SAP130, SAP30bp, SOX2, HCF-1, NCoR1, NCoR2 and Cnot2. There is evidence of a stable interaction between OGT and a subset of these targets, as OGT can be co-purified with Sin3a and HCF-1 (44). Moreover, Sin3a has been shown to recruit OGT to promoters to repress transcription (45). Our finding that several other Sin3a partners are also *O*-GlcNAc modified suggests that the role for this PTM in the regulation of Sin3a co-repressor complexes may be more extensive than previously appreciated. Mutational analysis will be necessary to determine the significance of *O*-GlcNAcylation of Sin3a and its interacting proteins for Sin3a-mediated transcriptional repression.

Loss of PRC2 Leads to Dysregulation of OGT and *O*-GlcNAcylation

We have shown that PRC2 mutant ESCs exhibit a notable decrease in OGT levels, while ESCs depleted for OGT exhibited normal PRC2 function. These data indicate that OGT functions downstream of PRC2 in ESCs, as is also observed in *Drosophila* (16). The decrease in OGT protein levels in PRC2 mutant ESCs is not a consequence of decreased *Ogt* transcription, as OGT mRNA levels are not significantly altered in *eed-/-* or *suz12* mutant ESCs (46, 47). It is also unlikely that PRC2 directly regulates OGT stability, since neither Suz12 nor EED form stable complexes with OGT as measured by co-immunoprecipitation (48, 49). However, it is possible that the expression of factors that affect the stability of OGT depends on PRC2.

Western blot analysis of *eed* mutant ESCs showed the drop in OGT levels coincided with decreased nuclear *O*-GlcNAcylation levels. The decrease in *O*-GlcNAcylation is not a consequence of up regulation of the sole *O*-GlcNAc hydrolase (50) (Figure 5). Our SILAC analyses revealed that not all of the *O*-GlcNAc sites were equally affected by this decrease in OGT levels. In fact, there were complex patterns of alterations in OGT target site usage in $eed^{-/-}$ mutant ESCs. For the NPC proteins, modified peptides were not as abundant in the *eed* mutant ESCs as in wild type ESCs, consistent with an overall decrease in *O*-GlcNAcylation. For other proteins, such as EMSY and HCF-1, some modified peptides were more abundant, some were less abundant, and some were

Figure 5 Analysis of whole cell and nuclear lysates for the *O*-GlcNAc hydrolase OGA/NCOAT/MGEA5 (OGA) in wild type and *eed*$^{-/-}$ mutant ESCs



unchanged in the mutant ESCs compared to wild type. For BPTF, the same peptide exhibited two patterns of *O*-GlcNAcylation, and loss of *eed* affected the two positional isomers differently. Finally, one HCF-1 peptide revealed evidence for an alteration in OGT processivity, as a decrease in abundance of triply and doubly modified glycopeptides was accompanied by an increase in abundance of the singly modified peptide (Table 2).

Our finding that certain *O*-GlcNAc sites are influenced by PRC2 while others are unaffected, along with previous studies showing the vast majority of the OGT substrates identified here are not directly, transcriptionally regulated by PRC2 (24, 51), leads us to believe that the connection between PRC2 and *O*-GlcNAc distribution is likely indirect. We speculate that the change in *O*-GlcNAcylation patterns in PRC2-null ESCs is due to transcriptional changes in proteins that influence OGT target site selection. PRC2-null ESCs show defects in transcriptional repression of proteins involved in lineage-specific differentiation (24). The aberrant transcription of these genes in PRC2-null ESCs may provide a different set of cofactors that regulate OGT specificity, resulting in inappropriate increases and decreases of *O*-GlcNAcylation at a subset of target sites. A change in the expression of cofactors that regulate OGT target site selection may also explain the decrease in OGT levels, if these factors also affect OGT stability or turnover. Protein-protein interaction studies aimed at identification of the subproteome that interacts with OGT in ESCs will help dissect the mechanism by which PRC2 influences OGT and *O*-GlcNAcylation.

This study unambiguously identifies 142 *O*-GlcNAc modification sites on 62 proteins. Most of these sites are novel (Table 3) We show that OGT modifies transcription factors and components of transcriptional regulatory complexes that are essential to maintain the ESC-specific expression profile and, therefore, are central to stem cell maintenance. In addition, this work extends the relationship between OGT and Polycomb from *Drosophila* to mammalian systems. These data provide evidence towards the nature of the relationship between PRC2 and OGT wherein OGT levels and *O*-GlcNAcylation patterns are influenced by PRC2 activity. Together these results indicate that *O*-GlcNAc is directly involved in the transcriptional regulation during early embryonic development.

Acknowledgements

We thank J. Trinidad, R. Chalkley and M. Ali for technical assistance. R. Chalkley, R. Bradshaw and T. Quan for critical review of the manuscript. B. Hamilton and J. Fistorino for discussion. This work was supported by NIH grant R01 GM085186 and NIH NCRR P41RR001614. S. Myers is supported by the NIH NIGMS T32 training grant and the Genentech Predoctoral Fellowship Program.

References

1. Torres CR & Hart GW (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J Biol Chem* 259(5): 3308-3317.
2. Yang X, Ongusaha PP, Miles PD, Havstad JC, Zhang F, So WV, Kudlow JE, Michell RH, Olefsky JM, Field SJ, Evans RM. (2008) Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. *Nature* 451(7181):964-969.
3. Tallent MK, Varghis N, Skorobogatko Y, Hernandez-Cuebas L, Whelan K, Vocadlo DJ, Vosseller K. (2009) In vivo modulation of O-GlcNAc levels regulates hippocampal synaptic plasticity through interplay with phosphorylation. *J Biol Chem* 284(1):174-181.
4. Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. (2010) Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci Signal* 3(104):ra2.
5. Hu P, Shimoji S, & Hart GW (2010) Site-specific interplay between O-GlcNAcylation and phosphorylation in cellular regulation. *FEBS Lett* 584(12):2526-2538.

6. Guinez C, Mir AM, Dehennaut V, Cacan R, Harduin-Lepers A, Michalski JC, Lefebvre T. (2008) Protein ubiquitination is modulated by O-GlcNAc glycosylation. *FASEB J* 22(8):2901-2911.

7. Butkinaree C, Park K, & Hart GW (2010) O-linked beta-N-acetylglucosamine (O-GlcNAc): Extensive crosstalk with phosphorylation to regulate signaling and transcription in response to nutrients and stress. *Biochim Biophys Acta* 1800(2):96-106.

8. Jínek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E. (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nat Struct Mol Biol* 11(10):1001-1007.

9. Clarke AJ, Hurtado-Guerrero R, Pathak S, Schüttelkopf AW, Borodkin V, Shepherd SM, Ibrahim AF, van Aalten DM. (2008) Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J* 27(20): 2780-2788.

10. Shafi R, Iyer SP, Ellies LG, O'Donnell N, Marek KW, Chui D, Hart GW, Marth JD. (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* 97(11):5735-5739.

11. Pardo M, Lang B, Yu L, Prosser H, Bradley A, Babu MM, Choudhary J. (2010) An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell* 6(4):382-395.

12. van den Berg DL, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. (2010) An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* 6(4):369-381.
13. Webster DM, Teo CF, Sun Y, Wloga D, Gay S, Klonowski KD, Wells L, Dougan ST. (2009) O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. *BMC Dev Biol* 9:28.
14. Lemischka IR (2010) Hooking up with Oct4. *Cell Stem Cell* 6(4):291-292.
15. Ingham PW (1984) A gene that regulates the bithorax complex differentially in larval and adult cells of Drosophila. *Cell* 37(3):815-823.
16. Gambetta MC, Oktaba K, & Muller J (2009) Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* 325(5936): 93-96.
17. Sinclair DA, Syrzycka M, Macauley MS, Rastgardani T, Komljenovic I, Vocadlo DJ, Brock HW, Honda BM. (2009) Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proc Natl Acad Sci U S A* 106(32):13427-13432.
18. Kerppola TK (2009) Polycomb group complexes--many combinations, many functions. *Trends Cell Biol* 19(12):692-704.
19. Chalkley RJ, Thalhammer A, Schoepfer R, & Burlingame AL (2009) Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci U S A* 106(22):8894-8899.

20. Hiromura M, Choi CH, Sabourin NA, Jones H, Bachvarov D, Usheva A. (2003) YY1 is regulated by O-linked N-acetylglucosaminylation (O-glcNAcylation). *J Biol Chem* 278(16):14046-14052.

21. Love DC, Krause MW, & Hanover JA (2010) O-GlcNAc cycling: emerging roles in development and epigenetics. *Semin Cell Dev Biol* 21(6):646-654.

22. Wray J, Kalkan T, & Smith AG (The ground state of pluripotency. *Biochem Soc Trans* 38(4):1027-1032.

23. O'Donnell N, Zachara NE, Hart GW, & Marth JD (2004) Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* 24(4):1680-1690.

24. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, Bell GW, Otte AP, Vidal M, Gifford DK, Young RA, Jaenisch R. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441(7091):349-353.

25. Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA, Schoepfer R, Burlingame AL. (2006) O-linked N-acetylglucosamine proteomics of postsynaptic density preparations using lectin weak affinity chromatography and mass spectrometry. *Mol Cell Proteomics* 5(5): 923-934.

26. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, & Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101(26):9528-9533.

27. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD. (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* 442(7098):86-90.

28. Landry J, Sharov AA, Piao Y, Sharova LV, Xiao H, Southon E, Matta J, Tessarollo L, Zhang YE, Ko MS, Kuehn MR, Yamaguchi TP, Wu C. (2008) Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genet* 4(10):e1000241.

29. Fazzio TG, Huff JT, & Panning B (2008) An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134(1):162-174.

30. Khidekel N, Ficarro SB, Peters EC, & Hsieh-Wilson LC (2004) Exploring the O-GlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci U S A* 101(36):13132-13137.

31. Mazars R, Gonzalez-de-Peredo A, Cayrol C, Lavigne AC, Vogel JL, Ortega N, Lacroix C, Gautier V, Huet G, Ray A, Monsarrat B, Kristie TM, Girard JP. The THAP-zinc finger protein THAP1 associates with coactivator HCF-1 and O-GlcNAc transferase: a link between DYT6 and DYT3 dystonias. *J Biol Chem* 285(18):13364-13371.

32. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1(5):376-386.

33. Hughes-Davies L, Huntsman D, Ruas M, Fuks F, Bye J, Chin SF, Milner J, Brown LA, Hsu F, Gilks B, Nielsen T, Schulzer M, Chia S, Ragaz J, Cahn A, Linger L, Ozdag H, Cattaneo E, Jordanova ES, Schuuring E, Yu DS, Venkitaraman A, Ponder B, Doherty A, Aparicio S, Bentley D, Theillet C, Ponting CP, Caldas C, Kouzarides T. (2003) EMSY links the BRCA2 pathway to sporadic breast and ovarian cancer. *Cell* 115(5):523-535.

34. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW. (2010) 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* 9(1):153-160.

35. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122(6):947-956.

36. Rizzino A (2009) Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. *Wiley Interdiscip Rev Syst Biol Med* 1(2):228-236.

37. Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, Barrandon O, Wray J, Yamanaka S, Chambers I, Smith A. (2009) Nanog is the gateway to the pluripotent ground state. *Cell* 138(4):722-737.

38. Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH.. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444(7117):364-368.

39. Wang ZX, Teh CH, Chan CM, Chu C, Rossbach M, Kunarso G, Allapitchay TB, Wong KY, Stanton LW.. (2008) The transcription factor Zfp281 controls embryonic stem cell pluripotency by direct activation and repression of target genes. *Stem Cells* 26(11):2791-2799.

40. Seth A, Alvarez E, Gupta S, & Davis RJ (1991) A phosphorylation site located in the NH2-terminal domain of c-Myc increases transactivation of gene expression. *J Biol Chem* 266(35):23521-23524.

41. Bode AM & Dong Z (2004) Post-translational modification of p53 in tumorigenesis. *Nat Rev Cancer* 4(10):793-805.

42. Swaney DL, Wenger CD, Thomson JA, & Coon JJ (2009) Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 106(4): 995-1000.

43. McDonel P, Costello I, & Hendrich B (2009) Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int J Biochem Cell Biol* 41(1):108-116.

44. Wysocka J, Myers MP, Laherty CD, Eisenman RN, & Herr W (2003) 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* 17(7):896-911.

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

46. Pasini D, Bracken AP, Hansen JB, Capillo M, & Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol Cell Biol* 27(10):3769-3779.

47. Schoeftner S, Sengupta AK, Kubicek S, Mechtler K, Spahn L, Koseki H, Jenuwein T, Wutz A. (2006) Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J* 25(13): 3110-3122.

48. Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139(7):1290-1302.

49. Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464(7286):306-310.

50. Love DC & Hanover JA (2005) The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci STKE* 2005(312):re13.

51. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, Koseki H, Fuchikami T, Abe K, Murray HL, Zucker JP, Yuan B, Bell GW, Herbolsheimer E, Hannett NM, Sun K, Odom DT, Otte AP, Volkert TL, Bartel DP, Melton DA, Gifford DK, Jaenisch R, Young RA. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125(2):301-313.

52. Fujimura Y, Isono K, Vidal M, Endoh M, Kajita H, Mizutani-Koseki Y, Takihara Y, van Lohuizen M, Otte A, Jenuwein T, Deschamps J, Koseki H.. (2006) Distinct roles of Polycomb group gene products in transcriptionally repressed and active domains of Hoxb8. *Development* 133(12): 2371-2381.

53. Montgomery ND, Yee D, Chen A, Kalantry S, Chamberlain SJ, Otte AP, Magnuson T.. (2005) The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Curr Biol* 15(10):942-947.

54. Kittler R, Surendranath V, Heninger AK, Slabicki M, Theis M, Putz G, Franke K, Caldarelli A, Grabner H, Kozak K, Wagner J, Rees E, Korn B, Frenzel C, Sachse C, Sönnichsen B, Guo J, Schelter J, Burchard J, Linsley PS, Jackson AL, Habermann B, Buchholz F. (2007) Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nat Methods* 4(4):337-344.

55. Fälth M, Savitski MM, Nielsen ML, Kjeldsen F, Andren PE, Zubarev RA.. (2008) Analytical utility of small neutral losses from reduced species in

electron capture dissociation studied using SwedECD database. *Anal Chem* 80(21):8089-8094.

Materials and Methods

Sample Preparation: Wild type E14 ESCs were grown without feeders under standard conditions. For SILAC experiments wild type ESCs were grown in $^{13}C_6$ arginine and $^{13}C_6$, $^{15}N_2$ lysine supplemented with 200mM proline (Invitrogen) to avoid arginine to proline conversion. *suz12*$^{-/-}$ (52) and *eed*$^{-/-}$ (53) ESC lines were grown on γ-irradiated mouse embryonic fibroblast feeder cells until the final passage where the feeders were depleted twice by preferential adherence to gelatinized culture dishes.

Cells were harvested, washed twice with cold PBS and combined 1:1 after counting with a haemocytomer. Combined cells were resuspended in nuclear prep buffer I (320 mM sucrose, 10 mM Tris pH 8.0, 3 mM $CaCl_2$, 2 mM $Mg(OAc)_2$, 0.1 mM EDTA, 0.1%, TritonX-100, 20 μM Thiamet G [Caymen Chemicals], Roche protease inhibitors and 1X Invitrogen phosphatase inhibitors) and dounce homogenized on ice until >95% of nuclei stained by Trypan blue. Two volumes of nuclear prep buffer II (2.0M Sucrose, Tris pH 8.0, 5mM $Mg(OAc)_2$, 5mM DTT, 20μM Thiamet G [Caymen Chemicals], Roche protease inhibitors and Invitrogen phosphatase inhibitors) were added to the nuclei suspension. Nuclei were pelleted by ultracentrifugation at 37,000 rpm at 4°C for 45 minutes. Pelleted nuclei were washed with cold PBS and stored at -80°C.

Nuclei, containing ~3 mg of protein, were resuspended in 6 M guanidine hydrochloride, 25mM ammonium bicarbonate, 20 μM Thiamet G [Caymen Chemicals], Roche protease inhibitors and Invitrogen phosphatase inhibitors),

sonicated and cleared by centrifugation. Soluble proteins were reduced using 2 mM tris(2-carboxyethyl)phosphine (TCEP, Thermo) for one hour at 55°C, alkylated using 10 mM iodoacetamide (IAM, Sigma) before digestion for 18 hours using modified trypsin at 37°C (Promega: Part# 9PIV5113). After digestion, the sample was acidified with formic acid and desalted using a C18 Sep-Pak (Waters), before drying down by vacuum centrifugation.

***O*-GlcNAcylated Peptide Chromatography-LWAC & high pH RP:** *O*-GlcNAc modified peptides were enriched as previously described (19). The *O*-GlcNAc enriched fractions of the chromatograph were pooled, desalted and dried. The *O*-GlcNAc enriched pool was resuspended in 2 mM NH_4OH pH 10 and 5% acetonitrile and separated on a Phenomenex Gemini® 3μm C18 reverse-phase column. The gradient of buffer A (2 mM NH_4OH and 5% acetonitrile) to buffer B (2 mM NH_4OH and 50% acetonitrile) increased from 0 to 100% over six mL at 50μL/ min. Twenty-four 250μL fractions were dried by vacuum centrifugation and resuspended in 0.1% formic acid for subsequent MS analysis.

LC-MS/MS LC-MS/MS of *O*-GlcNAcylated peptides was performed as previously described (19).

Data Analysis Fragment mass spectra were converted into peaklists using the in-house software PAVA. CAD and ETD data were searched separately using ProteinProspector version 5.6.0 against the UniProt database with a

concatenated database. Only mouse and human genomes were used for the database searching. Precursor tolerance was set to 20 ppm, whereas fragment mass error tolerance was set to 0.6 Da. Cysteine residues were assumed to be carbamidomethlyated while N-terminus acetylation, methionine oxidation, loss of N-terminal methionine and glutamate conversion to pyroglutamate were variable modifications. For ETD data, HexNAc modifications to serine and threonine residues and phosphorylation to serine/threonine/tyrosine was allowed as variable mass modifications. The same modifications were allowed for CAD data except the neutral loss of 203 Da (HexNAc) or 80 Da (phosphate) were considered. For SILAC experiments, the data was searched with arginine +6 Da and lysine +8 Da either as constant modifications or constantly absent. SILAC ratios were determined at a resolution of 30,000. All SILAC ratios and spectra of modified peptides were examined manually to assess spectral quality and site-assignment accuracy.

siRNA Knockdown siRNAs were created by *in vitro* cleavage of an *Ogt* double stranded RNA (54). A PCR product for in vitro transcription of *Ogt* RNA was generated using primers recommended by http://cluster-12.mpi-cbg.de/cgi-bin/riddle/search.

Western Blot Analysis Antibodies were purchased from Abcam (OCT4 ab19857, EZH2 ab3748, H3 ab1791), Millipore (H3K27me3 07-499), ProteinTech (OGA 14711-1-AP), Sigma (OGT SAB2101676 & *O*-GlcNAc CTD 110.6) and

BioRad (secondary antibodies 172-1019 & 172-1011) and used at recommended concentrations.

Appendix I ETD MS/MS spectra of the *O*-GlcNAcylated peptides identified in this study. Peptide sequence and site assignment of the *N*-acetylhexosamine (HexNAc) mass modification are assigned based on Protein Prospector results. Charged reduced species and common neutral losses (55) are not labeled. The data can be accessed at http://msf.ucsf.edu/personal_pages/myers/Myers_2011_Figure_S1.pdf

Table 1 All *O*-GlcNAc modified peptides identified in this study. The gene name and accession number is given. Black denotes the unambiguous assignment of the *O*-GlcNAc moiety (mass addition of a HexNAc). Red denotes the exact site is not assigned but can be localized to one of the red colored residues. Green and blue indicate a mixture of positional isomers, wherein the corresponding colors in the annotated mass spectrum (Appendix I) identify both positional isomers as *O*-GlcNAcylated.

Gene Name	Acc #	*O*-GlcNAc Peptides **Bold=confirmed** Red="ambiguous possibility", Blue/Green= mix, *=carbamidomethylation, Ac=acetlyation	Modified Residue
Agfg1	Q8K2K6	SSSADFGTFS**TS(HexNAc)**QSHQTASTVSK	T301 or S302
Atf7ip	Q7TT18	NSSTTAAPLG**T(HexNAc)T**LAVQAVPTAHSIVQATR	T902 or T903
Bnc2	Q2TBA4	AT**S(HexNAc)**GAATPVIASTK	S437
Bptf	A2A654	QTVV**S(HexNAc)S(HexNAc)**TENC*AR	S1742 & S1743
Bptf	A2A654	**S(HexNAc)**TVTTTTTTVTK	S1750
Bptf	A2A654	S**T(HexNAc)**VTTTTTTVTK	T1751
Bptf	A2A654	STV**T(HexNAc)**T**T(HexNAc)**TTTVTK	T1754 & T1756
Bptf	A2A654	STVTT**T(HexNAc)T(HexNAc)**TTVTK	T1755 & T1756
Bptf	A2A654	STVTT**T(HexNAc)**TTTV**T(HexNAc)**K(heavy)	T1755 & T1760
Bptf	A2A654	**T(HexNAc)**VI**T**EVTTMTSTVATESK	T1710 or T1713
Bptf (frag)	A2A653	IVAVNVPA**T(HexNAc)**QGGMVQVQQK	T225
C1orf88	Q9D9W1	VP**TST(HexNAc)**KGYAIGAR	T82 or S83 or T84
Cdk12	Q14AX6	**TS(HexNAc)T**LSSQTNSQPPVQVSMK	T588-T590
Cdk13	Q69ZA1	TENQHVPT**TS(HexNAc)**SSLTDPHAGVK	T1286 or S1287
Cnot2	Q8C5L3	SLSQGTQLPSHVTP**T(HexNAc)**TGVPTMSLHTPPSPSR	T113
Cnot2	Q8C5L3	SLSQGTQLPSHVTPTTGVP**T(HexNAc)**MSLHTPPSPSR	T118
Dido1	Q8C9B9	ILSSLKPGS**T(HexNAc)**STVTAP**T(HexNAc)T**AAITTTASPVTAATSK	T1280 and T1287 or T1288 mix
Ebp	P70245	Ac-**T(HexNAc)**TNTVPLHPYWPR	T2
Egr1	P08046	AMVET**S(HexNAc)**YPSQTTR	S117
Elf2	Q9JHC9	SPT**T(HexNAc)T(HexNAc)**APVSAAAAPR	T375 & T376
Emsy	Q8BMB0	II**S(HexNAc)**SNIVSGTTTK	S520
Emsy	Q8BMB0	ITF**T(HexNAc)**KPSTQTTNTTTQK	T228
Emsy	Q8BMB0	ITFTKP**S(HexNAc)**TQTTNTTTQK	S231
Emsy	Q8BMB0	ITFTKPSTQ**T(HexNAc)T**NTTTQK	T234 or T235
Emsy	Q8BMB0	LV**T(HexNAc)**TPTGTQATYTRPTVSPSLGR	T465
Emsy	Q8BMB0	TITVPV**S(HexNAc)**GSPK	S200
Emsy	Q8BMB0	T**T(HexNAc)S**GSIITVVPK	T499 and S500 mix
Emsy	Q8BMB0	VIIV**T(HexNAc)T(HexNAc)**SPSSTFVPNILSK	T246 & T247
Ep400	Q8CHI8	LASPVAPGTL**T(HexNAc)T**SGGSAPAQVVHTQQR	T2594 or T2595
Ep400	Q8CHI8	LASPVAPGTLTTSGG**S(HexNAc)**APAQVVHTQQR	S2599
Gata2b	Q5JP02	**S(HexNAc)**ISQSISGQK	S585
HcfC1	B1AUX1	APV**T(HexNAc)**VTSLPASVR	T515

HcfC1	B1AUX1	IPPS**S(HexNAc)**APTVLSVPAGTTIVK**T(HexNAc)**VAVTPGTTTLPATVK	S563 & T579
HcfC1	B1AUX1	I**T(HexNAc)**VAPGALER	T1148
HcfC1	B1AUX1	LVTPV**T(HexNAc)**VSAVKPAV**T(HexNAc)**TLVVK	T861 & T870
HcfC1	B1AUX1	LVTPV**T(HexNAc)**VSAVKPAVTTLVVK	T861
HcfC1	B1AUX1	QPETYH**T(HexNAc)**YTTNTPTTTR	T1238
HcfC1	B1AUX1	QPETYHTYTTNTPT**T(HexNAc)**TR	T1246
HcfC1	B1AUX1	RT**T(HexNAc)**NTPTVVR	T1139
HcfC1	B1AUX1	SGTVTVAQQAQVV**T(HexNAc)T(HexNAc)**VVGGVTK	T651 & T652
HcfC1	B1AUX1	SPITIIT**T(HexNAc)**KVMTSGTGAPAK	T801
HcfC1	B1AUX1	**T(HexNAc)**QGVPAVLK	T480
HcfC1	B1AUX1	**T(HexNAc)**VAVTPGTTTLPATVK	T579
HcfC1	B1AUX1	TAAAQVGT**S(HexNAc)**V**S(HexNAc)S(HexNAc)**AANTSTRPIITVHK	S620 & S622 & S623
HcfC1	B1AUX1	TAAAQVGT**S(HexNAc)**VS**S(HexNAc)**AANTSTRPIITVHK	S620 & S623
HcfC1	B1AUX1	TIPMSAII**T(HexNAc)**QAGATGVTSSPGIK	T779
HcfC1	Q61191	VMTSG**T(HexNAc)**GAPAK	T808
HcfC1	B1AUX1	VTGPQA**T(HexNAc)**TGTPLVTMRPASQAGK	T495
Hprt1	A8MSU4	VIGGDDLSTLTGK**S(HexNAc)**R	S129
Jmjd1c	Q69ZK6	HSVPQSLPQSNYFTTL**S(HexNAc)**NSVVNEPPR	S911
Jmjd1c	Q69ZK6	NSPSPWLHQP**T**PV**TS(HexNAc)**ADGIGLLSHIPVRPSSAEPHRPHK	T728 or T731 or S732
Jmjd1c	Q69ZK6	SVV**S(HexNAc)**QAVAQAK	S1250
Kdm3b	Q6ZPY7	VEHSPFSSFVSQASG**S(HexNAc)S**SSATSVTSK	S460 or S461
Kiaa1310	A2RSY1	**SSSS**EGGG**T(HexNAc)**ASTTPSVASSSATPNAIHTLQSR	Between S699-T707
Kiaa1310	A2RSY1	VP**T(HexNAc)**TITLTLR	T858
Kiaa1551	Q5DTW7	IST**T(HexNAc)**VVGSANPTNEVHVK(heavy)	T491
Kiaa1551	Q5DTW7	NNQLPTY**T(HexNAc)**QSLQSK	T40
Kiaa1551	Q5DTW7	QYSYILPA**T(HexNAc)T(HexNAc)**SLQVK	T232 & T233
Kiaa1551	Q5DTW7	QYSYILPA**T(HexNAc)T**SLQVK	T232 and T233 mix
Kiaa1551	Q5DTW7	YENQHVQNAQPV**S(HexNAc)**K	S291
Lman1	Q9D0F3	RGAG**T(HexNAc)**PGQPGQVSQQELDTVVK	T385
Mafk	O60675	VA**TT(HexNAc)**SVITIVK	T133 or T143
Med15	Q924H2	C*M**TS(HexNAc)**RLLQLPDK	T761 or S762
Myb	P06876	TPAIKR**S(HexNAc)**ILESSPR	S454
NCoR1	Q60974	RTPV**S(HexNAc)**YQNTISR	S1496
NCoR1	Q60974	SVQC*VC***T(HexNAc)S(HexNAc)**SALPSGK	T1899 & S1900
NCoR2	Q9WU42	GSPV**TT(HexNAc)**REPTPR	T1531 or T1532
Nfrkb	Q6PIJ4	IQTVPASHLQQG**T(HexNAc)**ASGSSK	T1270
Nfrkb	Q6PIJ4	QVPVNTTVVST**S(HexNAc**)QSGKLPTR	S1172
Nup53	Q8R4R6	**S(HexNAc**)ISGPSVGVMEMR	S53
Nup53	Q8R4R6	AST**S(HexNAc)**DYQVISDR	S297
Nup98	Q68G59	LPI**S(HexNAc)**ASHSSK	S426
Nup153	Q80WR0	FGIP**S(HexNAc)S**SSGLSQTLTSTGNFK	S898 or S899
Nup153	Q80WR0	QQEPVT**S(HexNAc)**TSLVFGK	S1102
Nup153	Q80WR0	**S(HexNAc)**VSVTPFTYK	S1046
Nup153	Q80WR0	SGFNFGTLD**T(HexNAc)**K	T1044
Nup153	Q80WR0	VDSPALQPT**T(HexNAc)T(HexNAc)**SSIVYTRPAISTFSSSGIEYGESLK	T627 & T628
Nup188	Q6ZQH8	AQRP**S(HexNAc)**TTTTTTTTTTALATPAGC*SSK	S1522
Nup214	Q80U93	AAPG**S(HexNAc**)GTSTFSFAPPSK	S513
Nup214	Q80U93	FTAVA**S(HexNAc)S**APVHSSTSTPSVLPFSSSPKPTASGPLSHPTPLPASSSSMPLK	S589 or S590
Nup214	Q80U93	GGGFFSGLGGKP**S(HexNAc)**QDAANKNPF**S**SAGGGFGSTAAPNTSNLFGNSGAK	S1889 or 1899

Nup214	Q80U93	HGAPGPSH**T(HexNAc)**VAAPQAAAAAALRR	T1091
Nup214	Q80U93	SSASVTGEPPLYPTG**S(HexNAc)**DSSR	S504
Nup214	Q80U93	VGQAEDSTKPV**S(HexNAc)**K	S1362
Phc3	Q8CHP6	SAGQ**T(HexNAc)**QSLTIC*HNK	T238
Phf21a	Q6ZPK0	FTPTTLPT**S(HexNAc)**QNSIHPVR	S286
Pogz	Q8BZH4	RPGV**T(HexNAc)**GENSNEVAK	T310
Pogz	Q8BZH4	V**T(HexNAc)**SSIPVFDLQDGGR	T358
Prdx6	O08709	LKLSILYPA**T(HexNAc)**TGR	T152
Qser1	A2BIE1	GTDVYVPY**TS(HexNAc)**P**S**LESC*DEGFQHQEK	T1077 or S1078 or S1080
Qser1	A2BIE1	T**S(HexNAc)**QGTVPTALAFER	S105
RanBP2	Q9ERU9	ALGTN**TS(HexNAc)**TAPNHTLR	T1306 & S1307 mix
RanBP2	Q9ERU9	SVFT**T(HexNAc)**AASELANK	T1138
Rbbp7	Q5JP02	Ac-DDH**T(HexNAc)**VC*LWDINAGPK	T4
Rbm14	Q8C2Q3	AQPSVS**(HexNAc)**LGAAYR	S244
Rbm14	Q8C2Q3	AQPSVS**(HexNAc)**LGAPYR	S280
Rprd2	Q3UY82	EKPVEKPAV**S(HexNAc)T(HexNAc)**GVPTK	S103 & T104
Rprd2	Q6NXI6	EKPVEKPAV**S(HexNAc)**TGVPTK	S103
Rprd2	Q6NXI6	SIFSSQSTLAAPAGHPPT**S(HexNAc)**GVEK(heavy)	S1110
Sal4	Q8BX22	LPHA**S(HexNAc)**VP**S**AASPLSSGLTSFTLKPDGTR	S299 or S302
Sap130	Q8BIH0	AQSPVIT**T(HexNAc)T(HexNAc)**AAHAADSTLSRPTLSIQHPPSAAISIQRPAQSR	T284 & T285
Sap130	Q8BIH0	I**T(HexNAc)**LPSHPALGTPK	T320
Sap30bp	Q02614	KGT**TT(HexNAc)**NATATSTSTASTAVADAQK	T232 or T233
Sin3a	Q60520	V**S(HexNAc)**KPSQLQAHTPASQQTPPLPPYASPR	S251
Snbo1	Q689Z5	FIQ**T(HexNAc)**TANTRPSVSAPAVR	T124
Sox2	P48432	**S(HexNAc)**EASSSPPVVTSSSHSR	S248
Sox2	P48432	SEASSSPPVV**T(HexNAc)**SSSHSR	T258
Sp2	Q9D2H6	**S(HexNAc)**STTTTPVQSGANVVK	S186
Spen	Q62504	TDRPSLEKPEPIHLSVS**T(HexNAc)**PVTQGGTVK	T2896
Spen	Q62504	TDRPSLEKPEPIHLSVSTPV**T(HexNAc)**QGGTVK	T2899
Taf4a	A2AC70	QV**S(HexNAc)**QAQTTVQPTTTLQR	S123
Taf6	Q62311	T**T(HexNAc)**LTITQPRPTLTLSQAPQPGPR	T49
Tcf20	Q9EPQ8	**T(HexNAc)**EPSKSPGSLR	T725
Tet1	D0VLQ4	VSI**T(HexNAc)**GSADVK	T535
Tle4	Q04727	**T(HexNAc)**DAPTPGSNSTPGLRPVPGKPPGVDPLASSLR	T340
Trim33	Q99PP7	QHSNPGHAGPFPVV**S(HexNAc)**AHNPINPTSPTTATMANANR	S350
Trnc6a	Q3UHK8	LTWSPGSVTN**TS(HexNAc)**LAHELWK	T1646 or S1647
Vim	Q3TWV0	TYSLGSALRPST**S(HexNAc)**R	S49
Yeats2	Q3TUF7	QEPGEAPHV**ST(HexNAc)**TGAASQSAFPQYVTVK	S600 and T601 mix
Yeats2	Q3TUF7	QEPGEAPHVST**T(HexNAc)**GAASQSAFPQYVTVK	T602
Ylpm1	Q9R0I7	QALLPTPV**S(HexNAc)**FGSTPPSPYHPPPQSEQVNSKPLNK	S636
Zfhx3	Q15911	NKNFQHPLV**S(HexNAc)**TANLIGPGHSFYGK	S353
Zfp281	Q3U063	VK**T(Phospho)**P**T(HexNAc)**SQSYR	T888
Zfp281	Q3U063	VKTPT**S(HexNAc)**QSYR	S889
Zfp281	Q3U063	STNAGFTLGHGFQFV**S(HexNAc)**LSSPLHNHTLFPEK	S691
Zfr	O88532	AGY**S(HexNAc)**QGATQYTQAQQAR	S195
Znf318	Q99PP2	TVIAH**T(HexNAc)S**PWTPVVTTSTQTK	T1185 and S1186 mix

Table 2 All *O*-GlcNAc modified peptides identified in the SILAC analyses, along with PRC2/WT ratios, as described in the text. Standard deviations are reported if the particular glycopeptide was identified between biological replicates.

Gene Name	O-GlcNAc Modified Peptides Identified from SILAC Analyses	eed$^{-/-}$/WT SILAC ratios	s. d. N=3
Agfg1	SSSADFGTFSTS(HexNAc)QSHQTASTVSK	0.67	
Atrx	IC(Carbamidomethyl)GS(HexNAc)GLNS(Phospho)DM(Oxidation)MEN NKEEGASTSEK	0.29	
Bnc2	ATS(HexNAc)GAATPVIASTK	0.27	
Bptf	QTVVS(HexNAc)S(HexNAc)TENC(Carbamidomethyl)AR	1.0	0.1
Bptf	S(HexNAc)TVTTTTTTVTK	0.52	0.03
Bptf	STVTTT(HexNAc)TTTVT(HexNAc)K	0.12	
Bptf	T(HexNAc)VITEVTTMTSTVATESK	1.00	
Bptf	TVIT(HexNAc)EVT(HexNAc)TMTSTVATESK	0.27	0.02
Bptf (fragment)	IVAVNVPAT(HexNAc)QGGMVQVQQK	0.91	
C1orf88 homolog	VPTST(HexNAc)KGYAIGAR	0.92	0.08
Cdk12	TS(HexNAc)TLSSQTNSQPPVQVSMK	2.42	
Cnot2	SLSQGTQLPSHVTPTT(HexNAc)GVPTMSLHTPPSPSR	1.12	
Dido1	ILSSLKPGST(HexNAc)STVTAPTT(HexNAc)AAITTTASPVTAATSK	1.12	
Egr1	AMVETS(HexNAc)YPSQTTR	2.03	
Emsy	IIS(HexNAc)SNIVSGTTTK	0.06	
Emsy	ITFT(HexNAc)KPSTQTTNTTTQK	1.37	0.09
Emsy	TITVPVS(HexNAc)GSPK	0.45	
Emsy	TTS(HexNAc)GSIITVVPK	1.1	0.3
Ep400	LASPVAPGTLTTSGGS(HexNAc)APAQVVHTQQR	0.59	
HcfC1	APVT(HexNAc)VTSLPASVR	1.72	
HcfC1	IPPSSAPTVLSVPAGT(HexNAc)TIVKTVAVT(HexNAc)PGTTTLPATVK	0.47	
HcfC1	IT(HexNAc)VAPGALER	0.97	0.06
HcfC1	LVTPVT(HexNAc)VSAVKPAVTTLVVK	6.93	
HcfC1	QPETYHT(HexNAc)YTTNTPTTTR	1.2	0.3
HcfC1	QPETYHTYTTNTPTTT(HexNAc)R	6.67	
HcfC1	SGTVTVAQQAQVVT(HexNAc)T(HexNAc)VVGGVTK	0.78	
HcfC1	SPITIITT(HexNAc)KVMTSGTGAPAK	1.0	0.1
HcfC1	T(HexNAc)VAVTPGTTTLPATVK	1.44	0.08
HcfC1	TAAAQVGTS(HexNAc)VS(HexNAc)S(HexNAc)AANTSTRPIITVHK	0.33	
HcfC1	TAAAQVGTS(HexNAc)VSS(HexNAc)AANTSTRPIITVHK	0.76	0.09
HcfC1	TAAAQVGTSVSS(HexNAc)AANTSTRPIITVHK	1.77	
HcfC1	TIPMSAIIT(HexNAc)QAGATGVTSSPGIK	0.9	0.3
HcfC1	VMTSGT(HexNAc)GAPAK	0.77	
HcfC1	VTGPQAT(HexNAc)TGTPLVTMRPASQAGK	0.70	
Jmjd1	NSPSPWLHQPTPVT(HexNAc)SADGIGLLSHIPVRPSSAEPHRPHK	0.47	
Kiaa1310	VPT(HexNAc)TITLTLR	13.87	
Kiaa1551	ISTT(HexNAc)VVGSANPTNEVHVK	0.41	
Kiaa1551	NNQLPTYT(HexNAc)QSLQSK	0.59	
Kiaa1551	YENQHVQNAQPVS(HexNAc)K	0.4	0.3
Lman1	RGAGT(HexNAc)PGQPGQVSQQELDTVVK	4.9	0.2

MafK	VATT(HexNAc)SVITIVK	1.13	
Myb	TPAIKRS(HexNAc)ILESSPR	6.45	
NCoR2	GSPVTT(HexNAc)REPTPR	0.85	0.03
Nfrkb	IQTVPASHLQQGT(HexNAc)ASGSSK	0.38	
Nfrkb	QVPVNTTVVSTS(HexNAc)QSGKLPTR	0.46	
Nup53	ASTS(HexNAc)DYQVISDR	0.45	
Nup53	S(HexNAc)ISGPSVGVMEMR	0.50	
Nup153	FGIPSSSSGLS(HexNAc)QTLTSTGNFK	0.28	
Nup153	QQEPVTS(HexNAc)TSLVFGK	0.41	
Nup153	S(HexNAc)VSVTPFTYK	0.40	
Nup214	AAPGS(HexNAc)GTSTFSFAPPSK	0.45	
Nup214	HGAPGPSHT(HexNAc)VAAPQAAAAAALRR	0.28	
Nup214	SSASVTGEPPLYPTGS(HexNAc)DSSR	0.72	
Nup214	VGQAEDSTKPVS(HexNAc)K	0.67	
Phc3	SAGQT(HexNAc)QSLTIC(Carbamidomethyl)HNK	5.10	
Phf21a	FTPTTLPTS(HexNAc)QNSIHPVR	1.31	
Pogz	RPGVT(HexNAc)GENSNEVAK	0.34	
Qser1	DSTQVS(HexNAc)NGVLPQK	0.96	
Qser1	TS(HexNAc)QGTVPTALAFER	0.65	
RanBP2	SVFTT(HexNAc)AASELANK	0.8	0.6
Rbm14	AQPSVS(HexNAc)LGAPYR	0.29	0.03
Rprd2	EKPVEKPAVS(HexNAc)T(HexNAc)GVPTK	0.6	0.1
Sap30bp	KGT(HexNAc)TTNATATSTSTASTAVADAQK	1.41	0.03
Sap130	AQSPVITT(HexNAc)T(HexNAc)AAHAADSTLSRPTLSIQHPPSAAISIQRPA QSR	0.82	
Sap130	IT(HexNAc)LPSHPALGTPK	0.21	
Sin3a	VS(HexNAc)KPSQLQAHTPASQQTPPLPPYASPR	0.69	
Snbo1	FIQT(HexNAc)TANTRPSVSAPAVR	0.45	
Sp2	S(HexNAc)STTTTPVQSGANVVK	0.37	
Spen	TDRPSLEKPEPIHLSVST(HexNAc)PVTQGGTVK	1.03	0.06
Taf4a	QVS(HexNAc)QAQTTVQPTTTLQR	0.33	
Taf6	TT(HexNAc)LTITQPRPTLTLSQAPQPGPR	2.23	
Tet1	VSIT(HexNAc)GSADVK	0.83	
Tle4	T(HexNAc)DAPTPGSNSTPGLRPVPGKPPGVDPLASSLR	1.2	0.7
Trim33	QHSNPGHAGPFPVVS(HexNAc)AHNPINPTSPTTATMANANR	0.91	
Yeats2	QEPGEAPHVSTT(HexNAc)GAASQSAFPQYVTVK	0.65	0.02
Zfhx3	NKNFQHPLVS(HexNAc)TANLIGPGHSFYGK	2.85	
Zfp281	STNAGFTLGHGFQFVS(HexNAc)LSSPLHNHTLFPEK	0.46	
Zfp281	VKT(Phospho)PT(HexNAc)SQSYR	0.47	0.02
Zfp281	VKTPTS(HexNAc)QSYR	0.6	0.2
Zfr	AGYS(HexNAc)QGATQYTQAQQAR	0.38	

Table 3 *O*-GlcNAcylated peptides identified from this study that overlap with those reported by Wang *et al.* 2010 (ref. 4) and Chalkley *et al.* 2009 (ref. 19). Red indicates differences in site assignments between studies.

Gene Name	Modification Position in Human, unless noted (UniProt)	UniProt ID (Wang *et al.*)	Modified Peptide Seqeunce from Wang *et al.* 2010 overlapping with this study	Also reported in Chalkely *et al.* 2009
HcfC1	T495	P51610	VTGPQAT(HexNAc)TGTPLVTMRPASQAGK	
HcfC1	S563	P51610	IPPSS(HexNAc)APTVLSVPAGTTIVK	
HcfC1	T779	P51610	TIPM-oxSAIIT(HexNAc)QAGATGVTSSPGIK	
HcfC1	T800	P51610	SPITIITT(HexNAc)K	
HcfC1	T808	P51610	VMTSGT(HexNAc)GAPAK	
HcfC1	T861	P51610	LVTPVT(HexNAc)VSAVKPAVTTLVVK	
Emsy	S236	Q7Z589	TITVPVS(HexNAc)GSPK	
Emsy	T501	Q7Z589	LVT(HexNAc)TPTGTQATYTRPTVSPSIGR	
Emsy	(T534/T535)	Q7Z589	(TT)(HexNAc)SGSIITVVPK	X
Emsy	S557	Q7Z589	IISS(HexNAc)NIVSGTTTK	
Nup153	S908	P49790	FGVS(HexNAc)SSSSGPSQTLTSTGNFK	
Nup153	S909	P49790	FGVSS(HexNAc)SSSGPSQTLTSTGNFK	
Bptf	S1749 & T1753	Q12830	S(HexNAc)TVTT(HexNAc)TTTTVTK	
Rbm14	S244	Q96PK6	AQPSVS(HexNAc)LGAAYR	
Rbm14	S280	Q96PK6	AQPSVS(HexNAc)LGAPYR	
Nfrkb	T1273	Q6P4R8	IQTVPASHLQQGT(HexNAc)ASGSSK	
Zfr	S195	Q96KR1	AGYS(HexNAc)QGATQYTQAQQTR	
Qser1	(T104/S105)	Q2KHR3	(TS)(HexNAc)QGTVPTALAFER	
Kiaa1310	T860	Q9P2N6	VPTT(HexNAc)ITLTLR	
NCoR1	S1487	O75376	TPVS(HexNAc)YQNTMSR	X
NCoR2	(T1569/T1570)	Q9Y618	GSPV(TT)(HexNAc)REPTPR	
Taf4a	S528	O00268	QVS(HexNAc)QAQTTVQPSATLQR	
Zfp281	S891	Q9Y2X9	TPTS(HexNAc)QSYR	

Chapter 4

O-GlcNAcylation inhibits the epigenetic priming activity of SOX2 during reprogramming

_O_-GlcNAcylation inhibits the epigenetic priming activity of SOX2 during reprogramming

Samuel A. Myers[1, 6], Sailaja Peddada[2, 6], Kiichiro Tomoda[3], Joshua L. Pollack[4], Shinya Yamanaka[3, 5], A.L. Burlingame[1] and Barbara Panning[2]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco,
600 16th Street, Genentech Hall Suite N472A, San Francisco, CA 94158

[2]Department of Biochemistry and Biophysics, University of California San Francisco, 600 16th Street, Genentech Hall Suite S372B, San Francisco, California 94158

[3]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA

[4]Lung Biology Center, Department of Medicine, University of California San Francisco, San Francisco, CA, USA

[5]Department of Reprogramming Science, Center for iPS Cell Research and Application, Kyoto University, Kyoto, 606-8507 Japan

[6]These authors contributed equally to this work

Abstract

The transcription factor SOX2 is central in establishing and maintaining pluripotent cells during development and somatic cell reprogramming. In embryonic stem cells, SOX2 is modified by the addition of *O*-linked N-acetylglucosamine (*O*-GlcNAc), a post-translational modification (PTM) that is emerging as a key player in the establishment and maintenance of pluripotency. The mechanisms by which this PTM regulates pluripotency are poorly understood. Here, we show that SOX2 is *O*-GlcNAc modified during somatic cell reprogramming. We find that a SOX2 mutation that eliminates *O*-GlcNAcylation results in increased reprogramming efficiency. The mutation did not change gene expression or SOX2 occupancy at early stages of reprogramming, but did result in epigenetic alterations at SOX2 targets, which may prime them for subsequent activation later in reprogramming. These results suggest that *O*-GlcNAcylation regulates epigenetic priming activity of SOX2, providing insight into the molecular role of this PTM.

Introduction

Reprogramming terminally differentiated cells to induced pluripotent stem cells (iPSCs) typically requires the ectopic expression of the transcription factors Pou5f1 (Oct4), Sox2, Klf4 and c-Myc (OSKM) (Takahashi and Yamanaka, 2006). Although this process is inefficient and stochastic, there are key cellular events that must occur for successful induction of pluripotency (Hanna et al., 2009). These events include a mesenchymal-epithelial transition (MET) (Li et al., 2010; Samavarchi-Tehrani et al., 2010), global changes in chromatin structure (Koche et al., 2011), activation of a deterministic subset of pluripotency genes (Buganim et al., 2012) and, ultimately, the activation of Nanog and the full pluripotency transcription network (Golipour et al., 2012; Silva et al., 2009).

Early during reprogramming O, S, and K are able to access their DNA target sites in chromatin prior to the time of transcriptional activation (Soufi et al., 2012). SOX2 also exhibits this pioneer transcription factor activity during neuronal or B-cell differentiation (Bergsland et al., 2011; Liber et al., 2010). When SOX2 acts as a pioneer factor, the chromatin at its target site exhibits histone modification patterns generally associated with transcriptionally active genes, priming the target for subsequent activation. Despite the potential importance of this epigenetic priming activity of SOX2 during somatic cell reprogramming, its regulation is poorly understood.

In embryonic stem cells (ESCs) SOX2 is modified by *O*-linked *N*-acetylglucosamine (O-GlcNAc) (Myers et al., 2011). *O*-GlcNAc is a dynamic and regulatory glycosylation of serine (S) or threonine (T) residues of nuclear, cytosolic and mitochondrial proteins (Torres and Hart, 1984). *O*-GlcNAc signaling is essential for establishing and maintaining pluripotent cells as deletion of the sole *O*-GlcNAc transferase (OGT) results in the loss of embryo and ESC viability (O'Donnell et al., 2004; Shafi et al., 2000). Although *O*-GlcNAc is emerging as central in regulating transcription, self-renewal and pluripotency, the specific role for transcription factor *O*-GlcNAcylation during reprogramming is not understood.

In this study we show that SOX2 is *O*-GlcNAc modified during early reprogramming, primarily at serine 248 (S248). Replacing wild type SOX2 ($SOX2^{WT}$) with S248A mutant SOX2 ($SOX2^{S248A}$), which cannot be *O*-GlcNAcylated at this residue, increases somatic cell reprogramming efficiency. We show that at early stages of reprogramming $SOX2^{WT}$ and $SOX2^{S248A}$ exhibit comparable association with promoter and enhancer sequences of genes important for reprogramming. Despite exhibiting occupancy comparable to $SOX2^{WT}$, $SOX2^{S248A}$ increases histone H3 lysine 4 dimethylation (H3K4me2) and DNA accessibility at these promoters and enhancers. These epigenetic alterations are not accompanied by detectable increases in expression early or later in reprogramming. Thus, association of $SOX2^{S248A}$ with its targets correlates with epigenetic alterations early during reprogramming and increased reprogramming efficiency. Together, these results suggest that *O*-GlcNAcylation

inhibits the epigenetic priming activity of SOX2 during somatic cell reprogramming, providing a mechanism by which this post-translational modification (PTM) functions to regulate gene expression.

Results

SOX2, not OCT4, is *O*-GlcNAcylated in ESCs

SOX2 and OCT4 have been reported to be *O*-GlcNAc modified in ESCs (Jang et al., 2012; Myers et al., 2011). While our previous work analyzing the murine ESC nuclear *O*-GlcNAcylated proteome identified the modification of SOX2 (Myers et al., 2011), we were unable to detect OCT4 *O*-GlcNAcylation. This discrepancy prompted us to examine OCT4 directly employing ZHBTc4 F-Oct4 cells, the same tagged OCT4 ESC line used by Jang et al. (Jang et al., 2012). No OCT4 *O*-GlcNAcylation was detected by Western blotting (Figure 1A). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis of purified FLAG-tagged OCT4 failed to detect *O*-GlcNAc modification of the site reported by Jang et al., T228, or of any other OCT4 peptide (Figure 1 and 2). In addition, the mass spectrum reported in Jang et al. identifying OCT4-T228 as *O*-GlcNAc modified exhibited many problems: 1) discrepancies with well-established fragmentation patterns for *O*-linked glycopeptides (Carr et al., 1993; Chalkley and Burlingame, 2001; Darula et al., 2010; Jebanathirajah et al., 2003; Medzihradszky et al., 1990; Medzihradszky et al., 1996; Myers et al., 2013; Seipert et al., 2009; Zhao et al., 2011); 2) misleading labeling of peaks; 3) prominent peaks are not assigned; 4) peptide sequence is incorrect; 5) peptide data is inconsistent with experimental methods reported; 6) data is uninterpretable because charge state and mass of the glycopeptide are not reported; and 7) contrary to standard procedures employed by the proteomics community, in which predicted peptide masses from the entire genome are used

Figure 1 OCT4 is not appreciably *O*-GlcNAc modified in ESCs. **A**, Two color Western blot against OCT4 (top), *O*-GlcNAc (middle) and the merged signal (bottom) from anti-FLAG immunoprecipitated material from ZHBTc4 F-OCT4 mESCs. OS$^{FLAG\text{-}WT}$KM infected MEFs were used as a control for the immunoprecipitation and the Western blotting (see Figure 2A). **B**, The mass spectrum from Jang *et al.* (top) compared to a synthetic *O*-GlcNAc modified OCT4 peptide of the reported amino acid sequence (bottom). * indicates H_2O loss, #, ammonia loss. Black labels indicate GlcNAc loss, red indicates GlcNAc retention.

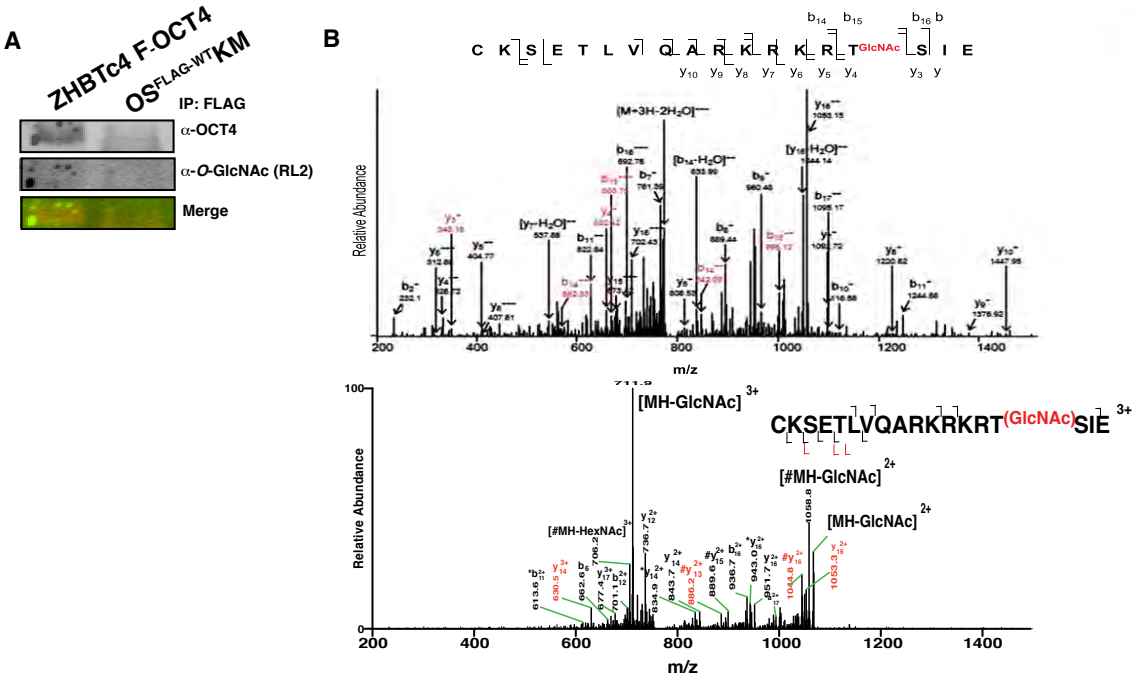
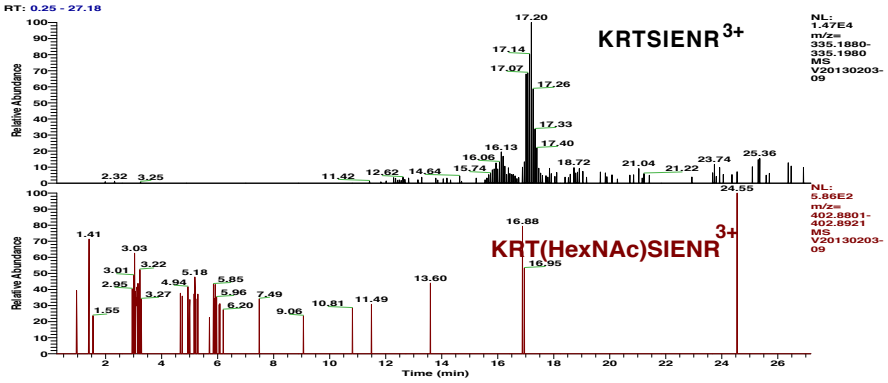


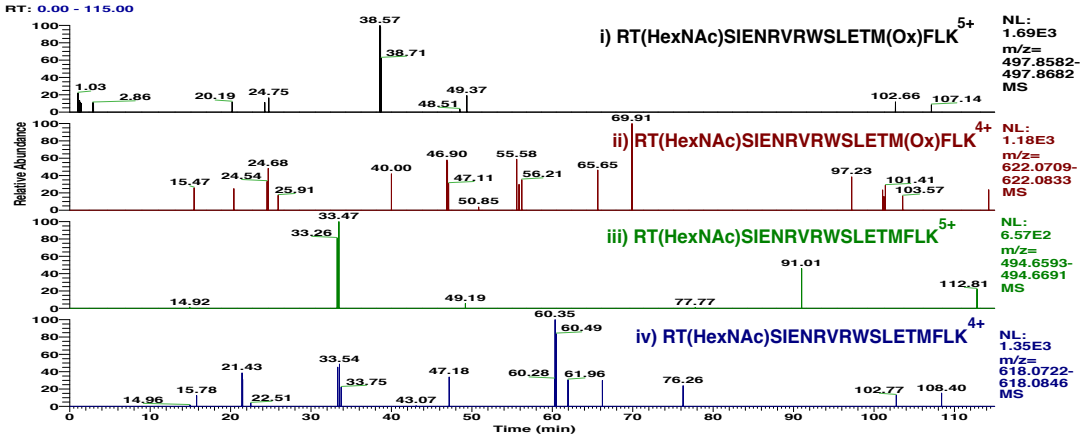
Figure 2 *O*-GlcNAcylation of OCT4 at T228 is undetectable. Analysis of OCT4 *O*-GlcNAcylation in ESCs. OCT4 was FLAG-affinity purified from ZHBTc4 F-Oct4 ESCs (the same cell line employed in (Jang et al., 2012)), which express FLAG-tagged OCT4. Peptides containing T228, the residue reported to be *O*-GlcNAcylated ((Jang et al., 2012)), were analyzed by LC-MS/MS. **A**, XICs (10 ppm) for predicted m/z of an unmodified tryptic peptide containing T228 (top, black panel) and the proposed *O*-GlcNAc modified peptide (bottom, red panel). XICs show the signal for the unmodified peptide, but no signal for the *O*-GlcNAc modified version. **B**, XICs (10 ppm) for m/z of predicted *O*-GlcNAc modified LysC-derived peptides containing T228 is not detectable (two charge states, with and without Met oxidation). **C**, Zoom of MS1 scans from Figure 2B where potential signal exists. No signal with the appropriate charge state was detected for any LysC derived peptide indicating all signal in B is background. Analysis of mass spectrum in (Jang et al., 2012) is incorrect and misleading. **D**, Several lines of evidence indicate the incorrect assignment of the mass spectrum reported in (Jang et al., 2012). 1) The MS/MS method employed by Jang et al uses ion trap collisional activation. Due to the lability of the glycosidic bond, this method leads to a prominent peak corresponding to the peptide minus the mass of the glycosylation (HexNAc). There is no such HexNAc-loss peak in the (Jang et al., 2012) spectrum. 2) Many of the assigned peaks are labeled in a misleading manner. Black arrows, which because of their size and labeling look like peaks, point to noise of the mass spectrum. In Panel **E**, Supplemental figure 1B is re-colored to show the differences. 3) Several prominent peaks are not assigned.

The masses for these peaks are not labeled, making the spectrum uninterpretable. 4) The peptide sequence is wrong. The GluC-derived peptide for mouse, rat, human, macaque, and cow, should be ICKSETLVQARKRKRTSIE. The sequence reported in Jang *et al.* is missing the first Ile, this type of non-specific cleavage with Glu-C has never been reported. 5) The Cys in the sequence is not reported as modified (alkylated by iodoacetamide), though the methods report the data was searched with it modified. 6) Finally, the charge state and the mass of the “glyco”peptide are not reported. Failing to report these parameters makes the data uninterpretable. 7) In contrast to the accepted procedures employed by the proteomics community the data was only searched against the OCT4 protein sequence instead of the entire proteome. This leads to high false positive rates and incorrect false discovery rate calculations. **F**, Figure 1B, lower panel, synthetic glycopeptide, for comparison to re-colored Jang *et al*. mass spectrum. The lack of alignment between peaks indicate that the Jang *et al.* spectrum is not attributed to the correct peptide.

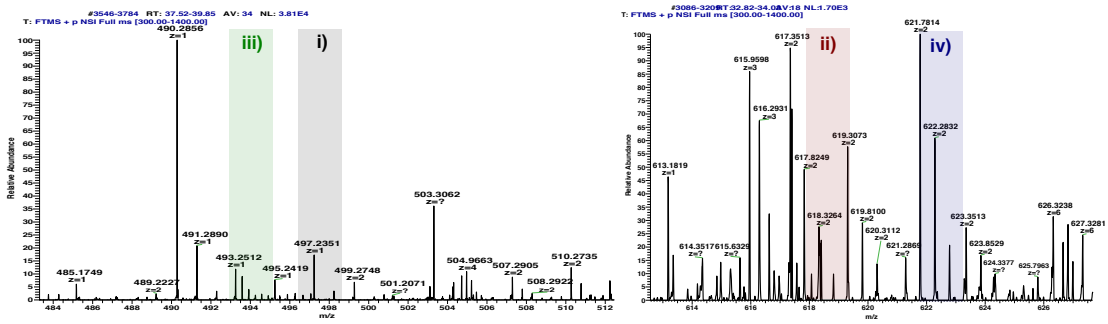
A

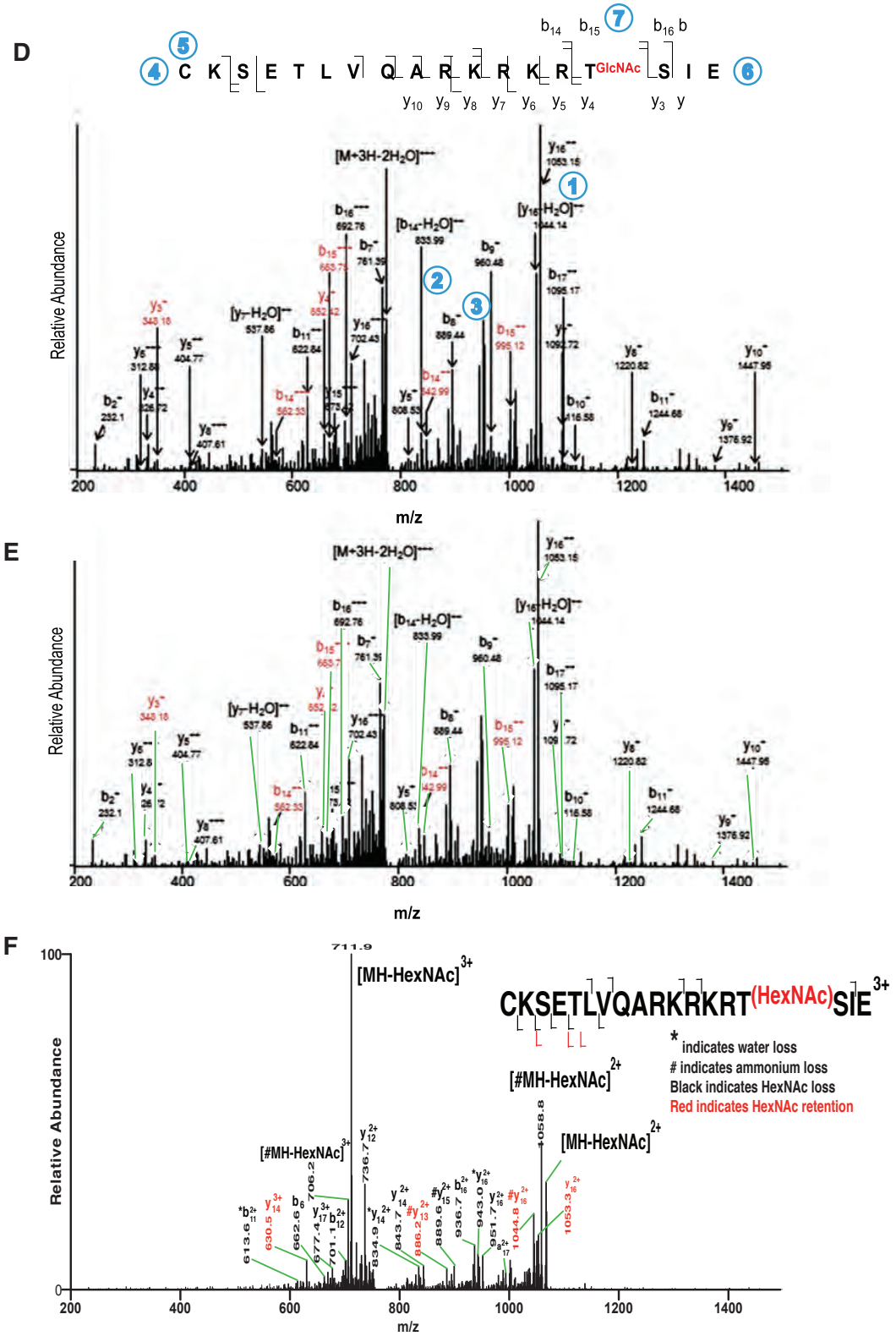

RT: 0.25 - 27.18
KRTSIENR 3+
NL: 1.47E4 m/z= 335.1880-335.1980 MS V20130203-09
KRT(HexNAc)SIENR 3+
NL: 5.86E2 m/z= 402.8901-402.8921 MS V20130203-09
Relative Abundance
Time (min)

B


RT: 0.00 - 115.00
i) RT(HexNAc)SIENRVRWSLETM(Ox)FLK 5+
NL: 1.69E3 m/z= 497.8582-497.8682 MS
ii) RT(HexNAc)SIENRVRWSLETM(Ox)FLK 4+
NL: 1.18E3 m/z= 622.0709-622.0833 MS
iii) RT(HexNAc)SIENRVRWSLETMFLK 5+
NL: 6.57E2 m/z= 494.6593-494.6691 MS
iv) RT(HexNAc)SIENRVRWSLETMFLK 4+
NL: 1.35E3 m/z= 618.0722-618.0846 MS
Relative Abundance
Time (min)

C


iii)
i)
ii)
iv)
Relative Abundance
m/z

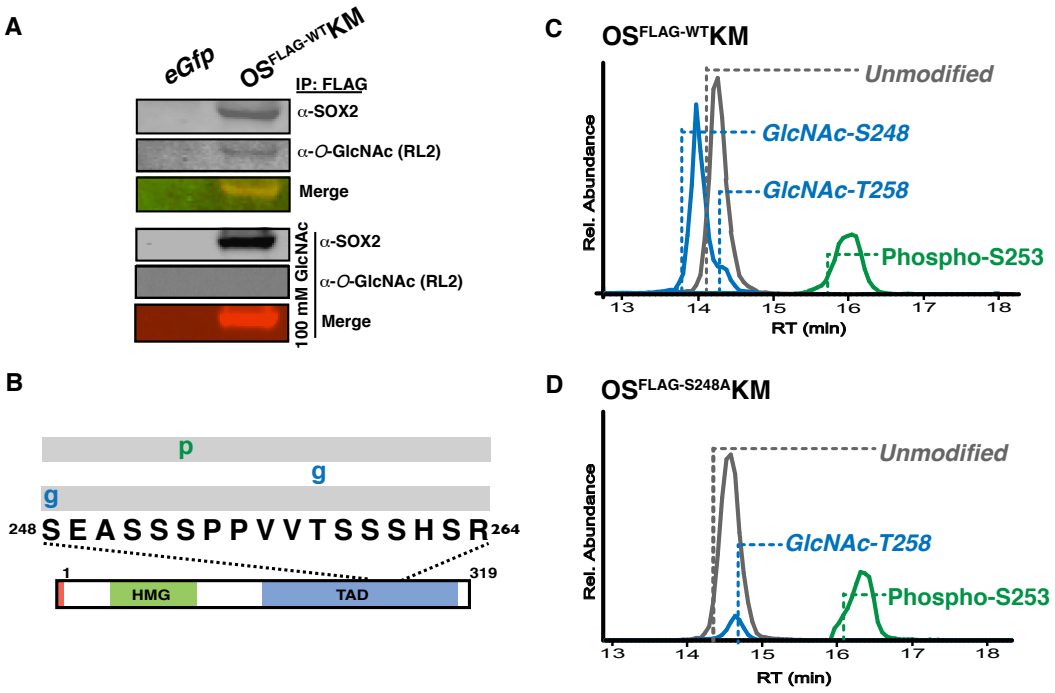

D
C K S E T L V Q A R K R K R T GlcNAc S I E
E
F
Relative Abundance
m/z
[M+3H-2H2O]3+
[MH-HexNAc]3+
[#MH-HexNAc]3+
[MH-HexNAc]2+
[#MH-HexNAc]2+
CKSETLVQARKRKRT(HexNAc)SIE3+
* indicates water loss
# indicates ammonium loss
Black indicates HexNAc loss
Red indicates HexNAc retention

to search MS data (which minimizes false assignments), Jang *et a*l. only queried the OCT4 protein sequence (Figure 2D). Comparison of the spectrum in Jang *1*. to a synthetic glycopeptide of the reported amino acid sequence shows that the original spectrum was misinterpreted (Figure 1B and Figure 2D-F). Our results indicate that OCT4 is not appreciably *O*-GlcNAcylated in ESCs, suggesting that this PTM is unlikely to regulate OCT4 function in pluripotency and self-renewal.

SOX2 is *O*-GlcNAc modified predominantly at S248 during early reprogramming

Because we could reliably and unambiguously identify SOX2 as *O*-GlcNAc modified in ESCs (Myers et al., 2011), we asked whether SOX2 was glycosylated during somatic cell reprogramming. Wild type mouse embryonic fibroblasts (MEFs) were infected with retroviruses containing $OS^{FLAG\text{-}WT}KM$, in which Sox2 contains a triple FLAG tag ($3xF\text{-}Sox2^{WT}$), or enhanced Green fluorescent protein (*eGfp*). FLAG immunoprecipitations from MEFs six days after infection were analyzed for *O*-GlcNAcylation. Western blotting against SOX2 and *O*-GlcNAc showed a signal specific to the $OS^{FLAG\text{-}WT}KM$ infected MEFs (Figure 3A). The anti-*O*-GlcNAc, but not the anti-SOX2 signal, was lost when GlcNAc was added as a competitor. These results indicate that SOX2 is *O*-GlcNAc modified during early reprogramming.

Figure 3 Characterization of SOX2 *O*-GlcNAcylation during somatic cell reprogramming **A**, Two color Western blot against SOX2 (top), *O*-GlcNAc (middle) and the merged signal (bottom) from anti-FLAG immunoprecipitated material from OS$^{FLAG\text{-}WT}$KM infected MEFs (top three panels). 3xF-SOX2 from the same purification was Western blotted with the same primary antibodies plus 100 mM GlcNAc (bottom three panels). **B**, Diagram of 3xF-SOX2 showing the TAD peptide PTMs detected in this study. Amino acid numbering from the Uniprot accession number P48432. *N*-terminal FLAG tag signified by the red. Sites of phosphorylation (p) and/or *O*-GlcNAcylation (g) are indicated on lines representing each PTM form identified. HMG, high-mobility group DNA binding domain; TAD, transactivation domain. **C**, XICs for 3xF-SOX2^{WT} TAD peptide. Dotted lines show time of identification for the unmodified (gray), phosphorylated (red), and *O*-GlcNAcylated (blue) peptides indicated. **D**, XICs for 3xF-SOX2^{S248A} TAD peptides as in **C**. ETD spectra can be viewed at **http://tinyurl.com/iPSC-3xF-SOX2-ETD**. HCD spectra can be viewed at **http://tinyurl.com/iPSC-3xF-SOX2-HCD**.


A
eGfp
OSFLAG-WTKM
IP: FLAG
α-SOX2
α-O-GlcNAc (RL2)
Merge
α-SOX2
α-O-GlcNAc (RL2)
Merge
100 mM GlcNAc
B
p
g
g
248 S E A S S S P P V V T S S S H S R 264
1
319
HMG
TAD
C
OSFLAG-WTKM
Unmodified
GlcNAc-S248
GlcNAc-T258
Phospho-S253
Rel. Abundance
13
14
15
16
17
18
RT (min)
D
OSFLAG-S248AKM
Unmodified
GlcNAc-T258
Phospho-S253
Rel. Abundance
13
14
15
16
17
18
RT (min)

LC-MS/MS of material isolated from $OS^{FLAG\text{-}WT}KM$-infected MEFs was used to identify sites of modification. Two *O*-GlcNAcylation sites (S248 and T258) and one phosphorylation site (S253) were found in a peptide from the transactivation domain (TAD) of SOX2 (Figure 3B, Table 1). Extracted ion chromatograms (XICs) showed chromatographic separation of *O*-GlcNAc positional isomer peptides, with the GlcNAc-S248 isomer correlating with the largest peak (Figure 3C). We next mutated S248 to alanine (S248A) and analyzed the mutant 3xF-SOX2 (3xF-$SOX2^{S248A}$) from MEFs infected with $OS^{FLAG\text{-}S248A}KM$ for six days. 3xF-$SOX2^{S248A}$ showed a dramatic decrease in the abundance of the *O*-GlcNAcylated TAD peptide, indicating that S248 is the major site of *O*-GlcNAcylation during early reprogramming (Figure 3D, Table 1).

Mutation of SOX2 S248 to alanine increases somatic cell reprogramming

To test whether *O*-GlcNAcylation of S248 affects SOX2 function, we asked whether this residue is necessary for SOX2 to induce pluripotency. We compared retroviral reprogramming efficiency of $Sox2^{WT}$ with $Sox2^{S248A}$ ($OS^{WT}KM$ and $OS^{S248A}KM$, respectively). MEFs carrying *Gfp* driven by *Nanog* regulatory sequences (Okita et al., 2007; Takahashi and Yamanaka, 2006) were used for visual detection of iPSCs. MEFs reprogrammed with $OS^{S248A}KM$ showed a five-fold increase ($p=0.0085$) in number of GFP^+ colonies 20 days after infection relative to $OS^{WT}KM$ (Figure 4A). $OS^{S248A}KM$-derived GFP^+ iPSC colonies exhibited standard colony morphology (Figure 4B) and activation of endogenous pluripotency transcription factor expression (Figure 5A). These iPSCs produced

Figure 4 SOX2^{S248A} increases somatic cell reprogramming efficiency. **A**, Number of GFP^{+} colonies from 1000 *Nanog-Gfp* MEFs that were infected with *dsRed*, OS^{WT}KM or OS^{S248A}KM and cultured on SNL feeders for 20 days after infection. **B**, Example of an OS^{S248A}KM derived colony in both phase and fluorescence microscopy. **C**, Chimaeric mouse derived from iPSCs obtained from infecting *Nanog-Gfp* MEFs with OS^{S248A}KM and his offspring, showing germline transmission.

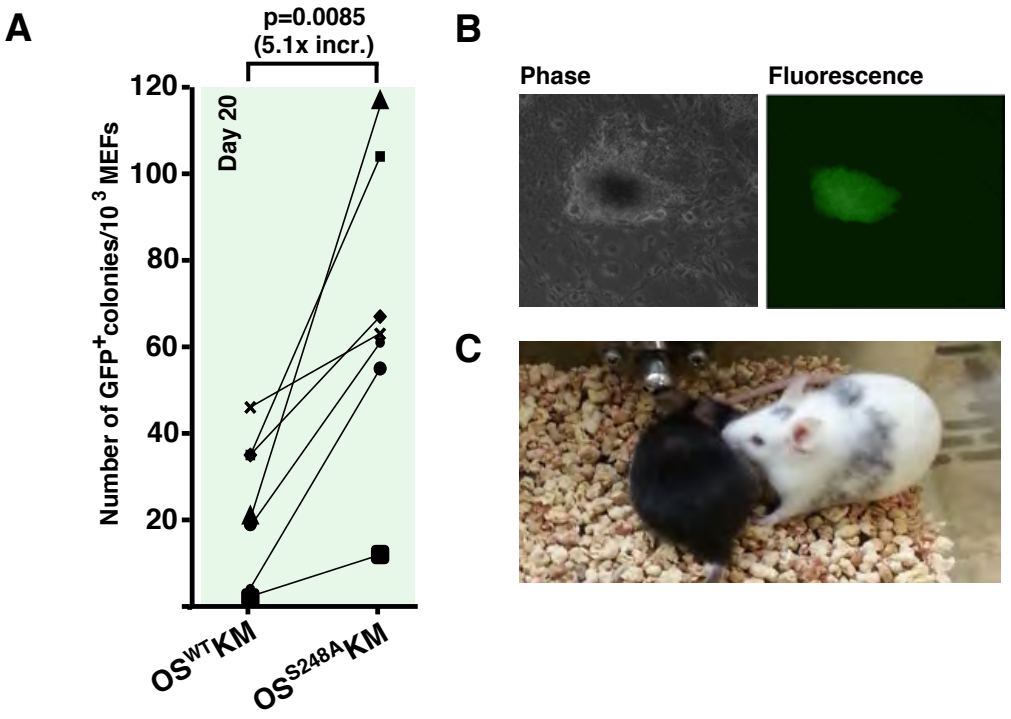
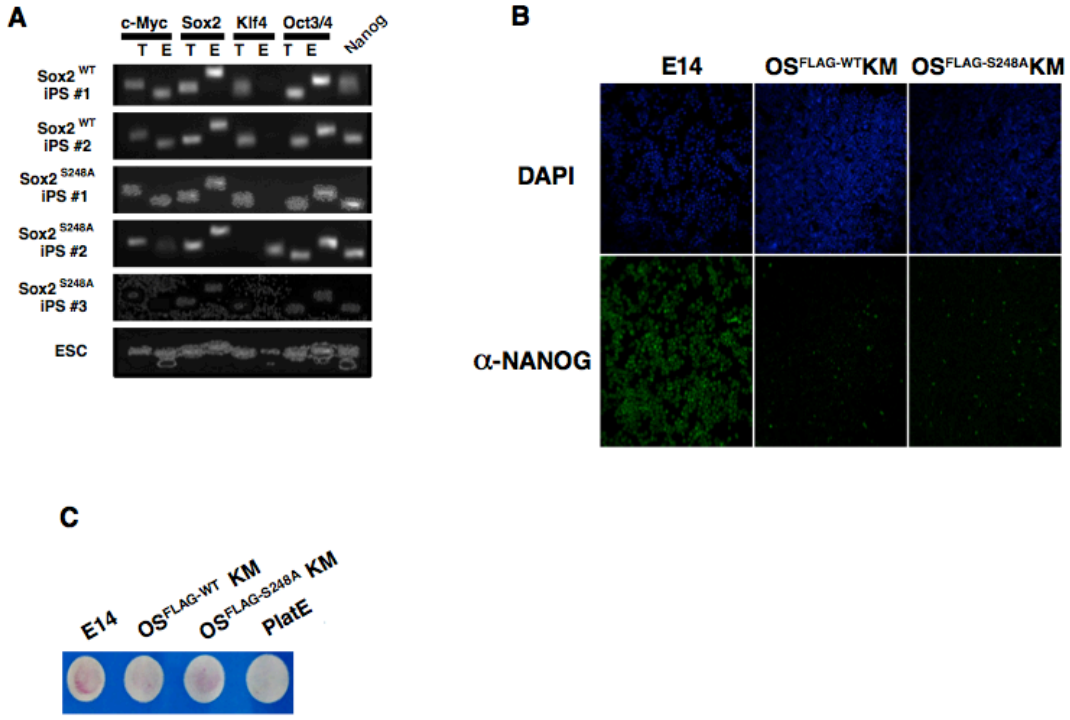


Figure 5 Analysis of retroviral and endogenous transcripts in iPSC lines. Two $OS^{WT}KM$ and three $OS^{S248A}KM$ iPSC lines were analyzed for the expression of exogenous and endogenous OSKM and Nanog transcripts. RT-PCR analysis of total (T) and endogenous (E) transcripts are shown compared to WT ESCs. Characterization of $OS^{FLAG\text{-}WT}KM$ and $OS^{FLAG\text{-}S248A}KM$ reprogramming. **B,** iPSCs 20 days after infection with $OS^{FLAG\text{-}WT}KM$ or $OS^{FLAG\text{-}S248A}KM$ were trypsinized, cytospun and immunofluorescent stained against NANOG. E14 ESCs were used as a control. **C**, iPSCs 16 days after infection with $OS^{FLAG\text{-}WT}KM$ or $OS^{FLAG\text{-}S248A}KM$ were trypsinized, cytospun and stained for alkaline phosphatase. E14 ESCs and PlatE cells were used as controls. A larger number of NANOG and alkaline phosphatase positive cells in $OS^{FLAG\text{-}S248A}KM$ shows that the addition of FLAG tag does not impair SOX2 function.

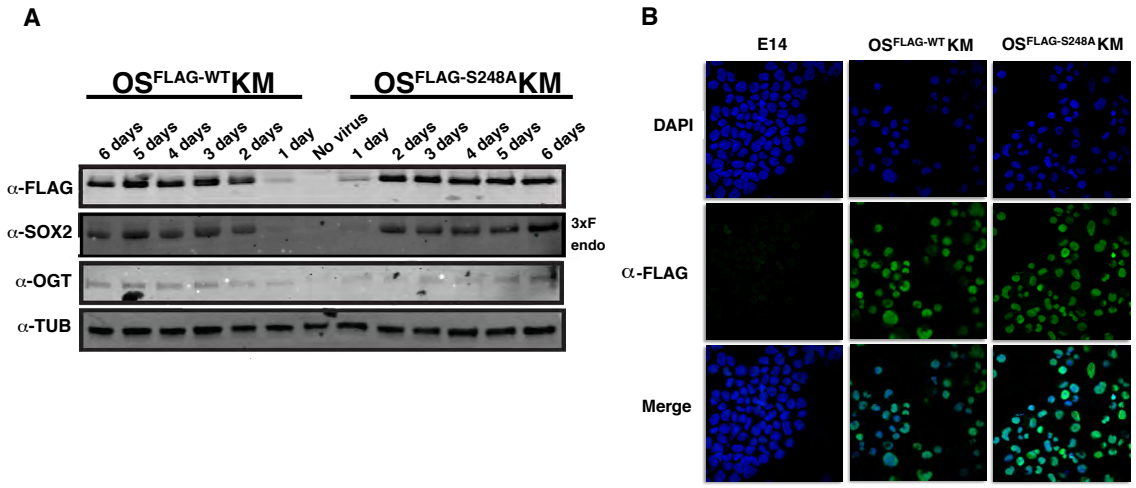


adult chimaeras that were capable of germline transmission (Figure 3C), consistent with pluripotency. Thus the S248A mutation promotes reprogramming without affecting the developmental potential of the resulting iPSCs.

The S248A mutation does not affect SOX2 localization or steady state levels

To gain insight as to how $SOX2^{S248A}$ increases reprogramming efficiency, we compared the properties of $SOX2^{WT}$ and $SOX2^{S248A}$. We used 3xF-$Sox2^{WT}$ or 3xF-$Sox2^{S248A}$ in the OSKM cocktails to distinguish ectopic SOX2 from endogenous SOX2. The FLAG tag did not impair SOX2 function as 3xF-$SOX2^{S248A}$ also exhibited increased reprogramming activity in wild type MEFs (Figure 5B-C). For the first six days of reprogramming, MEFs infected with $OS^{FLAG\text{-}WT}$KM or $OS^{FLAG\text{-}S248A}$KM showed similar levels of 3xF-SOX2 and no detectable expression of endogenous SOX2 (Figure 6A). Additionally, $OS^{FLAG\text{-}WT}$KM and $OS^{FLAG\text{-}S248A}$KM infected MEFs showed comparable increases in OGT levels over the first six days of reprogramming (Figure 6A), indicating that differences in OGT levels do not account for the altered *O*-GlcNAcylation of $SOX2^{S248A}$. Immunostaining at day six of reprogramming revealed no difference in subcellular localization between 3xF-$SOX2^{S248A}$ and 3xF-$SOX2^{WT}$ (Figure 6B). Together, these data demonstrate the increase in reprogramming efficiency mediated by the S248A mutation is not due to changes in SOX2 subcellular localization or steady state levels.

Figure 6 Characterization of 3xF-SOX2^{S248A} steady state levels and subcellular localization. **A**, Western blots against FLAG, SOX2, OGT and Tubulin from MEFs infected with OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM over a six day time course. **B**, Immunofluorescence against FLAG in mouse ESCs (control) or MEFs infected with OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM for six days.



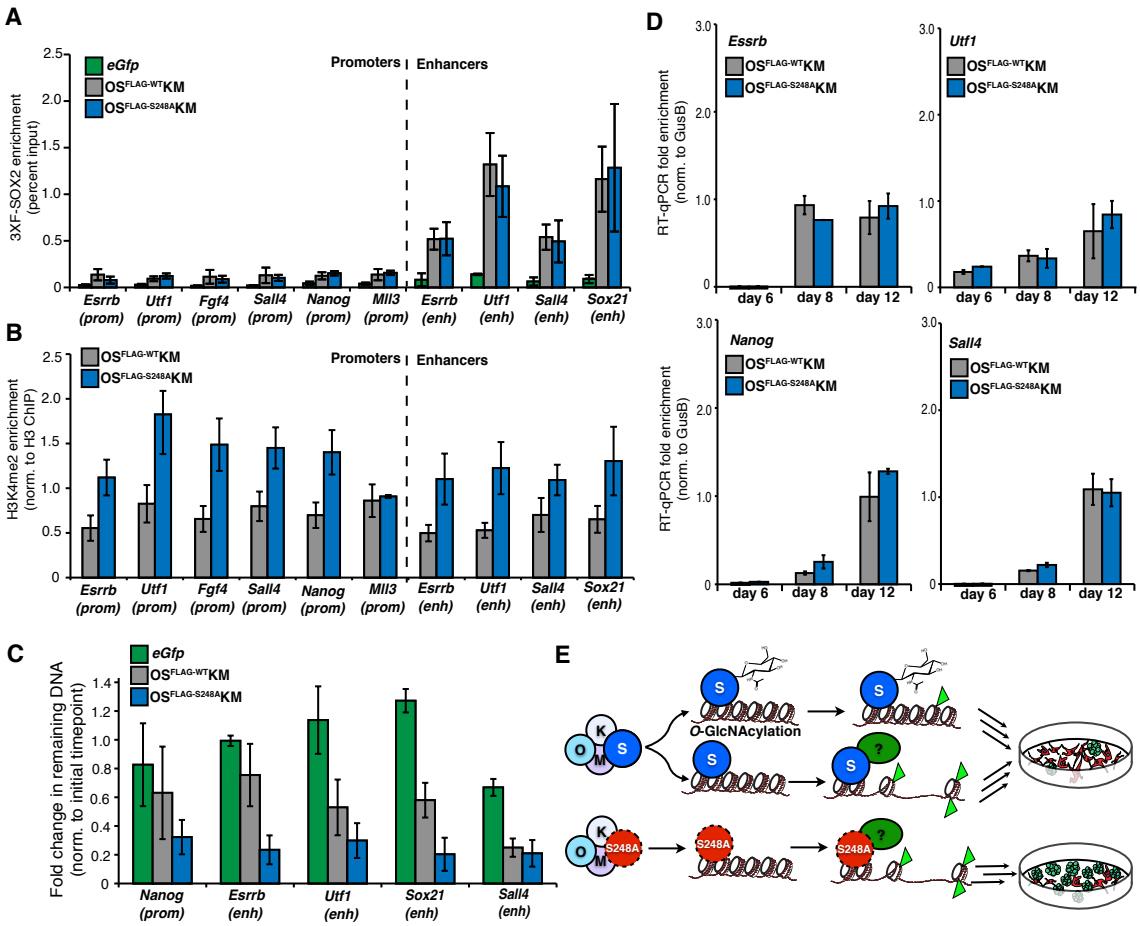
The S248A mutation does not affect SOX2 binding to targets

Since the S248A mutation did not change abundance or intracellular distribution of SOX2, we next asked whether this mutation affected association of SOX2 with its target sequences. We compared MEFs infected with $OS^{FLAG\text{-}WT}KM$, $OS^{FLAG\text{-}S248A}KM$, or *eGfp* viruses six days after infection using FLAG chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR). Promoters or enhancers of target genes that are bound and regulated by SOX2 in ESCs and that play an important role in reprogramming were selected for ChIP-qPCR (Figure 7A). At *Esrrb, Utf1, Sox21*, and *Sall4* enhancers $OS^{FLAG\text{-}WT}KM$ and $OS^{FLAG\text{-}S248A}KM$ infected MEFs exhibited increased enrichment of 3xF-SOX2 compared to MEFs infected with *eGfp*, and there was no difference between WT and S248A 3xF-SOX2. The promoters of *Esrrb, Utf1, Fgf4, Sox21, Sall4,* and *Nanog* exhibited the same pattern, but overall enrichment was 10-fold lower than with the enhancer ChIPs. The promoter of *Mll3*, a gene important for reprogramming but that does not have a SOX2 binding site in ESCs, shows a similar, low enrichment for 3xF-SOX2. These data suggest that enhancer binding is greater than promoter binding during early reprogramming and that the S248A mutation does not alter the association of SOX2 with target sequences.

The S248A mutation affects chromatin at SOX2 targets

Early in reprogramming OSKM mediates alterations in H3K4me2 patterns (Koche et al., 2011). To ask whether the S248A mutation affects H3K4me2 patterns, we performed ChIP-qPCR against H3K4me2. MEFs infected with

Figure 7 Characterization of 3xF-SOX2^{S248A} activity during somatic cell reprogramming. **A**, MEFs infected with *eGfp,* OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM for six days were analyzed by α-FLAG ChIP-qPCR at promoters or enhancers of genes indicated (n=4 ± S.E.M.). **B**, ChIP-qPCR against H3K4me2 from MEFs infected with OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM for six days. H3K4me2 signals were normalized to α-H3 ChIP-qPCR (n=4 ± S.E.M.). **C**, Fold change in DNA remaining after DNase 1 treatment from *eGfp,* OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM infected MEF nuclei six days after infection (n=3 ± S.E.M.). **D**, Relative expression of pluripotency genes from MEFs infected with OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM for six, eight and 12 days post OSKM infection (n=3 ± S.E.M.) **E,** *O*-GlcNAcylation regulates SOX2 epigenetic priming activity. During reprogramming with wild type OSKM, SOX2 (upper panel, blue circle) exists in two populations, *O*-GlcNAcylated (indicated by sugar molecule) or unmodified at S248. Unmodified SOX2 associates with additional factors (dark green oval) that mediate alterations in chromatin structure, indicated by changes in nucleosome (grey cylinders) density and modifications (green triangles). This epigenetic priming promotes iPSC formation. During reprogramming with mutant OSKM, SOX2^{S48A} (lower panel, red circle) is not *O*-GlcNAcylated and can mediate epigenetic priming in a greater number of cells and/or at a great number of loci, facilitating iPSC production.


A
eGfp
OSFLAG-WTKM
OSFLAG-S248AKM
Promoters
Enhancers
3XF-SOX2 enrichment (percent input)
Esrrb (prom)
Utf1 (prom)
Fgf4 (prom)
Sall4 (prom)
Nanog (prom)
Mll3 (prom)
Esrrb (enh)
Utf1 (enh)
Sall4 (enh)
Sox21 (enh)
B
H3K4me2 enrichment (norm. to H3 ChIP)
C
Fold change in remaining DNA (norm. to initial timepoint)
Nanog (prom)
Esrrb (enh)
Utf1 (enh)
Sox21 (enh)
Sall4 (enh)
D
Essrb
Utf1
Nanog
Sall4
RT-qPCR fold enrichment (norm. to GusB)
day 6
day 8
day 12
E
O-GlcNAcylation
S248A

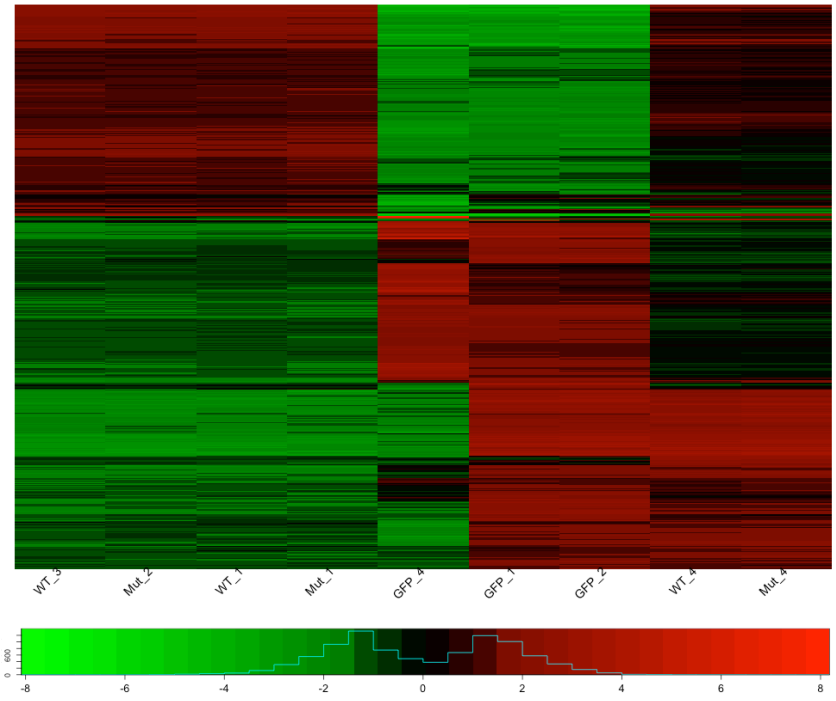
$OS^{FLAG\text{-}S248A}KM$ showed increased H3K4me2 compared to $OS^{FLAG\text{-}WT}KM$ at the majority of targets. The sole exception was the Mll3 promoter, at which H3K4me2 was unchanged between WT and S248A (Figure 7B). Since *Mll3* is not bound by SOX2 in ESCs, the higher H3K4me2 at *Esrrb, Utf1, Sox21, Fgf4, Sall4,* and *Nanog* during reprogramming with S248A suggests that this mutant alters the activity or abundance of histone-modifying enzymes at SOX2 targets associated with pluripotency.

In addition to enrichment for H3K4me2, SOX2 can also induce DNAse 1 hypersensitivity at regulatory regions of active genes or genes poised for transcription at later developmental stages (Liber et al., 2010). We therefore tested DNase 1 sensitivity in MEFs infected for six days with $OS^{FLAG\text{-}WT}KM$, $OS^{FLAG\text{-}S248A}KM$, or *eGfp* viruses using qPCR to determine the amount of DNA that remained after DNase 1 digestion (Figure 7C). In general both OSKM cocktails showed increased DNAse I sensitivity over eGFP. *Nanog, Essrb, Utf1,* and *Sox21* showed increased sensitivity with $OS^{FLAG\text{-}S248A}KM$ relative to $OS^{FLAG\text{-}WT}KM$, while *Sall4* exhibited comparable DNAse 1 sensitivity in both OSKM cocktails. These results indicate that 3xF-$SOX2^{S248A}$ binding increases DNA accessibility at a subset SOX2 targets early during reprogramming.

The S248A mutation does not alter pluripotency gene expression at early stages of reprogramming

The increased H3K4me2 and DNA accessibility at SOX2 binding sites six days after OS$^{FLAG\text{-}S248A}$KM infection suggested the possibility that the mutant may exhibit greater expression of *Esrrb, Utf1, Sox21, Sall4,* and *Nanog* than wild type at this time point. However, RNA-seq of three biological replicates showed no significant alterations in gene expression between OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM at this time point (Figure 8). To examine whether the mutation affected pluripotency gene expression at later stages of reprogramming, we measured transcript levels at six, eight, and 12 days after OSKM infection. We detected comparable levels of transcripts from *Esrrb, Utf1, Sall4*, and *Nanog* in MEFs infected with OS$^{FLAG\text{-}WT}$KM and OS$^{FLAG\text{-}S248A}$KM (Figure 7D). Together, these results suggest that the S248A mutation affects the ability of SOX2 to alter chromatin structure early during reprogramming without immediately affecting gene expression. This increased epigenetic priming of pluripotency genes may underlie the increased reprogramming efficiency of SOX2^{S248A}.

Figure 8 Gene expression analysis of MEFs infected with *eGfp*, OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM for six days. Heat maps show gene expression patterns of MEFs infected with OS$^{FLAG\text{-}WT}$KM or OS$^{FLAG\text{-}S248A}$KM are similar to each other but greatly different from *eGfp*.



Discussion

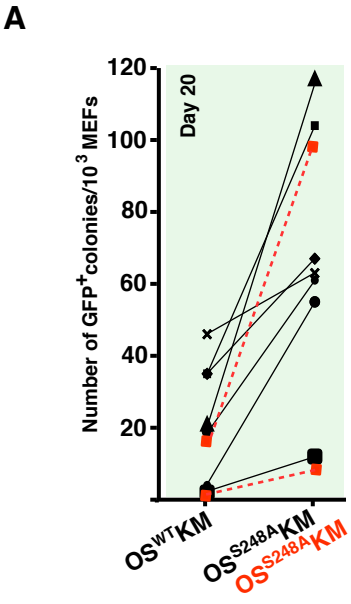
Analysis of the ESC nuclear *O*-GlcNAcylated proteome revealed that many OGT targets are transcription factors and chromatin regulatory proteins with roles in self-renewal (Myers et al., 2011), providing a connection between this PTM and the pluripotency transcription program. In addition blocking *O*-GlcNAcylation disrupts reprogramming (Jang et al., 2012). However, our understanding of how O-GlcNAc regulates its targets, particularly during reprogramming and self-renewal, is in its infancy. Here, we provide evidence that *O*-GlcNAcylation inhibits the epigenetic priming activity of SOX2 during early somatic cell reprogramming.

MS analysis of SOX2 demonstrated abundant *O*-GlcNAcylation of S248 during reprogramming. Substituting SOXS248A for SOX2WT in the OSKM cocktail lead to an increase in the efficiency of iPSC generation, suggesting *O*-GlcNAc inhibits SOX2 activity. The S248A mutation did not affect steady state levels or nucleo-cytoplasmic distribution of ectopic SOX2, indicating that the increase in reprogramming efficiency is not a consequence of altering amounts or distribution of SOX2. Increasing OGT levels promotes reprogramming (Jang et al., 2012). $OS^{FLAG\text{-}WT}KM$ and $OS^{FLAG\text{-}S248A}KM$ showed equal increases in OGT abundance during early reprogramming, indicating that the improved reprogramming efficiency with $OS^{FLAG\text{-}S248A}KM$ is not connected to alterations in OGT levels. ChIP-qPCR in MEFs infected with $OS^{FLAG\text{-}WT}KM$ or $OS^{FLAG\text{-}S248A}KM$ showed the S248A mutation did not affect the association of SOX2 with target

sequences, indicating that this mutation does not affect DNA or chromatin binding. Together these results suggest that the S248A mutation affects a SOX2 activity that lies downstream of its association with target regulatory sequences.

In some instances SOX2 functions as a pioneer transcription factor or promotes epigenetic priming, both examples of activities that alter chromatin modifications or DNA accessibility to regulate gene expression at a later stage. For example, in ESCs, SOX2 directs active histone marks and DNAse 1 hypersensitivy at a B-cell specific locus for subsequent activation upon differentiation into B-cells (Liber et al., 2010). SOX2 acts a pioneer transcription factor during human somatic cell reprogramming (Soufi et al., 2012), when early binding of SOX2 to cis-regulatory elements promotes subsequent activation of genes important for successful reprogramming. Our data suggest that SOX2 epigenetic priming during reprogramming is altered by the S248A mutation, as H3K4me2 at pluripotency gene promoters and enhancers, as well as DNase 1 sensitivity at enhancers, was increased with $SOX2^{S248A}$ relative to $SOX2^{WT}$. These epigenetic alterations were not accompanied by detectable changes in gene expression. These results suggest that the S248A mutation inhibits the ability of SOX2 to establish a permissive transcriptional state for subsequent activation of genes required for efficient reprogramming. While S248 is phosphorylated in some cell types (Swaney et al., 2009), we failed to detect S248 phosphorylation during reprogramming. Furthermore, a phosphomimetic S248D mutation also increased reprogramming activity (Figure 9), suggesting

Figure 9 SOX2^{S248D} also increases somatic cell reprogramming efficiency. Number of GFP^{+} colonies from 1000 *Nanog-Gfp* MEFs that were infected with *dsRed*, OS^{WT}KM, OS^{S248A}KM or the phosphomimetic OS^{S248D}KM and cultured on SNL feeders for 20 days after infection.



that loss of *O*-GlcNAcylation and not phosphorylation underlies the S248 mutant phenotype. Thus, our results imply that this *O*-GlcNAcylation regulates SOX2 epigenetic priming activity (Figure 7E).

The increased H3K4me2 during S248A reprogramming suggests that this mutation alters the abundance or activity of histone-modifying enzymes recruited by SOX2. S248 lies in the TAD, a region by which SOX2 associates with other transcription factors and chromatin regulators (Ambrosetti et al., 2000; Nowling et al., 2000), supporting a hypothesis that *O*-GlcNAcylation may regulate SOX2-centered protein-protein interactions. Additionally, the increased DNase 1 sensitivity exhibited at a subset of targets during early stages of reprogramming with $SOX2^{S248A}$ suggests the possibility that *O*-GlcNAcylation may influence association between SOX2 and nucleosome-remodeling complexes, such as SWI-SNF or CHD7, which associate with SOX2 in ESCs and neural stem cells, respectively (Engelen et al., 2011; Ho et al., 2009).

While both SOX2 and OCT4 have been reported to be *O*-GlcNAc modified in ESCs (Jang et al., 2012; Myers et al., 2011), we focused our efforts on SOX2 due to the lack of reliable data for OCT4 *O*-GlcNAcylation. Our lab has been able to unbiasedly detect SOX2 *O*-GlcNAcylation in mouse (Myers et al., 2011) and human ESCs (data not shown). However, we have yet to detect any evidence for the *O*-GlcNAcylation of OCT4

in ESCs. Moreover, the mass spectrum in Jang et al. claiming to have identified T228 as *O*-GlcNAc modified is chemically incorrect, exhibits experimental inconsistencies, and is labeled in a misleading fashion. *O*-GlcNAc modified peptide analysis with conventional mass spectrometry (Chalkley and Burlingame, 2003; Chalkley et al., 2009; Myers et al., 2013; Zhao et al., 2011) is challenging and there is poor predictive power of an OGT motif/consensus sequence (Trinidad et al., 2012). Our results highlight the importance of using appropriate MS methods for *O*-GlcNAc site assignment. Furthermore, they indicate that SOX2 and not OCT4 is the OGT target that ties *O*-GlcNAcylation to the pluripotency transcription factor network.

Manipulation of OGT levels and identification of OGT targets suggest that *O*-GlcNAcylation promotes pluripotency, self-renewal, and reprogramming (Jang et al., 2012; O'Donnell et al., 2004; Shafi et al., 2000; Yang et al., 2012)(Ding et al., 2012; Esch et al., 2013; Myers et al., 2011; Pardo et al., 2010; van den Berg et al., 2010). This work suggests that *O*-GlcNAcylation inhibits rather than promotes SOX2 function, illustrating that the connection between *O*-GlcNAcylation and pluripotency is more complex than previously appreciated. Thus, continued study of the relationship between *O*-GlcNAc and pluripotency is important to understand the function of this underappreciated PTM.

Finally, the SOX2^{S248A} mutation provides a minor alteration to the reprogramming protocol that increases efficiency of iPSC generation to levels

comparable with either Glis1 addition or p53/p21 pathway inhibition (Kawamura et al., 2009; Maekawa et al., 2011; Marion et al., 2009; Sato et al., 2004). Exploring whether combining the S248A mutation with these other perturbations results in further enhanced reprogramming efficiency may impact the ease with which these biomedically important cells can be produced.

Acknowledgements

We thank Thomas Fazzio, Katie Worringer, and LeRoy Robinson for technical assistance, and Robert Chalkley, Kati Medzihradszky, Tiffani Quan and Assen Roguev for discussion. We also acknowledge the UCSF Sandler Asthma Basic Research (SABRE) Center for Functional Genomics Core Facility. This work was supported by the Biomedical Technology Research Centers program of the NIH National Institute of General Medical Sciences, NIH NIGMS P41GM103481, 1S10RR019934, Howard Hughes Medical Institute and Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (A.L.B), NIH R01GM085186 and University of California San Francisco Program for Breakthrough Biomedical Research (B.P.). S.A.M. is supported by a National Institutes of Health National Institute of General Medical Sciences T32 training grant, the Genentech Predoctoral Fellowship Program and the QBC Fellowship for Interdisciplinary Research. The Cornell Stem Cell Core is supported by the Empire State Stem Cell fund through NYSDOH Contract # C024174, and opinions expressed here are solely those of the author and do not necessarily reflect those of the Empire State Stem Cell Fund, the NYSDOH, or the State of NY. The authors declare no competing financial interests.

References

Ambrosetti, D.C., Scholer, H.R., Dailey, L., and Basilico, C. (2000). Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. J Biol Chem *275*, 23387-23397.

Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. Genes Dev *25*, 2453-2464.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209-1222.

Carr, S.A., Huddleston, M.J., and Bean, M.F. (1993). Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. Protein Sci *2*, 183-196.

Chalkley, R.J., and Burlingame, A.L. (2001). Identification of GlcNAcylation sites of peptides and alpha-crystallin using Q-TOF mass spectrometry. J Am Soc Mass Spectrom *12*, 1106-1113.

Chalkley, R.J., and Burlingame, A.L. (2003). Identification of novel sites of O-N-acetylglucosamine modification of serum response factor using quadrupole time-of-flight mass spectrometry. Mol Cell Proteomics *2*, 182-190.

Chalkley, R.J., Thalhammer, A., Schoepfer, R., and Burlingame, A.L. (2009). Identification of protein O-GlcNAcylation sites using electron transfer dissociation

mass spectrometry on native peptides. Proc Natl Acad Sci U S A *106*, 8894-8899.

Darula, Z., Chalkley, R.J., Baker, P., Burlingame, A.L., and Medzihradszky, K.F. (2010). Mass spectrometric analysis, automated identification and complete annotation of O-linked glycopeptides. Eur J Mass Spectrom (Chichester, Eng) *16*, 421-428.

Ding, J., Xu, H., Faiola, F., Ma'ayan, A., and Wang, J. (2012). Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res *22*, 155-167.

Engelen, E., Akinci, U., Bryne, J.C., Hou, J., Gontan, C., Moen, M., Szumska, D., Kockx, C., van Ijcken, W., Dekkers, D.H.*, et al.* (2011). Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. Nat Genet *43*, 607-611.

Esch, D., Vahokoski, J., Groves, M.R., Pogenberg, V., Cojocaru, V., Vom Bruch, H., Han, D., Drexler, H.C., ArauzO-Bravo, M.J., Ng, C.K.*, et al.* (2013). A unique Oct4 interface is crucial for reprogramming to pluripotency. Nat Cell Biol *15*, 295-301.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. Cell Stem Cell *11*, 769-782.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595-601.

Ho, L., Ronan, J.L., Wu, J., Staahl, B.T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A.I., Ranish, J., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. Proc Natl Acad Sci U S A *106*, 5181-5186.

Jang, H., Kim, T.W., Yoon, S., Choi, S.Y., Kang, T.W., Kim, S.Y., Kwon, Y.W., Cho, E.J., and Youn, H.D. (2012). O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell *11*, 62-74.

Jebanathirajah, J., Steen, H., and Roepstorff, P. (2003). Using optimized collision energies and high resolution, high accuracy fragment ion selection to improve glycopeptide detection by precursor ion scanning. J Am Soc Mass Spectrom *14*, 777-784.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, A., Wahl, G.M., and Izpisua Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140-1144.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc *1*, 729-748.

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q*., et al.* (2010). A mesenchymal-tO-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell *7*, 51-63.

Liber, D., Domaschenz, R., Holmqvist, P.H., Mazzarella, L., Georgiou, A., Leleu, M., Fisher, A.G., Labosky, P.A., and Dillon, N. (2010). Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. Cell Stem Cell *7*, 114-126.

Maekawa, M., Yamaguchi, K., Nakamura, T., Shibukawa, R., Kodanaka, I., Ichisaka, T., Kawamura, Y., Mochizuki, H., Goshima, N., and Yamanaka, S. (2011). Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1. Nature *474*, 225-229.

Marion, R.M., Strati, K., Li, H., Murga, M., Blanco, R., Ortega, S., Fernandez-Capetillo, O., Serrano, M., and Blasco, M.A. (2009). A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. Nature *460*, 1149-1153.

Medzihradszky, K.F., Gillece-Castro, B.L., Settineri, C.A., Townsend, R.R., Masiarz, F.R., and Burlingame, A.L. (1990). Structure determination of O-linked glycopeptides by tandem mass spectrometry. Biomed Environ Mass Spectrom *19*, 777-781.

Medzihradszky, K.F., Gillece-Castro, B.L., Townsend, R.R., Burlingame, A.L., and Hardy, M.R. (1996). Structural elucidation of O-linked glycopeptides by high energy collision-induced dissociation. J Am Soc Mass Spectrom *7*, 319-328.

Myers, S.A., Daou, S., Affar el, B., and Burlingame, A. (2013). 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 *13*, 982-991.

Myers, S.A., Panning, B., and Burlingame, A.L. (2011). 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 *108*, 9490-9495.

Nowling, T.K., Johnson, L.R., Wiebe, M.S., and Rizzino, A. (2000). Identification of the transactivation domain of the transcription factor Sox-2 and an associated cO-activator. J Biol Chem *275*, 3810-3818.

O'Donnell, N., Zachara, N.E., Hart, G.W., and Marth, J.D. (2004). Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. Mol Cell Biol *24*, 1680-1690.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Pardo, M., Lang, B., Yu, L., Prosser, H., Bradley, A., Babu, M.M., and Choudhary, J. (2010). An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. Cell Stem Cell *6*, 382-395.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-tO-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64-77.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Seipert, R.R., Dodds, E.D., and Lebrilla, C.B. (2009). Exploiting differential dissociation chemistries of O-linked glycopeptide ions for the localization of mucin-type protein glycosylation. J Proteome Res *8*, 493-501.

Shafi, R., Iyer, S.P., Ellies, L.G., O'Donnell, N., Marek, K.W., Chui, D., Hart, G.W., and Marth, J.D. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc Natl Acad Sci U S A *97*, 5735-5739.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Swaney, D.L., Wenger, C.D., Thomson, J.A., and Coon, J.J. (2009). Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. Proc Natl Acad Sci U S A *106*, 995-1000.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Torres, C.R., and Hart, G.W. (1984). Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem *259*, 3308-3317.

Trinidad, J.C., Barkan, D.T., Gulledge, B.F., Thalhammer, A., Sali, A., Schoepfer, R., and Burlingame, A.L. (2012). Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. Mol Cell Proteomics.

van den Berg, D.L., Snoek, T., Mullin, N.P., Yates, A., Bezstarosti, K., Demmers, J., Chambers, I., and Poot, R.A. (2010). An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell *6*, 369-381.

Yang, Y.R., Song, M., Lee, H., Jeon, Y., Choi, E.J., Jang, H.J., Moon, H.Y., Byun, H.Y., Kim, E.K., Kim, D.H.*, et al.* (2012). O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. Aging Cell *11*, 439-448.

Zhao, P., Viner, R., Teo, C.F., Boons, G.J., Horn, D., and Wells, L. (2011). Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. J Proteome Res *10*, 4088-4104.

Materials and Methods

Cell Culture

Mouse embryonic stem cells (ESCs) were routinely passaged by standard methods in ESC media (KO-DMEM, 10% FBS, 2 mM glutamine, 1X non-essential amino acids, 1x pencillin/streptomycin, 0.1 mM β-mercaptoethanol and recombinant leukemia inhibitory factor). PlatE cells were cultured in ESC media without LIF and supplemented with 1 ug/uL puromycin (Invitrogen) and 10 ug/mL blasticidin S (Invitrogen). Mouse embryonic fibroblasts (MEFs) were cultured in ESC media without LIF for less than five passages.

Plasmids

pMXs-mouse *Sox2* wild-type or *S248A*, along with pMXs-*Oct4, Klf4* and *c-Myc* were used (Okita et al., 2007; Takahashi and Yamanaka, 2006). The pMXs-3xF-*Sox2* was created by cloning *Sox2* into pCMV-3xFLAG 7.1 (Sigma) and then subcloning 3xF-Sox2 via InFusion cloning (Clonetech) into pMXs. The S248A mutation was created with Quikchange site directed mutagenesis (Stratagene) (Primers, Table S2).

Antibodies and alkaline phosphatase

Antibodies were purchased from Abcam (SOX2 ab75179, TUBULIN GTU-88 ab11316, *O*-GlcNAc ab2739, OCT4 ab19857), Reprocell (NANOG, RCAB002P-F) and SIGMA (OGT DM-17 and FLAG A8592). Secondary antibodies were purchased from BioRad (172-1019 & 172-1011). Anti-FLAG magnetic beads

were purchased from SIGMA (M8823). Alkaline phosphatase activity staining was performed according to manufactures instructions (Stemgent).

Reprogramming Experiments

MEFs were derived from wild-type CD1 mice or the *Nanog-GFP*-IRES-*Puro*^{r} mice (Okita et al., 2007) and cultured in MEF media (KO-DMEM, 10% FBS, 2 mM glutamine, 1X non-essential amino acids, 1X pencillin/streptomycin, and 0.1 mM β-mercaptoethanol). pMXs vectors containing *Oct4*, *c-Myc*, *Klf4, eGfp*, *dsRed*, or wild-type or S248A *Sox2*, with or without FLAG, were transfected with Fugene 6 (Promega, E2691) into PlatE cells. Twenty-four hours after transfection, the media was changed. The next day, the retroviral supernatant was collected from transfected PlatE cells, filtered through 0.4 um filters and combined with each other at equal ratios. Polybrene (Millipore, TR-1003-G) was added to a final concentration of 4 ug/ml. The virus-containing media was added to *Nanog-Gfp* or wild-type MEFs that were passaged less than five times. Media was replaced the next day and every other until six days after infection. At day six, MEFs were trypsinized and either prepared for experiments or 1000 cells were plated onto gamma-irradiated SNL feeders. These 1000 MEFs were cultured in MEF media supplemented with recombinant leukemia inhibitory factor (ESC media) until GFP^{+} colonies were counted at day 20 (*Nanog-Gfp* MEFs) or cytospun onto glass slides and stained for NANOG, FLAG or alkaline phosphatase (CD1 MEFs).

Microinjection of iPSCs to generate chimera mice was conducted at Cornell University Stem Cell and transgenic core facility. iPSCs were grown on mouse embryonic fibroblasts (produced at the Cornell stem cell core) and mitotically inactivated by irradiation (3000 Rads). To produce donor embryos, wild-type albino mice of the strain http://jaxmice.jax.org/strain/000058.html were mated, embryos were flushed from the uterus at day 3.5, and the iPSCs were injected into the blastocyst of each embryo (15-30 cells per embryo). Injected embryos were then transferred to 2.5-day pseudo pregnant recipient animals and pup chimaerism was determined by coat color. Chimeras were mated to age-matched wild-type animals of the same albino strain used for embryo donors. iPSC contribution to the germline was determined by coat color of the resultant pups.

3xF-SOX2 Purification

Two 10 cm^2 dishes of CD1 MEFs infected with $OS^{FLAG\text{-}WT}KM$ or $OS^{FLAG\text{-}S248A}KM$ for six days were harvested by trypsinization, washed once with cold PBS and frozen in liquid N_2. Whole cell pellets were lysed in RIPA buffer without SDS, containing 500nM Thiamet G (Caymen Chemicals), 1X HALT protease and phosphatase inhibitors (Pierce), 2 mM TCEP (Sigma) and 20 mM *N*-ethylmaleimide (Sigma) and sonicated (with a probe sonicator on methanol ice for three rounds of pulses, 3 secs on, 2 off, 10 seconds total, at 35%). Anti-FLAG-based purifications were performed with anti-FLAG M2 Dynabeads (Sigma). Whole cell lysates were incubated with M2 beads at room temperature for 75 minutes, washed once with lysis buffer and three times with 25mM

ammonium bicarbonate with 150mM NaCl. Proteins were eluted with 100 mM glycine pH 4. Western blot and SDS-PAGE analysis was used to assess purification efficiency. OCT4 was purified in the same manner but from four 10 cm2 dishes of ZHBTc4 F-Oct4 cells grown on γ-irradiated feeder cells.

Mass spectrometric analysis

Silver stained SDS-PAGE gel bands were excised and digested in-gel with sequence grade trypsin (Roche). After 5% formic acid/50% acetonitrile extraction, peptides were dried by vacuum centrifugation, gel particulates were removed via C18 ZipTips (Millipore), dried, resuspended in 0.1% formic acid and analyzed by LC MS/MS. Chromatography was performed on a Nanoacquity HPLC (Waters) at 400 nL/min with a BEH130 C18 75 μm ID x150mm column (Waters). A 90-minute gradient from 98% solvent A (0.1% formic acid) to 22% solvent B (0.1% formic acid in acetonitrile) was used. Peptides were analyzed by an LTQ-Orbitrap Velos mass spectrometer (Thermo). After the survey scan of m/z 400-1,600 was measured in the Orbitrap at 30,000 resolution, the top three multiply charged ions were selected for both HCD and ETD. Automatic gain control for MS/MS was set to 2000. Normalized collision energy for HCD was set at 35 while the ETD activation time was charge state dependent, based on 100 ms for doubly charge precursors. Supplemental activation was implemented for ETD reactions. Dynamic exclusion of precursor selection was set for 25 seconds.

Data analysis

Fragment mass spectra were converted into peaklists using the in-house software PAVA. HCD and ETD data were searched separately using ProteinProspector version 5.10.0 against the UniProt database with a concatenated database. Only mouse and human genomes were used for the database searching. Precursor tolerance was set to 10 ppm, whereas fragment mass error tolerance was set to 0.6 Da for ETD and 20 ppm for HCD. *N*-terminal acetylation, methionine oxidation, loss of *N*-terminal methionine and glutamate conversion to pyroglutamate were allowed as variable modifications. For ETD data, HexNAc modifications to serine and threonine residues and phosphorylation to serine/threonine/tyrosine was allowed as variable mass modifications. For HCD, phosphorylation was searched the same way though HexNAc was considered as a neutral loss. Methylation (mono, di- and tri-) of K and R, monomethylation of D, E and H (artifact from MeOH fixing PAGE gels), acetylation of K and R, and ADP-ribosylation to C, E, K, N, S, and R were searched separately. SLIP scoring was used to distinguish possible positional isomers of HexNAc and/or phosphopeptides (Baker et al., 2011) and were verified manually (Medzihradszky, 2005). Relative abundances of each modified or unmodified peptide were calculated using the ICIS area calculated from XICs in Xcalibur (Thermo) at a 10 ppm mass tolerance.

RNA extraction and RT-PCR and RT-qPCR

Total RNA from CD1 MEFs infected with $OS^{FLAG\text{-}WT}KM$, $OS^{FLAG\text{-}S248A}KM$ or *eGfp* after 6, 8 and 12 days were extracted using Trizol reagent (Invitrogen,Carlsbad,

CA) according to the manufacturer's instructions.1 µg of total RNA was used for cDNA synthesis with oligo$(dT)_{18}$ primers using Tetro cDNA synthesis kit (Bioline, BIO-65042) and the resulting cDNA was diluted to 20ng/µl. Expression levels of *Nanog*, *Esrrb*, *Sall4* and *Utf1 were* analyzed using real-time, quantitative PCR carried out using CFX Connect™ Real-time PCR detection system (Biorad) and SensiFast™ SYBR Lo-ROX PCR master mix (Bioline, BIO-94020). Expression data were normalized to the mean Cq of housekeeping gene Glucuronidase, beta(*Gusb*) to control the variability in expression levels and the relative quantification in gene expression was determined using the $2^{-(\Delta Cq)}$ method (ΔCq= Cq(gene)-Cq(*Gusb*) (Livak and Schmittgen 2001).The primer sequences used are provided in Table S2.

Total RNA from iPSCs reprogrammed with $OS^{FLAG\text{-}WT}KM$ ($SOX2^{WT}$) or $OS^{FLAG\text{-}S248A}KM$ ($SOX2^{S248A}$) or ESC control was extracted using Trizol reagent (Invitrogen,Carlsbad, CA) according to the manufacturer's instructions.1 µg of total RNA was used for cDNA synthesis with random oligo$(dT)_{18}$ primers using Tetro cDNA synthesis kit (Bioline, BIO-65042) and the resulting cDNA was diluted to 20ng/µl. RT-PCR reactions were optimized to 95ºC for 1 min, 28 amplification cycles at 95ºC for 15 s, 58ºC for 15 s, 72ºC for 15 s, and a final extension of 5 min at 72ºC using 2x MyFi PCR master mix (Bioline, BIO-25049). Amplified products were resolved on 2% agarose gels and visualized by ethidium bromide staining. The primer sequences used are provided in Table S2.

RNA-seq library preparation

Sample library preparation was performed according to standard protocols from the SABRE Functional Genomics Core (http://www.arrays.ucsf.edu). Total RNA quality was assessed by spectrophotometer (NanoDrop, Thermo Fisher Scientific Inc., Waltham, MA) and the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Cytoplasmic and mitochondrial ribosomal RNA was removed from the samples using Epicentre Ribo-Zero Gold (Human/Mouse/Rat) kit (Epicentre® Biotechnologies, Madison, WI). 2.3) and RNA sequencing libraries were generated using the TruSeq RNA sample prep kits with multiplexing primers, according to the manufacturer's protocol (Illumina, San Diego, CA). Fragment size distribution was assessed using the Bioanalyzer 2100 and the DNA high-sensitivity chip. Library concentrations were measured using KAPA Library Quantification Kits (Kapa Biosystems, Inc., Woburn, MA) and equal amounts of indexed libraries were pooled and sequenced on the Illumina HiSeq 2500.

RNA-seq analysis

We obtained 172,287,640 single-end, 50 base pair reads across the 6 samples (n=2 per group) from 2 lanes of the HiSeq 2500. Sequence alignment and splice junction mapping were performed using Bowtie2 (Langmead and Salzberg, 2012) and TopHat (Trapnell et al., 2009) respectively, using the latest version of the mouse genome (mm10) provided by UCSC (Waterston et al., 2002). The subset of transcriptomic mappings were tabulated on a per-gene basis and subsequently served as raw input for DESeq (Anders and Huber, 2010). DESeq

normalizes by size factor (reads per sample) as well as by library complexity and then utilizes an exact test type statistic (employing a negative binomial model of dispersion, consistent with the type of variance seen in RNAseq datasets) to assess differential expression.

Chromatin Immunoprecipitation (ChIP)

$1x10^7$ CD1-MEFs infected with $OS^{FLAG\text{-}WT}KM$, $OS^{FLAG\text{-}S248A}KM$ or *eGfp* after six days were crosslinked with 1% formaldehyde for 10 min at 37^{o}C. After quenching with 125 mM glycine for 5 min, the cells were washed twice with ice cold 1xPBS. Cell pellets were resuspended in 1ml lysis buffer 1 (50 mM HEPES·KOH (pH 7.6),140 mM NaCl,1 mM EDTA ,10 %(v/v) Glycerol ,0.5 % NP-40,0.25 % Triton X-100, 1X complete protease inhibitor cocktail (PIC) (Roche,11697498001) for 10 min at 4 ^{o}C, followed by spin at 1,350 x g for 5 min at 4^{o}C. The supernatant was discarded and the pellet was resuspended in 1ml of lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1X PIC) for 10 min in at 4 ^{o}C, followed by spin at 1,350 x g for 5 min at 4 ^{o}C. Finally, the cell pellet was resuspended in 700 µl of lysis buffer 3 (10 mM Tris·HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Sodium deoxycholate, 0.5% N-Lauroylsarcosine, 1X PIC) and sonicated to desired length using Bioruptor (Diagenode) for 30 min [10 minutes x 3 times (each cycle consists of 30s on, 30s off) at high settings]. Lysates were clarified by centrifugation at 10,000 rpm in a microcentrifuge for 10min at 4 ^{o}C. The supernatant was transferred to a new tube and stored at -80 ^{o}C or used immediately for ChIP. 30 µl of Dynabeads Protein G

(Invitrogen, 10004D) were blocked with 5mg/ml BSA in 1X PBS followed by overnight coupling with 5 µg of FLAG antibody (Sigma, F1804) or 5 µg H3 antibody (abcam 1791) or 5ug of H3K4me2 antibody (abcam 7766). Antibody beads were added to 20µg of sonicated chromatin diluted three fold in ChIP dilution buffer +1% Triton X-100 (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1X PIC) and immunoprecipitated overnight at 4 ^{o}C. As 3xF-SOX2 was used in the reprogramming cocktail, FLAG antibody was used to isolate 3xF-SOX2 bound chromatin fragments, and ChIP against histone H3 on the same set of chromatin samples was performed as an internal control for FLAG and H3K4me2 ChIP. 10% of the undiluted chromatin from each sample is saved as input DNA. The beads were removed by placing the tubes in a magnetic separation rack and carefully remove the unbound chromatin. The beads were washed with 1ml of each of the following buffers at 4 ^{o}C, twice with ChIP dilution buffer, once with ChIP dilution buffer + 500 mM NaCl, four times with RIPA buffer (10 mM Tris-HCl (pH 8.0), 0.25 M LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Sodium deoxycholate) and two washes in 1x TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA). DNA was eluted from the beads twice in 100µl of ChIP elution buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 20 mM EDTA, 1% SDS) by incubating at 65 ^{o}C for 15 min. The eluted DNA and the input samples were then reverse crosslinked at 65 ^{o}C for overnight, followed by RNase A (0.2 mg/ml) digestion at 37 ^{o}C for 2 h and Proteinase K (0.2 µg/µl) (NEB, P8102S) digestion at 55 ^{o}C for 1 h. The ChIP and input DNA was recovered by phenol-chloroform extraction and ethanol precipitation. Enrichment of SOX2, or H3K4me2 and H3 bound DNA was

assayed by ChIP-quantitative PCR (ChIP-qPCR) using CFX Connect™ Real-time PCR detection system (Biorad) and SensiFast™ SYBR Lo-ROX PCR master mix (Bioline, BIO-94020). Fold enrichment of immunoprecipitated DNA over input were calculated using $2^{-(\Delta Cq)}$ method (ΔCq= Cq(IP)-Cq(Input)) (Livak and Schmittgen, 2001). Primer pairs used in FLAG-SOX2 and H3K4me2 ChIPs were designed using published SOX2 and H3K4me2 ChIP-seq data sets (GSE 35496 (Lodato et al., 2013) GSM 640755 (Koche et al., 2011). Sequences of the primers used in ChIP-qPCR are listed in Table S2.

DNase qPCR

After six days of infection, $3x10^6$ *eGfp*, $OS^{WT}KM$ or $OS^{248A}KM$ MEFs from the same preparation used for the final three ChIP samples were washed once in PBS and transferred to cold dounce homogenizers with cold swelling buffer (10 mM TRIS, pH 8, 85 mM KCl, 0.5% NP-40, 1X HALT and 500 nM Thiamet G). Cells were dounced and nuclei were examined for Trypan blue exclusion. Three volumes of DNase buffer (15 mM TRIS, pH 7.5, 300 mM sucrose, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 1X HALT and 500 nM Thiamet G) was added to transfer the nuclei to 1.5 mL tubes. The nuclei were pelleted at 1000 rpm for 5 minutes at 4°C and resuspended in 200 uL DNase buffer with 5 U DNase I (Invitrogen). Tubes were incubated for 2 or 16 minutes at 37°C. The reaction was stopped by addition of EDTA to 10 mM and SDS to 1%. Nuclei were treated with RNase A (0.2 mg/ml) for one hour at 37°C followed by Proteinase K (Bioline) for 2 hours at 55°C. DNA was phenol:choloroform extracted, ethanol precipitated and

further cleaned up with Qiagen PCR clean up columns. qPCR Cq values were measured the same as with ChIP. Melting curves were checked and only primer set yielding a single PCR product were used for analysis. Primer sets used are listed in Table S2. Fold changes was calculated using $2^{-(\Delta Cq)}$ method (ΔCq= Cq(16 min time point)-Cq(two min time point)).

OCT4 glycopeptide synthesis and analysis

The glycopeptide sequence reported in (Jang et al., 2012). was synthesized to 70% purity by Pierce Biotechnology. The lyophilized peptide was resuspended in LC grade water. The peptide was diluted to 500 attomols/μl with 0.1% formic acid and five μl was analyzed by LC MS/MS. The precursor m/z was measured in the Orbitrap while CAD was performed and measured in the ion trap. The data was interpreted manually.

References associated with Materials and Methods

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

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

Jang, H., Kim, T.W., Yoon, S., Choi, S.Y., Kang, T.W., Kim, S.Y., Kwon, Y.W., Cho, E.J., and Youn, H.D. (2012). O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell *11*, 62-74.

Koche, R.P., Smith, Z.D., Adli, M., Gu, H., Ku, M., Gnirke, A., Bernstein, B.E., and Meissner, A. (2011). Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell *8*, 96-105.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. PLoS Genet *9*, e1003288.

Medzihradszky, K.F. (2005). Peptide sequence analysis. Methods Enzymol *402*, 209-244.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313-317.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell *151*, 994-1004.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

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

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Table 1 SOX2 TAD peptides from MEFs six days after infection with $OS^{FLAG\text{-}WT}KM$ or $OS^{FLAG\text{-}S248A}KM$. Unmodified and all unambiguous positional isomers of *O*-GlcNAc modified or phosphorylated TAD peptides are reported. Scoring and expectation values are reported for all peptides. Direct HCD link: *http://prospector2.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=g85xarndhe&search_name=msviewer* . Direct ETD link: http://prospector2.ucsf.edu/prospector/cgi-bin/mssearch.cgi?report_title=MS-Viewer&search_key=pabb5khwbl&search_name=msviewer

Gene	m/z	z	ppm	SOX2 TAD Peptide	Score	Expect
SOX2	567.9413	3	3.2	SEASSSPPVVTSSSHSR	59.1	3.4e-11
SOX2	594.5953	3	0.66	SEASSS(Phospho)PPVVTSSSHSR	53.7	4.2e-11
SOX2	635.6326	3	0.022	S(HexNAc)EASSSPPVVTSSSHSR	51.5	3.9e-11
SOX2	635.6338	3	1.9	SEASSSPPVVT(HexNAc)SSSHSR	47.0	1.5e-9
S248A	562.6106	3	4.9	AEASSSPPVVTSSSHSR	50.6	1.2e-9
S248A	589.2652	3	3.3	AEASSS(Phospho)PPVVTSSSHSR	51.8	1.8e-9
S248A	630.3024	3	4.9	AEASSSPPVVT(HexNAc)SSSHSR	50.0	4.8e-9

Table 2 Primers used in this study

Gene	Primer	5'-3' Primer Sequence	Notes
ChIP primers			
Esrrb	F	TTGGGCCTCGAGTTCCTGTA	H3K4me2 ChIP
	R	CCCAGTCACCCTCTCAGGTA	
Esrrb	F	CGGCTGGTATCACCTGATTT	FLAG-SOX2 ChIP ,
	R	TGATCCTTTGGAGTGGAGGA	
Utf1	F	GGTCCTTAGAGCGTGTGTGG	H3K4me2 ChIP
	R	CATCCCTGAGCCAGGTAGAG	
Utf1	F	GGCTTAGGTGCAGGTAGAGG	FLAG-SOX2 ChIP
	R	GATGGGCCCAGAATTTGTAA	
Fgf4	F	TTGCGTCCCTATTTGCTCTC	H3K4me2 ChIP
	R	AGCAGCGTCCCTGTGGTC	
Nanog	F	AGGATGCCCCCTAAGCTTTC	FLAG-SOX2 ChIP
	R	TGAATTCACAGTTAATCCCACCT	
Sall4	F	GTAGCTGCTGAGGCTGCTC	H3K4me2 ChIP
	R	GCAAGTCACCAGGGCTCTT	
Sall4	F	GAATCCCCATTCCAGAATCGGC	FLAG-SOX2 ChIP
	R	CAATGCGCGCACCATGTCG	
Sox21	F	GAGGAGAGGGGAGAGCAGAA	H3K4me2 ChIP
	R	CCGGGACCCTGAAACTTCTC	
Sox21	F	TCTGCCTTGTAGATTGCATTCTAATGC	FLAG-SOX2 ChIP
	R	GGAATGAATAGCGGTGCAATGTGT	
Mll3	F	TGGCTCTCCTTGTCACCCTA	H3K4me2 ChIP
	R	ACCCATCCAGAGGAACAGGA	
Mll5	F	GTGCGGGGAGGGAGTAAATC	H3K4me2 ChIP
	R	GCCAGCGATTGCATAAGGTG	
Mutagenesis			
Sox2S248Af	F	ggtcGCCtccgaggccagctcca	
Sox2S248Ar	R	cggaGGCgaccacagagcccatggag	

Gene	Primer	5'-3' Primer Sequence	Notes
RT-PCR primers to detect viral and endogenous transcripts			
cMyc-total	F	CAGAGGAGGAACGAGCTGAAGCGC	Total (Viral +Endo)
	R	TTATGCACCAGAGTTTCGAAGCTGTTCG	
cMyc-endo	F	TGACCTAACTCGAGGAGGAGCTGGAATC	Endogenous
	R	AAGTTTGAGGCAGTTAAAATTATGGCTGAAGC	
Sox2-total	F	GGTTACCTCTTCCTCCCACTCCAG	Total (Viral +Endo)
	R	TCACATGTGCGACAGGGGCAG	
Sox2-endo	F	TAGAGCTAGACTCCGGGCGATGA	Endogenous
	R	TTGCCTTAAACAAGACCACGAAA	
Klf4-total	F	GTGCAGCTTGCAGCAGTAAC	Total (Viral +Endo)
	R	GGAAGACGAGGATGAAGCTG	
Klf4-endo	F	GATGGGAGCGGAGTTGTCCTA	Endogenous
	R	AAGGAAGGCGTTCCAGATTT	
Oct4-total	F	CAAATCGGAGACCCTGGTG	Total (Viral +Endo)
	R	CCTTCTCTAGCCCAAGCTGA	
Oct4-endo	F	TCTTTCCACCAGGCCCCCGGCTC	Endogenous
	R	TGCGGGCGGACATGGGGAGATCC	
RT-PCR primers to detect viral transcripts			
pMX-AS3200	R	TTATCGTCGACCACTGTGCTGCTG	used w/ all For primers
RT-qPCR primers for gene expression			
Esrrb-RT	F	ATCAATCACGGAGTCTGGATCT	
	R	TTCGGTTCAGCAGCATGGTT	
Utf1-RT	F	GTGGAGCAAGAGGCCGAG	
	R	CAGAGTGTCGGTGCTCGTAA	
Nanog-RT	F	CCAGTCCCAAACAAAAGCTC	
	R	ATCTGCTGGAGGCTGAGGTA	
Sall4-RT	F	GAACCCGGTGCTCCAGTGAA	
	R	GAACTCGGCACAGCATTTGT	
Gusb-RT	F	CCTTGGCTTTGTGAACTCTTG	
	R	GAACCTCCAGTAGAACAGTCA	

Chapter 5

O-GlcNAc regulates SOX2 activity in embryonic stem cells by altering protein-SOX2 interactions

O-GlcNAc regulates SOX2 activity in embryonic stem cells by altering protein-SOX2 interactions

Samuel A. Myers[1], Sailaja Peddada[2], Sean Thomas[3], Gregor Krings[2], Michael Lopez[1], Marena Trinidad[1], Barbara Panning[2] and A.L. Burlingame[1]

[1]Department of Pharmaceutical Chemistry, University of California San Francisco,
600 16th Street, Genentech Hall Suite N472A, San Francisco, CA 94158

[2]Department of Biochemistry and Biophysics, University of California San Francisco, 600 16th Street, Genentech Hall Suite S372B, San Francisco, California 94158

[3]Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA

Abstract

SOX2 is a versatile transcription factor that maintains embryonic stem cell (ESC) pluripotency and self-renewal, and is important for proper lineage specification and adult stem cell maintenance. This versatility is likely due to post-translational modifications (PTMs) as SOX2 has been reported to be modified by numerous chemical moieties in a variety of cell types. One such PTM is *O*-GlcNAc, the dynamic and regulatory glycosylation of intracellular proteins. Global *O*-GlcNAc is essential for ESC self-renewal though the function of SOX2 *O*-GlcNAcylation in ESC is not understood. Here, we show that SOX2 is highly abundant as *O*-GlcNAc modified in the transactivation domain at serine 248. Replacement of wild-type SOX2 with an *O*-GlcNAc deficient mutant SOX2 ($SOX2^{S248A}$) increases the pluripotency transcriptional network while down-regulating the differentiation network in ESCs. Removal of self-renewing signals induces changes in SOX2 *O*-GlcNAc stoichiometry, while OCT4 depletion triggers a stronger differentiation transcriptional response in $SOX2^{S248A}$ ESCs compared to WT. Analysis of SOX2-interacting proteins from ESCs revealed that the WT and mutant SOX2 interact with distinct subsets of transcriptional regulatory complexes. Thus, SOX2 *O*-GlcNAcylation controls the transcriptional landscape of a cell by modulating SOX2 activity and recruiting different regulatory complexes.

Introduction

SOX2 (sex determining region Y-box 2) is a transcription factor that is necessary for embryonic stem cell (ESC) self-renewal and is expressed in progenitor and adult stem cells (1, 2). Precise control of SOX2 is critical for embryonic stem cell (ESC) maintenance, since increased or decreased expression of SOX2 interferes with self-renewal and pluripotency (1, 3). Post-translational modification (PTM) of SOX2 may play a role in its regulation, as SOX2 has been reported to be phosphorylated, ubiquitinylated, SUMOylated, acetylated, methylated, and PARPylated (4-12).

We have previously identified SOX2 as modified by *O*-linked N-acetlyglucosamine (O-GlcNAc) in ESCs (13). *O*-GlcNAc is the dynamic and regulatory glycosylation of nucleocytosolic proteins and is catalyzed by a single *O*-GlcNAc transferase (OGT) and a single hydrolase (OGA). *O*-GlcNAc signaling is essential for embryo viability (14-16) and ESC self-renewal (17). While OGT and OGA are critical for the establishment and maintenance of pluripotency and self-renewal, the protein- and site-specific functions of *O*-GlcNAcylation in ESCs have yet to be explored in any substantial way.

Here, we show that SOX2 exists in several post-translationally modified states in ESCs and breast cancer cells. In depth characterization of SOX2 in ESCs reveals SOX2 is highly abundant as *O*-GlcNAc modified at serine 248. Replacement of wild-type SOX2 ($SOX2^{WT}$) with an *O*-GlcNAc deficient mutant

SOX2 ($SOX2^{S248A}$) in ESCs causes an increase in pluripotency gene expression, a down-regulation of genes involved in differentiation and morphological changes consistent with a more naive pluripotent state (18, 19). $SOX2^{S248A}$ shows only minor changes in its genomic distribution compared to $SOX2^{WT}$ whereas there is an increase in $SOX2^{S248A}$ occupancy at distal regulatory regions for genes differentially expressed between the WT and mutant ESCs. This change in gene expression can be attributed to the differential recruitment of transcriptional regulatory complexes between WT and mutant SOX2. While the self-renewal pathway is upregulated in unperturbed $SOX2^{S248A}$ mutant ESCs, we show that OCT4 depletion can simultaneously activate differentiation genes to a greater extent when compared to $SOX2^{WT}$ ESCs. This study reveals a new mechanism by which SOX2 is post-translationally regulated, and provides new insight as to how *O*-GlcNAc modulates pluripotency and transcription factor function.

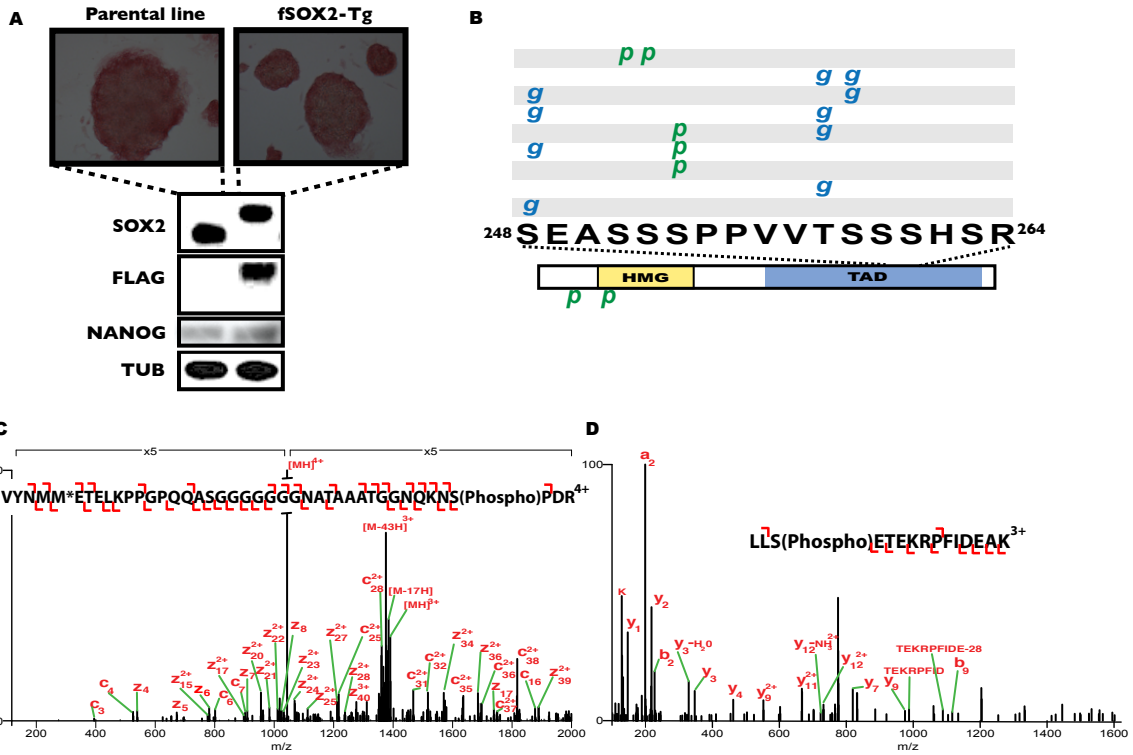
Results

GlcNAc-S248 is, among many, the predominant SOX2 PTM form

SOX2 has been reported to be heavily post-translationally modified in a variety of cell types (4-13). To characterize the SOX2 PTMs, we created a murine ESC line that expresses a triply FLAG tagged SOX2 (*3xF-Sox2*) from a cDNA transgene as the sole source of *Sox2* (Figure 1A). These mESCs, denoted as fSOX2-Tg, facilitated SOX2 purification unbiased by the presence of PTMs. fSOX2-Tg cells exhibit normal ESC colony morphology and alkaline phosphatase (AP) staining, and have SOX2 and NANOG protein levels comparable to the parental control line (1), indicating the tagged, transgenic SOX2 exhibits normal function (Figure 1A).

We characterized purified SOX2 using liquid chromatography tandem mass spectrometry (LC-MS/MS), employing both higher energy collisional dissociation (HCD) and electron transfer dissociation (ETD). We identified a complex PTM pattern in the transactivation domain (TAD) of 3xF-SOX2, with multiple *O*-GlcNAcylation sites (S248, T258, and S259) in single or pairwise combinations, and two phosphorylation sites (S252 and S253) occurring singly or in combination and present with and without *O*-GlcNAcylation (Figure 1B). We also identified two phosphorylation sites (S39 and S85) in the high mobilitiy group (HMG) DNA binding domain of SOX2 (Figure 1 C-D**)**. We focused our analyses on the TAD of SOX2 because several reports have identified various PTMs in this

Figure 1 SOX2 possesses a diverse set of PTMs in ESCs. **A**, fSOX2-Tg cells were created by introducing the 3xF-SOX2^{WT} into the parental, Tet-off Sox2 TS22C mESC line (1). AP staining and colony morphology of fSOX2-Tg cells compared to parental line. Western blots against SOX2, FLAG, NANOG and TUBULIN. **B**, Diagram of SOX2 and the PTMs identified from ESCs. Each line (above) represents a TAD PTM form identified. g, *O*-GlcNAc; p, phosphorylation; HMG, high-mobility group DNA binding domain; TAD, transactivation domain. **C**, ETD mass spectrum of SOX2 phospho-S39. * indicates oxidation. **D**, HCD mass spectrum of SOX2 phospho-S85.



region in ESCs (4, 13, 20) or in other heterologous cell types (8, 12). We did not detect acetylation, methylation or PARPylation of SOX2 (5, 6, 10, 11). We also were unable to detect any evidence of SUMOylation by LC-MS/MS or Western blotting (8, 12) (not shown).

To determine the stoichiometries of the TAD PTM states of SOX2, we compared extracted ion chromatograms (XICs) from the LC-MS measurements. The singly *O*-GlcNAcylated form was the predominant PTM form (42 ± 2 %) of the total SOX2 TAD peptide (Figure 2A). The next most abundant form was the unmodified peptide (36 ± 2 %), followed by the singly phosphorylated peptides (11 ± 2 %) and the peptides containing a single *O*-GlcNAcylation and single phosphorylation (7 ± 1 %). Finally, the doubly *O*-GlcNAcylated (3.7 ± 0.5%) and the doubly phosphorylated (0.5 ± 0.2%) peptides were the least abundant. This high stoichiometry of the singly *O*-GlcNAcylated species was also observed in an independently-derived HA tagged-SOX2 ESC line (haSOX2-Tg), indicating that abundant *O*-GlcNAcylation of the SOX2 TAD peptide occurs across ESC lines (Figure 3A-B). Singly *O*-GlcNAcylated and unmodified synthetic TAD peptides chromatographically separated and had comparable ionization efficiencies (Figure 3C), indicating that accurate stoichiometry of the singly *O*-GlcNAcylated SOX2 peptides can be determined by LC-MS.

The SOX2 TAD peptide population contained three residues identified as *O*-GlcNAc modified, S248, T258, or S259. To determine at which site *O*-GlcNAc

Figure 2 Stoichiometry determination of SOX2 GlcNAc-S248 in ESCs. **A**, Fractions of total TAD peptide PTM forms detected and quantified from 3xF-SOX2, n=4 +/- S.E.M**. B**, Summed ETD mass spectra for targeted analysis of the singly O-GlcNAcylated, triply charged TAD peptide. Mass spectra containing the charged reduced species were summed. XICs of the $z^{\bullet}+1_{15}^{+}$ product ions for the positional isomers (below) show the GlcNAc-S248 position approximately three times more abundant that the other two positional isomers combined.

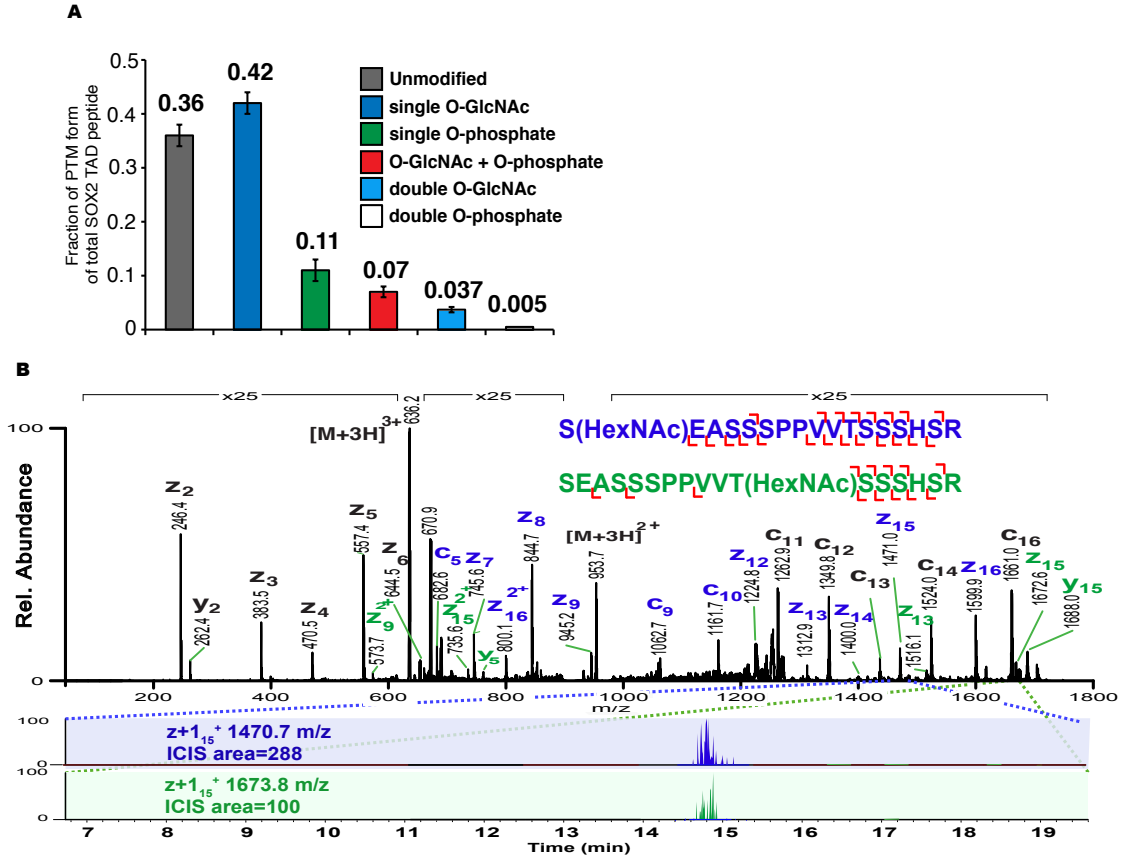
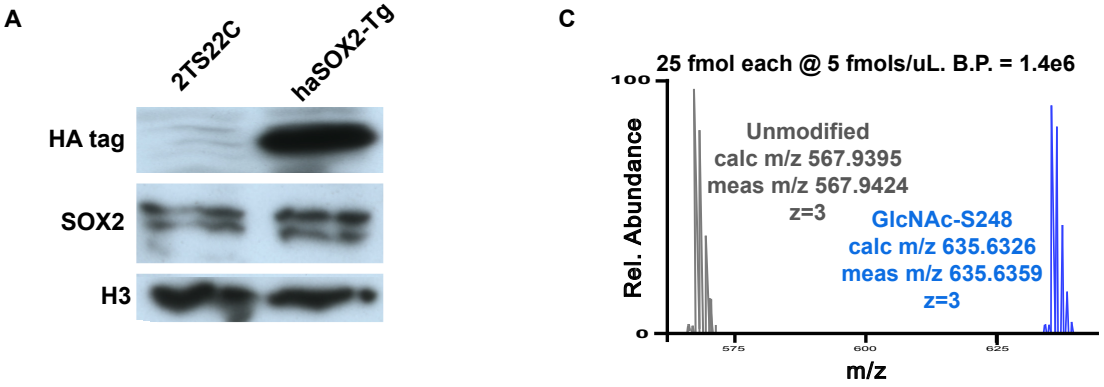


Figure 3 High *O*-GlcNAc stoichiometry is consistent across independently derived ESC lines. **A,** Characterization of haSOX2-Tg cells. Western blots against HA tag, SOX2 and TUBULIN. **B,** Peptide identifications and corresponding precursor intensities of the GlcNAc-S248 and unmodified SOX2 TAD peptide purified from haSOX2-Tg cells. Intesities of XICs show a similar stoichiometry to Figure 2A. **C,** Example of equal amounts of synthetic unmodified and GlcNAc-S248 SOX2 TAD peptides analyzed by LC-MS. Summed MS1 scans show that the modified and unmodified electrospray with equal efficiencies.



B

SOX2-P48431

m/z	z	ppm	DB Peptide	Variable Mods	Intensity	Area	Score	Expect	# in DB
635.6363	3	5.8	SEASSSPPVVTSSSHSR	HexNAc@1=18	27646	798.8	46.1	1.9e-9	1
567.9421	3	4.7	SEASSSPPVVTSSSHSR		24720	699.8	42.8	1.9e-7	1

is the most abundant, we performed targeted ETD of the singly *O*-GlcNAc modified SOX2 TAD peptide (Figure 2B). XICs of the $z^{\bullet}+1_{15}^{+}$ product ion, which distinguishes positional isomers, allowed us to estimate that GlcNAc-S248 is approximately three times more abundant than combined total of *O*-GlcNAcylation at either T258 or S259. *In vitro* glycosylation of recombinant mouse SOX2 by recombinant human OGT (99% identical to mouse OGT) resulted in TAD peptide *O*-GlcNAc modified exclusively at S248 (data not shown). Together, these data indicate that SOX2 is a direct substrate of OGT and that S248 is the major site of SOX2 *O*-GlcNAcylation in ESCs.

SOX2 PTMs respond to extracellular cues and are altered in breast cancer cells

Although *Sox2* is expressed during initial differentiation stages and in progenitor/multipotent stem cells, these cells receive different extracellular cues and possess a different transcriptional landscape compared to ESCs (2, 21, 22). To test whether SOX2 PTMs are responsive to cytokines, we transferred fSOX2-Tg cells to chemically defined growth media lacking the self-renewal cues leukemia inhibitory factor (LIF) and bone-morphogenic protein (BMP), and analyzed the PTM profile of the TAD peptide. 3xF-SOX2 purified from fSOX2-Tg cells cultured in N2B27 without LIF or BMP for 48 hours showed decreases in the abundance of the *O*-GlcNAc modified species while simultaneously showing increases in the unmodified and singly phopshorylated versions (Figure 4).

These results reveal that SOX2 PTMs respond to extracellular cues, and suggesting a mode of regulation of SOX2 during differentiation (23).

Sox2 expression is also associated with the dedifferentiation phenotype of cancer cells and cancer stem cells, and is often associated with poor prognosis of various tumors (24-26). We, therefore, asked whether SOX2 is *O*-GlcNAc modified in breast cancer cell lines. Analysis of 3xF-SOX2 purified from transiently transfected MCF7 or MDA-MB-453 breast cancer cells showed that SOX2 was indeed *O*-GlcNAc modified and/or phosphorylated within the TAD peptide (Figure 5). Comparison of XICs of the SOX2 TAD between the two breast cancer cell lines showed no notable differences in the relative abundances of any modified species (not shown). However, comparison of the SOX2 TAD PTM profiles from 3xF-SOX2 purified from either breast cancer cell line to the tagged mESCs showed differences in the relative abundances between PTM forms (Figure 5). SOX2 expressed in breast cancer cells showed higher relative levels of the singly *O*-GlcNAcylated, singly phosphorylated and the co-modified singly *O*-GlcNAc/singly phosphorylated TAD peptides when compared to fSOX2-Tg cells (Figure 5). These data indicate SOX2 is *O*-GlcNAc modified in breast cancer cells, suggesting that SOX2 is differentially regulated in cancer cells and ESCs.

Figure 4 Removal of self-renewing cytokines alters SOX2 TAD PTM profile. **A**, XICs comparing the ratios between the singly *O*-GlcNAcylated and the singly phosphorylated SOX2 TAD peptides (top) or the singly O-GlcNAcylated and the unmodified TAD peptides from fSOX2-Tg cells grown in standard culture conditions. **B**, XICs comparing the ratios between the singly O-GlcNAcylated and the singly phosphorylated SOX2 TAD peptides (top) or the singly O-GlcNAcylated and the unmodified TAD peptides from fSOX2-Tg cells grown in in N2B27 media for 48 hours.

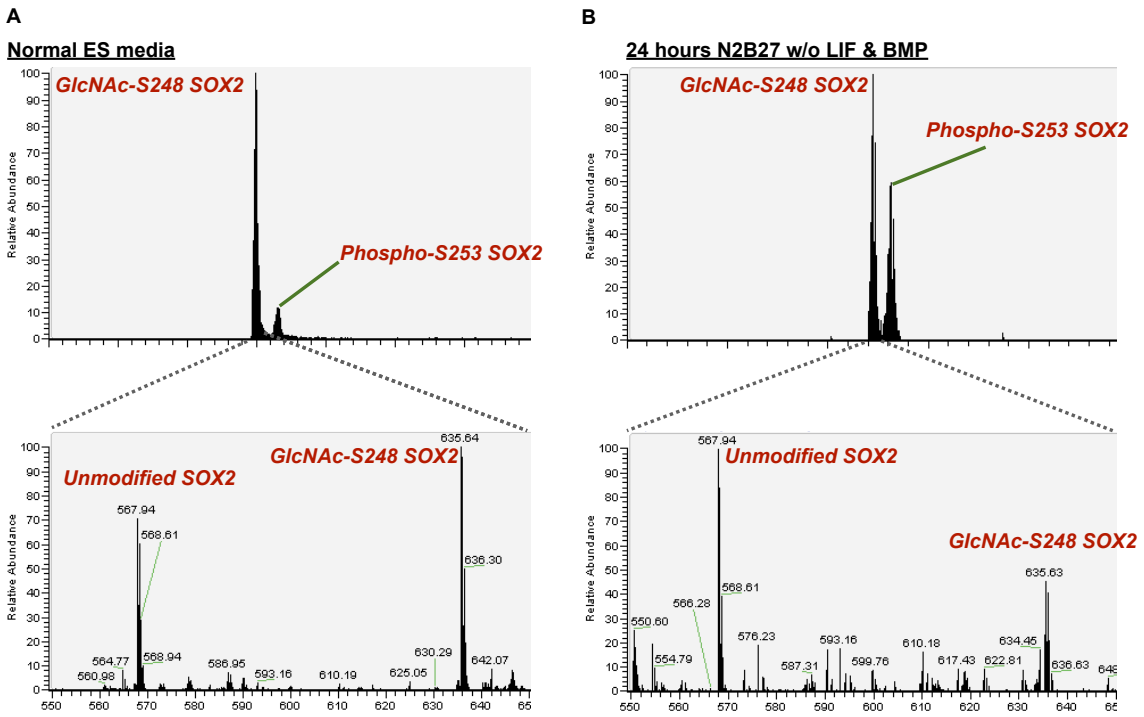
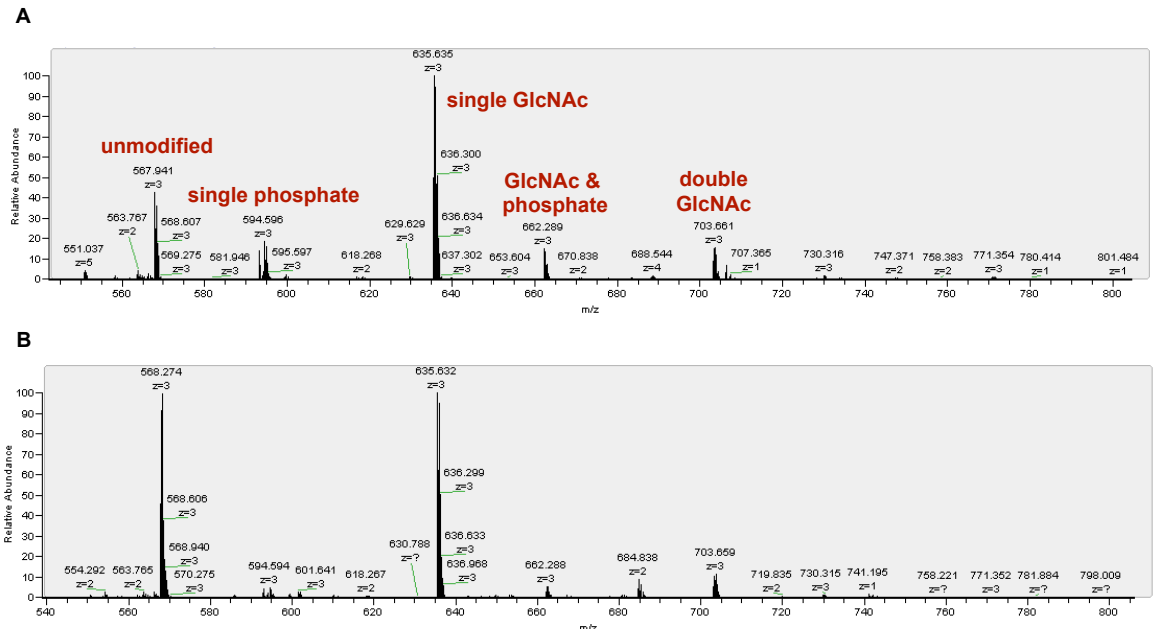


Figure 5 SOX2 TAD PTM patterns differ between ESCs and breast cancer cells. **A**, Summed MS1 scans over the elution window of the five most abundant SOX2 TAD post-translationally modified peptide forms. Analysis of the SOX2 TAD peptides purified from breast cancer cells (top) or from fSOX2-Tg cells (bottom) shows differences in PTM profiles.



An *O*-GlcNAc-deficient SOX2, SOX2^{S248A}, can replace WT in ESCs

Since the SOX2 TAD possesses a diverse set of PTMs, where GlcNAc-S248 is the most abundant, we investigated if this residue affected SOX2 function. We generated an S248 to alanine mutant SOX2 ESC line (fS248A-Tg cells) using the same strategy as that used to create fSOX2-Tg line. fS248A-Tg cells posses ESC-like morphology and stain positive for AP activity (Figure 6A) indicating that SOX2^{S248A} can replace WT in ESCs.

Mass spectrometric analysis showed the S248A mutation resulted in the loss of the bulk of TAD *O*-GlcNAcylation (Figure 6B), confirming that S248 is the major *O*-GlcNAc site in ESCs. The S248A mutation did not affect subcellular localization of SOX2 (Figure 7A)**.** Knockdown of OGT, which causes loss of cellular *O*-GlcNAc and major changes in ESC colony morphology, did not affect the steady state levels of SOX2 (Figure 7B). The loss of the major *O*-GlcNAcylated SOX2 species in fS248A-Tg cells also showed a dramatic increase of S253 phosphorylation (Figure 6B). These data suggest there is a crosstalk between GlcNAc-S248 and phospho-S253 and that the subcellular distribution and steady state levels of SOX2 are not affected PTM changes in this region.

SOX2^{S248A} alters ESC gene expression

While both transgenic SOX2 ESC lines exhibited standard colony morphology and stained positive for AP activity, the fS248A-Tg cells consistently stained darker and formed more circular colonies than the WT or parental

Figure 6 SOX2^{S248A} mutant ESCs show altered gene expression. **A**, fS248A-Tg ESCs, where 3xF-SOX2^{S248A} is the sole source of SOX2, compared to parental and fSOX2-Tg ESCs. AP staining and Western blots against SOX, FLAG and TUBULIN show comparable proteins levels though fS248A-Tg cells have higher AP staining and more circular colony morphology. **B**, XICs, within 10 ppm, of the TAD PTM peptides from fSOX2-Tg cells (top) compared with XICs of the analogous TAD PTM peptides from fS248A-Tg cells. **C**, Volcano plot showing genes differentially expressed between fSOX2-Tg and fS248A-Tg cells as assessed by microarray. **D**, RT-qPCR validation of select genes differentially expressed (* indicates $p < 0.05$) between fSOX2-Tg and fS248A-Tg cells (n=3, +/- S.E.M.).

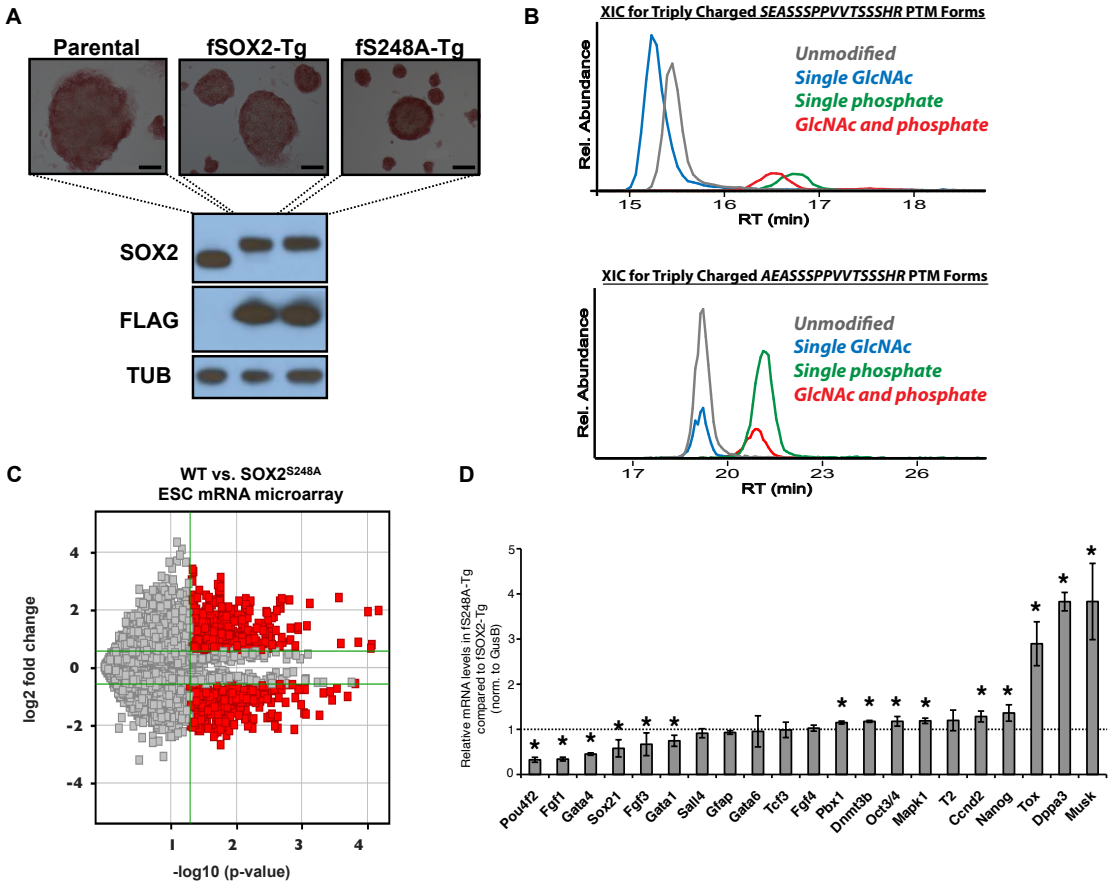
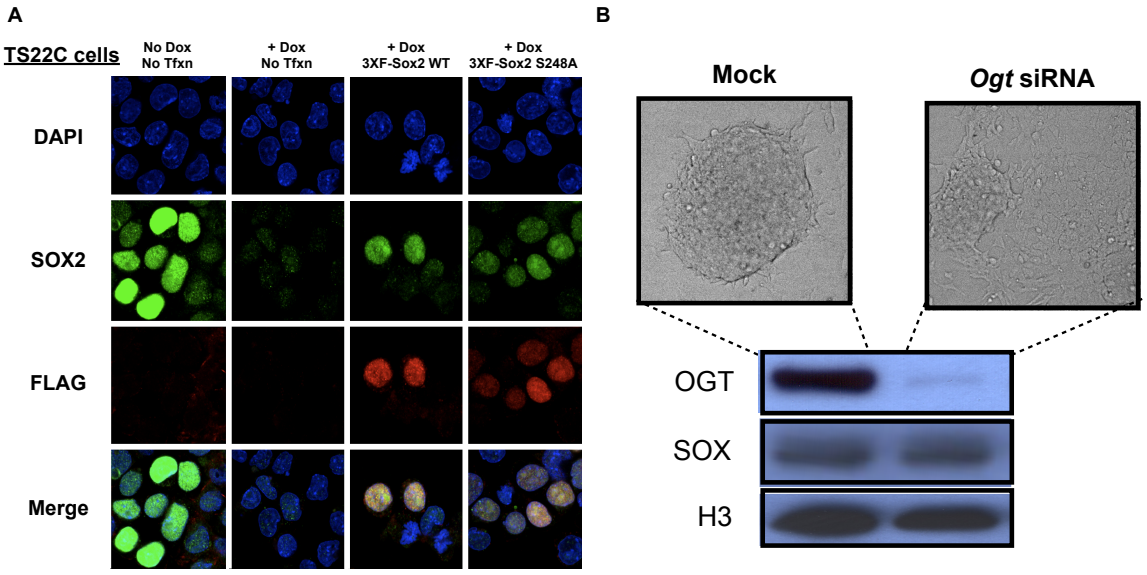


Figure 7 *O*-GlcNAcylation does not affect subcellular localization or steady state protein levels of SOX2. **A**, Immunofluorescence (IF) staining against SOX2 in TS22C cells transiently transfected with no plasmid, 3xF-SOX2^{WT} or 3xF-SOX2^{S248A} and treated with doxycylcline. **B**, Knockdown of OGT does not affect SOX2 protein levels. E14 ESCs were depleted of OGT by siRNA for three days. Western blotting against OGT, SOX2 and TUBULIN shows comparable levels of SOX2 compared to mock (*Gfp*).

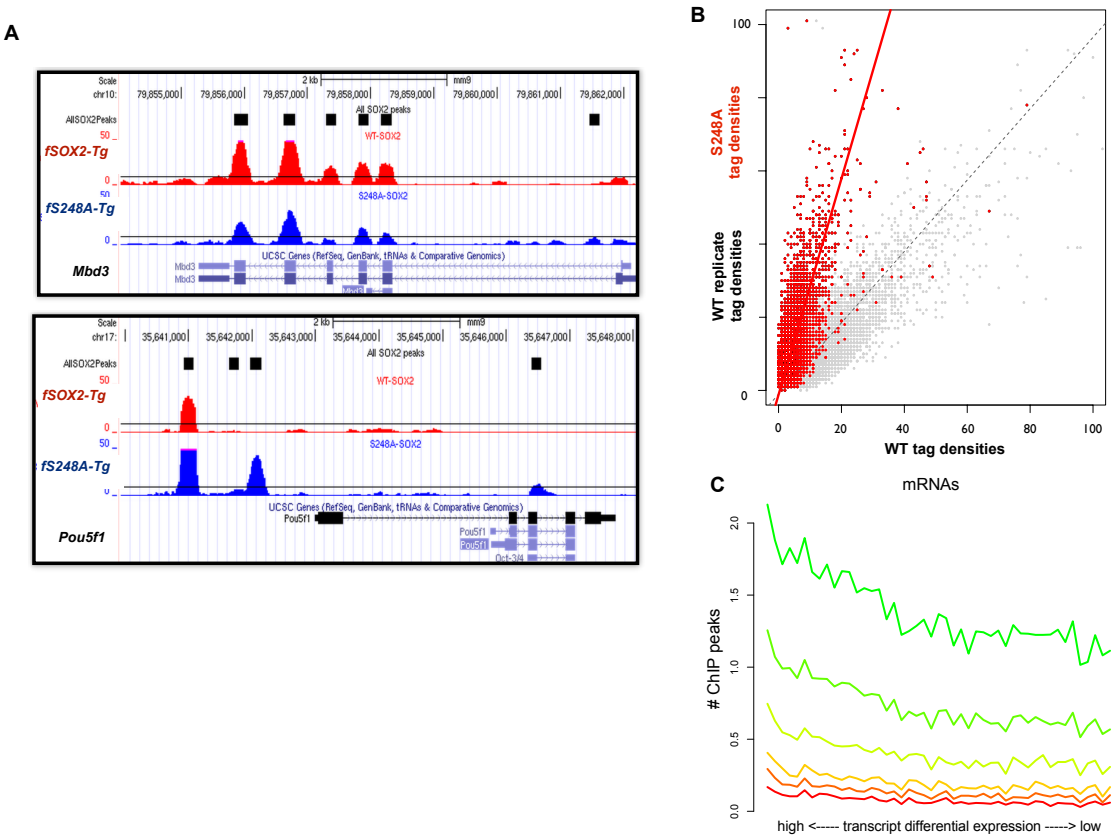


counterparts (Figure 6A). This suggested that the fS248A-Tg were more undifferentiated than WT cells (18, 19). We, therefore, used microarrays to compare gene expression between fSOX2-Tg and fS248A-Tg cells. There were significant changes in gene expression, with 320 genes up-regulated and 344 genes down-regulated (Figure 6C), several of which were validated by RT-qPCR (Figure 6D). Several of the genes with increased expression in fS248A-Tg cells are associated with pluripotency and self-renewal (*Dppa3, Nanog, Pou5f1/ Oct3/4*), while many of the down-regulated genes are associated with differentiation (*Pou4f2, Fgf1, Fgf3, Gata4, Gata1*). These data indicate that the S248A mutation affects SOX2 activity in ESCs and that this mutation alters the balance between self-renewal and differentiation gene expression.

SOX2^{S248A} exhibits altered genomic occupancy but not distribution

To examine whether the changes in gene expression associated with S248A were accompanied by changes in SOX2 genomic occupancy or distribution, we performed anti-FLAG chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) to compare SOX2 genomic distribution in fSOX2-Tg and fS248A-Tg ESCs. The S248A mutation altered SOX2 occupancy in two ways. First, while the majority of SOX2 bound loci were common between WT and S248A SOX2, several new sites emerged, while some sites no longer exhibited SOX2 association (Figure 8A). Second, for the commonly bound SOX2 sites, we were able to ChIP four to five-fold more DNA from fS248A-Tg cells compared to fSOX2-Tg cells (Figure 8B). This increase in SOX2 occupancy

Figure 8 fS248A-Tg ESCs show minor changes in SOX2 distribution but large increases in DNA occupancy at distal regulatory regions. **A**, Browser tracks of rare events showing changes in genomic SOX2 distribution between fSOX2-Tg and fS248A-Tg cells. New 3xF-SOX2^{WT} binding site in exon of *Mbd3* (top panel). Bottom panel shows new binding site of 3xF-SOX2^{S248A} at the promoter of *Pou5f1*. **B,** Number of sequencing reads per SOX2 bound locus between fSOX2-Tg and fS248A-Tg cells. fS248A-Tg cells show four to five times more DNA immunoprecipitated for the same regions bound in fSOX2-Tg cells. **C**, Correlation between number of SOX2 binding events and genes differentially expressed between fSOX2-Tg and fS248A-Tg cells within varying distances from the transcription start site (TSS). Genes more differentially expressed between fSOX2-Tg and fS248A-Tg cells have a higher number of SOX2 binding events as the distance from the TSS is increased.

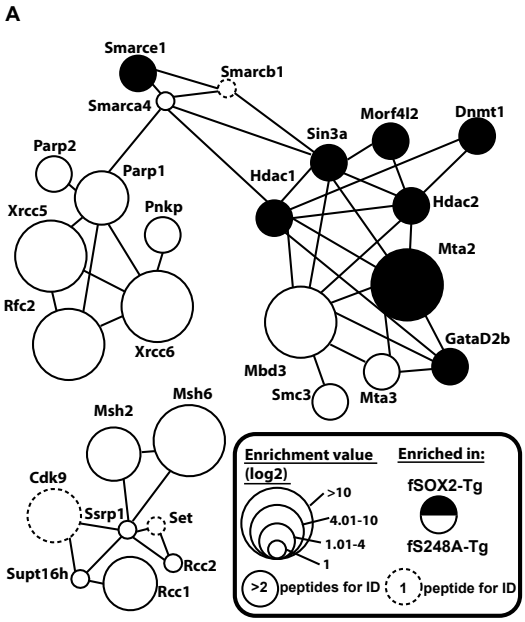

A
Scale
chr10:
79,855,000
79,856,000
79,857,000
79,858,000
79,859,000
79,860,000
79,861,000
79,862,000
2 kb
mm9
All SOX2 peaks
AllSOX2Peaks
WT-SOX2
fSOX2-Tg
S248A-SOX2
fS248A-Tg
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Mbd3
Mbd3
Scale
chr17:
35,641,000
35,642,000
35,643,000
35,644,000
35,645,000
35,646,000
35,647,000
35,648,000
2 kb
mm9
All SOX2 peaks
AllSOX2Peaks
WT-SOX2
fSOX2-Tg
S248A-SOX2
fS248A-Tg
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Pou5f1
Pou5f1
Pou5f1
Oct-3/4
B
S248A tag densities
WT replicate tag densities
WT tag densities
0
20
40
60
80
100
C
mRNAs
ChIP peaks
0.0
0.5
1.0
1.5
2.0
high <----- transcript differential expression -----> low

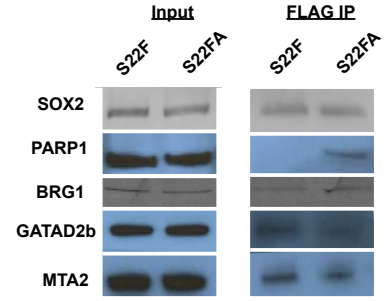
correlated with differential gene expression between fS248A-Tg and fSOX2-Tg cells. Genes more differentially expressed between fS248A-Tg and fSOX2-Tg cells showed more SOX2 binding within 100 kbp window of the gene with altered expression, though this correlation decreased as the distance from the transcriptional start site decreased (Figure 8C). Together, these results suggest changes in the gene expression are due to increased SOX2 occupancy at distal regulatory regions.

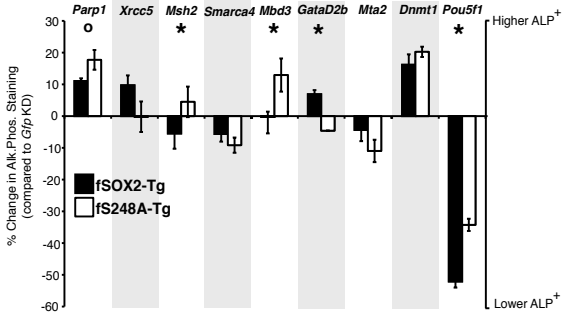
The S248A mutation alters protein-protein interactions

The TAD of SOX2 is responsible for interacting with transcriptional regulatory machinery, and truncations of this region have different affects on transcriptional activation (27-29). We tested whether PTMs of the SOX2 TAD affect the protein-protein interactions of SOX2. We performed co-IP against FLAG from nuclear extracts of either fS248A-Tg or fSOX2-Tg cells and used quantitative mass spectrometry to identify proteins that were enriched with either 3xF-SOX2^{WT} or 3xF-SOX2^{S248A}. We identified 73 and 50 proteins (>2 peptides for ID) that were enriched at least two fold with 3xF-SOX2^{S248A} and 3xF-SOX2^{WT}, respectively (Figure 9A). Western blotting corroborated the mass spectrometric data, exhibiting differential interactions between 3xF-SOX2^{WT} or 3xF-SOX2^{S248A} and the queried protein (Figure 9B). Many of the proteins enriched by either 3xF-SOX2^{S248A} or 3xF-SOX2^{WT} cells are in transcriptional regulatory complexes and are known to co-purify with SOX2 (Figure 9A) (30-33). In general, known protein complexes were enriched as a unit with either 3xF-SOX2^{WT} or with 3xF-

Figure 9 $SOX2^{S248A}$ interacts with a different subset of proteins than $SOX2^{WT}$**. A**, Interaction diagram of a subset of proteins that differentiatlly interact with 3xF-$SOX2^{WT}$ and 3xF-$SOX2^{S248A}$**.** Proteins (circles) are grouped and connected by lines according to String. Size and color of circle indicates the enrichment score and with which SOX2 the protein interacts, respectively. Dotted lines indicate the protein was identified and quantitated with a single peptide. **B,** Western blots against a subset of differentially interacting proteins corroborate the MS data. **C**, Depletion of proteins differentially interacting with SOX2 show cell-line specific contributions to self-renewal. Protiens identified to differentially interact with 3xF-$SOX2^{WT}$ or 3xF-$SOX2^{S248A}$ were depleted with siRNAs and assayed for AP staining three days after transfection. Loss or gain of AP activity was measured as described in Methods. *, ($p < 0.05$, T-test) differences in changes of AP staining between fSOX2-Tg and fS248A-Tg cells when the gene of interest is depleted compared to *Gfp*. º indicates $p < 0.1$.


A
Smarce1
Smarcb1
Smarca4
Morf4l2
Dnmt1
Sin3a
Parp2
Parp1
Hdac1
Hdac2
Xrcc5
Pnkp
Mta2
Rfc2
Xrcc6
GataD2b
Mbd3
Smc3
Mta3
Msh6
Msh2
Cdk9
Ssrp1
Set
Rcc2
Supt16h
Rcc1
Enrichment value (log2)
>10
4.01-10
1.01-4
1
Enriched in:
fSOX2-Tg
fS248A-Tg
>2 peptides for ID
1 peptide for ID


B
Input
FLAG IP
S22F
S22FA
S22F
S22FA
SOX2
PARP1
BRG1
GATAD2b
MTA2


C
Parp1
Xrcc5
Msh2
Smarca4
Mbd3
GataD2b
Mta2
Dnmt1
Pou5f1
o
*
*
*
*
Higher ALP+
Lower ALP+
% Change in Alk.Phos. Staining (compared to Gfp KD)
30
20
10
0
-10
-20
-30
-40
-50
-60
fSOX2-Tg
fS248A-Tg

SOX2^{S248A}. For example, several components of the PARP-XRCC and the DNA mismatch repair (MMR) system were enriched by 3xF-SOX2^{S248A}, whereas components of the NuRD complex were significantly enriched in 3xF-SOX2^{WT}. We were able, however, to identify subunits that behaved discordantly. MBD3 and MTA3, both of which can be part of the NuRD complex were enriched in SOX2^{S248A} rather than SOX2^{WT}. Similarly, some subunits of the SWI/SNF complex, a nucleosome remodeling complex known to associate with SOX2 (34), associated more with 3xF-SOX2^{S248A} while others with 3F-SOX2^{WT}. Together, these results demonstrate that the S248A mutation alters the protein-protein interactions of SOX2, suggesting *O*-GlcNAc regulates the transcriptional activity of SOX2 by recruiting different chromatin modifying complexes.

Differential SOX2 interactors have different contributions towards maintaining self-renewal in fSOX2-Tg and fS248A-Tg cells

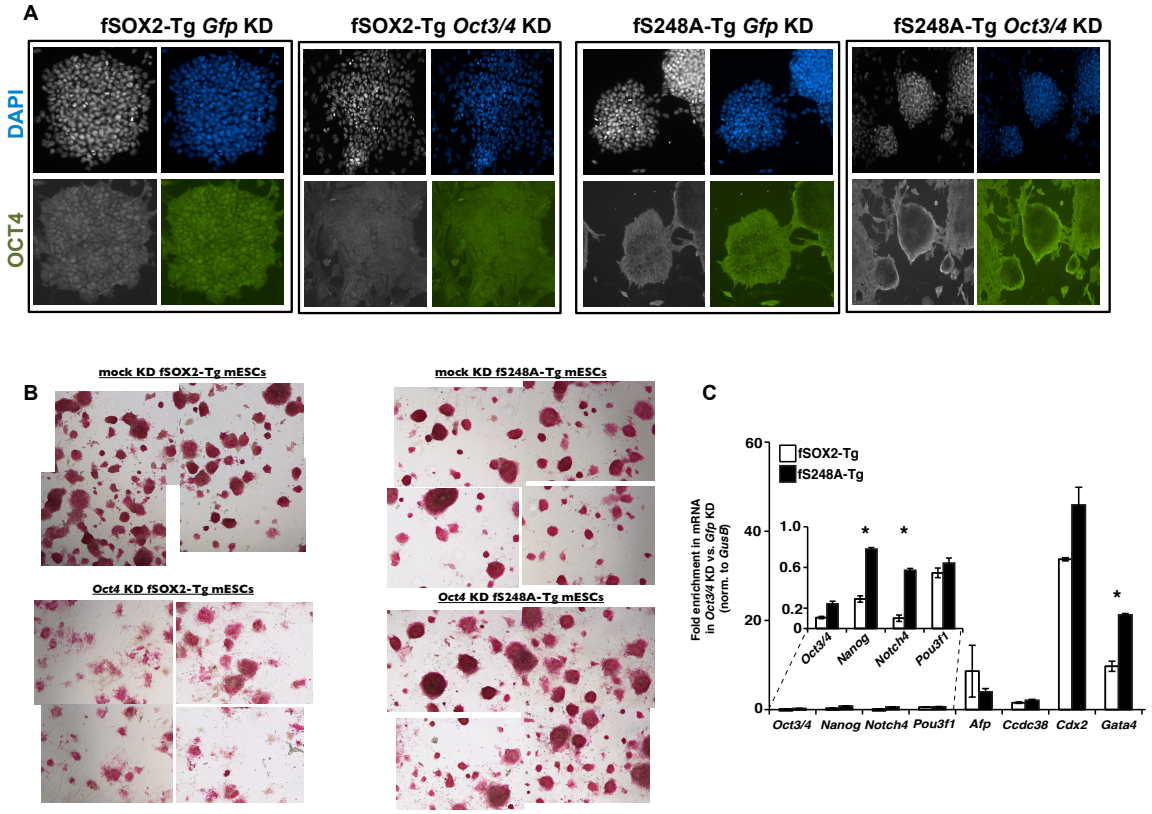
Because S248A and WT SOX2 co-purified with different transcriptional complexes, we asked whether these SOX2-form-specific interactions had different functional consequences in fSOX2-Tg or fS248A-Tg cells. We used siRNA pools to knock down mRNA levels of proteins enriched with either 3xF-SOX2^{WT} (*GataD2b, Mta2, Dnmt1, Pou5f1/Oct3/4*) or 3xF-SOX2^{S248A} (*Parp1, Xrcc5, Smarca4, Mbd3*) and analyzed AP activity. Quantitation of AP staining between fSOX2-Tg and fS248A-Tg cells showed that knockdown of *Msh2, Mbd3, GataD2b* and *Pou5f1/Oct4* showed significant differences ($p < 0.05$) in the percent change in AP activity when compared to *Gfp* knockdown (**Figure 9C**).

Parp1 and *Xrcc5* also showed differences in changes of AP activity between fSOX2-Tg and fS248A-Tg cells though they were less statistically significant ($p < 0.1$). *Smarca4*, *Mta2* and *Dnmt1* knockdown showed no difference in AP activity between the two cell lines. Qualitative assessment of ESC colony morphology also supported these findings (Appendix II). These data suggest that the protein complex-specific contributions towards the maintenance of pluripotency and self-renewal is regulated by certain SOX2 PTM states.

SOX2^{S248A} alters transcriptional and/or cellular landscape

The altered gene expression profile of fS248A-Tg cells (Figure 6D), and the resistance to *Pou5f1/Oct4* knockdown (Figure 9C) suggested that this mutation may promote self-renewal at the expense of differentiation. Therefore, we further examined the effects of OCT4 depletion, which promotes ESC differentiation (35). siRNA-mediated knockdown of OCT4 showed a smaller decrease in AP activity staining (Figure 9C) and almost no change in colony morphology when compared to comparable knockdown in fSOX2-Tg cells (Figure 10A-B). We next measured the mRNA levels of the pluripotency marker *Nanog*, along with several genes associated with OCT4 depletion-mediated differentiation (*Notch4, Pou3f1, Afp, Ccdc38, Cdx2,* and *Gata4*). As expected, OCT4 depletion had a smaller effect on the repression of *Nanog* transcription in fS248A-Tg cells than in fSOX2-Tg cells. Both cell lines showed activation of differentiation genes three days after OCT4 depletion. However, the transcripts for *Notch4* and *Gata4*, and to a lesser extent *Cdx2*, were activated to a greater

Figure 10 fS248A-Tg cells are more refractory to self-renewal loss while simultaneously activating differentation genes. **A**, fSOX2-Tg and fS248A-Tg were depeleted of OCT4 with siRNAs for three days and IF stained against OCT4. Images show efficient depletion of OCT4 in both fSOX2-Tg and fS248A-Tg cells. **B**, Colony morphology and AP staining of fSOX2-Tg and fS248A-Tg cells three days after OCT4 depletion. Four fields of view were acquired. **C**, RT-qPCR measurements of genes expressesed during OCT4-knockdown mediated trophectoderm differentiation comparing fSOX2-Tg and fS248A-Tg cells.



extent in fS248A-Tg cells than in the fSOX2-Tg line (Figure 10C). Together, these results indicate that colony morphology and AP activity is uncoupled from the transcriptional state of ESCs. These results also demonstrate that SOX2^{S248A} can simultaneously promote self-renewal and differentiation transcriptional networks, which suggests that *O*-GlcNAc inhibits SOX2 activity independent of cellular identity.

Discussion

SOX2 is a versatile transcription factor in that it is essential to establish and maintain stem cell self-renewal and is also important for lineage specification. The broad developmental potential of SOX2 is, in part, regulated by its extensive PTMs. However, the post-translationally modified states of SOX2, and how they regulate certain transcriptional networks, are not understood.

Here, we unbiasedly characterized the PTMs of SOX2 in ESCs. We found that SOX2 primarily exists as modified by *O*-GlcNAc, a regulatory glycosylation that is poorly understood. We used ETD MS to directly measure the stoichiometries of the *O*-GlcNAcylation and the nearby phosphorylation that occur within the TAD. Over one-third of SOX2 molecules in ESCs are glycosylated at S248. Mutational analysis of the predominant *O*-GlcNAc site showed that substitution of $SOX2^{WT}$ with an *O*-GlcNAc deficient version ($SOX2^{S248A}$) increased the ability of SOX2 to up-regulate the pluripotency transcriptional network and down-regulate genes associated with differentiation. These alterations in gene expression was accompanied by increased association of $SOX2^{S248A}$ with distal regulatory elements but not at promoter regions. These results suggest that *O*-GlcNAc inhibits the long-range regulatory activities of SOX2.

The TAD is the region of SOX2 responsible for its interaction with other transcription factors and transcriptional machinery. We reasoned that because

the TAD possessed a diverse set of PTMs, changes in these modifications might regulate the transcriptional activity of SOX2. We purified 3xF-SOX2^{WT} and 3xF-SOX2^{S248A} from ESCs and identified proteins that were enriched with WT or mutant. We identified several transcriptional regulatory complexes that were specifically enriched with either SOX2^{WT} or SOX2^{S248A}. SOX2^{S248A} specifically interacts with two complexes involved in DNA damage, the PARP/XRCC and the MMR complex. SOX2, along with OCT4, have been previously shown to interact with another DNA damage complex that activates *Nanog* transcription (36). Our data suggests that the PARP/XRCC and the MMR complex may regulate ESC self-renewal, in addition to genome integrity. Further studies will be needed to determine which protein of the respective complexes directly interacts with SOX2, and to which PTM form the complex is recruited (Figure 2A).

We identified the nucleosome remodeling/deacetylase (NuRD) complex enriched with SOX2^{WT}. The NuRD complex is a chromatin modifying complex known to interact with pluripotency transcription factors (37), inhibit pluripotency gene expression in ESCs (38, 39), and suppress somatic cell reprogramming (40, 41). The capacity of SOX2^{WT} to recruit the NuRD complex may account for the transcriptional suppressive affects of SOX2^{WT} in both ESC and in somatic cell reprogramming (see previous chapter). The specific enrichment of the NuRD complex with SOX2^{WT} also implies that a subunit of the complex may be able to specifically recognize GlcNAc-S248. There are currently no known *O*-GlcNAc binding domains, besides the *O*-GlcNAc hydrolase. Future work will be

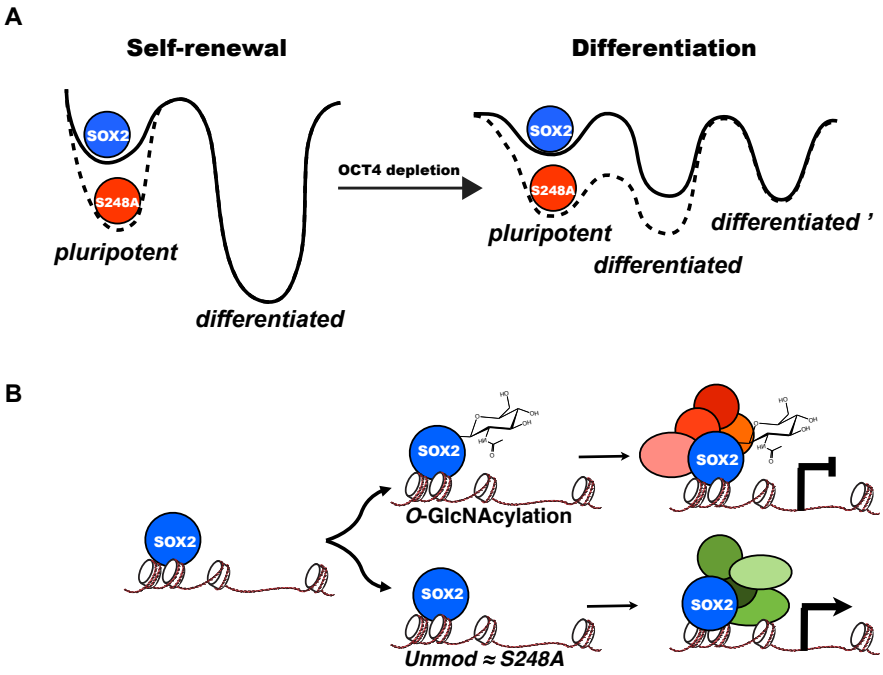
necessary to dissect the mechanism by which the NuRD complex is recruited to $SOX2^{WT}$ but not $SOX2^{S248A}$.

O-GlcNAc signaling is essential for embryo and ESC viability (14-17). However, *O*-GlcNAc site identification, as well as investigation into the function of these glycosylation events in ESCs, is in its infancy. This study shows that despite the necessity for global *O*-GlcNAc in ESC self-renewal, key pillars of pluripotency are inhibited by *O*-GlcNAc. The *O*-GlcNAcylation of SOX2 may serve to inhibit the transcription factor in an unperturbed state while allowing for rapid activation upon proper developmental cues (Figure 4) or abrupt changes in DNA binding partners (42) (Figure 10C). The ability of $SOX2^{S248A}$ to simultaneously promote pluripotency and trophectoderm transcriptional networks during OCT4 depletion (Figure 10C**)** suggests that the PTM state of SOX2 alters the "energy potential landscape" of ESCs both during self-renewing conditions and during differentiation (Figure 11A). Whether this propensity for differentiation is specific for OCT4 depletion, extends to the commitment of specific lineages (neural or primitive ectoderm lineages) or is a general phenomenon remains to be determined.

This work provides a new insights into the regulation of SOX2 through the *O*-GlcNAcylation of the TAD and the differential recruitment of transcriptional regulatory complexes (Figure 11B)**.** The ability of $SOX2^{S248A}$ to induce a more naive pluripotent state in ESCs while possessing the ability to promote

differentiation illustrates that TAD PTMs have a profound affect on controlling transcriptional networks. Finally, this study demonstrates the role of *O*-GlcNAc in pluripotency and and transcriptional regulation is more complex than previously appreciated.

Figure 11 The S248A mutation alters the “energy landscape” of cellular identity and alters interactions with transcriptional regulatory machinery. **A**, $SOX2^{S248A}$ changes in the energy landscape, as in chemical reactions, of the cell state and differentitation. Under self-renewing conditions (left) $SOX2^{WT}$ (blue circle) keeps cells in a potential well of a certain depth (solid line), keeping the bulk of cells in the pluripotent state and out of a differentiated state. $SOX2^{S248A}$ (red circle) deepens the pluripotency well, preventing spurious differentiation under self-renewing conditions (dashed line). Under differentiation conditions, as in loss of OCT4, the energy landscape changes increasing the potential of towards various differentation states. $SOX2^{S248A}$ can simultaneously deepen the pluripotency well while lowering the “kinetics” of differentiation. **B**, SOX2 *O*-GlcNAcylation regulates transcription by altering protein-protein interactions. SOX2 exists in two major PTM states, unmodified (bottom) and *O*-GlcNAc modified (top). The unmodified state, simulated by S248A, recruits transcriptional activating complexes (green circles), while the *O*-GlcNAc modified SOX2 recruits transcriptionally repressive complexes (red circles).


A
Self-renewal
Differentiation
SOX2
S248A
pluripotent
differentiated
OCT4 depletion
SOX2
S248A
pluripotent
differentiated
differentiated '
B
SOX2
SOX2
O-GlcNAcylation
SOX2
SOX2
Unmod ≈ S248A
SOX2

Materials and Methods

Cell Culture

Mouse embryonic stem cells (ESCs) were routinely passaged by standard methods in ESC media (KO-DMEM, 10% FBS, 2 mM glutamine, 1X non-essential amino acids, 1x pencillin/streptomycin, 0.1 mM β-mercaptoethanol and recombinant leukemia inhibitory factor). Tagged SOX2 cell lines were cultured in ESC media with 1ug/mL doxycycline hyclate (Sigma). MCF7 and MDA-MB-453 breast cancer cells were cultured using standard protocols. Transient transfections were done with Lipofectamine 2000 (Invitrogen) according to manufactures instructions.

Cell Line Derivation

2TS22C mouse ESCs (accession number AES0125), the parental cell line for derivation of tagged SOX2 lines (1), were obtained through Riken BioResource Center. To create the fSOX2-Tg ESC line, 4 ug of the plasmid CAG-3xF-Sox2 was transfected with Lipofectamine 2000 (Invitrogen) into a 6-well plate containing 2TS22C cells. 2TS22C mouse ESCs express SOX2 under control of a tetracycline-repressible system. Twenty-four hours after transfection 1 ug/mL doxycycline was added to silence expression of the TetO SOX2. Fourty-eight hours after transfection, 7 ug/mL puromycin was added, to select for the integration and expression of CAG-3xF-Sox2. After about two weeks, colonies exhibiting the typical ESC morphology were expanded and tested via western

blot, morphology and alkaline phosphatase staining (Clontech). The same strategy was used to generate the fS248A-Tg and haSOX2-Tg ESC lines

Plasmids

CAG-3xF-Sox2 was constructed by inserting mouse *Sox2* cDNA into pCMV-3xFLAG 7.1 (Sigma) and then subcloning 3xF-Sox2 via InFusion cloning (Clonetech) into CAG-HA-Sox2-IP (Addgene plasmid 13459), replacing the HA tag with the triply FLAG tagged. S248A mutations in CAG-3xF-Sox2 were performed with Quickchange site directed mutagenesis (Stratagene) (primers available upon request).

Antibodies and alkaline phosphatase

Antibodies were purchased from Abcam (NANOG ab70482, SOX2 ab75179, TUBULIN GTU-88 ab11316, Histone H3 ab1791, HA tag ab13834), Developmental Studies Hybridoma Bank (SUMO-1 21C7 and SUMO-2 8A2), Reprocell (for IF, NANOG, RCAB002P-F), SIGMA (OGT DM-17 and FLAG A8592) Bethyl (BRG1 A300-813A, PARP1 A301-375A, GATAD2B 301-281A) and Abbomax (MTA2 500-10034). Secondary antibodies were purchased from BioRad (172-1019 & 172-1011). Anti-FLAG magnetic beads were purchased from SIGMA (M8823). Alkaline phosphatase activity staining was performed according to manufactures instructions (Stemgent).

SOX2 Purification for PTM characterization

fSOX2-Tg cells were expanded to one to three 15 cm^2 dishes depending on the experiment. Cells were harvested by trypsinization, washed once with cold PBS and frozen in liquid N_2. Whole cell pellets were lysed in RIPA buffer without SDS, containing 500 nM Thiamet G (Caymen Chemicals), 1X HALT protease and phosphatase inhibitors (Pierce), 2 mM TCEP (Sigma) and 20 mM *N*-ethylmaleimide (Sigma) and sonicated (with a probe sonicator on methanol ice for three rounds of pulses, 3 secs on, 2 off, 10 seconds total, at 35%). Anti-FLAG-based purifications were performed with anti-FLAG M2 Dynabeads (Sigma). Whole cell lysates were incubated with M2 beads at room temperature for 75 minutes, washed once with lysis buffer and three times with 25mM ammonium bicarbonate (ABC) with 150 mM NaCl. Proteins were eluted with 100 mM glycine pH 4. Western blot and SDS-PAGE analysis was used to assess purification efficiency. For HA-SOX2 purification, anti-HA antibodies were coupled to aldehyde-coated magnetic beads (BioClone) via reductive amination in 20 mM bicine pH 7.8 and purified as above.

Mass spectrometric analysis

Silver or Coomassie stained SDS-PAGE gel bands were excised and digested in-gel with sequence grade trypsin (Roche). After 5% formic acid/50% acetonitrile extraction, peptides were dried by vacuum centrifugation, gel particulates were removed via C18 Zip Tips (Millipore), dried, resuspended in 0.1% formic acid and analyzed by LC MS/MS. Chromatography was performed on a Nanoacquity HPLC (Waters) at 400 nl/min with a BEH130 C18 75 μm ID x150mm column

(Waters). A 90- or 120-minute gradient from 98% solvent A (0.1% formic acid) to 22% solvent B (0.1% formic acid in acetonitrile) was used. Peptides were analyzed by an LTQ-Orbitrap Velos mass spectrometer (Thermo). After the survey scan of m/z 400-1,600 was measured in the Orbitrap at 30,000 resolution, the top three multiply charged ions were selected for both HCD and ETD. Automatic gain control for MS/MS was set to 2000. Normalized collision energy for HCD was set at 35 while the ETD activation time was charge state dependent, based on 100 ms for doubly charge precursors. Supplemental activation was implemented for ETD reactions. Dynamic exclusion of precursor selection was set for 25 seconds.

Data analysis

Fragment mass spectra were converted into peaklists using the in-house software PAVA. HCD and ETD data were searched separately using ProteinProspector version 5.10.0 against the UniProt database with a concatenated database. Only mouse and human genomes were used for the database searching. Precursor tolerance was set to 10 ppm, whereas fragment mass error tolerance was set to 0.6 Da for ETD and 20 ppm for HCD. *N*-terminal acetylation, methionine oxidation, loss of *N*-terminal methionine and glutamate conversion to pyroglutamate were allowed as variable modifications. For ETD data, HexNAc modifications to serine and threonine residues and phosphorylation to serine/threonine/tyrosine was allowed as variable mass modifications. For HCD, phosphorylation was searched the same way though

HexNAc was considered as a neutral loss. Methylation (mono, di- and tri-) of K and R, monomethylation of D, E and H (artifact from MeOH fixing PAGE gels), acetylation of K and R, and ADP-ribosylation to C, E, K, N, S, and R were searched separately. SLIP scoring was used to distinguish possible positional isomers of HexNAc and/or phosphopeptides (43). Relative abundances of each modified or unmodified peptide were calculated using the ICIS area calculated from XICs in Xcalibur (Thermo) at a 10 ppm mass tolerance.

Targeted ETD analysis

For targeted analysis, the mass spectrometer was set to constantly perform ETD (66.6 ms reaction time) on any precursor ions of 635.6 m/z within an isolation window of 3 Da. Product ions were measured in the ion trap. All spectra containing the charged reduced species ($M^{\bullet}+3H^{2+}$=935.7 m/z) within 500 ppm were summed together and used to estimate product ion peak area. ICIS calculated areas of product ion XICs in Xcalibur were used for relative quantitation.

Synthetic (glyco)peptide synthesis and analysis

Materials obtained commercially were reagent grade and were used without further purification. ^{1}H NMR and ^{13}C NMR spectra were recorded on a Varian 400 spectrometer at 400 and 100 MHz, respectively. LC MS was used to confirm mass and purity of >95%. FMOC-Ser(β-D-GlcNAc(Ac)3)OH was synthesized as previously described (44), FMOC-Ser(β-D-GlcNAc(Ac)3)OH) containing peptide

was synthesized by solid state peptide synthesis (Pierce), and the *O*-acetyl groups were removed from the *O*-GlcNAc containing peptide as previously described (45).

FMOC-Ser(β-D-GlcNAc(Ac)3)OH). White powder; 1H NMR (d6-DMSO, 400 MHz) δ 1.70 (s, 3H), 1.88 (s, 3H), 1.94 (s, 3H), 1.97 (s, 3H), 3.63-3.76 (m, 3H), 3.80-3.93 (m,2H), 4.14-4.26 (m, 6H), 4.68 (d,1H, J = 8Hz), 4.78 (t, 1H, J = 12Hz), 5.06 (t,1H,=8Hz), 7.24-7.41 (m, 5H), 7.70 (d, 2H, J = 8Hz), 7.86 (m, 3H); 13C NMR (d6-DMSO, 100 MHz) δ 21.01, 21.08, 21.17, 21.72, 23.26, 47.26, 53.72, 54.75, 62.42, 62.44 66.52, 69.13, 71.51, 73.18, 100.92, 120.71, 120.77, 125.89, 125.94, 127.71, 127.75,128.31, 141.36, 141.38, 144.48, 156.59, 169.92, 170.20, 170.31, 170.72, 171.85; LCMS molecular ion predicted for C32H37N2O13 = 657.23, found 657.21.

Unmodified and *O*-GlcNAc modified SOX2 peptides were dissolved in 0.1% formic acid and measured individually, and then combined at a 1:1 ratio. LC-MS measurements were conducted on an LTQ-FT-ICR mass spectrometer (Thermo) at a resolution of 30,000. 25 fmols of each peptide were combined and measured.

Analysis of 3xF-SOX2 interactors

Six to ten 10 cm^2 plates worth of haSOX2-Tg, fSOX2-Tg or fS248A-Tg cells were harvested for nuclear extract preparations. Nuclear extracts were as previously

described with minor modifications (46). Buffers A, C and D were supplemented with 2 μM Thiamet G, 2 μM PUGNAc (Tocris) and 1X HALT protease, and phosphatase inhibitors and instead of dialysis of extracts, two volumes of buffer D were used to dilute salt concentration. Seven μL of M2 beads per $10cm^2$ plate were used per co-IP and samples were nutated at 4^oC for two hours. Beads were washed once with Buffer D plus inhibitors, then twice with 50 mM ABC with 150 mM NaCl. Each wash was only as long as it took to transfer the beads to a new, cold tube and place on the magnetic rack. Beads were then resuspended in 50 μL 100 mM ABC with 500 ng trypsin and shaken at 37C for one hour. Supernatant was transferred to a new tube, the beads were washed once with 50 μL ABC and combined to digest overnight. Digestions were desalted with one or two Zip Tips, depending on size of experiment, and dried via vacuum centrifugation. IP-WB experiments were performed similarly except proteins were eluted with 2X SDS-PAGE loading buffer without reducing agent.

Chromatography was performed on a Nanoacquity HPLC (Waters) at 400 nl/min with an EASY-spray C18, 3 μm particle size, 75 μm ID x15 cm column and source (Thermo). A 90-minute gradient from 98% solvent A (0.1% formic acid) to 30% solvent B (0.1% formic acid in acetonitrile) was used. Peptides were analyzed on a Q-Exactive Plus mass spectrometer (Thermo). After the survey scan of m/z 400-1,600 was measured in the Orbitrap at 60,000 resolution, the top ten multiply charged ions were selected for HCD and measured at 17,500

resolution. Normalized collision energy for HCD was set at 35, and the dynamic exclusion of precursor selection was set for 25 seconds.

XIC for all peptides identified in at least one sample were extracted using Skyline (47) and the MS1 filter settings. The intensity for each peptide was normalized to the median intensity of all SOX2 peptides in the respective run, excluding the TAD peptide and any PTM forms. The normalized peptide intensities were compared by $\log_2$ transforming the ratio of the normalized peptides from the fSOX2-Tg samples to the normalized peptides from the fS248A-Tg samples (**Figure peptide distb)**. The median protein intensity from all (at least two) peptides was used to determine the enrichment score, which is the absolute value of the $\log_2$ value. An arbitrary enrichment score cutoff of 1 (two-fold enrichment) was used as the minimum enrichment. The distribution of median protein enrichment scores was used to determine proteins enriched by their respective SOX2 forms greater than two standard deviations.

Microarray and RT-qPCR analysis

Total RNA was extracted with Trizol (Invitrogen) according to manufactures instructions. Arraystar (http://www.arraystar.com) prepped and hybridized the samples, and performed the data analysis.

For RT-qPCR, 1 µg total RNA was reverse transcribed to cDNA with iScript (Biorad), diluted 1:20 or 1:50, depending on the abundance of the transcript, and 4 µL was used. Quantitative PCR was performed on a CFX Connect™ Real-time

PCR detection system (Biorad) with SensiFast™ SYBR Lo-ROX PCR master mix (Bioline, BIO-94020). Fold enrichment was determined by $2^{-(\Delta Cq)}$ method (ΔCq= Cq(gene)-Cq(*GusB*)).

In vitro SOX2-OGT reaction

The recombinant OGT expression plasmid was a generous gift from Suzanne Walker. OGT was expressed as previously described (48). Mouse *Sox2* was cloned into pPRO EX Hta (Invitrogen) and expressed the same way as OGT. Both poly-His tagged proteins were purified by Ni-NTA agarose resin (Qiagen), eluted and buffer exchanged into 50 mM TRIS, pH 7.8, 300 mM NaCl. GlcNAc or UDP-GlcNAc was added to 10 mM and reactions were carried out at room temperature overnight. Reactions were run on and excised from an SDS-PAGE gel and analyzed by MS.

siRNA knockdown

siRNAs were created by as previously described (49). Five thousand cells per 6-well were plated the night prior to transfection. PCR products for the *in vitro* transcription were chosen from the Riddle database. Primers available upon request.

ChIP-seq

ChIP was performed as previously described (50).

Quantitation of AP staining

Cells were stained for AP activity three days post transfection of siRNA pools. Images were taken on a Nikon Ecplise Ti microscope. Automatic colony identification and calculation of “mean brightness” of each individual colony in a field of view was performed using the NIS-Elements Br software was used. Two images from opposing quadrants of the well were used for mean brightness measurements per biological replicate. The percent change was calculated using the difference in mean brightness between the knockdown of the gene of interest and the *Gfp* knockdown cells, and was normalized to the mean brightness of *Gfp* treated cells.

References

1. Masui S*, et al.* (2007) Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat Cell Biol* 9(6):625-635.

2. Arnold K*, et al.* (2011) Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* 9(4): 317-329.

3. Kopp JL, Ormsbee BD, Desler M, & Rizzino A (2008) Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells* 26(4):903-911.

4. Brumbaugh J*, et al.* (2012) Phosphorylation regulates human OCT4. *Proc Natl Acad Sci U S A* 109(19):7162-7168.

5. Gao F, Kwon SW, Zhao Y, & Jin Y (2009) PARP1 poly(ADP-ribosyl)ates Sox2 to control Sox2 protein levels and FGF4 expression during embryonic stem cell differentiation. *J Biol Chem* 284(33):22263-22273.

6. Lai YS*, et al.* (2012) SRY (sex determining region Y)-box2 (Sox2)/poly ADP-ribose polymerase 1 (Parp1) complexes regulate pluripotency. *Proc Natl Acad Sci U S A* 109(10):3772-3777.

7. Swaney DL, Wenger CD, Thomson JA, & Coon JJ (2009) Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 106(4):995-1000.

8. Tsuruzoe S*, et al.* (2006) Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem Biophys Res Commun* 351(4):920-926.

9. Van Hoof D*, et al.* (2009) Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* 5(2):214-226.

10. Zhao HY, Zhang YJ, Dai H, Zhang Y, & Shen YF (2011) CARM1 mediates modulation of Sox2. *PLoS One* 6(10):e27026.

11. Baltus GA*, et al.* (2009) Acetylation of sox2 induces its nuclear export in embryonic stem cells. *Stem Cells* 27(9):2175-2184.

12. Tahmasebi S*, et al.* (2013) Sumoylation of Kruppel-like factor 4 inhibits pluripotency induction but promotes adipocyte differentiation. *J Biol Chem* .

13. Myers SA, Panning B, & Burlingame AL (2011) 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* 108(23):9490-9495.

14. O'Donnell N, Zachara NE, Hart GW, & Marth JD (2004) Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol Cell Biol* 24(4):1680-1690.

15. Shafi R*, et al.* (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc Natl Acad Sci U S A* 97(11):5735-5739.

16. Yang YR*, et al.* (2012) O-GlcNAcase is essential for embryonic development and maintenance of genomic stability. *Aging Cell* 11(3):439-448.

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

18. Wray J, Kalkan T, & Smith AG (The ground state of pluripotency. *Biochem Soc Trans* 38(4):1027-1032.

19. Ying QL*, et al.* (2008) The ground state of embryonic stem cell self-renewal. *Nature* 453(7194):519-523.

20. Swaney DL, McAlister GC, & Coon JJ (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat Methods* 5(11):959-964.

21. Aksoy I*, et al.* (2013) Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm. *EMBO J* 32(7):938-953.

22. Sarkar A & Hochedlinger K (2013) The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* 12(1):15-30.

23. Thomson M*, et al.* (2011) Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* 145(6):875-889.

24. Schoenhals M*, et al.* (2009) Embryonic stem cell markers expression in cancers. *Biochem Biophys Res Commun* 383(2):157-162.

25. Herreros-Villanueva M*, et al.* (2013) SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis* 2:e61.

26. Wasik AM*, et al.* (2014) Reprogramming and carcinogenesis--parallels and distinctions. *Int Rev Cell Mol Biol* 308:167-203.

27. Ambrosetti DC, Scholer HR, Dailey L, & Basilico C (2000) Modulation of the activity of multiple transcriptional activation domains by the DNA binding domains mediates the synergistic action of Sox2 and Oct-3 on the fibroblast growth factor-4 enhancer. *J Biol Chem* 275(30):23387-23397.

28. Nowling TK, Johnson LR, Wiebe MS, & Rizzino A (2000) Identification of the transactivation domain of the transcription factor Sox-2 and an associated co-activator. *J Biol Chem* 275(6):3810-3818.

29. Yuan H, Corbi N, Basilico C, & Dailey L (1995) Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* 9(21):2635-2645.

30. Engelen E*, et al.* (2011) Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nat Genet* 43(6):607-611.

31. Cox JL*, et al.* (2013) The SOX2-Interactome in Brain Cancer Cells Identifies the Requirement of MSI2 and USP9X for the Growth of Brain Tumor Cells. *PLoS One* 8(5):e62857.

32. Gao Z*, et al.* (2012) Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *J Biol Chem* 287(14):11384-11397.

33. Mallanna SK*, et al.* (2010) Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells* 28(10):1715-1727.

34. Ho L*, et al.* (2009) An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. *Proc Natl Acad Sci U S A* 106(13):5187-5191.

35. Hough SR, Clements I, Welch PJ, & Wiederholt KA (2006) Differentiation of mouse embryonic stem cells after RNA interference-mediated silencing of OCT4 and Nanog. *Stem Cells* 24(6):1467-1475.

36. Fong YW*, et al.* (2011) A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* 147(1):120-131.

37. Liang J*, et al.* (2008) Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* 10(6):731-739.

38. Yildirim O*, et al.* (2011) Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147(7): 1498-1510.

39. Reynolds N*, et al.* (2012) NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* 10(5):583-594.

40. Luo M*, et al.* (2013) NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* 31(7):1278-1286.

41. Rais Y*, et al.* (2013) Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* 502(7469):65-70.

42. Kondoh H & Kamachi Y (2010) SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int J Biochem Cell Biol* 42(3):391-399.

43. Baker PR, Trinidad JC, & Chalkley RJ (2011) Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 10(7):M111 008078.

44. Aguilar PS*, et al.* (A plasma-membrane E-MAP reveals links of the eisosome with sphingolipid metabolism and endosomal trafficking. *Nat Struct Mol Biol* 17(7):901-908.

45. Chen YX*, et al.* (2006) Alternative O-GlcNAcylation/O-phosphorylation of Ser16 induce different conformational disturbances to the N terminus of murine estrogen receptor beta. *Chem Biol* 13(9):937-944.

46. Dignam JD, Lebovitz RM, & Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11(5):1475-1489.

47. Schilling B*, et al.* (2012) Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol Cell Proteomics* 11(5):202-214.

48. Gross BJ, Kraybill BC, & Walker S (2005) Discovery of O-GlcNAc transferase inhibitors. *J Am Chem Soc* 127(42):14588-14589.

49. Fazzio TG, Huff JT, & Panning B (2008) An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell* 134(1):162-174.

50. Lee J, Kim HK, Han YM, & Kim J (2008) Pyruvate kinase isozyme type M2 (PKM2) interacts and cooperates with Oct-4 in regulating transcription. *Int J Biochem Cell Biol* 40(5):1043-1054.

Chapter 6

Characterization of O-GlcNAcylation of the C-terminal domain of RPB1

Introduction

Human RNA polymerase II (pol II) possesses a unique C-terminal domain (CTD) that consists of 52 imperfect repeats of the amino acid sequence $Y_1S_2P_3T_4S_5P_6S_7$ (Figure 1). This heptad repeat is well-characterized to exist in an umodified form (pol IIA), or phosphorylated (pol IIO) at serines 2, 5 and 7. The current model of the transcription cycle proposes that pol II forms a preinitiation complex (PIC) in its unmodified form, where its CTD, lacks any post-translational modifications. The CTD is then phosphorylated during intiation and again during elongation. These phosphorylation events of the CTD do not affect enzymatic activity of RPB1 but rather alter the recruitment of various post-transcriptional mRNA processing factors during elongation. While the phosphorylation of a single heptad repeat has been characterized, the phosphorylation patterns of CTD purified from cells has not been characterized.

In 1993, Kelly, Dahmus and Hart showed that pol IIA contained a population that was O-GlcNAcylated, calling into question the theory that pol IIA is truly unmodified (1). Recent studies have showed that O-GlcNAc cycling is necessary for *in vitro* transcription, and that inhibition of OGA inhibited formation (2). Edman degradation of the O-GlcNAc modified pol IIA species suggested that T_4 and S_5 were glycosylated (1). Serines are labile during Edman degradation (3) and may have underrepresented serine 5 O-GlcNAcylation. Moreover, mutational analysis (S/T to A) of the CTD and subsequent enzymatic reaction with UDP-GlcNAc and recombinant OGT (rOGT) suggested that S5 and S7, but not T4, are

the major O-GlcNAcylated residues. However, no direct evidence of CTD O-GlcNAcylation (pol IIγ) has been reported.

Results

We performed LC-MS/MS of a glutathione S-transferase fused to a CTD containing the human repeats 27-52 (GST-CTD) treated with rOGT and UDP-GlcNAc. We chose the 27-52 repeats because this half of the CTD yields tryptic peptides amenable for MS/MS (Figure 1). Collisional activation (HCD) of GST-CTD tryptic peptides provided evidence of a HexNAc modification on the 33rd and the 43rd repeat of the CTD. Electron transfer dissociation (ETD) MS/MS identified O-GlcNAcylation of serine 5 in CTD repeat 33 and S2 of repeat 43 (Figure 2). This indicates that serine 5 is a direct target of OGT. In contrast, we did not observe any *O*-GlcNAcylation of T4 *in vitro*. These data suggest that the previously observed block to TFIIH-dependent phosphorylation of the serine 5 residue of the CTD (2, 4) is likely due to *O*-GlcNAcylation of serine 5.

We next asked whether we could identify CTD *O*-GlcNAc sites on pol II purified from cells. To detect pol IIγ from cells we first immunoprecipitated pol II, *O*-GlcNAc (with the RL2 antibody), or TBP from chromatin preps. We then reversed the formaldehyde crosslinks and performed a second IP on the de-crosslinked material using an anti-*O*-GlcNAc antibody (110.6). These “ChIP-IPs” were then analyzed by western blots probing for RNA pol II (Figure 3).

Figure 1 Heptad peptide repeats of the human pol II CTD. Adapted from (9).

1	YSPTSPA	**19**	YSPTSPS	**37**	YSPTSPE
2	YEPRSPGG	**20**	YSPTSPS	**38**	YSPTSPK
3	YTPQSPS	**21**	YSPTSPS	**39**	YSPTSPK
4	YSPTSPS	**22**	YSPTSPN	**40**	YSPTSPK
5	YSPTSPS	**23**	YSPTSPN	**41**	YSPTSPT
6	YSPTSPN	**24**	YSPTSPS	**42**	YSPTTPK
7	YSPTSPS	**25**	YSPTSPS	**43**	YSPTSPT
8	YSPTSPS	**26**	YSPTSPN	**44**	YSPTSPV
9	YSPTSPS	**27**	YSPTSPN	**45**	YTPTSPK
10	YSPTSPS	**28**	YSPTSPS	**46**	YSPTSPT
11	YSPTSPS	**29**	YSPTSPS	**47**	YSPTSPK
12	YSPTSPS	**30**	YSPTSPS	**48**	YSPTSPT
13	YSPTSPS	**31**	YSPSSPS	**49**	YSPTSPKGST
14	YSPTSPS	**32**	YTPQSPT	**50**	YSPTSPG
15	YSPTSPS	**33**	YTPSSPS	**51**	YSPTSPT
16	YSPTSPS	**34**	YSPSSPS	**52**	YSLTSPAISPDDSDEEN
17	YSPTSPS	**35**	YSPTSPK		
18	YSPTSPS	**36**	YTPTSPS		

$Y_1S_2P_3T_4S_5P_6S_7$

Consensus sequence

Figure 2 Identification of CTD *O*-GlcNAcylation. ETD mass spectrum of the GST-CTD repeat 33 modified by *N*-acetylglucosamine (HexNAc) at the serine 5 position. b) ETD mass spectrum of the GST-CTD repeat 43 modified by *N*-acetylglucosamine (HexNAc) at the serine 2 position. Inset labeled “HCD zoom” indicates the detection of the HexNAc oxonium ion and fragments thereof (denoted by *) by higher energy collisional dissociation (HCD) from the same precursor. b.p., base peak.

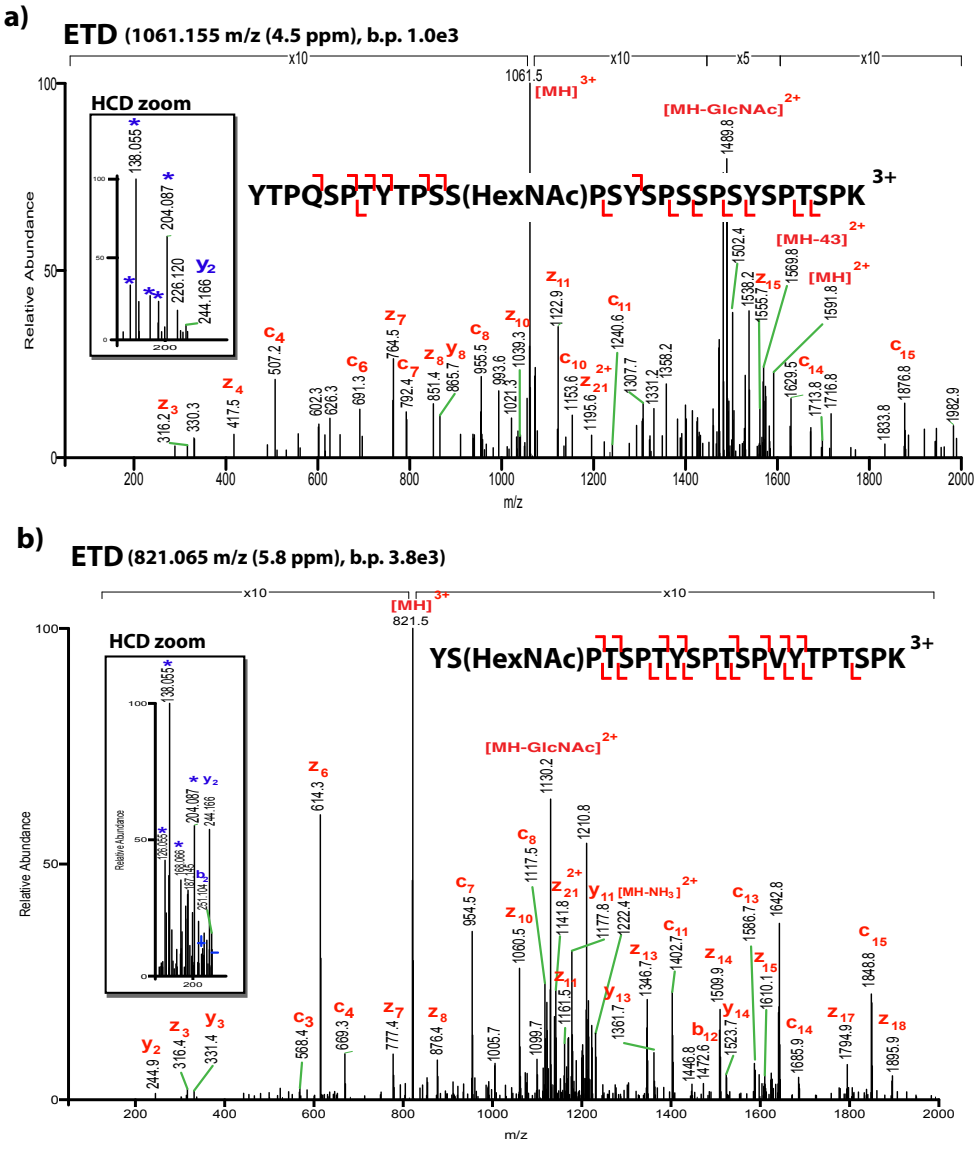
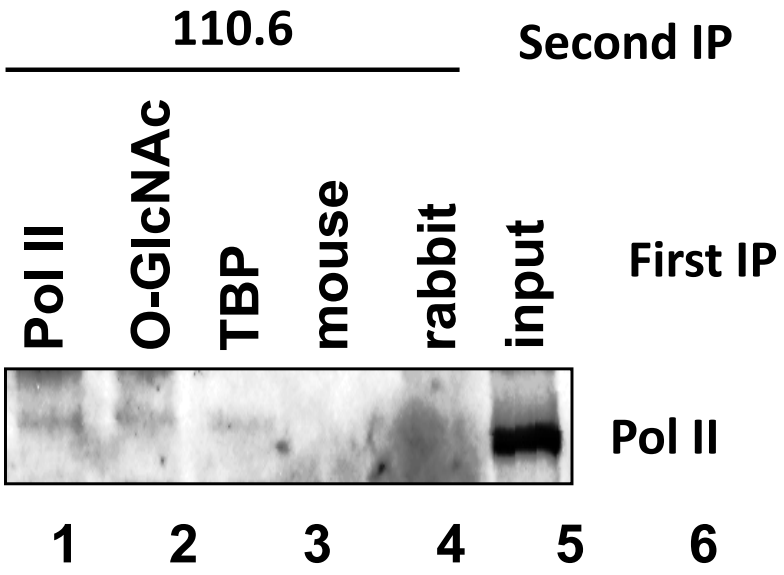


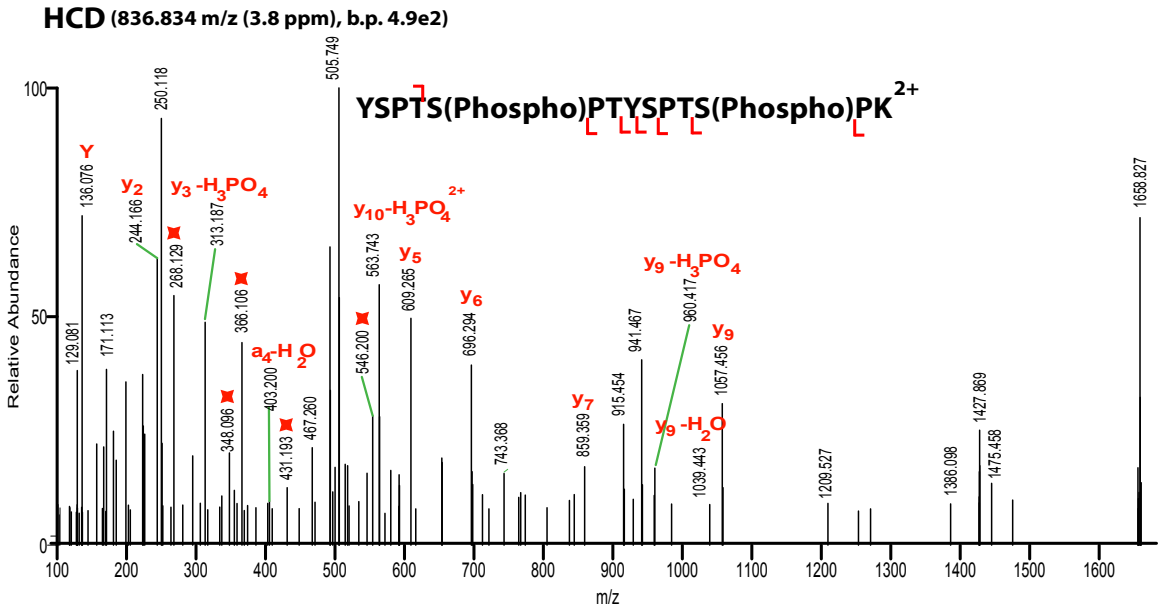
Figure 3 Detection of pol IIγ in cells. Western blot for pol II following double immunoprecipitation for a variable first factor and O-GlcNAc as the second factor shows an O-GlcNAcylated Pol II species at TBP-bound regions.



Lane 1 shows that pol IIγ was detected after a second IP with the anti-O-GlcNAc antibody, indicating that pol IIγ exists on chromatin in vivo. Lane 2 shows that 110.6 can specifically recover *O*-GlcNAc-modified pol II, as does TBP in lane 3. TBP IPs of chromatin are specifically pulling out promoters since TBP is a promoter-specific factor and part of the PIC. The reversal of crosslinking of the TBP IP followed by the *O*-GlcNAc IP and detection of pol II in that IP indicates that pol IIγ is present on promoters in vivo.

To identify sites of *O*-GlcNAcylation on the CTD, we analyzed RPB1 purified from cells by LC-MS/MS. RPB1 was immunoprecipitated from HEK293 cells stably expressing a FLAG-tagged RPB1, digested with trypsin and analyzed by LC-MS/MS. We were able to detect four phospho-CTD peptides, suggesting an appreciable stoichiometry of CTD phosphorylation (Table 1). Figure 4 shows the identification of a doubly phosphorylated CTD peptide where both serine 5 residues (according to the heptad numbering) were modified. We also identified a GlyGly modified CTD peptide, suggesting ubiquitinylation of the CTD. RPB1 ubiquitinylation has been reported previously and can have roles for both degradation and for pol II subunit composition (5-7). We were unable to detect the *O*-GlcNAc modification on CTD peptides. This could be due to the low stoichiometry of the modification or that *O*-GlcNAcylation that occurs in cells resides on the half of the CTD that does not yield tryptic peptides with appropriate charge density to be identified by MS/MS.

Figure 4 HCD mass spectrum of a doubly phosphorylated CTD peptide from RPB1 purified from cells. b.p., base peak. Modified residues are S1913 and S1920 and correspond to S5 in the heptad repeat. Stars indicate internal fragment ions.



Conclusions

Here, we showed that OGT can biochemically *O*-GlcNAcylate the CTD of RPB1 at serine 5 of CTD repeat 33 and at serine 2 of repeat 43. The identification of S5 O-GlcNAcylation is consistent with previous reports that used Edman degradation (4). S2 *O*-GlcNAcylation has not been reported previously. Mutational analysis of CTD *O*-GlcNAcylation did not propose S2 as a potential site of modification (2). This suggests that all heptad repeats in the CTD may not be equal and that OGT recognition of the CTD may be complex. Together, this data also suggest crosstalk between phosphorylation and *O*-GlcNAcylation to regulate transcription initiation.

ChIP-IP of RPB1 indicates that the CTD is O-GlcNAc modified in cells. However, we were unable to detect *O*-GlcNAc modifications by LC MS/MS. This is most likely due to the substiochiometric levels of pol IIγ in cells. Further effort will be dedicated to enriching the pol IIγ fraction to identify the *O*-GlcNAcylation state of the CTD.

The existence of pol IIγ as part of the PIC is both unexpected and outside the paradigms of transcriptional regulation. The presence of additional enzymatic processes in the PIC and their regulation by the nutrient state of the cell foreshadows a complex regulatory environment where nutrients play a direct role in transcriptional output by directly affecting the levels of RNA polymerase II and possibly the number of functional PICs, PIC stability, or initiation frequency and

length. Pol IIγ establishes a direct link between the promoter and its contents to the cellular physiological state, a link previously shown to be the purview only of DNA binding specific activators and repressors.

Experimental methods

ChIP-IP

Hela S3 cells were grown in DMEM/10% FBS/pen-strep to approximately 70 to 80% confluency and were crosslinked with 1% formaldehyde for 10 minutes. The reaction was quenched with 125 mM glycine for 10 mins. Cells were harvested and 1 x 10^7 cells were resuspended in 1 ml lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM EDTA, 10% glycerol, 1% Triton X-100, and Complete mini EDTA-free protease inhibitor cocktail tablet [Roche]). Cells were sonicated with a Misonic Sonicator (20 seconds on, 40 seconds off) for a total of 1 minute and then rotated for 30 minutes with DNase I, 3 mM $MgCl_2$ and 3 mM $CaCl_2$ at room temperature. Lysate was incubated overnight with initial antibodies and then protein G beads for 2 hrs. The crosslinks were reversed in BC100 buffer (20 mM Tris pH 7.9, 100 mM KCl, 20% glycerol, 0.2 mM EDTA) overnight at 68˚C. Supernatant was incubated with 110.6 (anti-O-GlcNac), 110.6 with 0.5 M GlcNAc, or IgG mouse control overnight then with IgM beads for 2 hrs. Proteins were eluted from beads with sample buffer and run on a 4-12% gradient gel. Proteins were transferred to PVDF membrane and probed with 8WG16 (anti-pol II).

LC-MS/MS and data analysis

OGT labeling of GST-CTD was done as described (2). GST-CTD fusion containing CTD repeats 27-52 were digested with 1:50 trypsin:substrate ratio at 37^{o}C for three hours. Peptides were desalted and analyzed as previously

described (8) except the 90-minute LC gradient only reached 20% acetonitrile. Data was searched using Protein Prospector v 5.10.0. against the SwissProt database (21 March, 2012), where all 535248 entries, including the user added protein, GST-CTD amino acid sequence, was searched. Manual inspection for 204.087 m/z detection in HCD spectra was used to identify HexNAc modified peptides. In these instances, data was interpreted manually.

References

1. Kelly WG, Dahmus ME, Hart GW. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. J Biol Chem. 1993;268(14):10416-24.

2. Ranuncolo SM, Ghosh S, Hanover JA, Hart GW, Lewis BA. 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):23549-61. PMCID: 3390630.

3. Reason AJ, Dell A, Romero PA, Herscovics A. Specificity of the mannosyltransferase which initiates outer chain formation in Saccharomyces cerevisiae. Glycobiology. 1991;1(4):387-91.

4. Comer FI, Hart GW. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. Biochemistry. 2001;40(26): 7845-52.

5. Li H, Zhang Z, Wang B, Zhang J, Zhao Y, Jin Y. Wwp2-mediated ubiquitination of the RNA polymerase II large subunit in mouse embryonic pluripotent stem cells. Mol Cell Biol. 2007;27(15):5296-305. PMCID: 1952083.

6. Daulny A, Geng F, Muratani M, Geisinger JM, Salghetti SE, Tansey WP. Modulation of RNA polymerase II subunit composition by ubiquitylation. Proc Natl Acad Sci U S A. 2008;105(50):19649-54. PMCID: 2604917.

7. Chen X, Ruggiero C, Li S. Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage. Mol Cell Biol. 2007;27(13):4617-25. PMCID: 1951484.

8. Myers SA, Daou S, Affar el B, Burlingame A. 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):982-91.

9. Palancade B, Bensaude O. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. Eur J Biochem. 2003;270(19):3859-70.

Table 1 Modified CTD peptides identified from RPB1 purified from cells

Gene	m/z	z	ppm	DB Peptide	Variable Mods	Score	Expect
Polr2a	796.8541	2	8.1	YSPTSPTYSPTSPK	Phospho@5=10	36	1.10E-12
Polr2a	770.394	2	7.6	YSPTSPKYSPTSPK		39.8	1.70E-12
Polr2a	836.8365	2	6.8	YSPTSPTYSPTSPK	Phospho@5=7;Phospho@12=7	24.4	1.80E-12
Polr2a	796.8541	2	8.1	YSPTSPTYSPTSPK	Phospho@7|9|11	35.4	6.40E-12
Polr2a	827.4133	2	4.5	YSPTSPKYSPTSPK	GlyGly@7	16.2	7.90E-08

Chapter 7

Characterization of TET2 post-translational modifications

Abstract

The TET proteins are emerging as important epigenetic regulators through the metabolism of methylated DNA. Hydroxylation of 5-methlycytosine by TETs is associated with activation of genes previously silenced by DNA methylation. In mammalian nuclei, TET proteins form stable complexes with OGT, which is also known to have functional roles in transcriptional regulation. While the OGT-TET complexes have been characterized, the *O*-GlcNAcylation state of TET proteins remains unclear. Here, we show that TET2 is heavily *O*-GlcNAcylated, possessing more that 50 *O*-GlcNAc sites. This work will allow functional studies for individual *O*-GlcNAc sites to be carried out.

Introduction

The Ten-eleven translocation family of enzymes (TET1/2/3) are dioxygenases that hydroxylate 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). Both 5mC and 5hmC are DNA modifications that epigenetically regulate gene transcription in mammals. TET1/2/3 have been shown to physically interact with OGT and influence substrate *O*-GlcNAcylation (1, 2).

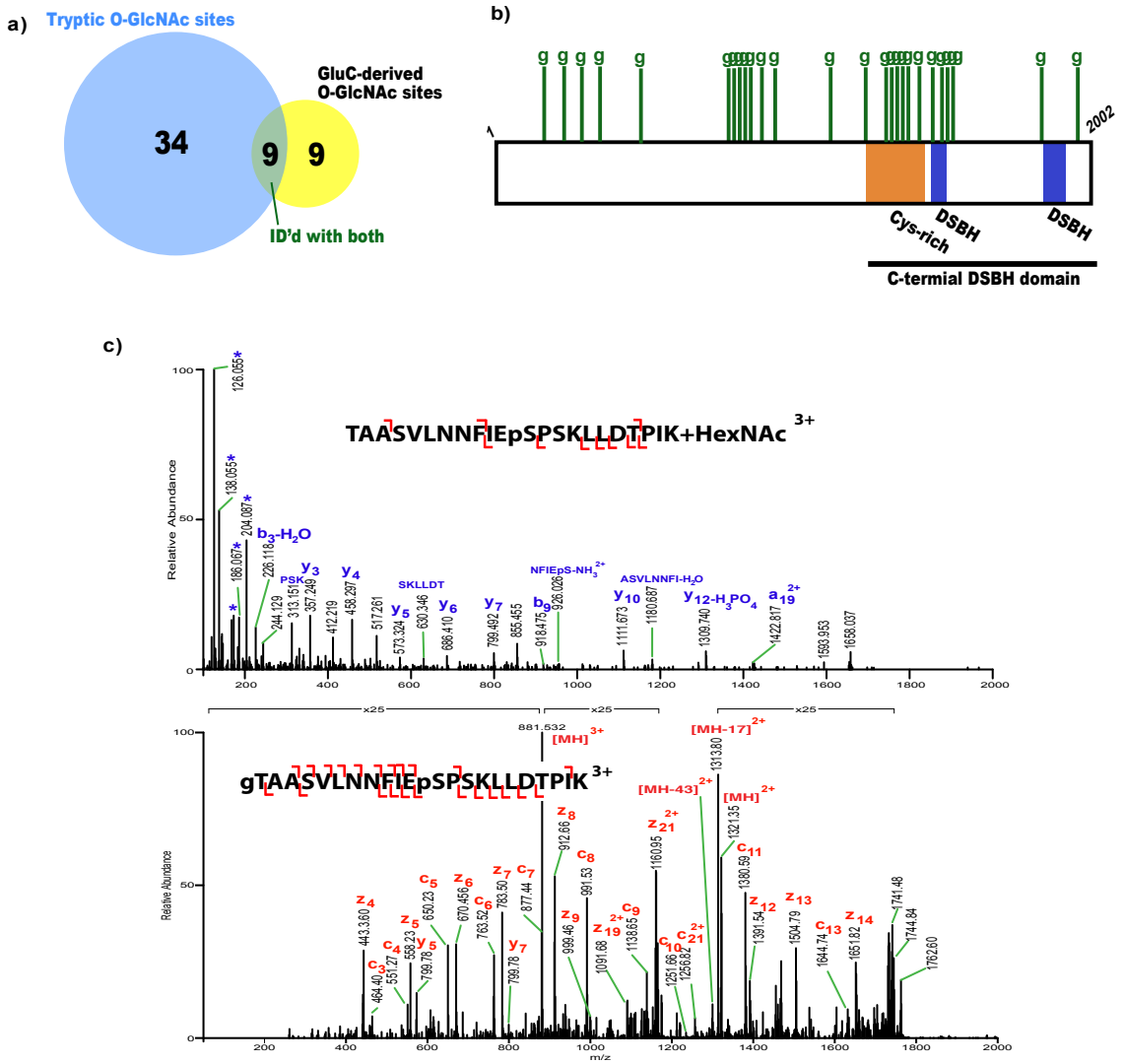
Recent studies have shown that the TET proteins act as scaffolding proteins to recruit OGT to chromatin. Once recruited, OGT glycosylates different sets of substrates depending on the presence or absence of TETs. While the function of the TET-OGT complex has been extensively studied, confusion remains about the *O*-GlcNAcylation state of the TET proteins. In mouse embryonic stem cells, both expressed TET proteins (TET1/2) are *O*-GlcNAc modified, though only on TET1 has an *O*-GlcNAc site been identified (3, 4). In HEK293 cells, where all three TET proteins are expressed, one study failed to detect TET2/3 *O*-GlcNAcylation (1), whereas another study identified all three TET proteins as *O*-GlcNAc modified (5), though no sites of *O*-GlcNAcylation were reported. Therefore, we sought to identify sites of *O*-GlcNAcylation on TET2 purified from HEK293 cells using liquid chromatography tandem mass spectrometry.

Results and Discussion

TET2 was purified from HEK293 cells and subjected to proteolytic digestion using either trypsin or GluC. Both digests were analyzed by UPLC coupled to an ESI LTQ Orbitrap Velos mass spectrometer using both HCD and ETD. We identified 52 *O*-GlcNAc sites between the two digestion methods (Figure 1a). We also identified three dimethylarginine sites, six phosphorylation sites and two lysine acetylation sites (Table 1). The majority of the TET2 *O*-GlcNAcylation was localized to two regions, between amino acids 460-890 and 1015-1677 (Figure 1b). One co-modified peptide containing a single *O*-GlcNAc and a single phosphate was detected (Figure 1c).

Although OGT is known to form a stable complex with TET2 in several cell types, its *O*-GlcNAcylation status is poorly understood. Here, we show that TET2 is heavily glycosylated, consistent with its stable interaction with OGT. Of the two regions possessing the majority of TET2 *O*-GlcNAcylation, one is the double-strand β-helix domain (DSBH) and is responsible for the interaction between OGT and TET2 (6). This data suggests two non-mutually exclusive outcomes. First, the extensive *O*-GlcNAcylation of TET2 may be a result of forming a stable complex with OGT. Second, the *O*-GlcNAcylation of TET2 may functionally alter the activity of TET2 or alter the biomolecular interactions of TET2 with DNA or other proteins.

Figure 1 Identification of TET2 *O*-GlcNAcylation. a) Overlap of *O*-GlcNAc sites on TET2 identified using trypsin or GluC. b) Diagram of TET2. Green "g" indicate approximately where *O*-GlcNAc sites were identified though not all sites are diagramed. Protein domains are indicated. Numbering according to Uniprot accession number Q6N021. DSBH, double-stranded β-helix. c) HCD (top) and ETD (bottom) mass spectra of a TET2 peptide co-modified by *O*-GlcNAc and phosphate.



This study shows unambiguously that TET2 expressed in HEK293 cells is *O*-GlcNAc modified, and that TET2 is heavily glycosylated. The site-specific assignment of *O*-GlcNAc sites on TET2 will allow functional characterization of the modification to determine whether catalytic activity or protein association of TET2 is affected by OGT. This work will enable further studies to understand the function of the TET-OGT complex and its role in transcriptional regulation.

References

1. Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, Dawson MA, Volkmar M, Putmans P, Calonne E, Shih AH, Levine RL, Bernard O, Mercher T, Solary E, Urh M, Daniels DL, Fuks F. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J. 2013;32(5):645-55. PMCID: 3590984.

2. Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K, Pasini D. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013;49(4):645-56.

3. Myers SA, Panning B, Burlingame AL. 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):9490-5. PMCID: 3111310.

4. Shi FT, Kim H, Lu W, He Q, Liu D, Goodell MA, Wan M, Songyang Z. 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):20776-84. PMCID: 3774349.

5. Zhang Q, Liu X, Gao W, Li P, Hou J, Li J, Wong J. Differential Regulation of Ten-Eleven Translocation Family of Dioxygenases by O-Linked beta-N-Acetylglucosamine Transferase OGT. J Biol Chem. 2014;289:5986-5996.

6. Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013;493(7433):561-4. PMCID: 3684361.

7. Guan S, Price JC, Prusiner SB, Ghaemmaghami S, Burlingame AL: A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. *Mol Cell Proteomics* 2011, 10(12):M111 010728.

8. Baker PR, Trinidad JC, Chalkley RJ: Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 2011, 10(7):M111 008078.

9. Medzihradszky KF: Peptide sequence analysis. *Methods Enzymol* 2005, 402:209-244.

Materials and Methods

Purification and preparation of TET2

HEK293T (293T) cells were maintained in DMEM with 5 % FBS. The cells were plated at 70 % confluency (10 X 10 cm dishes) and transfected with 7 µg of pGCN-HA-HCF-1 FL using Polyethylenimine (2 mg/ml) in serum free media. Next day, the transfected cells were plated into 15X15cm dishes. Three days post-transfection, cells were harvested for HCF-1 purification. Total cell extracts were prepared using the lysis buffer (50 mM Tris-HCl, pH 7.3; 5 mM EDTA; 300 mM Nacl; 10 mM NAF; 1% NP-40; 1 mM phenylmethylsulfonyl fluoride (PMSF); 1 mM dithiothreitol; protease inhibitors cocktail (Sigma), and 5 µM PUGNAC (TRC)). Following centrifugation at 15000 rpm/30 min, the supernatant was incubated overnight at 4 ºC with 200 µl of anti-HA-agarose beads (Sigma) (100 µl packed beads for 10 ml extract). The beads were washed several times with the lysis buffer and then transferred into a 800 µl chromatography columns (Biorad) for elution. Bound proteins were eluted with 200 µg/ml of HA peptide (sigma). The purified proteins were precipitated with cold TCA (30%), washed with cold acetone and resuspended in SDS-PAGE sample buffer for western blotting and Coomassie staining. Anti-HCF-1 antibodies, a generous gift from Winship Herr's laboratory, and the anti-O-GlcNAc antibody RL-2 (Santa Cruz Biotechnology) were used as previously described [29]. Following gel destaining, HCF-1 bands were cuts from the gel, reduced with 2 mM TCEP (Thermo) for 30 minutes at 56°C, alkylated with iodoacetamide (Sigma) for 10 minutes at room temperature

in the dark and digested in-gel with 1:100 trypsin or GluC (Roche). Peptides were extracted with 5% formic acid and 50% acetonitrile, concentrated using C18 Ziptips (Millipore) and vacuum centrifugation followed by LC-MS/MS analysis.

LC MS/MS analysis

Chromatography was performed on a Nanoacquity HPLC (Waters) at 600 nl/min with a BEH130 C18 75 μM ID x150mm column (Waters). A 120-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 three most highly abundant, multiply charged ions within a 3 Da isolation window for subsequent HCD and ETD. Precursor scans and HCD product ions were measured in the Orbitrap at a resolution 30,000 and 7,500, respectively. For ions measured in the Orbitrap, one microscan of 250 ms injection time was used. 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 30 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 [7]. 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. 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 and methionine oxidation. Trypsin was allowed one missed cleavage, while GluC was allowed up to two. Proteins identified from this analysis were searched again allowing for HexNAc (neutral loss for HCD data) and phosphorylation of serine and threonines, acetylation, methylation and carbamidomethylation of lysines though only the former two for arginine residues. SLIP scoring was reported for all modification site localization [8] unless the modification site was ambiguous (a false localization rate of less than 5%), then manual interpretation was employed [9].

Table 1 Modified peptides Protein Prospector output for TET2 peptides identified using trypsin and ETD. Bold indicates the residue was identified by hand. Strikethrough indicates that residue was ruled out.

m/z	z	ppm	DB Peptide	Variable Mods	Protein Mods	RT	Score	Expect
830.3967	3	5.7	CFQEQEQKSQQASVLQGYK	HexNAc@9=30	HexNAc@890=30	42.132	36.9	1.10E-07
775.8845	2	4.7	ESVSSVAQENAVK	HexNAc@5=9	HexNAc@153=9	38.394	27.4	4.30E-05
1153.2903	4	6.1	ESVSSVAQENAVKDFTSFSTHNCSGPENPELQILNEQEGK	HexNAc@2=7	HexNAc@150=7	59.28	19.4	2.10E-04
583.8274	2	5.1	FIKSLAER	HexNAc@4	HexNAc@1970	36.649	23.3	4.40E-05
1080.2626	4	8.4	IEGKPEAPPSQSPNPSTHVCSPSPMLSERPQNNCVNR	Oxidation@25;HexNAc@10|16|17	Oxidation@475;HexNAc@460|466|467	41.57	20.4	7.80E-05
678.3285	3	5.2	IQVSCSNNTHLVSENK	HexNAc@4=6	HexNAc@833=6	38.037	38.2	2.70E-07
746.02	3	2.5	IQVSCSNNTHLVSENK	HexNAc@4=12;HexNAc@6|9|13	HexNAc@833=12;HexNAc@835|838|842	37.811	30	7.70E-06
738.5575	5	5.1	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@19=8;HexNAc@4|13	HexNAc@848=8;HexNAc@**833**|842	42.935	23.9	3.90E-05
779.1729	5	4.2	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@4=15;HexNAc@6&19|6&20|9&19|13&19	HexNAc@833=15;HexNAc@835&848|835&849|838&848|8	42.699	37.8	5.40E-08
872.1726	4	2.4	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@4|6|9|13	HexNAc@833|835|838|842	43.112	40.1	5.80E-07
922.9462	4	6.4	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@9&13|9&19|13&19	HexNAc@838&842|838&848|842&848	42.855	36.4	3.60E-07
973.7173	4	7.3	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@4=9;HexNAc@6=6;HexNAc@9|13|19|20	HexNAc@833=9;HexNAc@835=6;HexNAc@838|842|848|8	42.655	25.8	8.80E-05
1162.5649	3	5.7	IQVSCSNNTHLVSENKEQTTHPELFAGNK	HexNAc@4=7	HexNAc@833=7	43.184	16.5	4.80E-04
758.7104	3	6	KVTKQENPPASCDNVQQK	HexNAc@11=32	HexNAc@1015=32	27.077	32.3	3.50E-06
826.4027	3	4.5	KVTKQENPPASCDNVQQK	HexNAc@3;HexNAc@11	HexNAc@1007;HexNAc@1015	27.524	18.2	9.40E-05
793.7346	3	4.1	LAAMLNTCSFQKPEQLQQQK	Oxidation@4	Oxidation@310	46.188	40.6	9.90E-09
713.1311	4	6.8	LSLPPIHTLYQPRFGNSQSFTSK	HexNAc@8=15;Dimethyl@13	HexNAc@1677=15;Dimethyl@1682	53.171	44.3	2.40E-09
763.9018	4	7.5	LSLPPIHTLYQPRFGNSQSFTSK	HexNAc@8=10;Dimethyl@13;HexNAc@22=11	HexNAc@1677=10;Dimethyl@1682;HexNAc@1691=11	52.564	29.1	1.80E-06
627.6337	3	7.3	NDIQTAGTMTVPLCSEK	Oxidation@9	Oxidation@496	44.983	41	4.10E-07
894.7752	3	6.8	NEASLPSILQYQPNLSNQMTSK	HexNAc@4=7;Oxidation@19	HexNAc@585=7;Oxidation@600	57.026	37.2	7.70E-09
834.1581	4	5.3	NPGESSQPNVSDLSDKKESVSSVAQENAVK	HexNAc@19|21|22	HexNAc@150|152|153	43.526	37.8	5.90E-07
637.7854	2	5.2	NSPYSQTMK	HexNAc@5=20;Oxidation@8	HexNAc@820=20;Oxidation@823	25.531	16.8	4.40E-05
603.27	3	6.5	NSPYSQTMKSSACK	Oxidation@8;HexNAc@10=12	Oxidation@823;HexNAc@825=12	25.262	39	3.40E-08
670.9626	3	5.1	NSPYSQTMKSSACK	HexNAc@5=9;Oxidation@8;HexNAc@10=14	HexNAc@820=9;Oxidation@823;HexNAc@825=14	25.045	27.7	1.90E-07
674.0822	4	7.3	QAAQTQPSQSSHLPQNQQQQQK	HexNAc@5=17	HexNAc@730=17	33.182	42	8.60E-08
966.133	3	6.1	QAAQTQPSQSSHLPQNQQQQQK	HexNAc@5=12;HexNAc@10|11	HexNAc@730=12;HexNAc@735|**736**	32.803	25	8.40E-06
661.334	3	8.2	QAYTQKTTQLEHK	HexNAc@7=21;HexNAc@8=34	HexNAc@624=21;HexNAc@625=34	30.344	30.9	1.80E-06
889.955	2	6.1	QAYTQKTTQLEHK	HexNAc@7|8	HexNAc@624|625	30.567	21.9	4.40E-06
606.606	3	3.5	QENPPASCDNVQQK	HexNAc@7	HexNAc@1015	30.886	29.2	5.60E-06
764.104	4	4.7	QEQQQTQQPQTESCHSQMHRPIK	HexNAc@11=7;Oxidation@18	HexNAc@970=7;Oxidation@977	29.414	18.1	0.0012
785.3917	4	3.1	QHLNQQASETEPFSNSHLLQHKPHK	HexNAc@8|10|14|16	HexNAc@708|710|714|716	38.243	35.1	2.00E-07
919.0314	5	5.2	QPPQPQQQQRPQQQQPHHPQTESVNSYSASGSTNPYMR	Oxidation@37;HexNAc@21|23|26|28|30|32	Oxidation@1570;HexNAc@1554|1556|1559|1561|1563|15	38.966	32.7	4.00E-07
959.6485	5	6.3	QPPQPQQQQRPQQQQPHHPQTESVNSYSASGSTNPYMR	HexNAc@21=9;Oxidation@37;HexNAc@**28**|30|32	HexNAc@1554=9;Oxidation@1570;HexNAc@**1561**|1563|1	38.412	25.7	9.30E-06
500.2558	2	5	QSSVFTK	HexNAc@2|3	HexNAc@384|385	32.783	15.2	0.0028
960.4526	3	5.4	QTTAAELDSHTPALEQQTTSSEK	HexNAc@2=8;HexNAc@18|19|20	HexNAc@1069=8;HexNAc@1085|1086|1087	41.718	32.2	2.00E-07
1028.14	3	-0.5	QTTAAELDSHTPALEQQTTSSEK	HexNAc@2&3&19|2&3&21|2&9&21|3&9&21	HexNAc@1069&1070&1086|1069&1070&1088|1069&1076	41.238	17.4	4.40E-05
693.3397	5	-0.7	QTTAAELDSHTPALEQQTTSSEKTPTKR	HexNAc@2=7;HexNAc@18|19	HexNAc@1069=7;HexNAc@1085|1086	38.98	34.5	6.00E-07
815.6573	4	4.6	QTTAAELDSHTPALEQQTTSSEKTPTKR	HexNAc@9|18|19|20	HexNAc@1076|1085|1086|1087	39.16	37	3.00E-07
932.0043	2	6.5	QVKVEMSGPVTVLTR	Oxidation@6;HexNAc@7=26	Oxidation@1058;HexNAc@1059=26	43.666	31.7	1.00E-07
855.8953	2	4.8	QYTGNSNMPGGLPR	HexNAc@3=25;Oxidation@8	HexNAc@606=25;Oxidation@611	38.445	30.3	5.00E-08
820.0656	3	6.3	QYTGNSNMPGGLPRQAYTQK	HexNAc@3=28;Oxidation@8;Dimethyl@14	HexNAc@606=28;Oxidation@611;Dimethyl@617	39.82	24.8	1.30E-05
946.8058	3	3.2	RNEASLPSILQYQPNLSNQMTSK	HexNAc@5=13;Oxidation@20	HexNAc@585=13;Oxidation@600	52.739	29	6.40E-07
646.3414	3	6.3	RTAASVLNNFIESPSK	HexNAc@2=19	HexNAc@1096=19	48.06	42.2	3.80E-07
653.7129	3	6.2	RTVSEPSLSGLLQIKK	HexNAc@2=14	HexNAc@97=14	48.664	45.8	1.70E-08
947.7941	3	7.7	SLAERTMSVTTDSTVTTSPYAFTR	Oxidation@7;HexNAc@1|6	Oxidation@1976;HexNAc@1970|1975	48.728	36.2	2.80E-08
819.7264	6	4.4	SQMYQVEMNQGQSQGTVDQHLQFQKPSHQVHFSKTDHLPK	Oxidation@3;Oxidation@8;HexNAc@33|**35**	Oxidation@633;Oxidation@638;HexNAc@663|**665**	43.331	19.4	4.30E-04
706.3635	2	9.5	SQQASVLQGYK	HexNAc@1=33	HexNAc@890=33	40.073	23.2	2.50E-05
890.957	2	5.8	TAASVLNNFIESPSK	HexNAc@1=25	HexNAc@1096=25	53.253	31.9	2.30E-07
854.4674	3	6.5	TAASVLNNFIESPSKLLDTPIK	HexNAc@1|4	HexNAc@1096|1099	61.224	21.1	1.70E-05
881.1246	3	8.3	TAASVLNNFIESPSKLLDTPIK	HexNAc@1=8;Phospho@12|14	HexNAc@1096=8;Phospho@1107|1109	66.35	42.3	1.40E-09
881.2217	4	7.4	TAASVLNNFIESPSKLLDTPIKNLLDTPVK	HexNAc@1=7;Phospho@**12**|14	HexNAc@1096=7;Phospho@**1107**|1109	71.748	47.6	1.90E-10
1174.624	3	5.2	TAASVLNNFIESPSKLLDTPIKNLLDTPVK	HexNAc@1&Phospho@14|HexNAc@1&Phospho@19|HexNAc@4&Phospho@19	HexNAc@1096&Phospho@1109|HexNAc@1096&Phospho@	69.486	18.2	3.00E-05
692.3434	3	7.5	TLSQYYPDCVSIAVQK	HexNAc@11=43	HexNAc@239=43	52.493	36.8	2.00E-07
439.89	3	6.6	TRPMSEHLK	Oxidation@4;HexNAc@1|5	Oxidation@508;HexNAc@505|509	24.723	17.9	0.0035
558.9801	3	5.8	TVSEPSLSGLLQIK	HexNAc@1=14	HexNAc@97=14	59.043	43.6	3.40E-07
939.508	2	7.1	TVSEPSLSGLLQIK	HexNAc@1=11;HexNAc@3=7	HexNAc@97=11;HexNAc@99=7	58.501	19.6	8.50E-04
610.5953	3	3.6	VEECFHGENQYSK	HexNAc@12	HexNAc@792	35.676	43.3	3.00E-09
754.3915	2	5.6	VEMSGPVTVLTR	Oxidation@3;HexNAc@4=24	Oxidation@1058;HexNAc@1059=24	44.379	28	1.70E-06
684.8207	2	4.6	VSPDFTQESR	HexNAc@6=24	HexNAc@79=24	38.32	25.2	9.30E-05
716.0118	3	5.9	VTKQENPPASCDNVQQK	HexNAc@10=24	HexNAc@1015=24	27.989	27.9	5.50E-06
783.7054	3	6	VTKQENPPASCDNVQQK	HexNAc@2;HexNAc@10	HexNAc@1007;HexNAc@1015	30.411	22.3	1.40E-04

Table 2 Modified peptides Protein Prospector output for TET2 peptides identified using GluC and ETD. Bold indicates the residue was identified by hand. Strikethrough indicates that residue was ruled out.

m/z	z	ppm	DB Peptide	Variable Mods	Protein Mods	RT	Score	Expect
888.3612	2	-1.5	CFHGENQYSKSSE	HexNAc@9=14	HexNAc@792=14	31.619	40.8	5.20E-09
684.6146	3	0.27	CFHGENQYSKSSEFE	HexNAc@9=13	HexNAc@792=13	51.786	32.4	1.40E-08
716.3361	2	0.75	GSFFGQTKVEE	HexNAc@7=20	HexNAc@779=20	39.55	36.4	2.30E-08
752.3623	3	1.7	ILQTFPHPQSNNDQQRE	HexNAc@10=17	HexNAc@765=17	43.232	37.1	2.80E-08
847.6735	4	1.7	ITHPSHTSGQINSAQTSNSELPPKPAAVVSE	HexNAc@30=11	HexNAc@293=11	40.636	34.3	1.10E-07
684.1048	4	-1.4	KEKSAPSRTKQTENASQAKQLAE	HexNAc@12|16	HexNAc@1502|**1506**	52.456	34.4	1.20E-07
697.8532	2	-1.3	KKESVSSVAQE	HexNAc@7=11	HexNAc@153=11	54.086	29.1	1.90E-07
799.3941	2	0.39	KKESVSSVAQE	HexNAc@4&6|4&7|6&7	HexNAc@150&152|150&153|152&153	53.492	25.1	2.20E-07
718.3724	2	-0.87	KSAPSRTKQTE	HexNAc@2|5	HexNAc@**1494**|1497	32.215	29.8	8.60E-07
619.8215	4	0.093	KSAPSRTKQTENASQAKQLAE	HexNAc@10=15	HexNAc@1502=15	39.358	29.6	8.70E-07
1262.2836	3	0.72	KTLSQYYPDCVSIAVQKTTSHINAINSQATNE	HexNAc@12=20	HexNAc@239=20	44.449	26.3	1.00E-06
946.9646	4	0.8	KTLSQYYPDCVSIAVQKTTSHINAINSQATNE	HexNAc@18|19|20|27|30	HexNAc@**245**|246|247|*254*|257	54.126	21.4	1.30E-06
709.1375	4	-0.79	KTPTKRTAASVLNNFIESPSKLLD	HexNAc@7=7	HexNAc@1096=7	31.125	25.3	1.60E-06
932.7681	4	0.65	KTPTKRTAASVLNNFIESPSKLLDTPIKNLLD	HexNAc@7|10	HexNAc@**1096**|1099	61.98	26	1.70E-06
940.967	2	-0.12	MSGPVTVLTRQTTAAE	Oxidation@1;HexNAc@12|13	Oxidation@1058;HexNAc@1069|1070	50.638	21.6	2.40E-06
948.8032	3	0.63	MSGPVTVLTRQTTAAELDSHTPALE	Oxidation@1;HexNAc@12=12	Oxidation@1058;HexNAc@1069=12	53.601	17.4	6.10E-06
1016.4963	3	0.59	MSGPVTVLTRQTTAAELDSHTPALE	Oxidation@1;HexNAc@13=8;HexNAc@9|12	Oxidation@1058;HexNAc@1070=8;HexNAc@1066|**1069**	45.095	27.4	7.80E-06
847.7006	3	1.5	NAVKDFTSFSTHNCSGPENPE	HexNAc@10|11	HexNAc@167|168	45.411	30.5	8.40E-06
643.7959	2	-1.1	NKEQTTHPE	HexNAc@5|6	HexNAc@848|**849**	40.501	20.2	8.80E-06
711.3069	2	-0.092	NQYSKSSEFE	HexNAc@4|6	HexNAc@792|794	40.789	24.5	1.60E-05
927.7002	4	1.2	RTMSVTTDSTVTTSPYAFTRVTGPYNRYI	Oxidation@3;HexNAc@6&7|6&9|6&10|6&12|6&13|7&9|7&10|7&12|7&13|9&1	Oxidation@1976;HexNAc@1979&1980|1979&1982|1979&1983|1979&1985|1979&1986|1980&19	46.07	23.2	2.10E-05
1236.5984	3	1.7	RTMSVTTDSTVTTSPYAFTRVTGPYNRYI	Oxidation@3;HexNAc@2&19|4&19|6&19|6&22	Oxidation@1976;HexNAc@1975&1992|1977&1992|1979&1992|1979&1995	50.348	21.3	2.90E-05
1168.9044	3	1	RTMSVTTDSTVTTSPYAFTRVTGPYNRYI	Oxidation@3;HexNAc@9|10|12	Oxidation@1976;HexNAc@**1982**|1983|1985	41.964	20.1	5.30E-05
924.2538	4	1	TMEQHLKQFHAKSLFDHKALTLKSQKQVKVE	Oxidation@2	Oxidation@1028	23.909	24	1.20E-04
892.8231	5	-0.3	VQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSE	Oxidation@15;HexNAc@12&14|12&17|12&18|17&18|17&25|17&27|17&30|18	Oxidation@823;HexNAc@820&822|820&825|820&826|825&826|825&833|825&835|825&838|826	40.958	17.7	4.60E-04
1115.7778	4	0.41	VQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSE	Oxidation@15;HexNAc@17&18|17&25|17&27|18&25|18&27	Oxidation@823;HexNAc@825&826|825&833|825&**835**|826&833|826&**835**	28.617	18.7	5.70E-04
1065.0053	4	-2	VQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSE	Oxidation@15;HexNAc@17|18|25|27	Oxidation@823;HexNAc@825|826|833|835	24.32	17.1	5.80E-04
710.3397	6	-1.4	VQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSE	Oxidation@15;HexNAc@14|17|18|25|30|34	Oxidation@823;HexNAc@822|825|826|833|838|842	36.167	20.4	7.20E-04
967.0605	5	1.2	VQNINRRNSPYSQTMKSSACKIQVSCSNNTHLVSENKE	Oxidation@15;HexNAc@14&17|14&18|14&25|14&27|14&30|17&18|17&25|17	Oxidation@823;HexNAc@822&825|822&826|822&833|822&835|822&838|825&826|825&833|825	29.54	18	7.80E-04

Appendix I Annotated ETD mass spectra for *O*-GlcNAc modified peptides from Mouse Embryonic Stem Cells

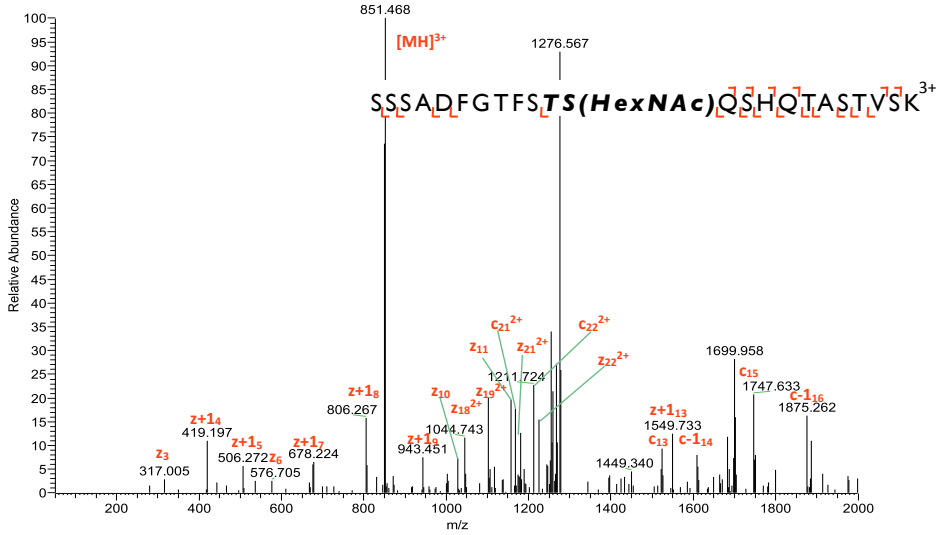
Mouse Embryonic Stem Cell *O*-GlcNAc ETD Mass Spectra

The gene name and the modified residues are given for *Mus musculus*. Bold signifies ambiguous site assginment though modification can be localized to one of the bolded residues. Green and blue identifies mixtures of modification sites where the colors correspond to the residue being modified. Astericks identify carbamidomethylated cystiene residues. Charged reduced species and common neutral losses are not labeled (Falth *et al.* (2008) *Anal Chem* 80 (21) 8089-8094).

Agfg1 GlcNAc-T301 ***or*** S302

T20100618-14 #2046 RT: 22.46 AV: 1 NL: 5.60E2

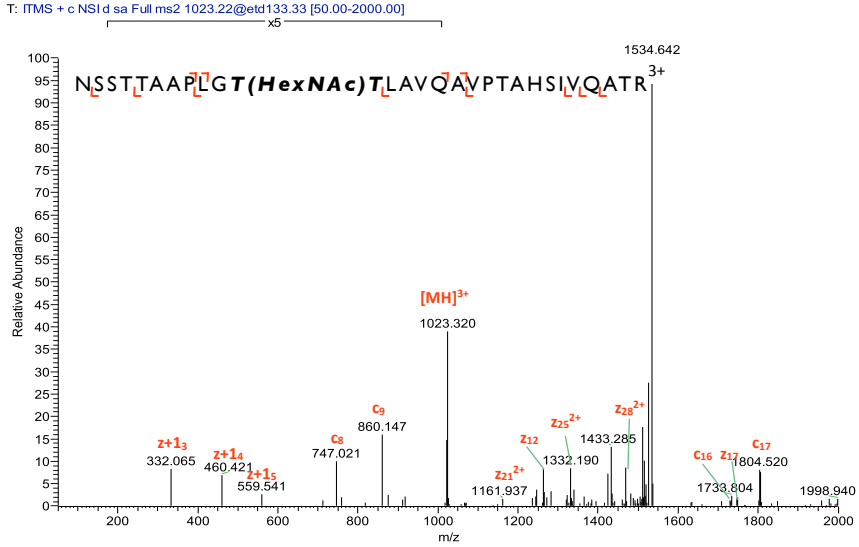
T: ITMS + c NSI d sa Full ms2 851.39@etd133.33 [50.00-2000.00]



Atf7ip GlcNAc-T902 ***or*** T903

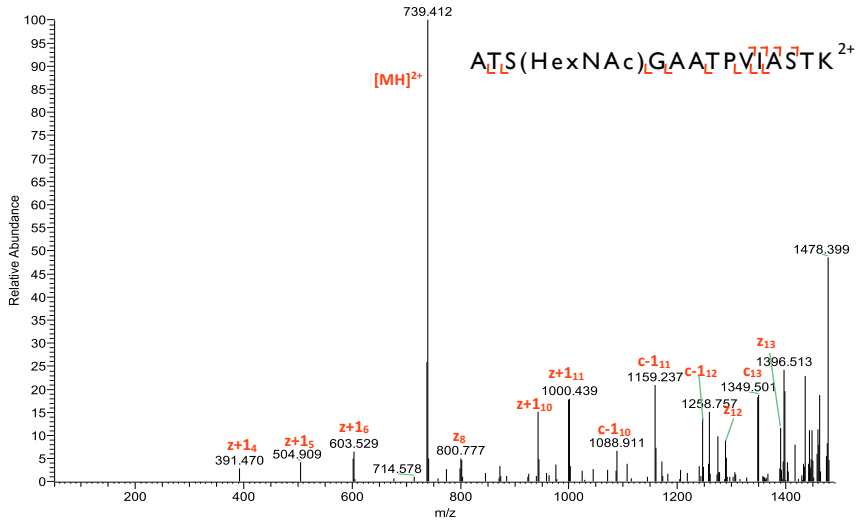
T20100409-24 #3856 RT: 38.48 AV: 1 NL: 1.87E3

T: ITMS + c NSI d sa Full ms2 1023.22@etd133.33 [50.00-2000.00]



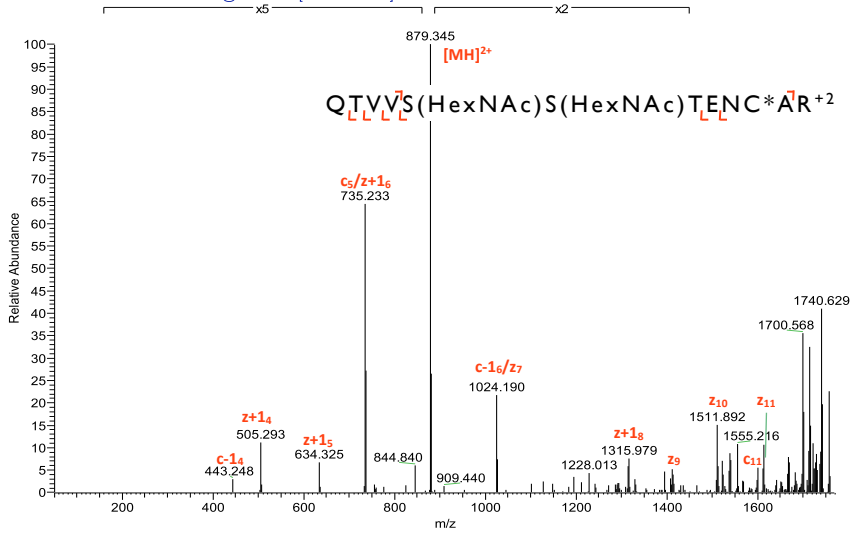
Bnc2 GlcNAc-S515

T20100618-12 #1511 RT: 17.91 AV: 1 NL: 6.57E2
T: ITMS + c NSI d sa Full ms2 739.39@etd200.00 [50.00-1490.00]



Bptf GlcNAc-S1742 & S1743

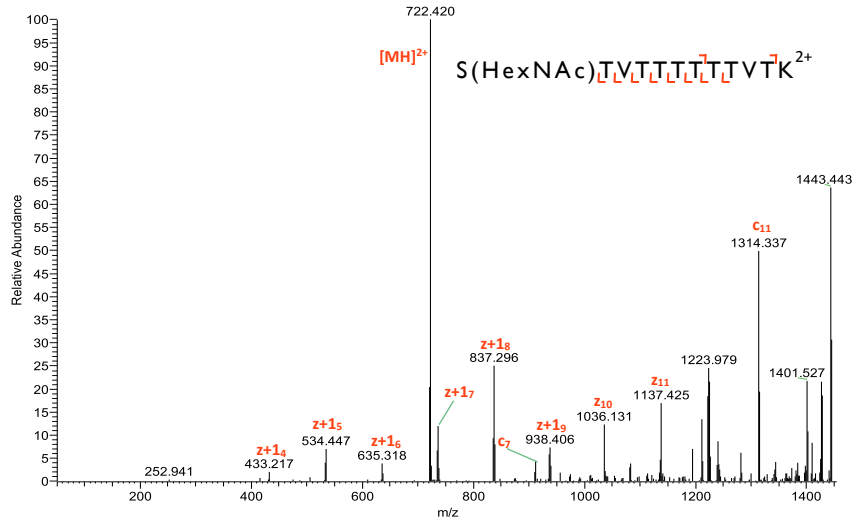
T20101031-16 #1382 RT: 15.99 AV: 1 NL: 1.70E3
T: ITMS + c NSI d sa Full ms2 879.40@etd200.00 [50.00-1770.00]



5

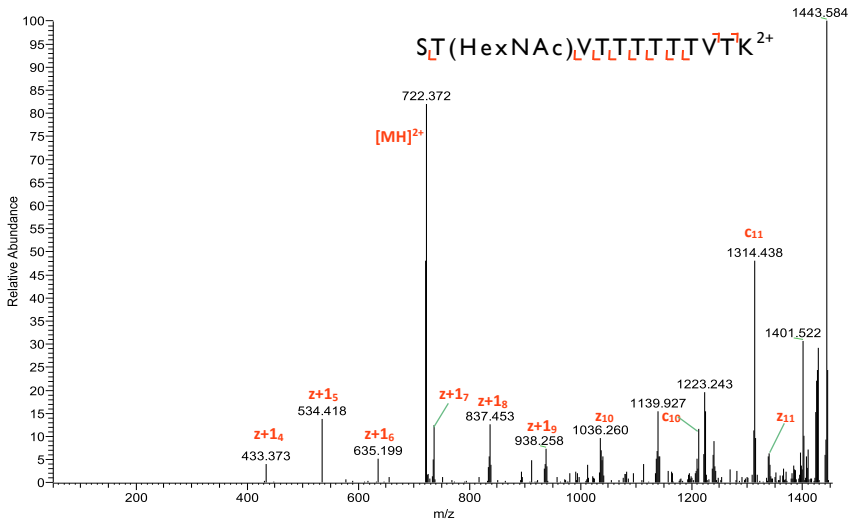
Bptf GlcNAc-S1750

T9100414 #986 RT: 14.59 AV: 1 NL: 1.81E3
T: ITMS + c NSI d sa Full ms2 722.38@etd200.00 [50.00-1455.00]



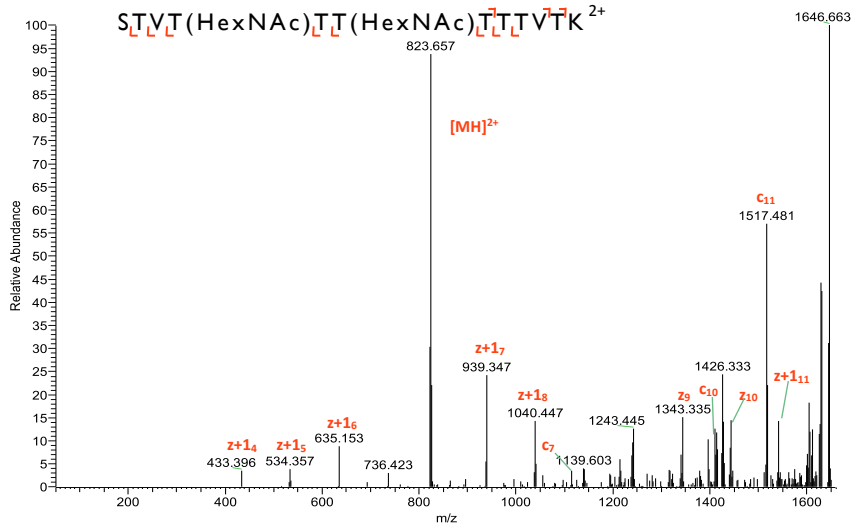
Bptf GlcNAc-T1751

T20100409-13 #1151 RT: 13.28 AV: 1 NL: 1.24E3
T: ITMS + c NSI d sa Full ms2 722.38@etd200.00 [50.00-1455.00]



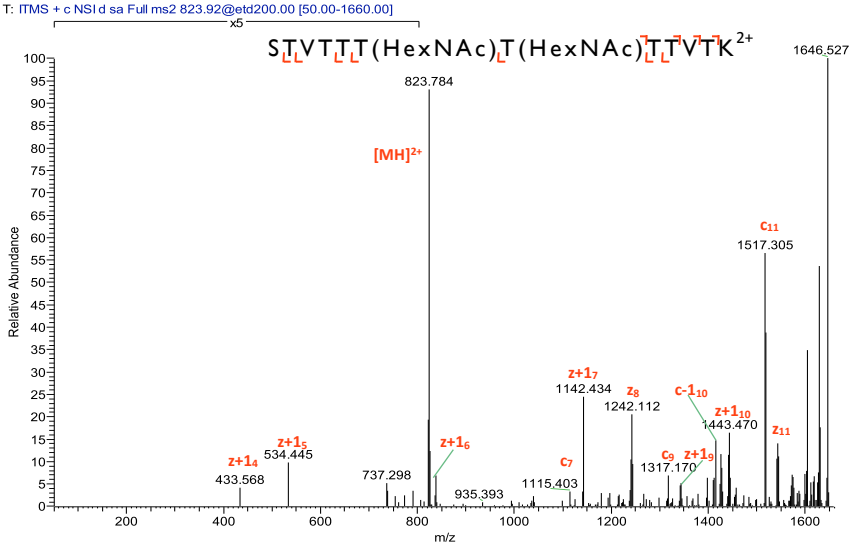
Bptf GlcNAc-T1753 & T155

T20100409-13 #1139 RT: 13.17 AV: 1 NL: 1.22E3
T: ITMS + c NSI d sa Full ms2 823.92@etd200.00 [50.00-1660.00]



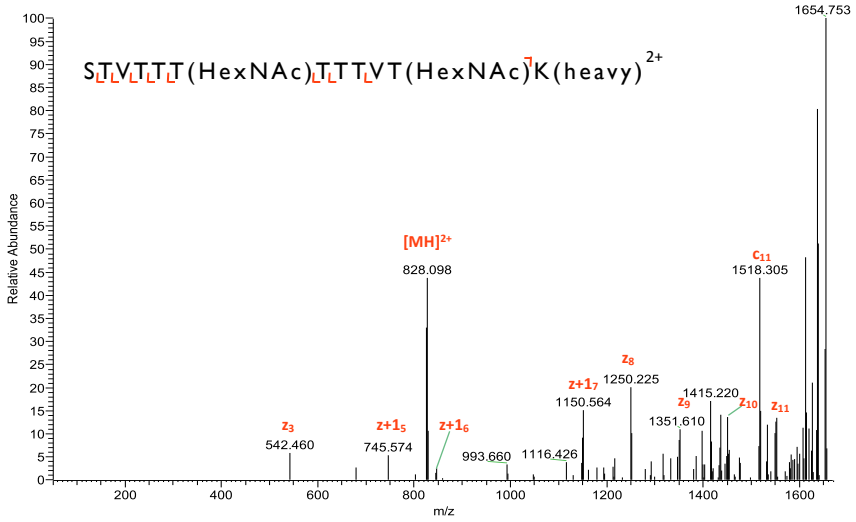
Bptf GlcNAc-T1755 & T1756

T20100409-12 #1081 RT: 12.53 AV: 1 NL: 1.36E3
T: ITMS + c NSI d sa Full ms2 823.92@etd200.00 [50.00-1660.00]



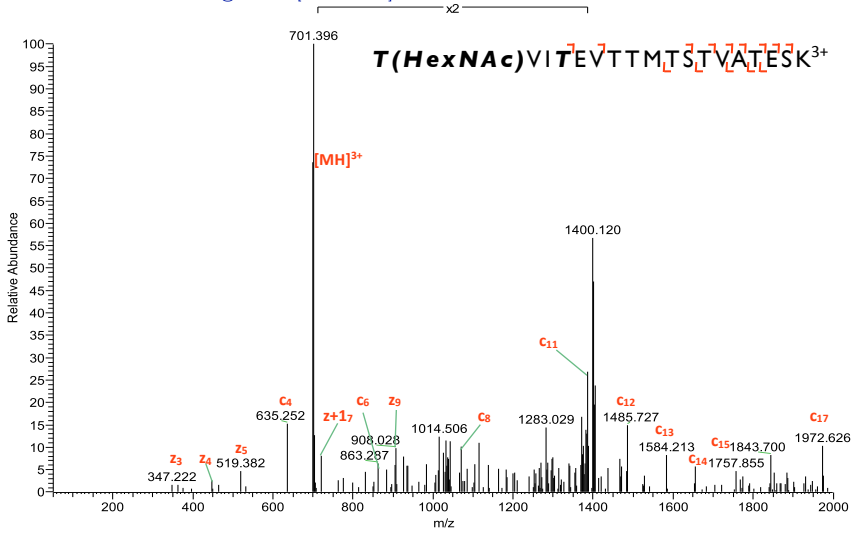
Bptf GlcNAc-T1755 & T1760

T20100618-09 #994 RT: 12.62 AV: 1 NL: 3.85E2
T: ITMS + c NSI d sa Full ms2 827.92@etd200.00 [50.00-1670.00]

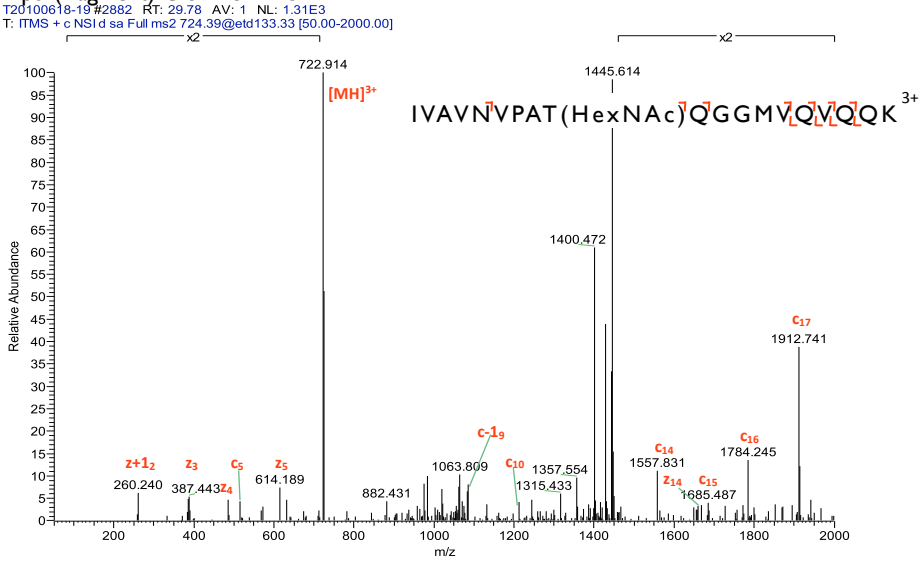


Bptf GlcNAc-T1710 *or* T1713

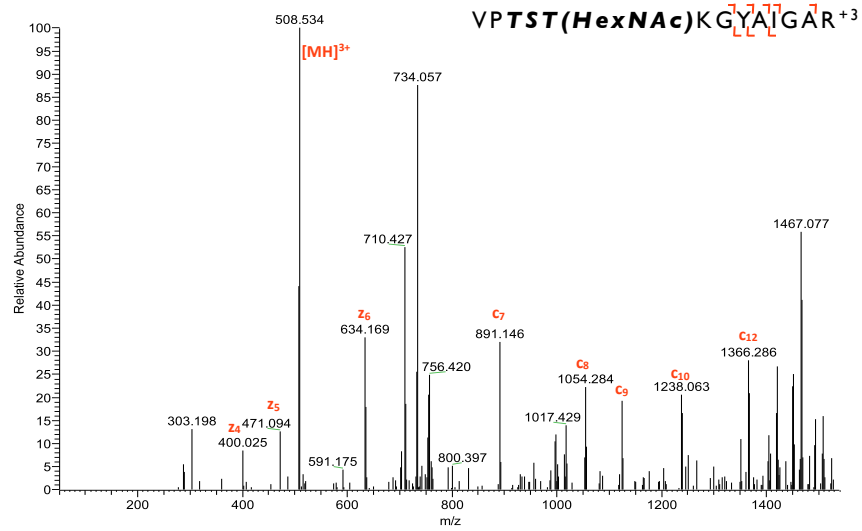
T20100618-14 #3538 RT: 35.22 AV: 1 NL: 5.76E2
T: ITMS + c NSI d sa Full ms2 701.69@etd133.33 [50.00-2000.00]



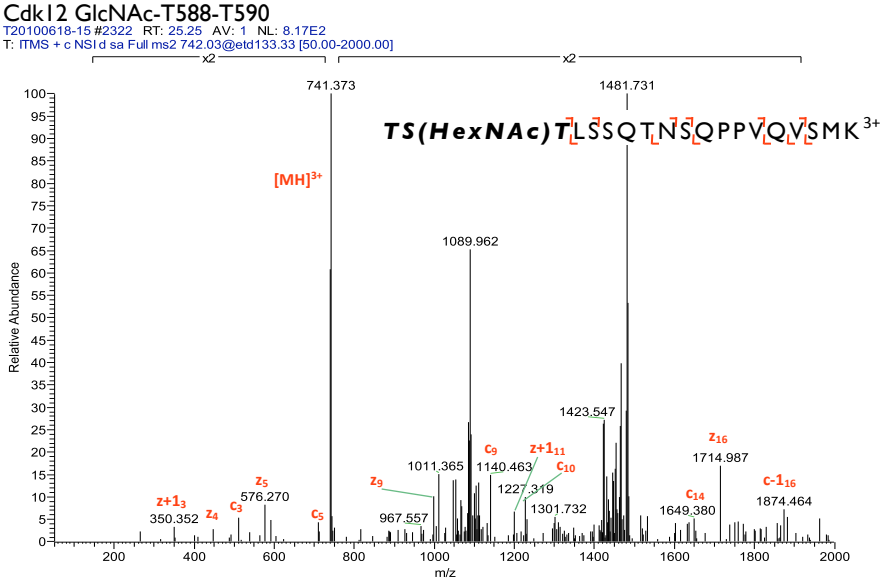
Bptf (fragment) GlcNAc-T225



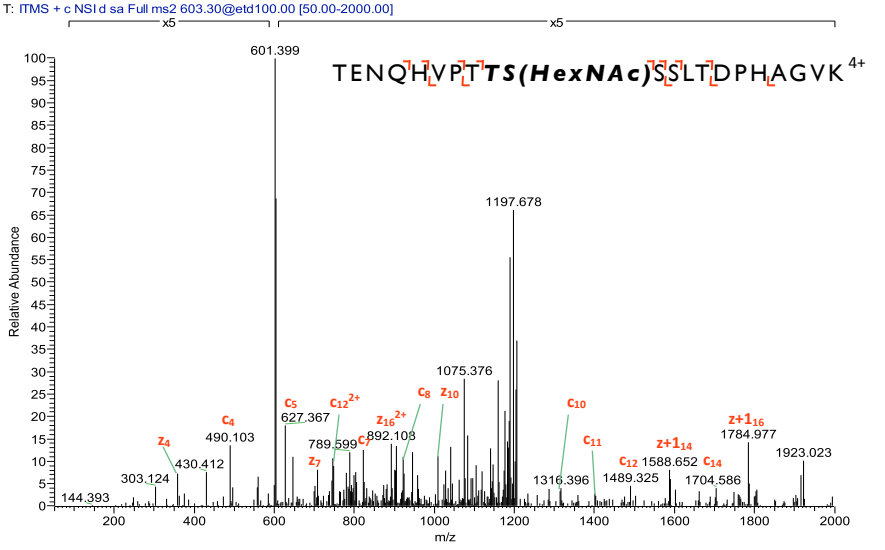
C1orf88 GlcNAc-T82 ***or*** S83 ***or*** T84



Cdk12 GlcNAc-T588-T590

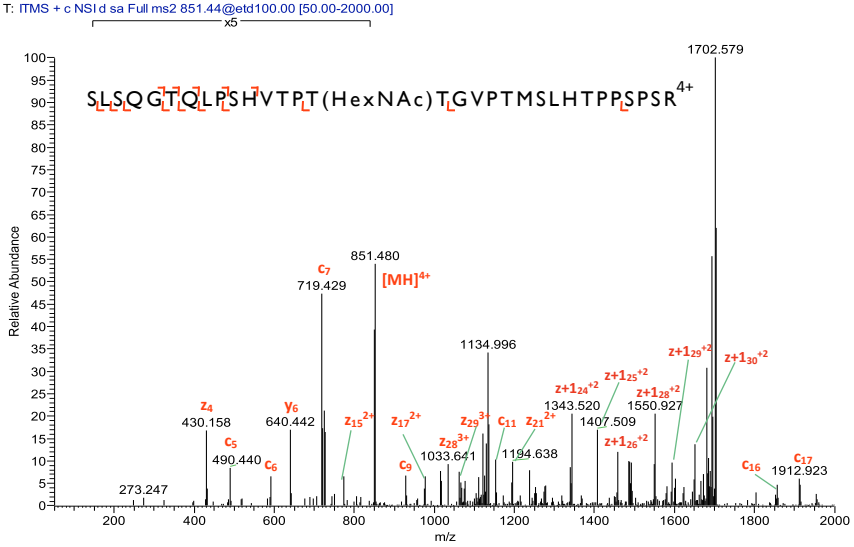


Cdk13 GlcNAc-residue T1286 **or** *S1287*



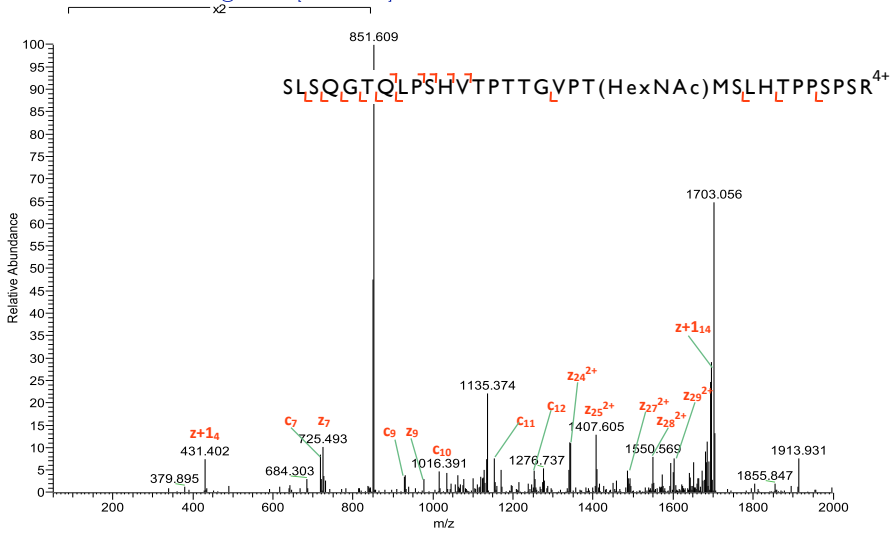
Cnot2 GlcNAc-T113

T20100409-23 #3308 RT: 33.06 AV: 1 NL: 2.91E3
T: ITMS + c NSI d sa Full ms2 851.44@etd100.00 [50.00-2000.00]

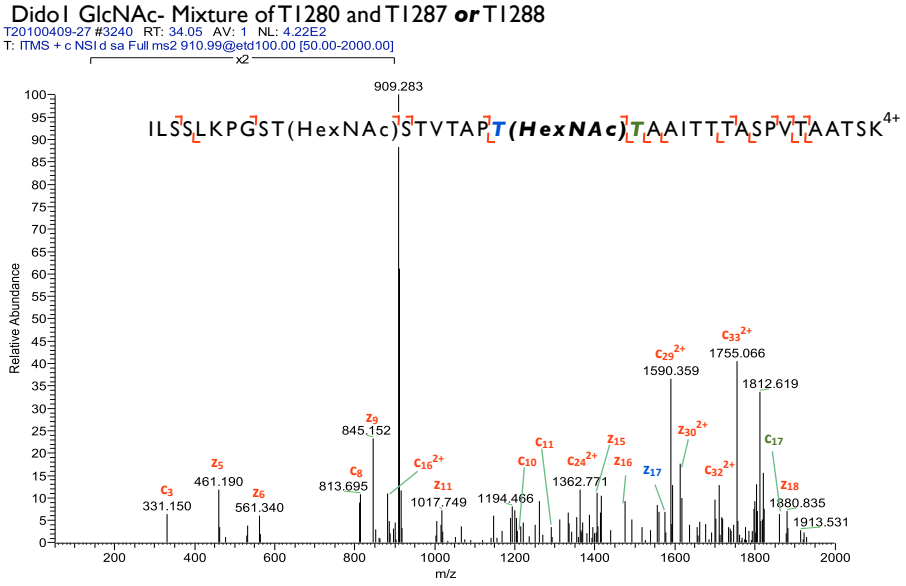


Cnot2 GlcNAc-T118

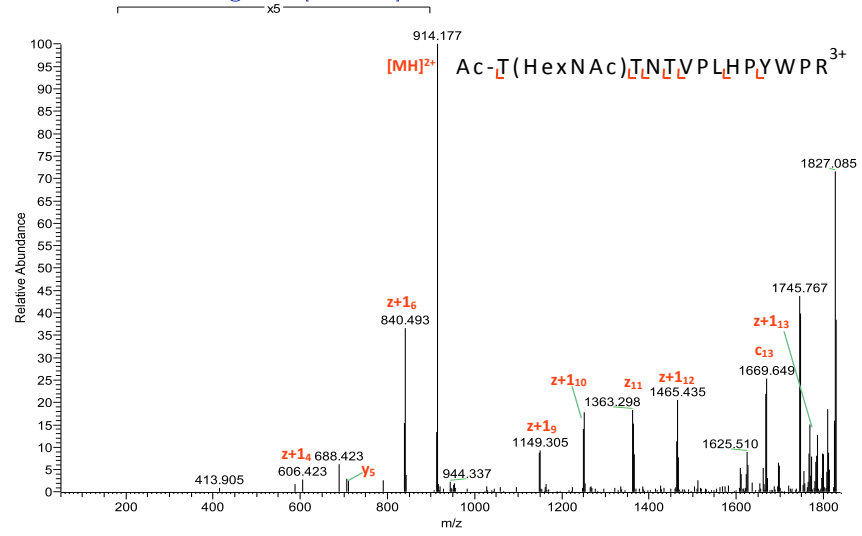
T20100409-24 #3240 RT: 33.18 AV: 1 NL: 2.01E3
T: ITMS + c NSI d sa Full ms2 851.43@etd100.00 [50.00-2000.00]



Dido1 GlcNAc- Mixture of T1280 and T1287 ***or*** T1288

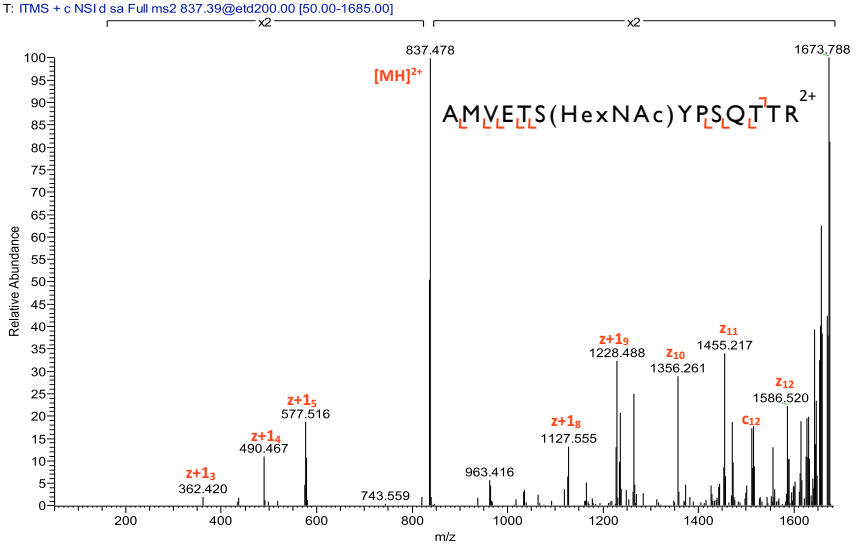


Ebp GlcNAc-T2



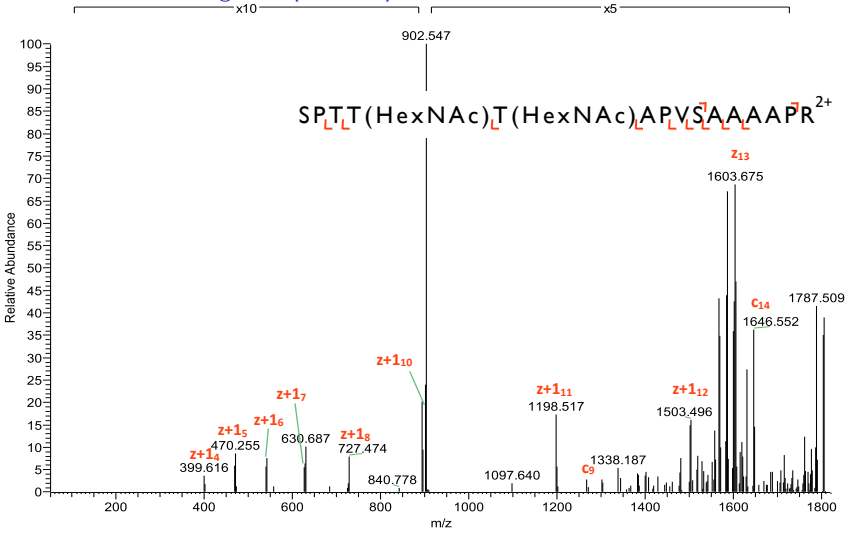
Egr1 GlcNAc-S117

T20100618-10 #1724 RT: 19.20 AV: 1 NL: 1.92E3
T: ITMS + c NSI d sa Full ms2 837.39@etd200.00 [50.00-1685.00]

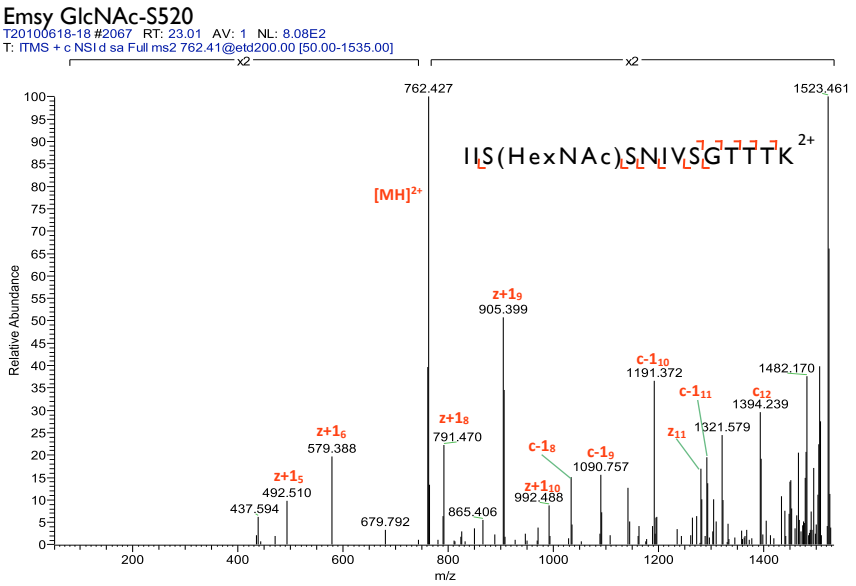


Elf2 GlcNAc-T376 & T376

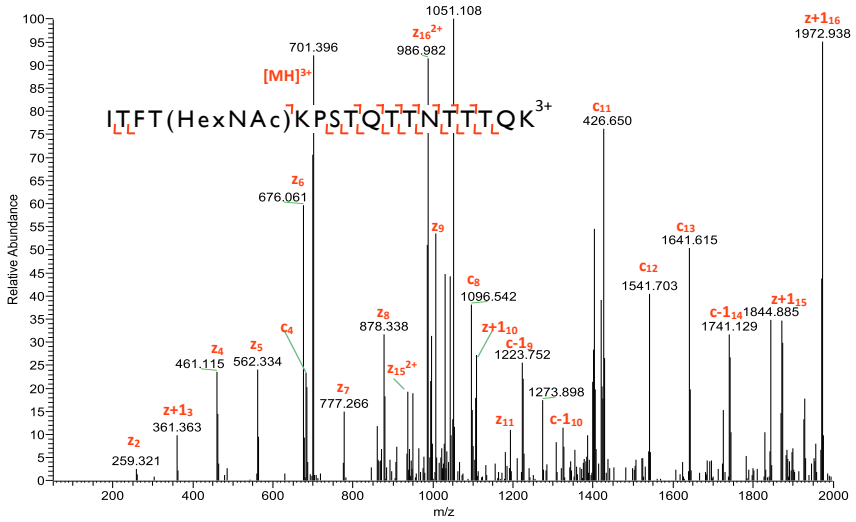
T20100409-15 #1602 RT: 17.49 AV: 1 NL: 2.25E3
T: ITMS + c NSI d sa Full ms2 902.96@etd200.00 [50.00-1820.00]



Emsy GlcNAc-S520

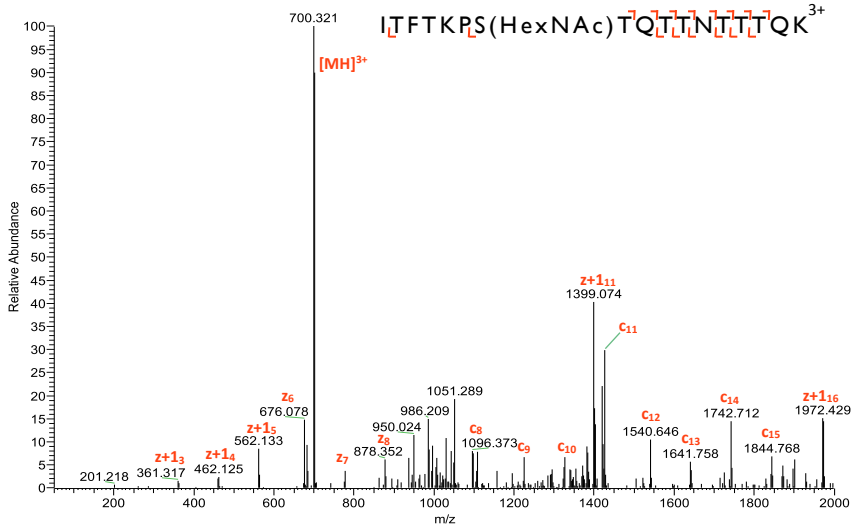


Emsy GlcNAc-T228



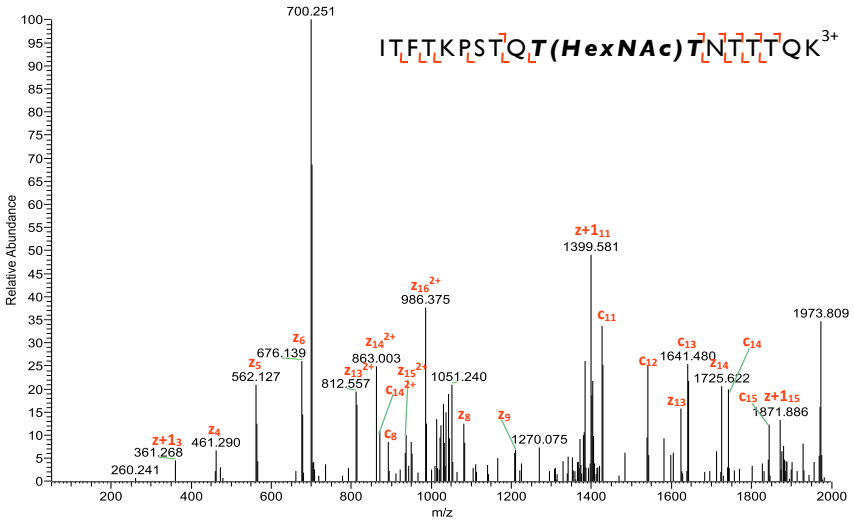
Emsy GlcNAc-S231

T20100409-23 #1710 RT: 19.35 AV: 1 NL: 1.26E3
T: ITMS + c NSI d sa Full ms2 701.36@etd133.33 [50.00-2000.00]

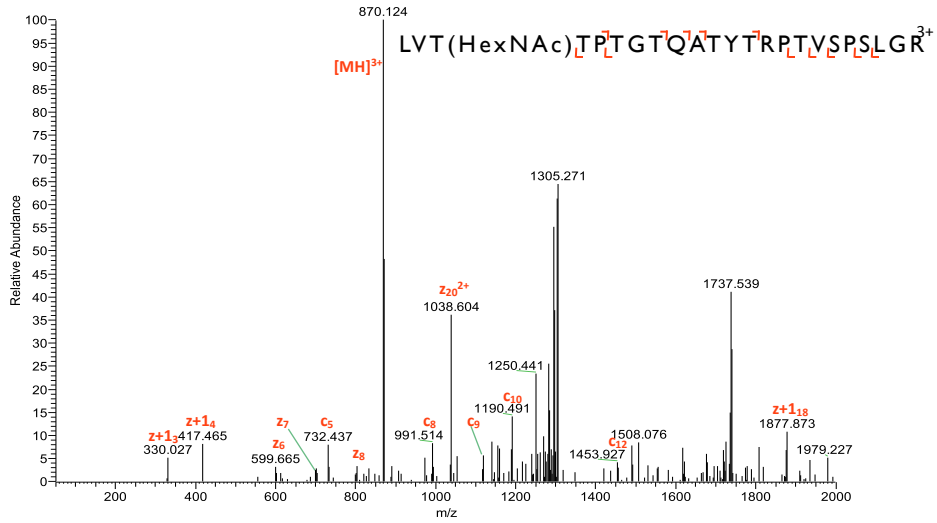


Emsy GlcNAc-T234 ***or*** T235

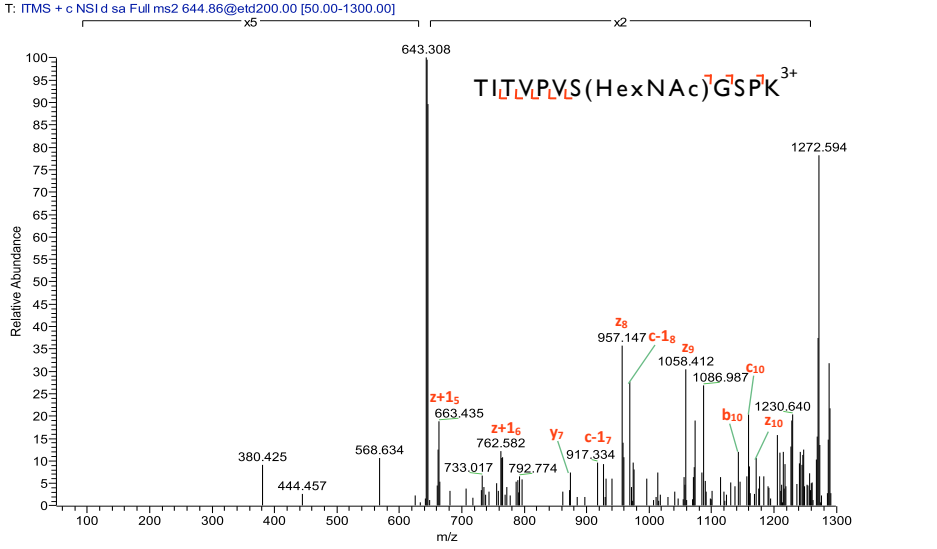
T20100618-20 #1543 RT: 18.59 AV: 1 NL: 3.85E2
T: ITMS + c NSI d sa Full ms2 701.36@etd133.33 [50.00-2000.00]



Emsy GlcNAc-T465

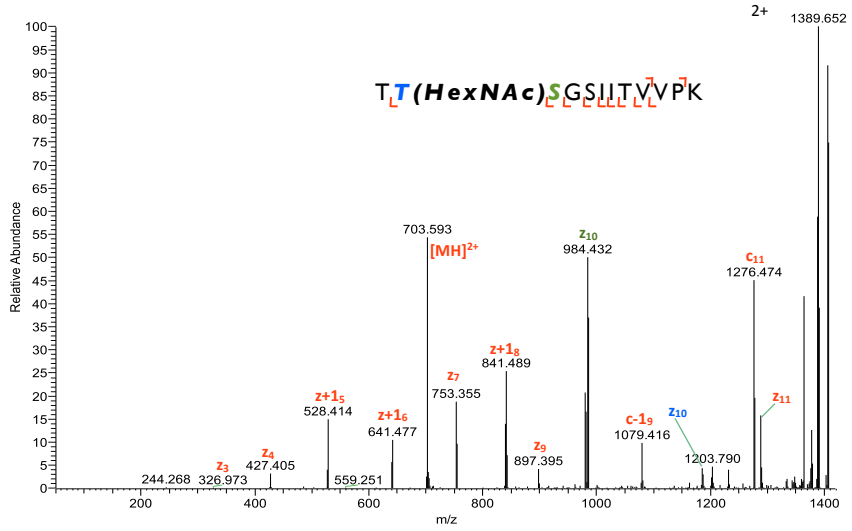


Emsy GlcNAc-S200



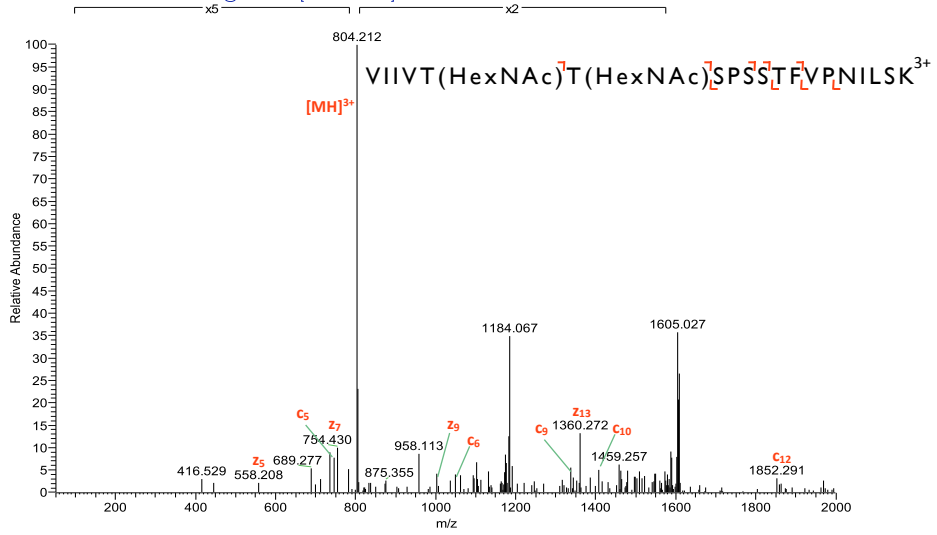
Emsy GlcNAc-Mixture of T499 and S500

T9081017 #1317 RT: 30.69 AV: 1 NL: 1.04E3
T: ITMS + c NSI d sa Full ms2 703.39@etd200.00 [50.00-1420.00]

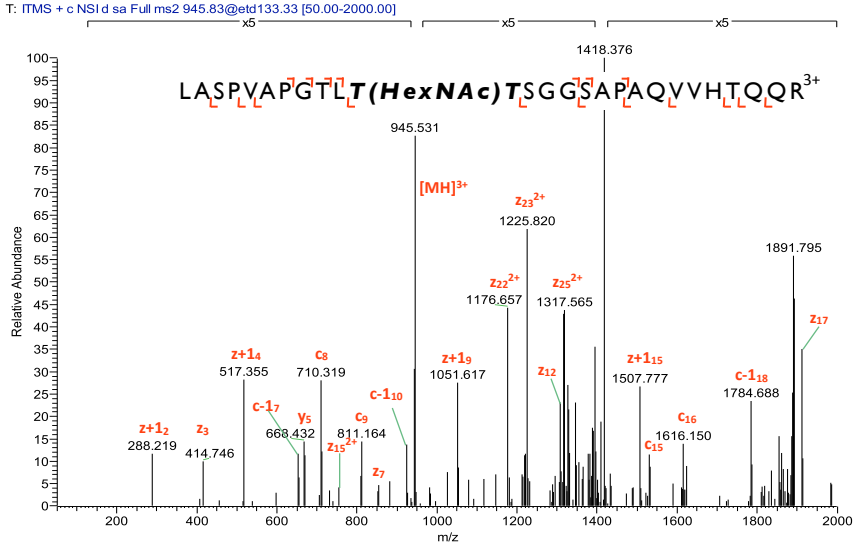


Emsy GlcNAc-T246 & T247

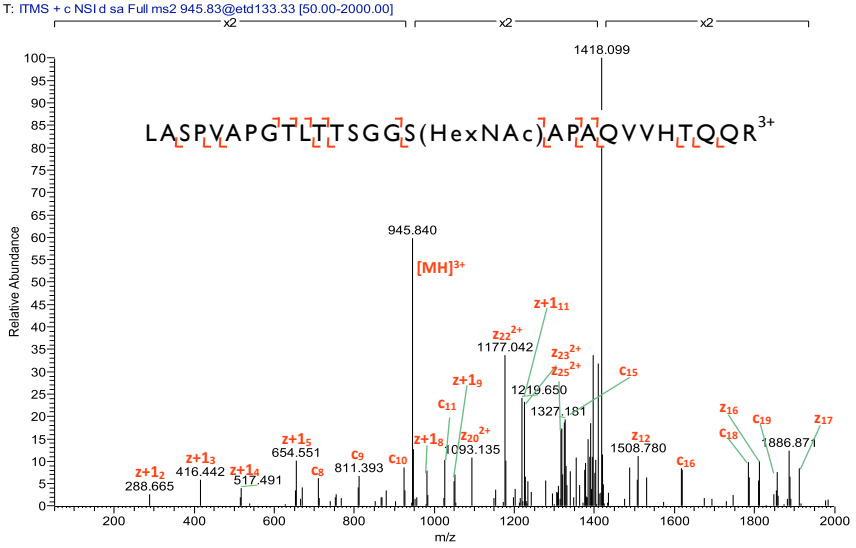
T20100409-26 #4440 RT: 44.08 AV: 1 NL: 9.11E2
T: ITMS + c NSI d sa Full ms2 804.11@etd133.33 [50.00-2000.00]



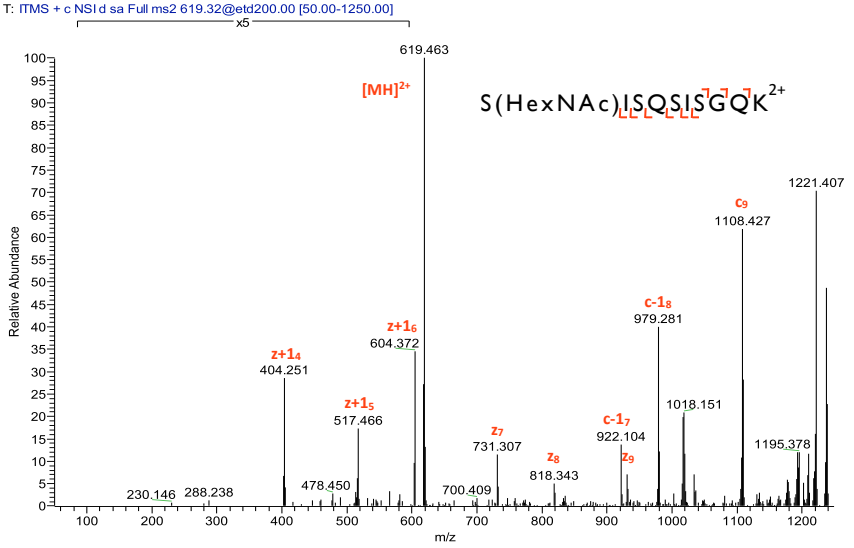
Ep400 GlcNAc-T2594 ***or*** T2595



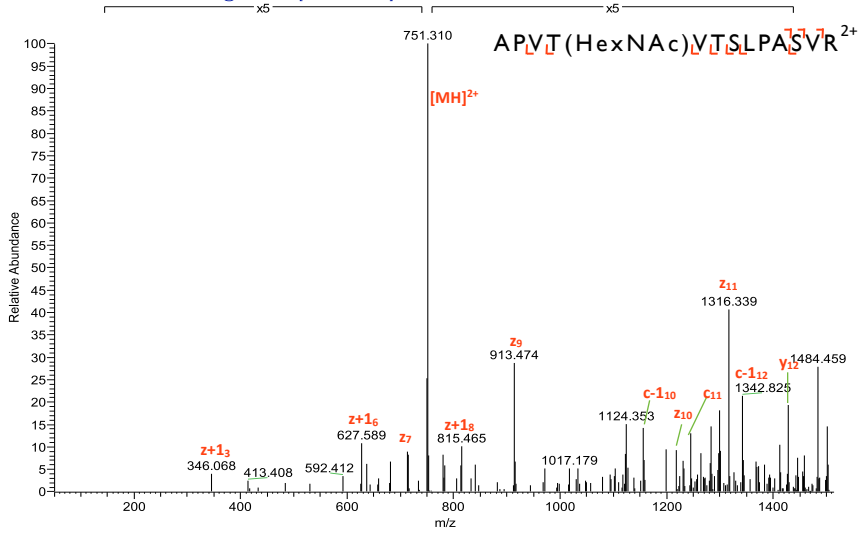
Ep400 GlcNAc-S2599



Gata2b GlcNAc-S585

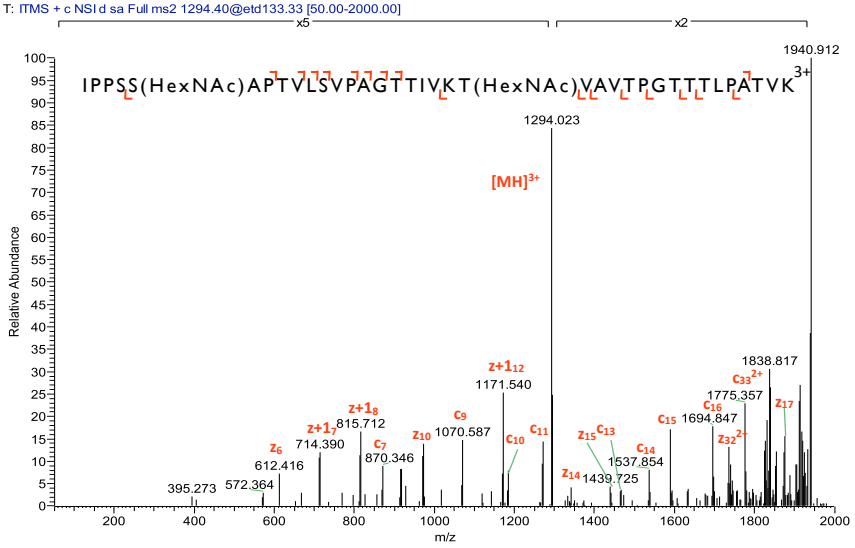


HcfC1 GlcNAc-T515



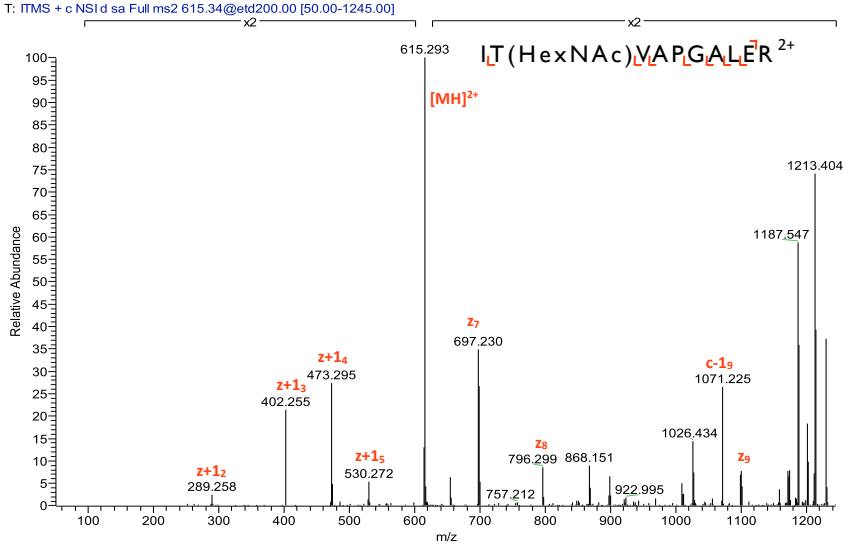
HcfC1 GlcNAc-S563 & T579

T20100409-27 #4075 RT: 41.47 AV: 1 NL: 1.67E3
T: ITMS + c NSI d sa Full ms2 1294.40@etd133.33 [50.00-2000.00]



HcfC1 GlcNAc-T1148

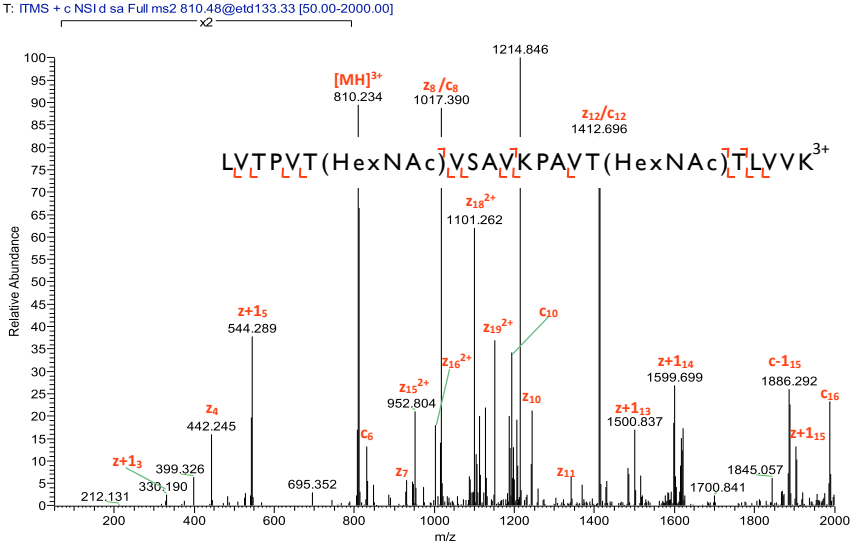
T9100416 #2221 RT: 25.69 AV: 1 NL: 1.24E4
T: ITMS + c NSI d sa Full ms2 615.34@etd200.00 [50.00-1245.00]



HcfC1 GlcNAc-T861 & T870

T20100409-27 #3475 RT: 36.14 AV: 1 NL: 1.64E3

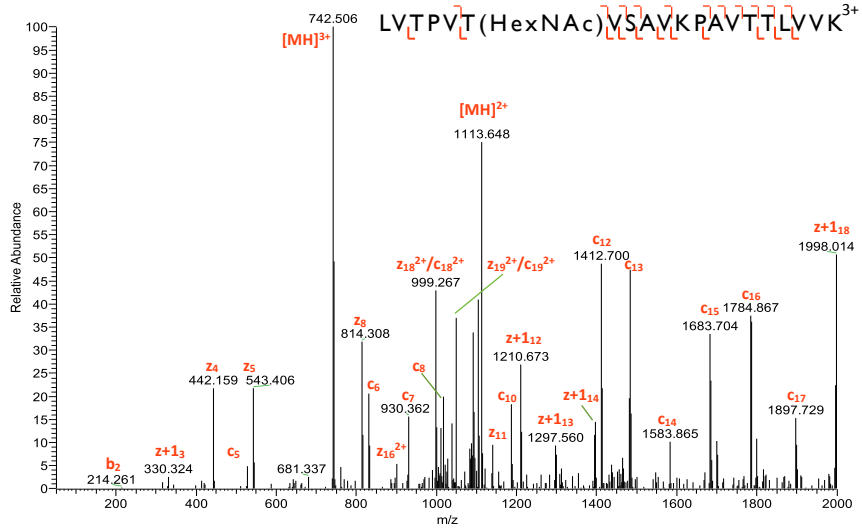
T: ITMS + c NSI d sa Full ms2 810.48@etd133.33 [50.00-2000.00]



HcfC1 GlcNAc-T861

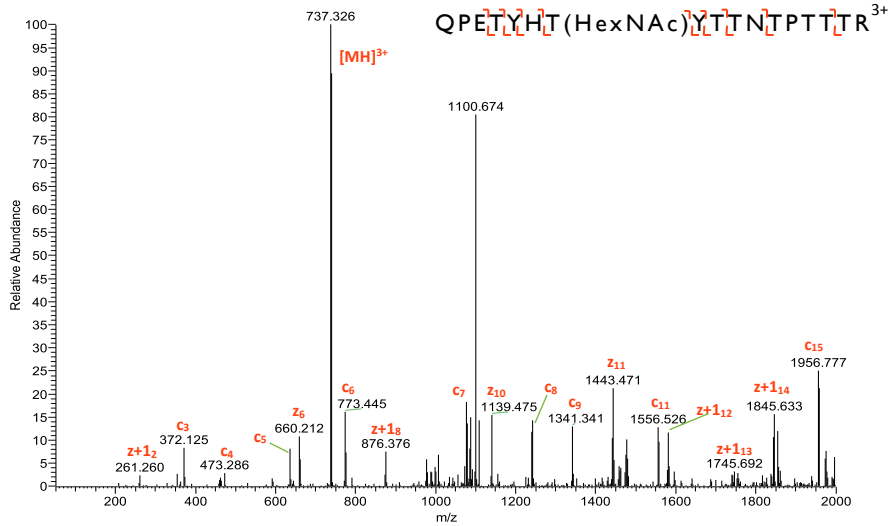
T20100409-28 #3323 RT: 37.86 AV: 1 NL: 8.03E2

T: ITMS + c NSI d sa Full ms2 742.79@etd133.33 [50.00-2000.00]



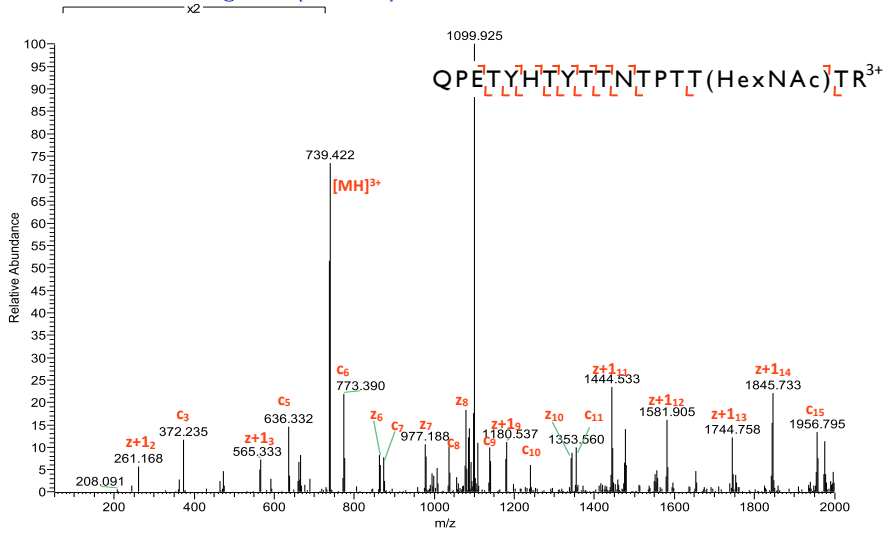
HcfC1 GlcNAc-T1238

T20100409-14 #1485 RT: 16.33 AV: 1 NL: 2.66E3
T: ITMS + c NSI d sa Full ms2 739.35@etd133.33 [50.00-2000.00]

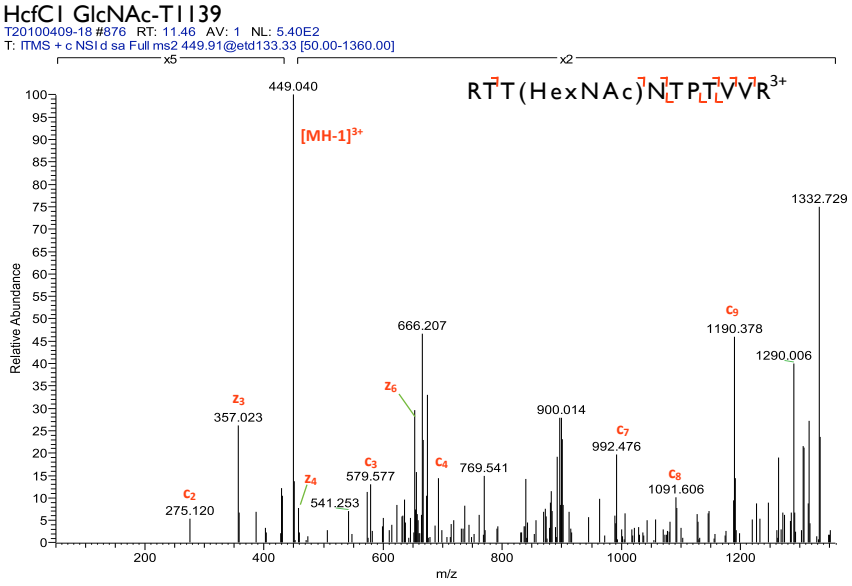


HcfC1 GlcNAc-T1246

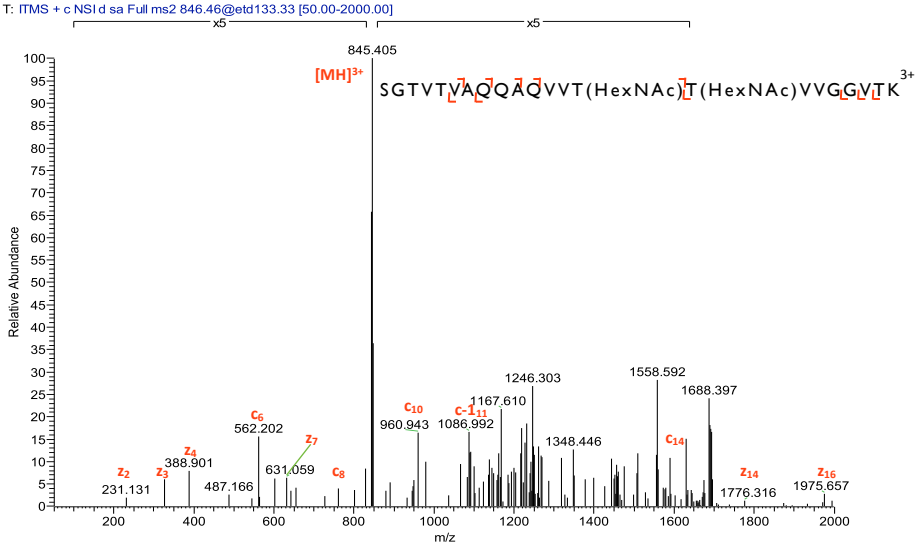
T20100409-15 #1600 RT: 17.47 AV: 1 NL: 4.35E3
T: ITMS + c NSI d sa Full ms2 739.35@etd133.33 [50.00-2000.00]



HcfC1 GlcNAc-T1139



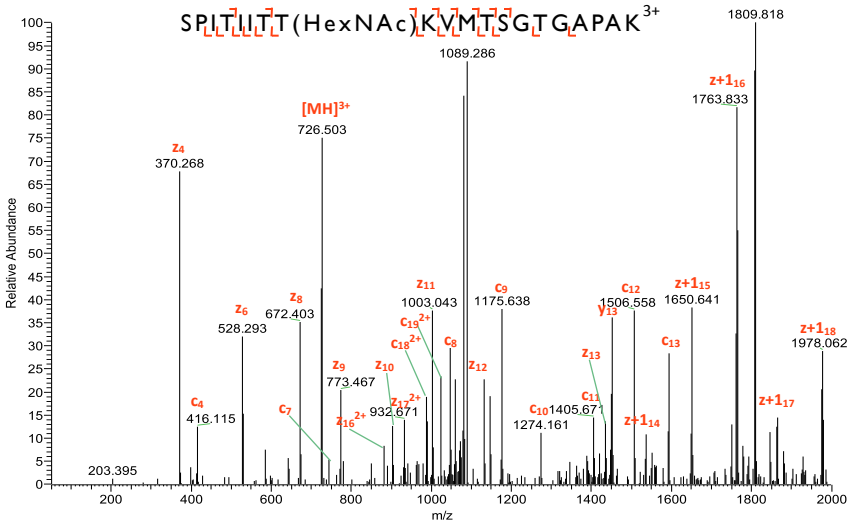
HcfC1 GlcNAc-T651 & T652



HcfC1 GlcNAc-T801

T20100618-24 #2866 RT: 32.10 AV: 1 NL: 7.20E2

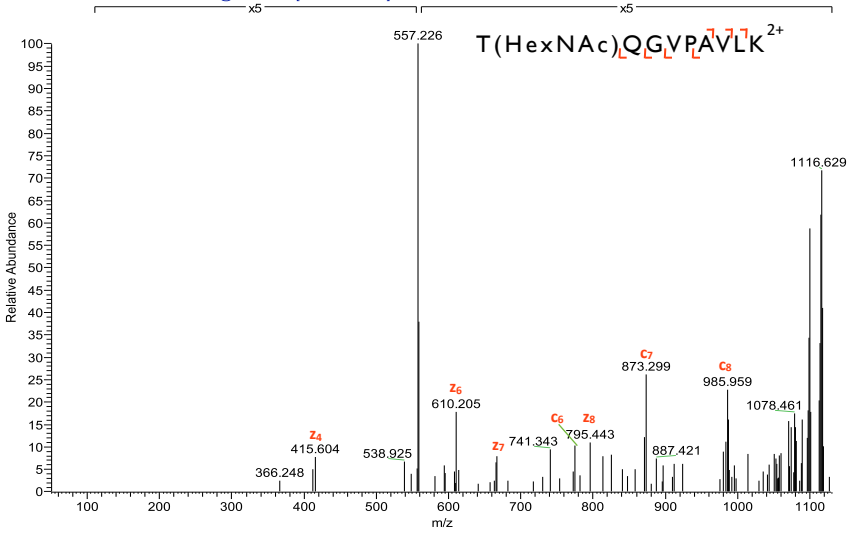
T: ITMS + c NSI d sa Full ms2 726.73@etd133.33 [50.00-2000.00]



HcfC1 GlcNAc-T480

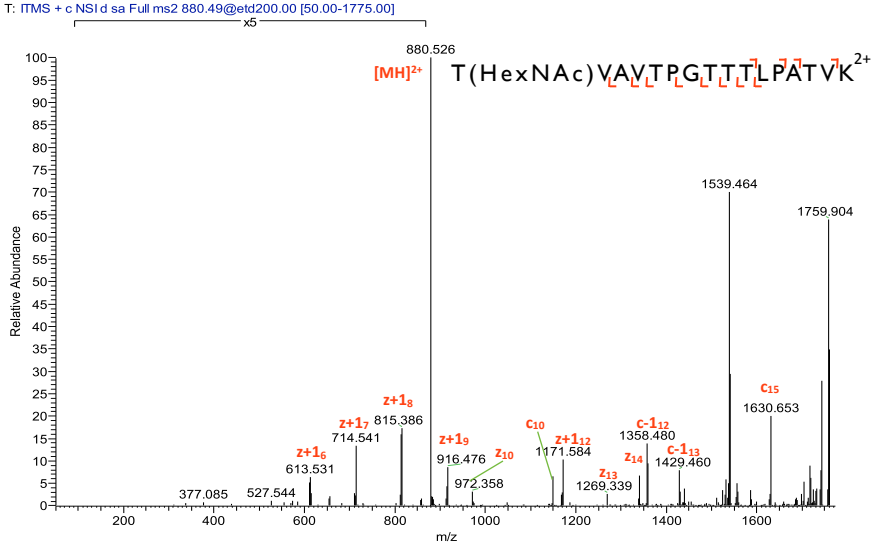
T9070411 #1344 RT: 22.70 AV: 1 NL: 5.64E2

T: ITMS + c NSI d sa Full ms2 558.32@etd200.00 [50.00-1130.00]



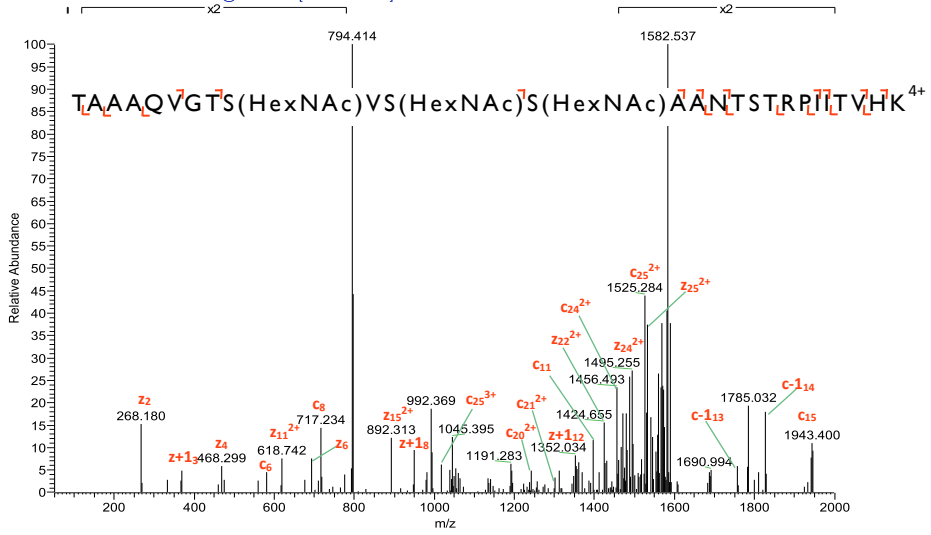
HcfC1 GlcNAc-T579

T20100409-20 #2798 RT: 28.48 AV: 1 NL: 1.32E4
T: ITMS + c NSI d sa Full ms2 880.49@etd200.00 [50.00-1775.00]



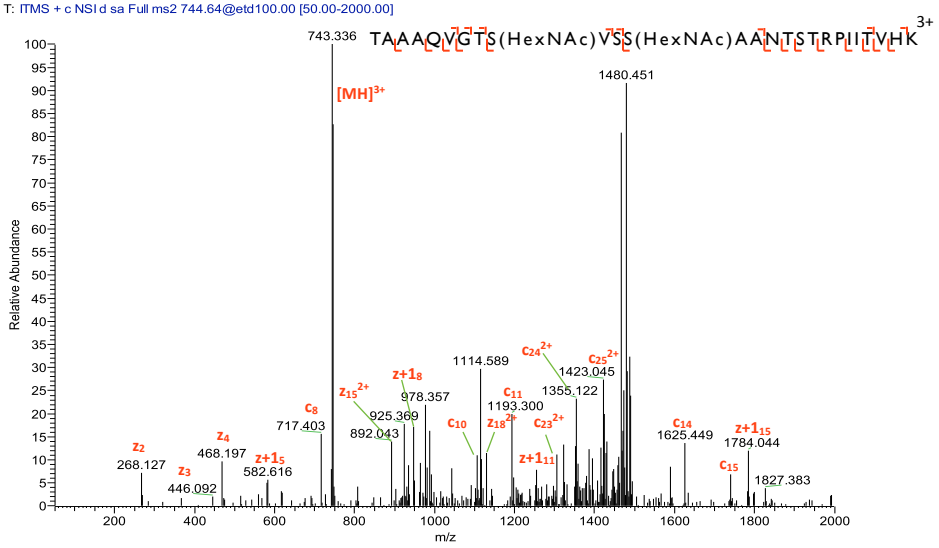
HcfC1 GlcNAc-S620 & S622 & S623

T20100618-19 #2035 RT: 22.83 AV: 1 NL: 7.72E2
T: ITMS + c NSI d sa Full ms2 795.66@etd100.00 [50.00-2000.00]



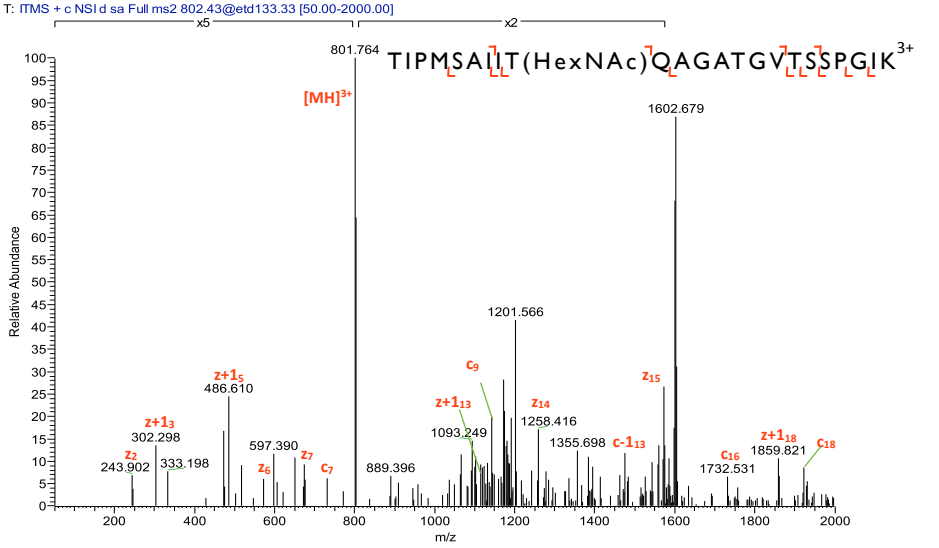
HcfC1 GlcNAc-S620 & S623

T20100409-25 #2370 RT: 25.90 AV: 1 NL: 8.47E2
T: ITMS + c NSI d sa Full ms2 744.64@etd100.00 [50.00-2000.00]



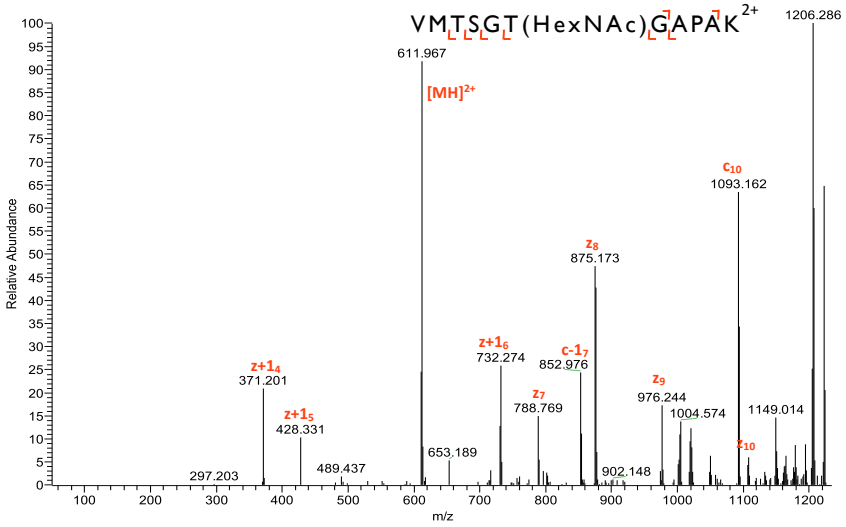
HcfC1 GlcNAc-T779

T20100409-23 #3880 RT: 38.04 AV: 1 NL: 5.53E2
T: ITMS + c NSI d sa Full ms2 802.43@etd133.33 [50.00-2000.00]



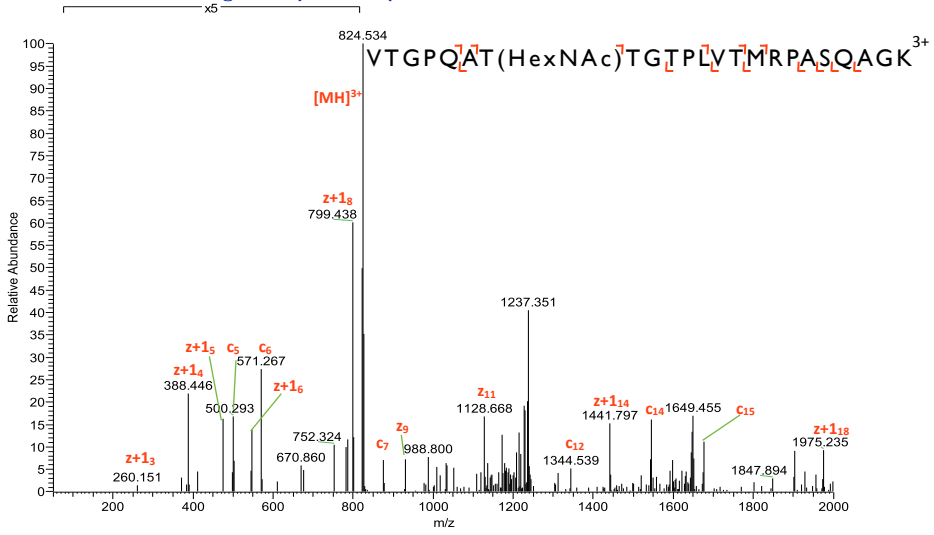
HcfC1 GlcNAc-T808

T20101031-17 #1110 RT: 13.96 AV: 1 NL: 1.13E3
T: ITMS + c NSI d sa Full ms2 611.80@etd200.00 [50.00-1235.00]



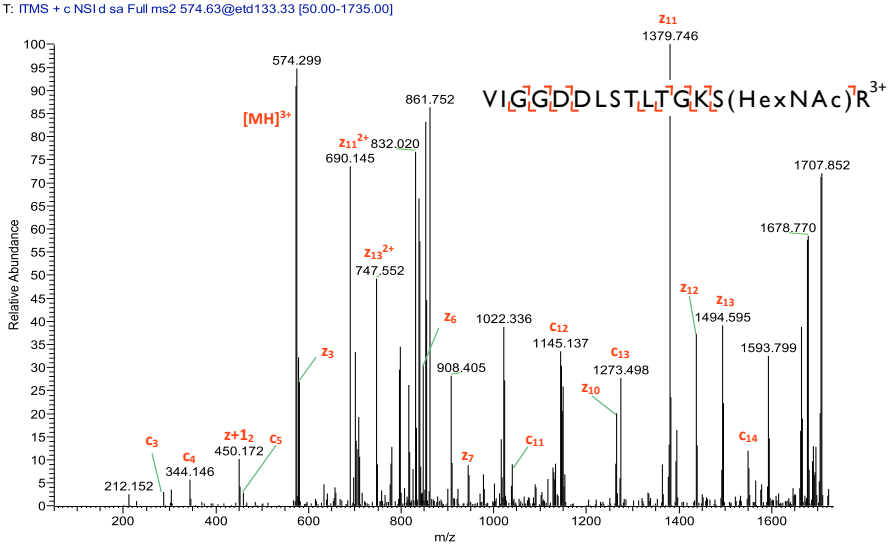
HcfC1 GlcNAc-T495

T20100409-25 #2517 RT: 27.24 AV: 1 NL: 7.63E2
T: ITMS + c NSI d sa Full ms2 825.10@etd133.33 [50.00-2000.00]



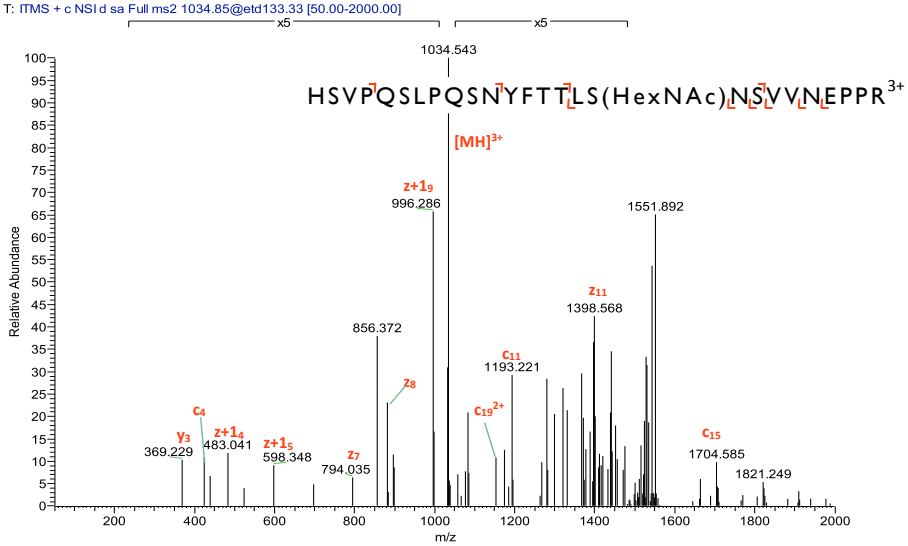
Hprt1 GlcNAc-S129

T20100409-13 #2296 RT: 22.94 AV: 1 NL: 3.11E3
T: ITMS + c NSI d sa Full ms2 574.63@etd133.33 [50.00-1735.00]



Jmjd1c GlcNAc-S911

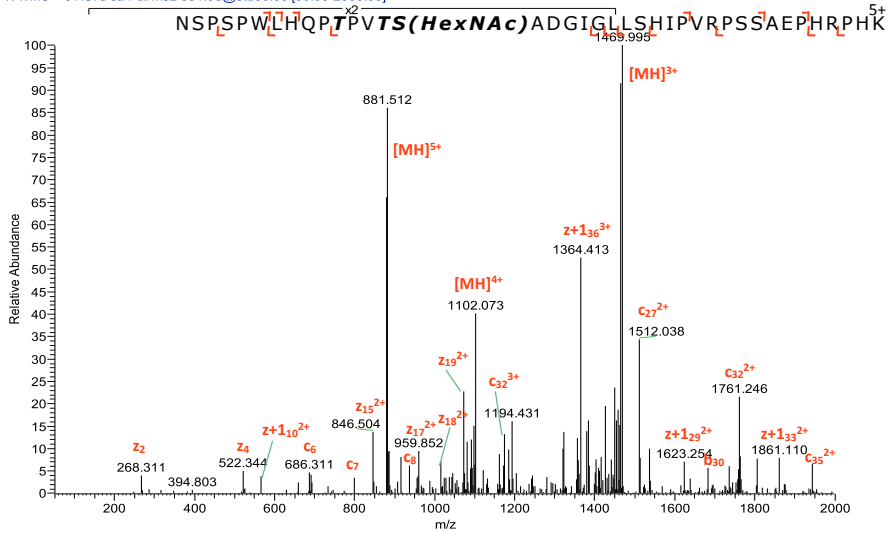
T20100409-22 #4088 RT: 39.73 AV: 1 NL: 7.67E2
T: ITMS + c NSI d sa Full ms2 1034.85@etd133.33 [50.00-2000.00]



Jmjd1c GlcNAc-T728 ***or*** T731 ***or*** S732

T20100409-26 #4195 RT: 41.93 AV: 1 NL: 2.37E3

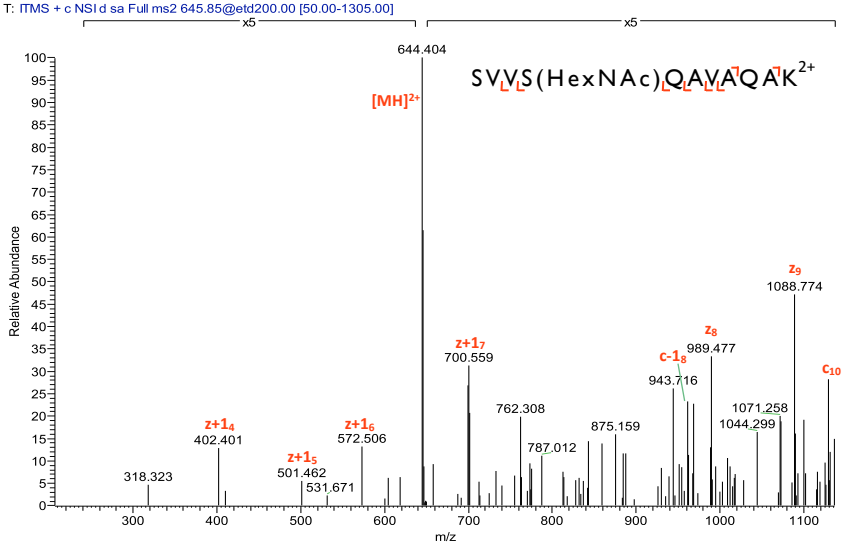
T: ITMS + c NSI d sa Full ms2 881.86@etd80.00 [50.00-2000.00]



Jmjd1c GlcNAc-S1250

T20100409-17 #1969 RT: 21.12 AV: 1 NL: 6.63E2

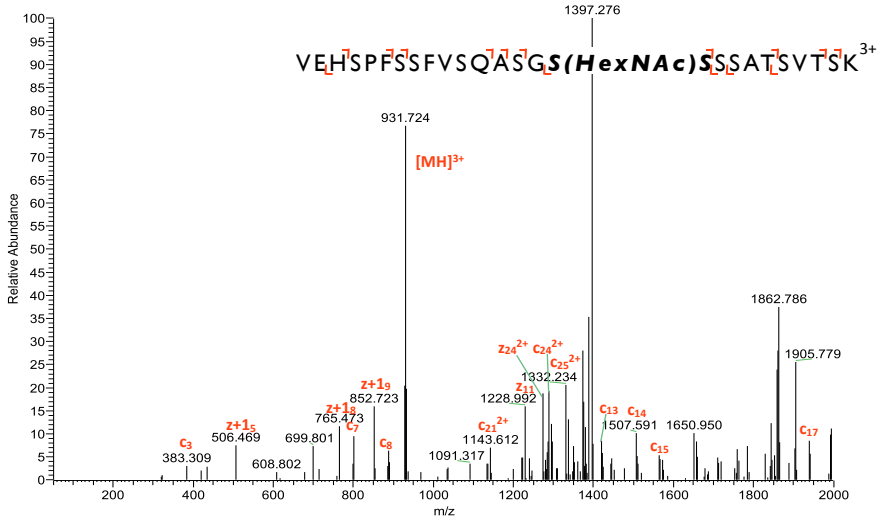
T: ITMS + c NSI d sa Full ms2 645.85@etd200.00 [50.00-1305.00]



Kdm3b GlcNAc-S460 ***or*** S461

T20100618-19 #3120 RT: 31.65 AV: 1 NL: 4.83E2

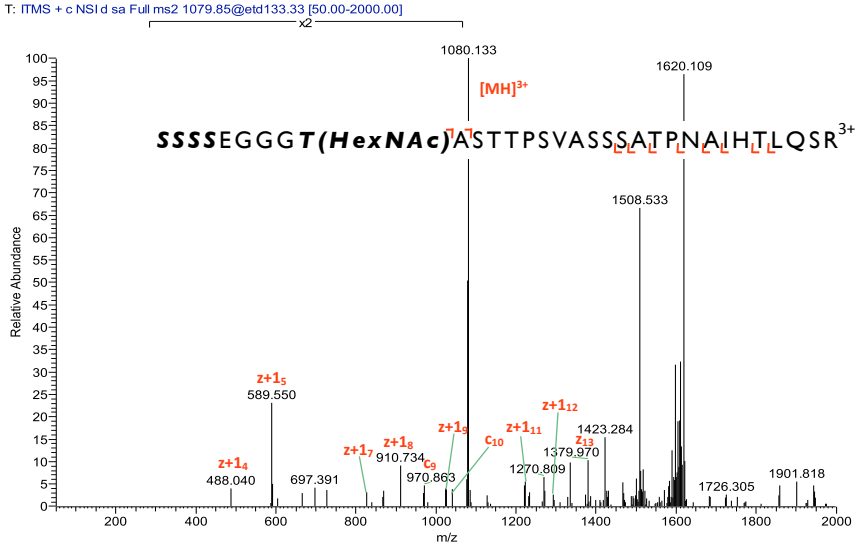
T: ITMS + c NSI d sa Full ms2 931.44@etd133.33 [50.00-2000.00]



Kia1310 Between GlcNAc-S699-T707

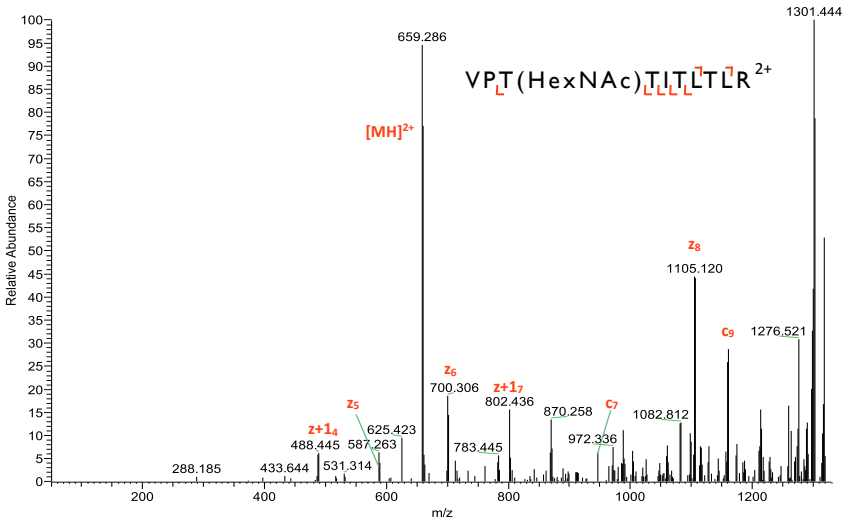
T20100409-18 #2638 RT: 26.92 AV: 1 NL: 7.90E2

T: ITMS + c NSI d sa Full ms2 1079.85@etd133.33 [50.00-2000.00]



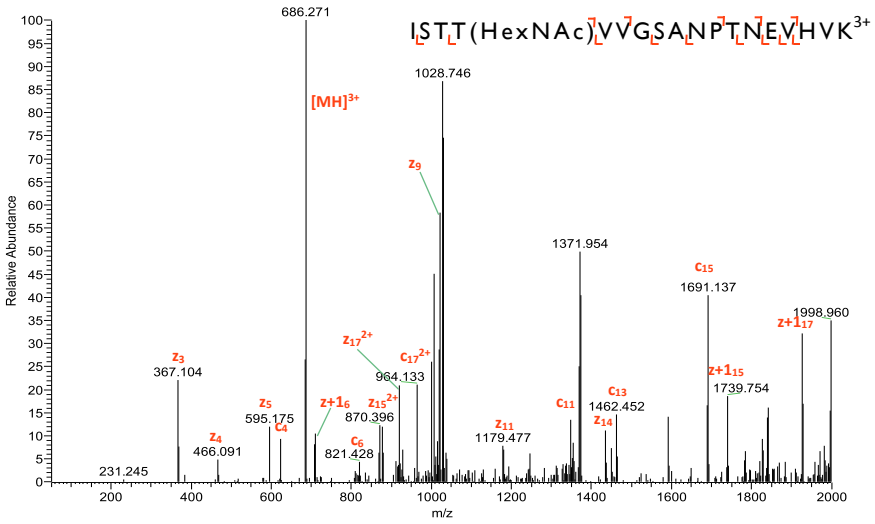
Kia1310 GlcNAc-T858

T20100409-23 #3345 RT: 33.38 AV: 1 NL: 6.71E2
T: ITMS + c NSI d sa Full ms2 659.39@etd200.00 [50.00-1330.00]



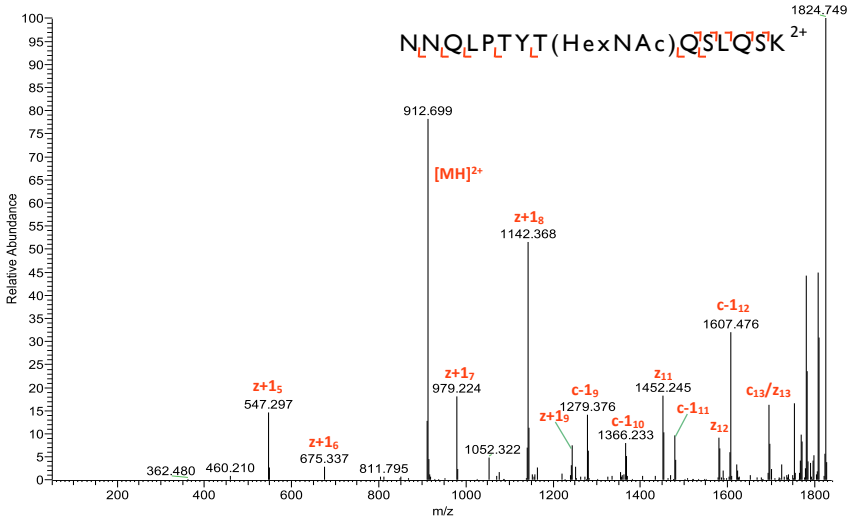
Kia1551 GlcNAc-T491

T20100618-14 #2301 RT: 24.69 AV: 1 NL: 1.23E3
T: ITMS + c NSI d sa Full ms2 686.36@etd133.33 [50.00-2000.00]



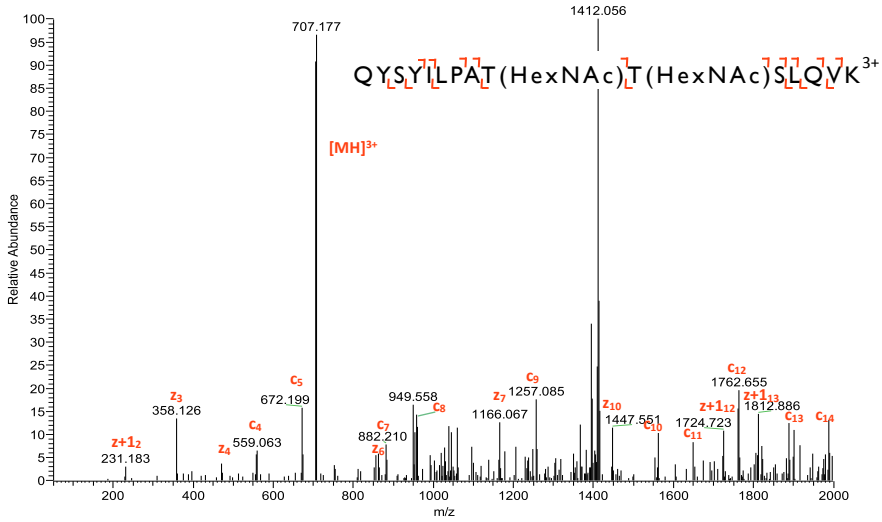
Kia1551 GlcNAc-T40

T20100409-19 #2356 RT: 24.56 AV: 1 NL: 1.72E3
T: ITMS + c NSI d sa Full ms2 912.96@etd200.00 [50.00-1840.00]



Kia1551 GlcNAc-T232 & T233

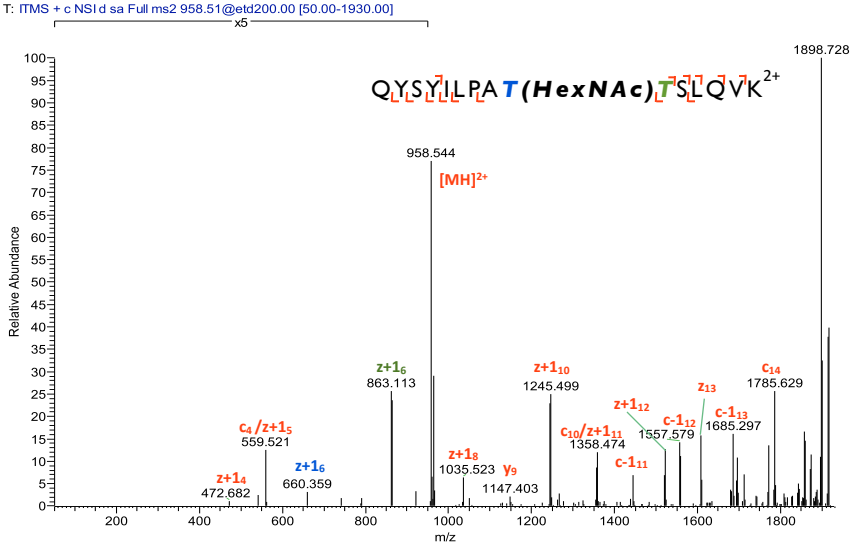
T20100618-19 #3519 RT: 34.86 AV: 1 NL: 5.97E2
T: ITMS + c NSI d sa Full ms2 707.04@etd133.33 [50.00-2000.00]



Kiaa1551 Mixture of GlcNAc-T232 and T233

T20100409-23 #3861 RT: 37.88 AV: 1 NL: 2.14E3

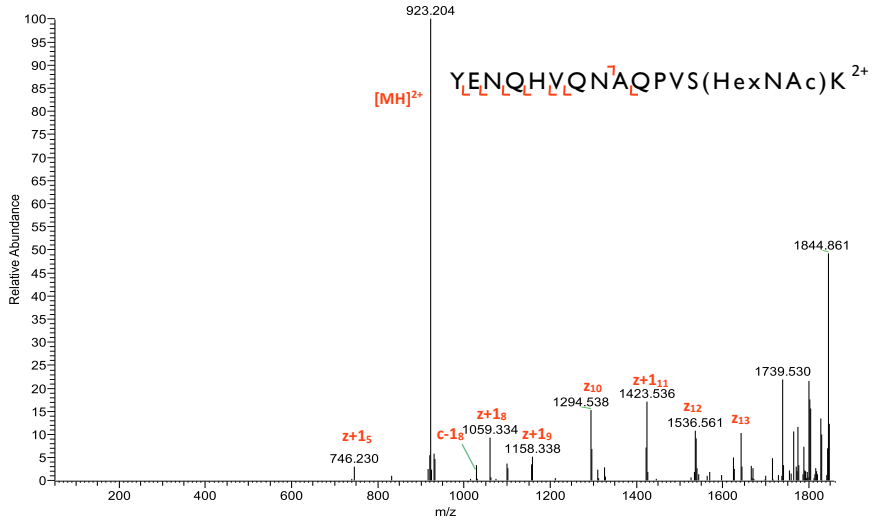
T: ITMS + c NSI d sa Full ms2 958.51@etd200.00 [50.00-1930.00]



Kia1551 GlcNAc-S291

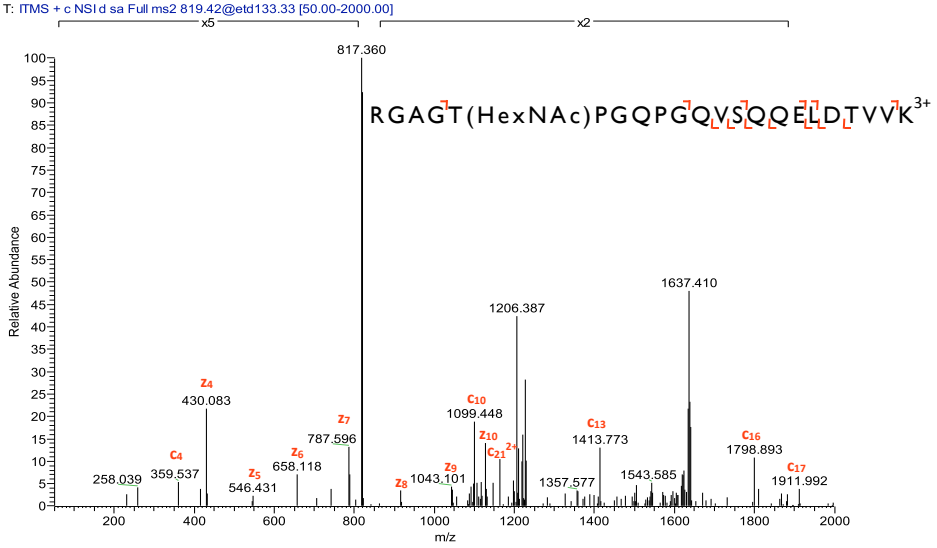
T20100409-12 #931 RT: 11.19 AV: 1 NL: 7.34E2

T: ITMS + c NSI d sa Full ms2 922.95@etd200.00 [50.00-1860.00]



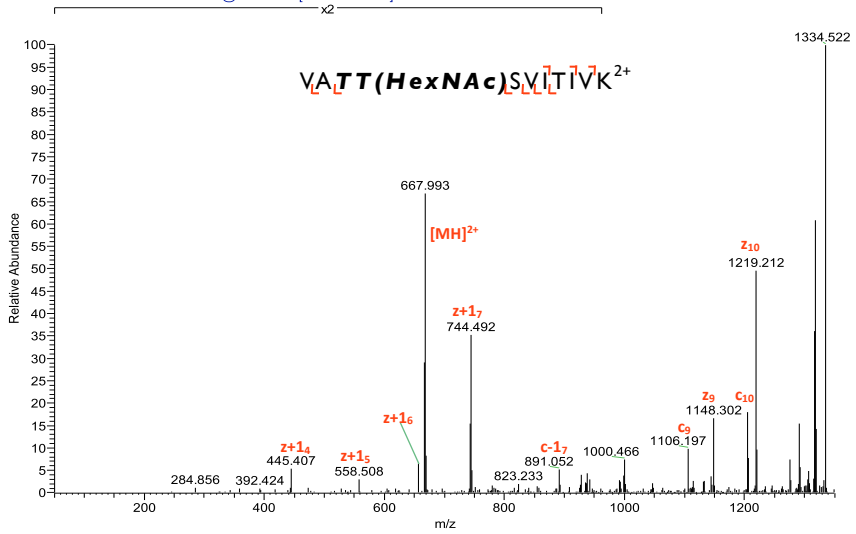
Lman1 GlcNAc-T385

T20100409-15 #2815 RT: 27.65 AV: 1 NL: 1.99E3
T: ITMS + c NSI d sa Full ms2 819.42@etd133.33 [50.00-2000.00]



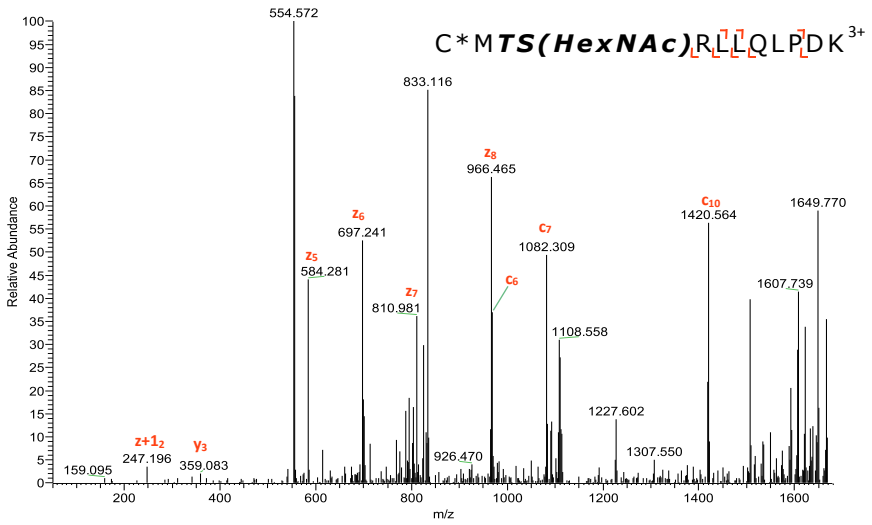
Mafk GlcNAc-T133 *or* T143

T20100618-18 #2938 RT: 30.15 AV: 1 NL: 4.09E3
T: ITMS + c NSI d sa Full ms2 667.90@etd200.00 [50.00-1350.00]



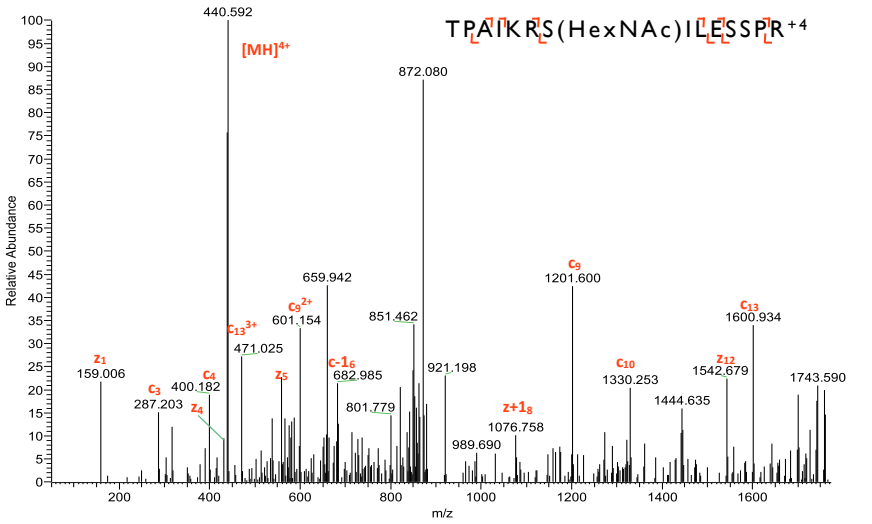
Med15 GlcNAc-T761 ***or*** S762

T20100409-25 #3828 RT: 38.74 AV: 1 NL: 1.18E3
T: ITMS + c NSI d sa Full ms2 555.61@etd133.33 [50.00-1680.00]



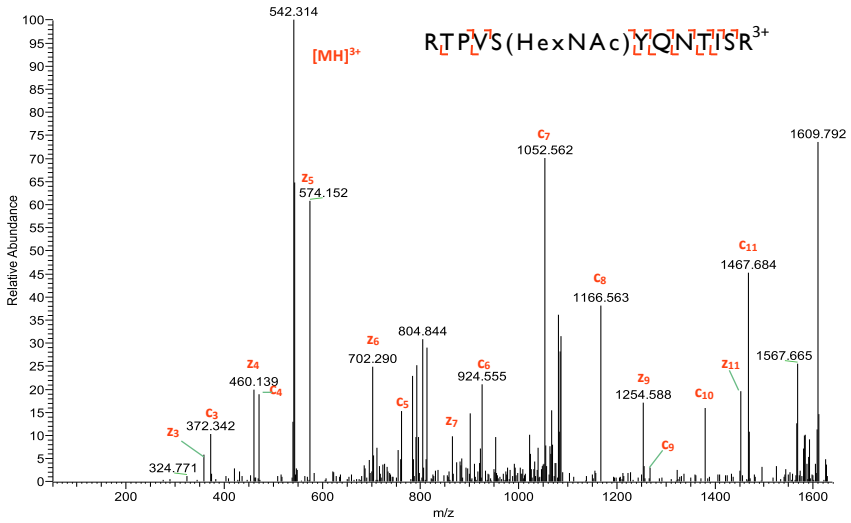
Myb GlcNAc-S454

T20101031-22 #2052 RT: 24.26 AV: 1 NL: 4.05E2
T: ITMS + c NSI d sa Full ms2 440.51@etd100.00 [50.00-1775.00]



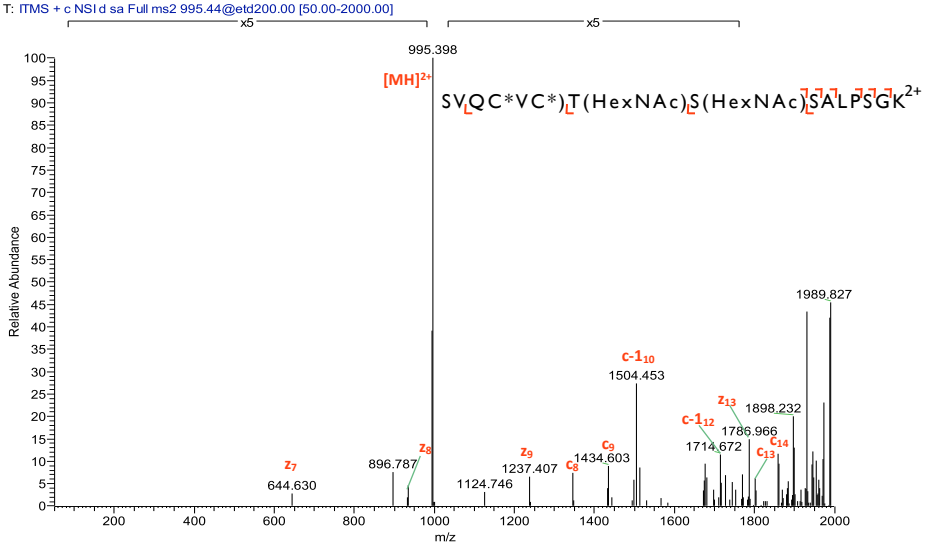
NCoR1 GlcNAc-S1496

T20100409-21 #1502 RT: 17.42 AV: 1 NL: 8.42E2
T: ITMS + c NSI d sa Full ms2 542.28@etd133.33 [50.00-1640.00]



NCoR1 GlcNAc-T1899 & S1900

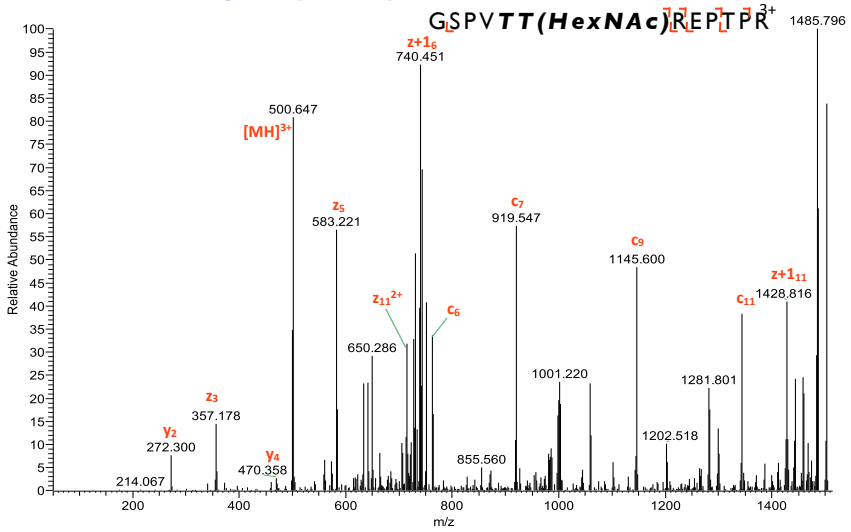
T20100409-11 #2188 RT: 22.10 AV: 1 NL: 1.18E3
T: ITMS + c NSI d sa Full ms2 995.44@etd200.00 [50.00-2000.00]



NCoR2 GlcNAc-T1531 ***or*** T1532

T20100409-13 #1160 RT: 13.35 AV: 1 NL: 3.85E3

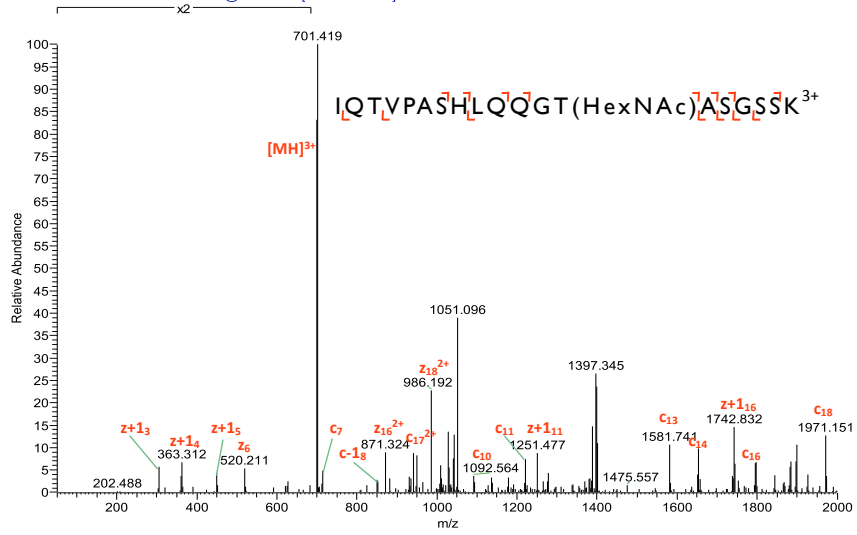
T: ITMS + c NSI d sa Full ms2 500.93@etd133.33 [50.00-1515.00]



Nfrkb GlcNAc-T1270

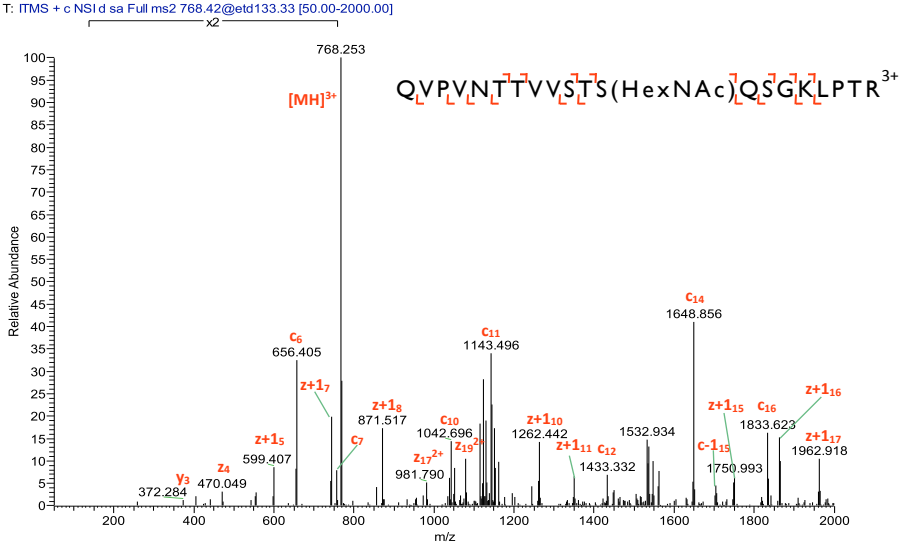
T20100409-19 #1525 RT: 17.44 AV: 1 NL: 1.63E3

T: ITMS + c NSI d sa Full ms2 700.69@etd133.33 [50.00-2000.00]



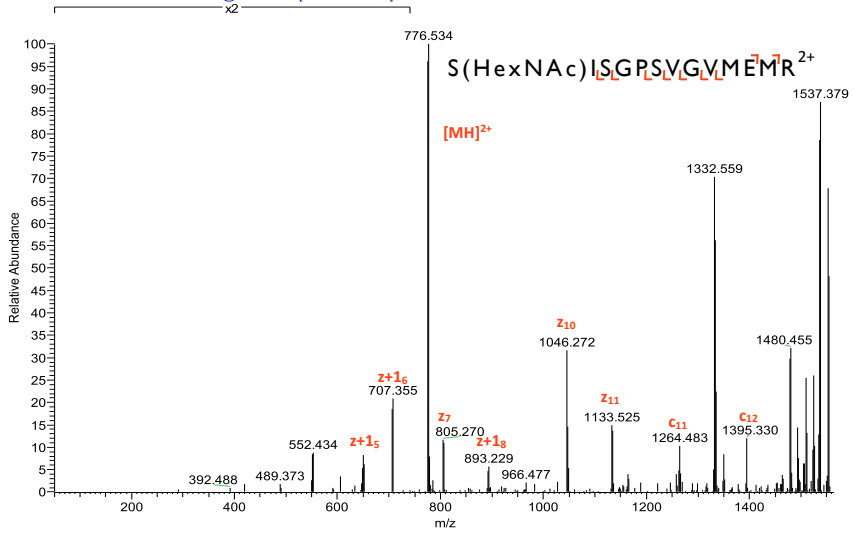
Nfrkb GlcNAc-S1172

T20100409-23 #2340 RT: 24.89 AV: 1 NL: 1.55E3
T: ITMS + c NSI d sa Full ms2 768.42@etd133.33 [50.00-2000.00]

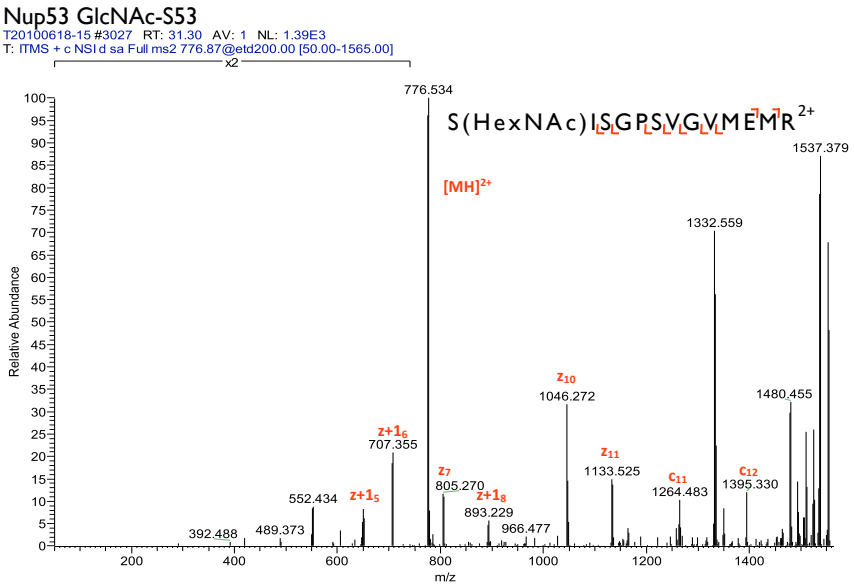


Nup53 GlcNAc-S53

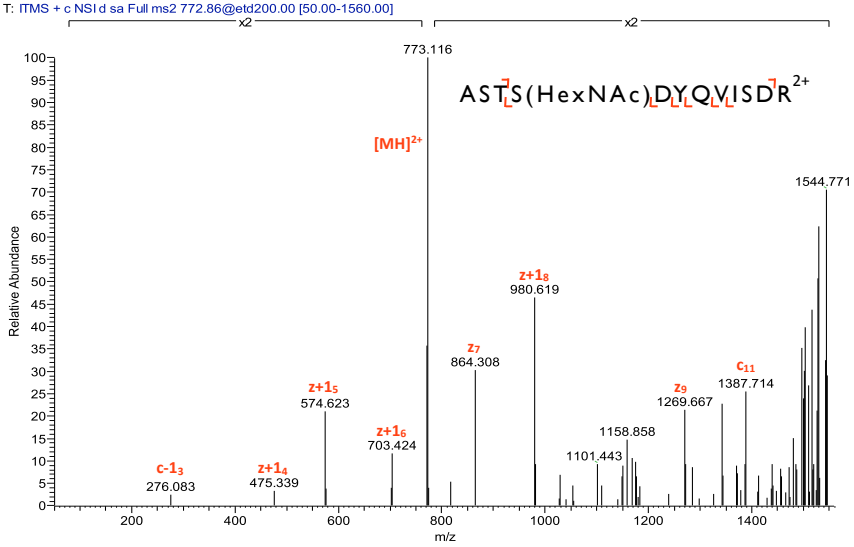
T20100618-15 #3027 RT: 31.30 AV: 1 NL: 1.39E3
T: ITMS + c NSI d sa Full ms2 776.87@etd200.00 [50.00-1565.00]



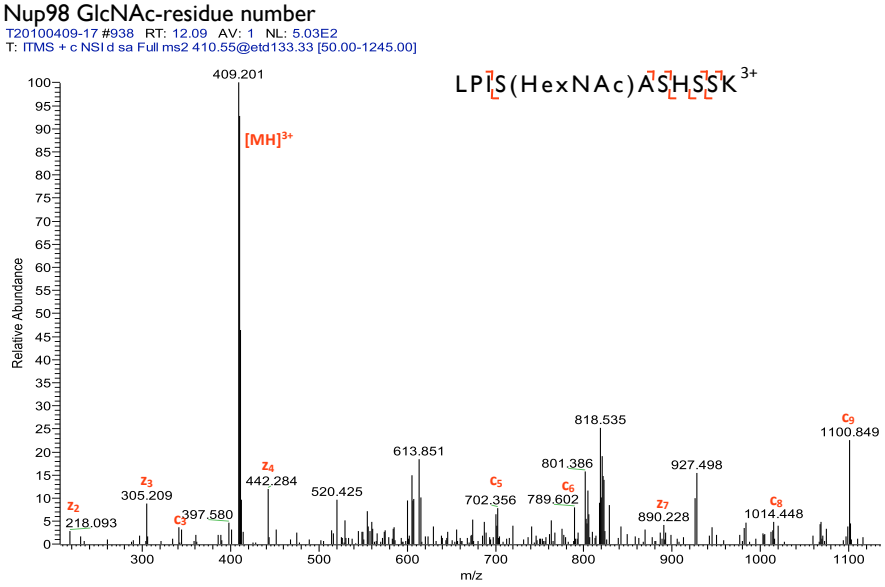
Nup53 GlcNAc-S53



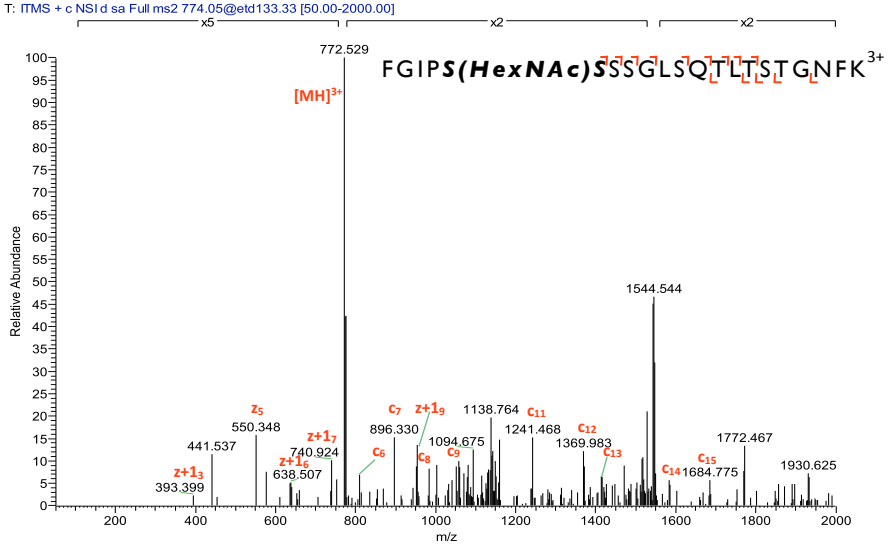
Nup53 GlcNAc-S297



Nup98 GlcNAc-residue number

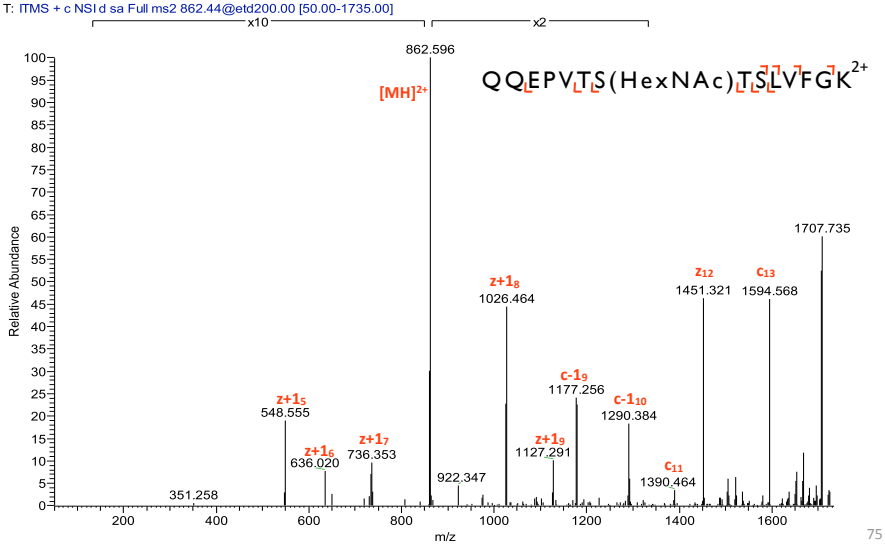


Nup153 GlcNAc-S898 ***or*** S899



Nup153 GlcNAc-S1102

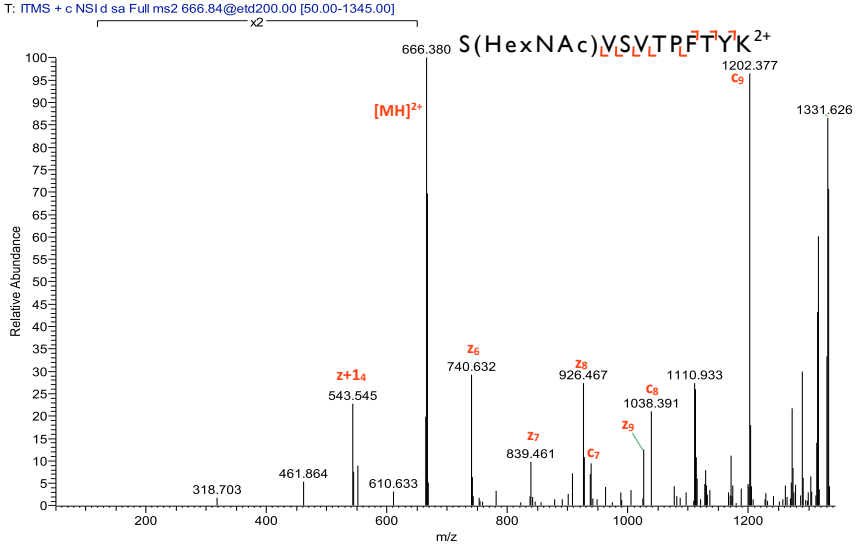
T20100409-17 #3078 RT: 30.73 AV: 1 NL: 3.38E3
T: ITMS + c NSI d sa Full ms2 862.44@etd200.00 [50.00-1735.00]



75

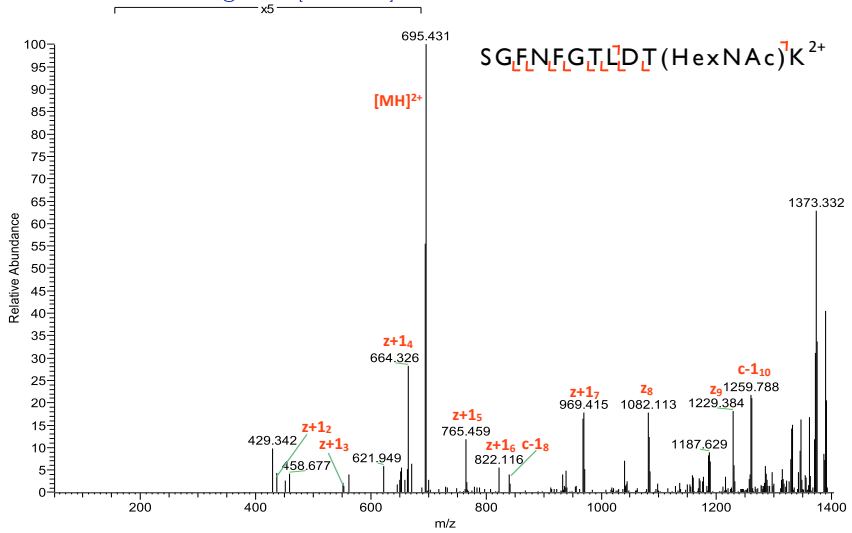
Nup153 GlcNAc-S1046

T9070413 #1859 RT: 27.19 AV: 1 NL: 2.83E2
T: ITMS + c NSI d sa Full ms2 666.84@etd200.00 [50.00-1345.00]



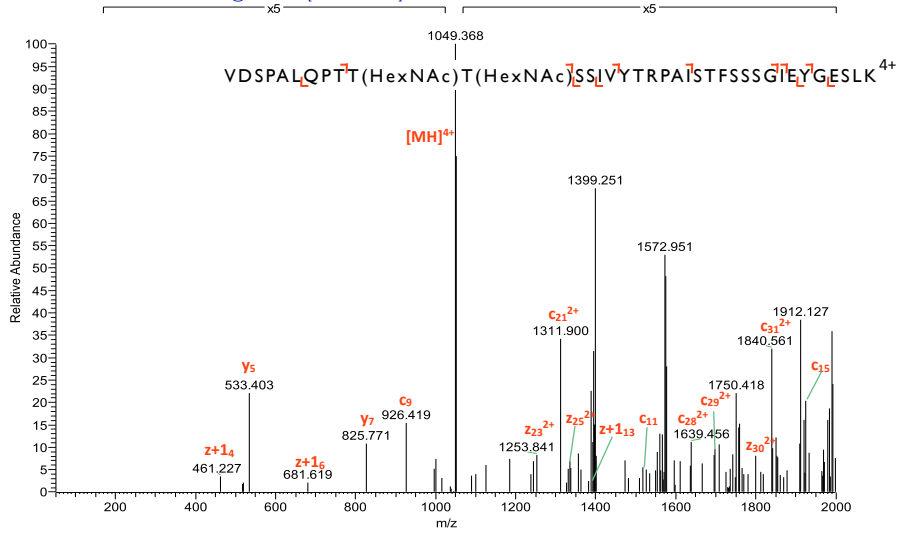
Nup153 GlcNAc-T1044

T20100409-18 #3389 RT: 33.45 AV: 1 NL: 1.23E3
T: ITMS + c NSI d sa Full ms2 695.33@etd200.00 [50.00-1405.00]



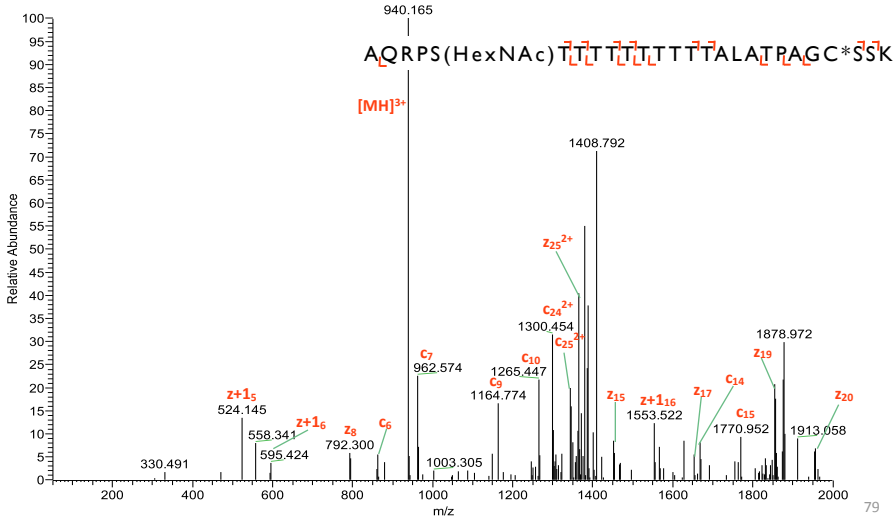
Nup153 GlcNAc-T627 & T628

T20100409-21 #4839 RT: 46.55 AV: 1 NL: 9.68E2
T: ITMS + c NSI d sa Full ms2 1050.28@etd100.00 [50.00-2000.00]



Nup188 GlcNAc-S1522

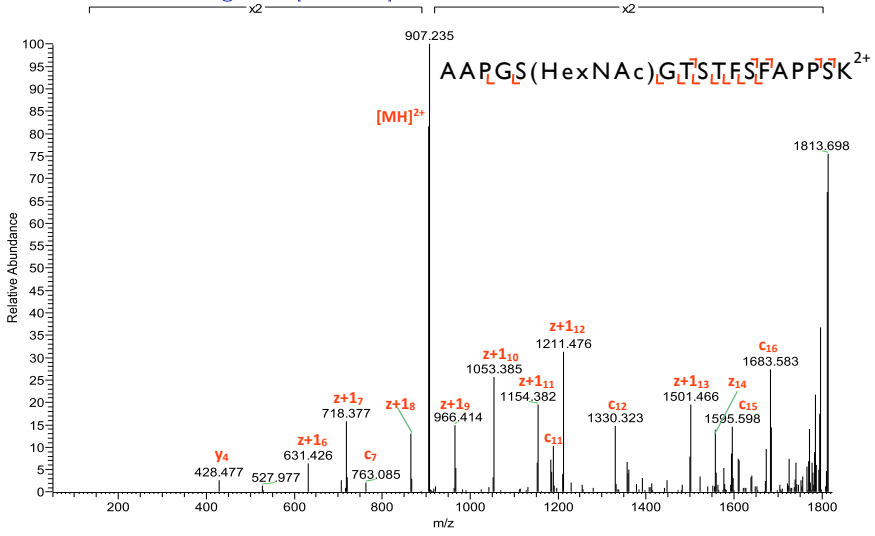
T20100409-17 #1974 RT: 21.16 AV: 1 NL: 3.95E2
T: ITMS + c NSI d sa Full ms2 939.79@etd133.33 [50.00-2000.00]



79

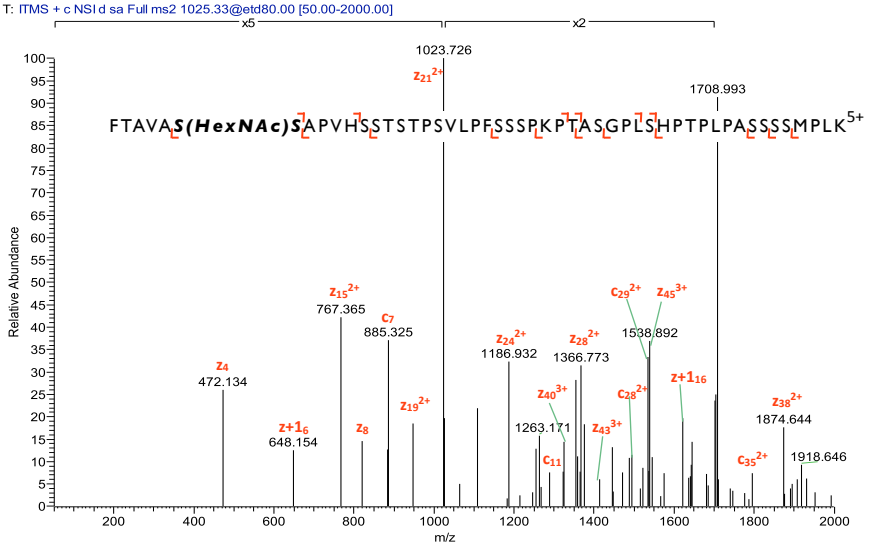
Nup214 GlcNAc-S513

T20100409-21 #2939 RT: 30.00 AV: 1 NL: 1.77E3
T: ITMS + c NSI d sa Full ms2 906.94@etd200.00 [50.00-1825.00]



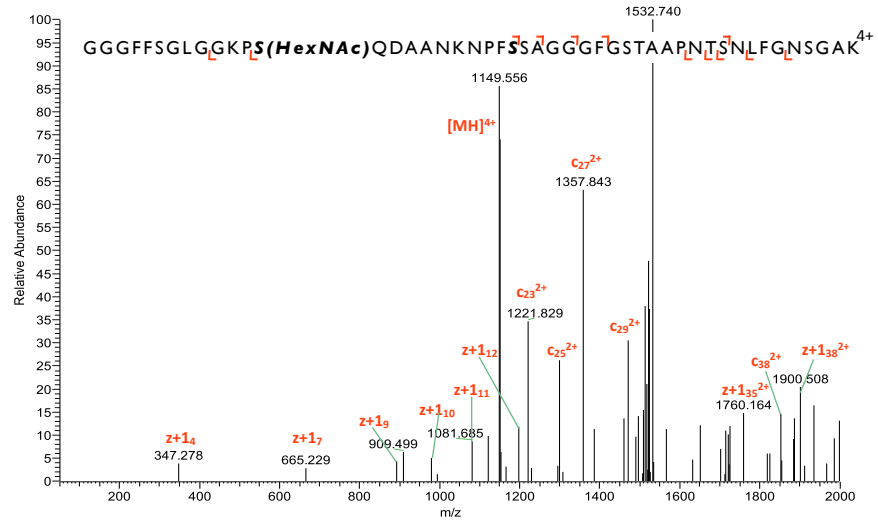
Nup214 GlcNAc-S589 ***or*** S590

T20100409-28 #3542 RT: 39.88 AV: 1 NL: 2.87E2
T: ITMS + c NSI d sa Full ms2 1025.33@etd80.00 [50.00-2000.00]



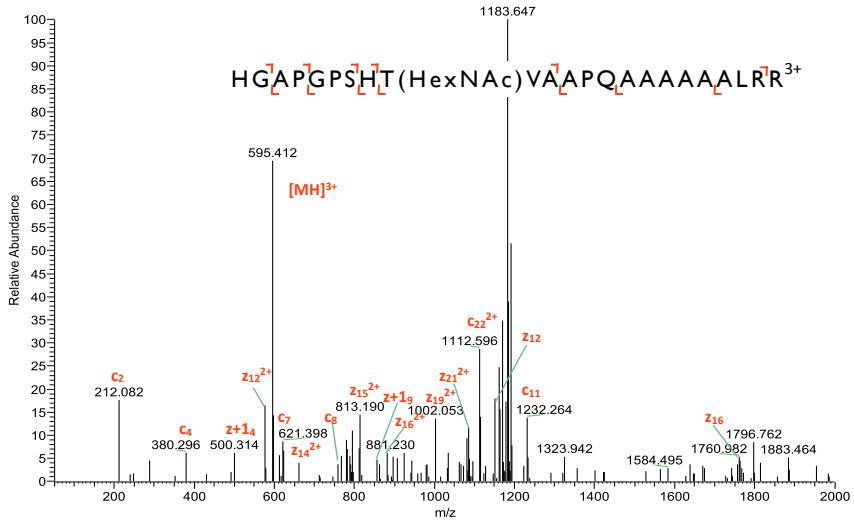
Nup214 GlcNAc-S1889 ***or*** S1899

T20100409-29 #4275 RT: 45.33 AV: 1 NL: 1.34E2
T: ITMS + c NSI d sa Full ms2 1150.05@etd100.00 [50.00-2000.00]



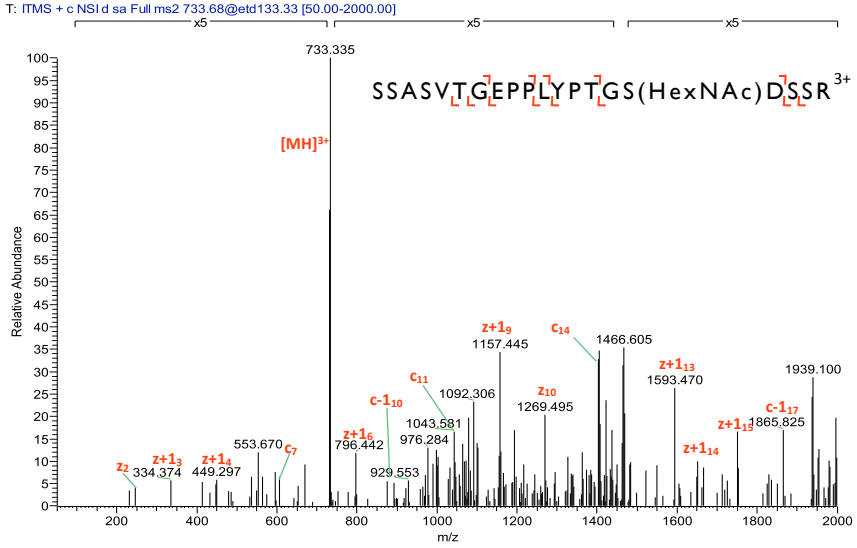
Nup214 GlcNAc-T1091

T20101031-22 #1961 RT: 23.41 AV: 1 NL: 3.86E2
T: ITMS + c NSI d sa Full ms2 596.32@etd100.00 [50.00-2000.00]



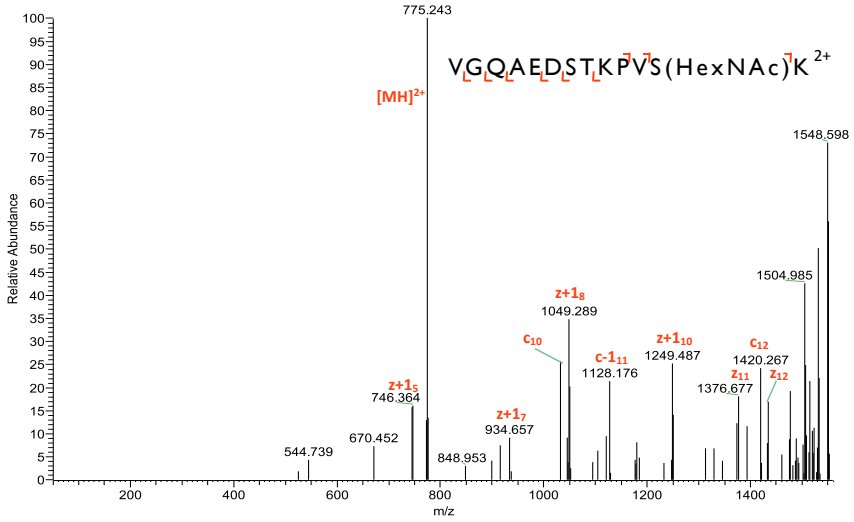
Nup214 GlcNAc-S504

T20100409-12 #2649 RT: 26.01 AV: 1 NL: 1.40E3
T: ITMS + c NSI d sa Full ms2 733.68@etd133.33 [50.00-2000.00]



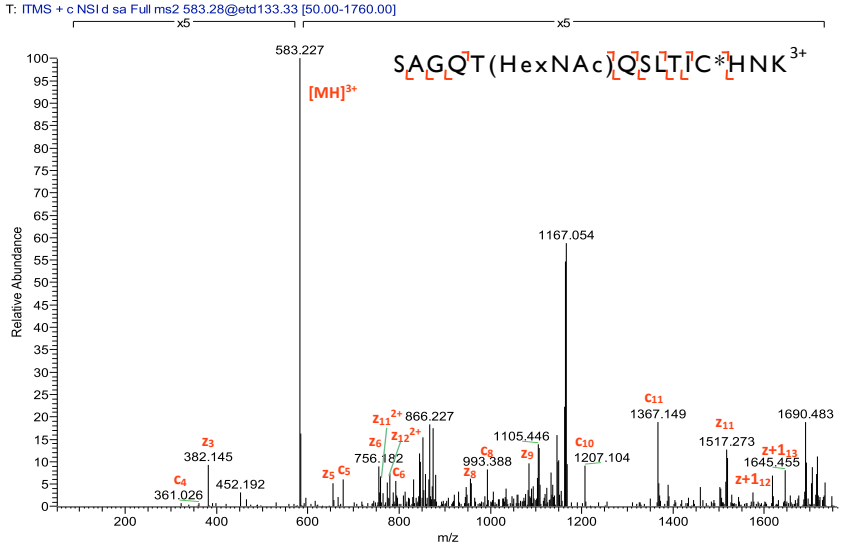
Nup214 GlcNAc-S1362

T20100409-09 #612 RT: 8.00 AV: 1 NL: 1.64E2
T: ITMS + c NSI d sa Full ms2 774.89@etd200.00 [50.00-1560.00]



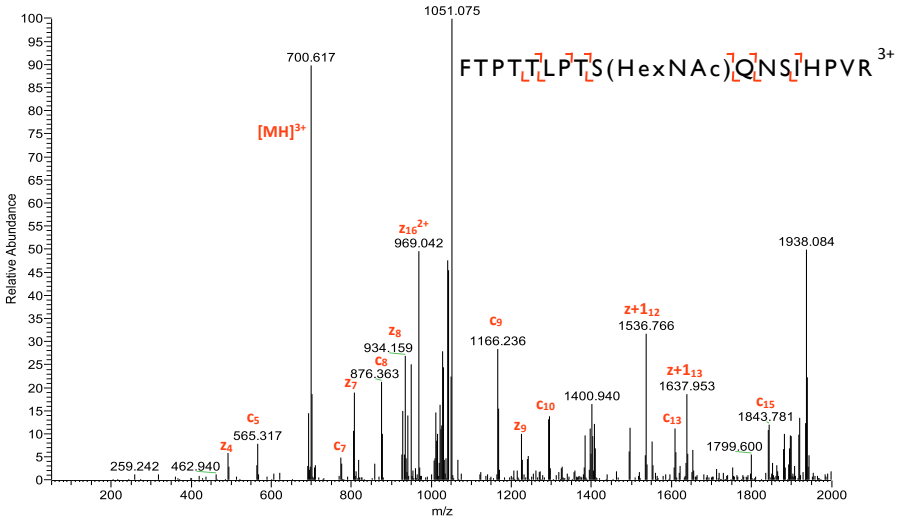
Phc3 GlcNAc-T238

T20101031-18 #1516 RT: 17.81 AV: 1 NL: 7.03E3
T: ITMS + c NSI d sa Full ms2 583.28@etd133.33 [50.00-1760.00]



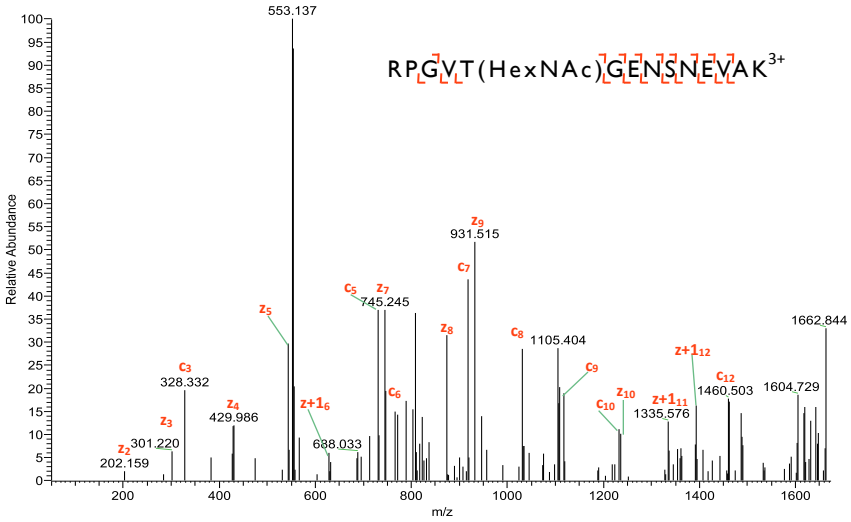
Phf21a GlcNAc-S286

T20100618-19 #2511 RT: 26.82 AV: 1 NL: 4.06E3
T: ITMS + c NSI d sa Full ms2 700.70@etd133.33 [50.00-2000.00]



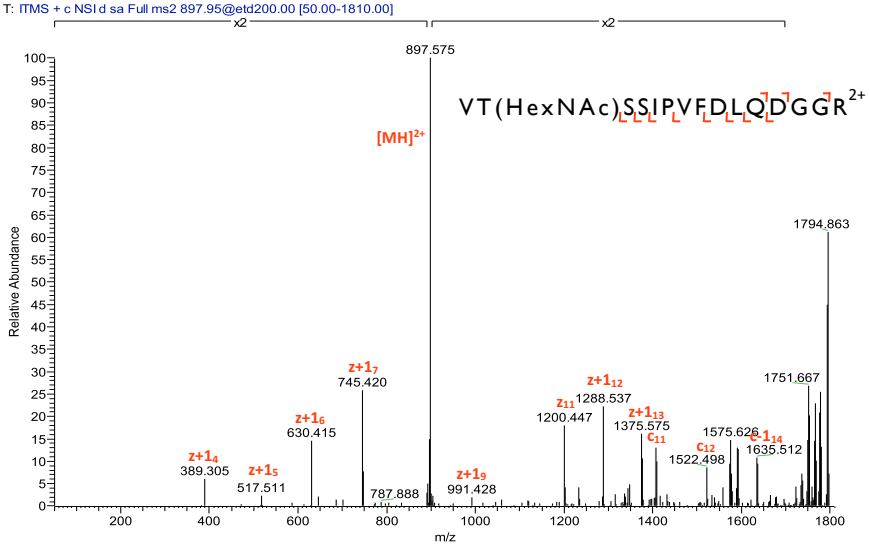
Pogz GlcNAc-T310

T20100618-06 #863 RT: 11.13 AV: 1 NL: 2.17E2
T: ITMS + c NSI d sa Full ms2 554.28@etd133.33 [50.00-1675.00]



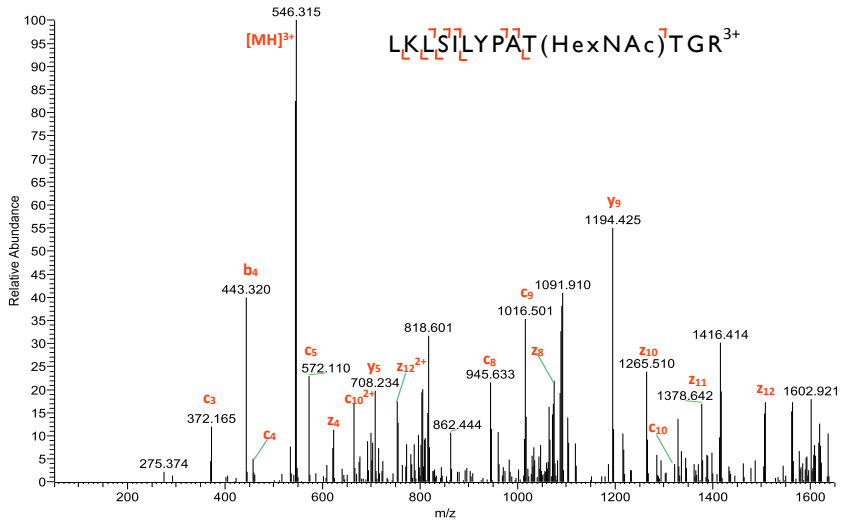
Pogz GlcNAc-T358

T20100409-18 #3938 RT: 38.27 AV: 1 NL: 2.82E3
T: ITMS + c NSI d sa Full ms2 897.95@etd200.00 [50.00-1810.00]

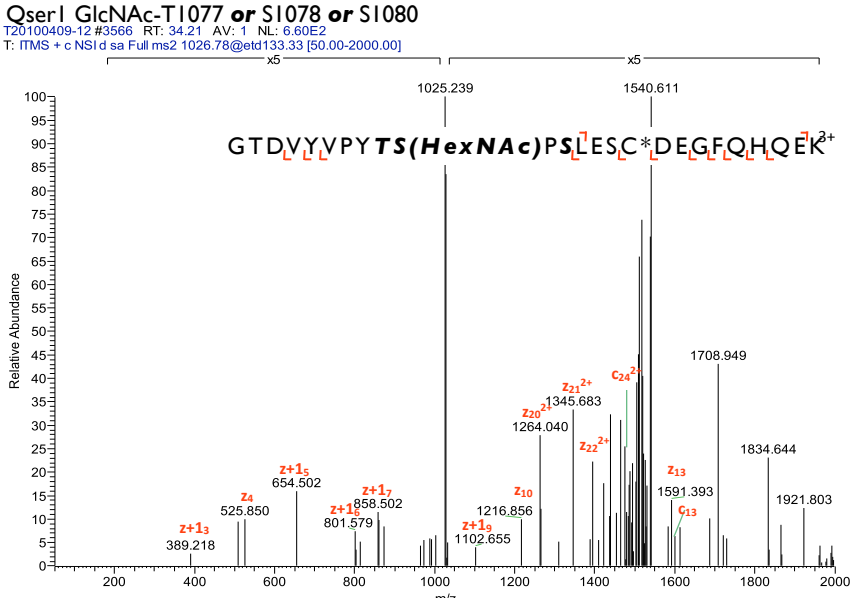


Prdx6 GlcNAc-T152

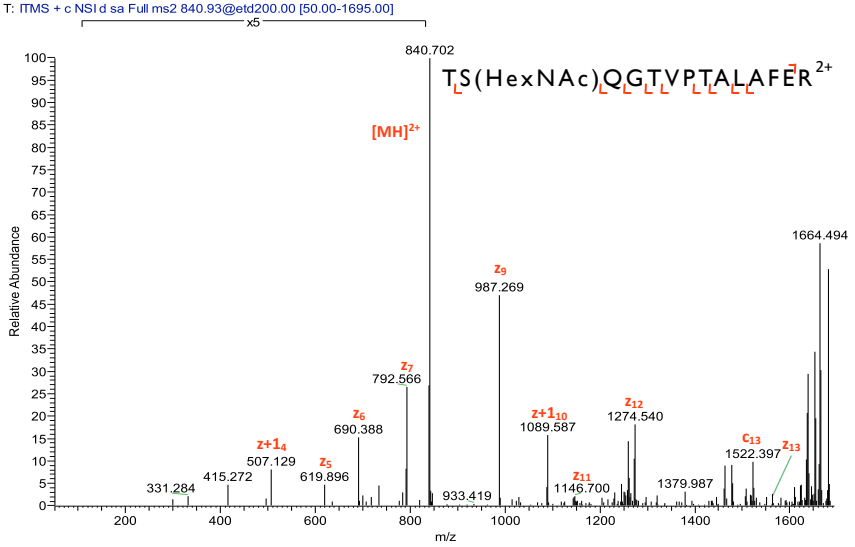
T20100409-18 #3023 RT: 30.22 AV: 1 NL: 1.48E3
T: ITMS + c NSI d sa Full ms2 545.98@etd133.33 [50.00-1650.00]



Qser1 GlcNAc-T1077 ***or*** S1078 ***or*** S1080



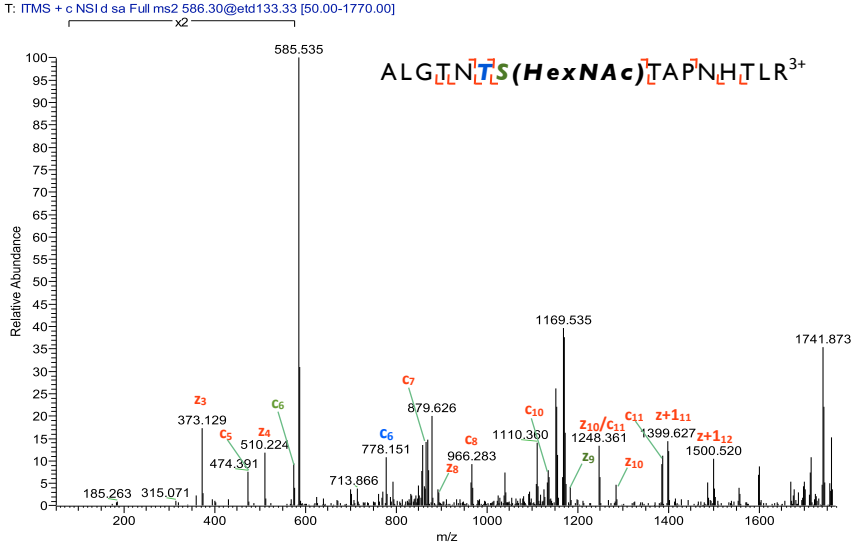
Qser1 GlcNAc-S105



RanBP2 Mixture of GlcNAc-T1306 and S1307

T20100409-18 #1429 RT: 16.50 AV: 1 NL: 1.84E3

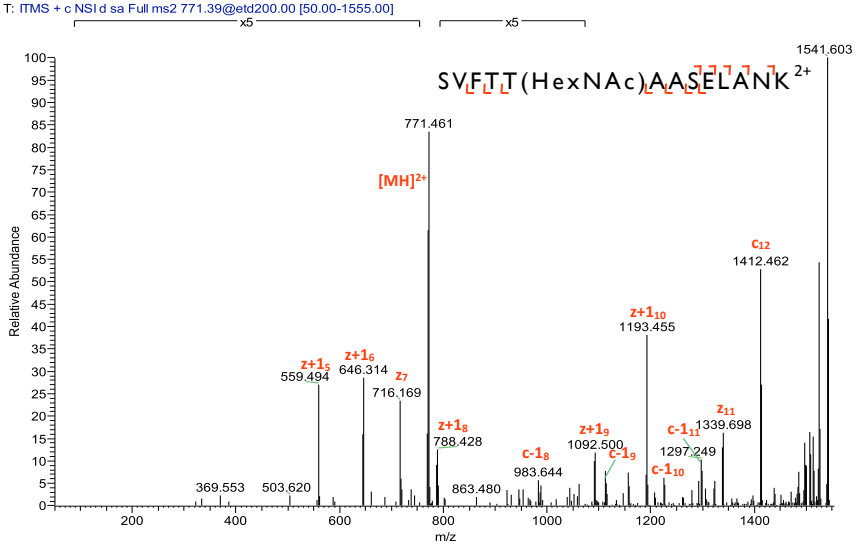
T: ITMS + c NSI d sa Full ms2 586.30@etd133.33 [50.00-1770.00]



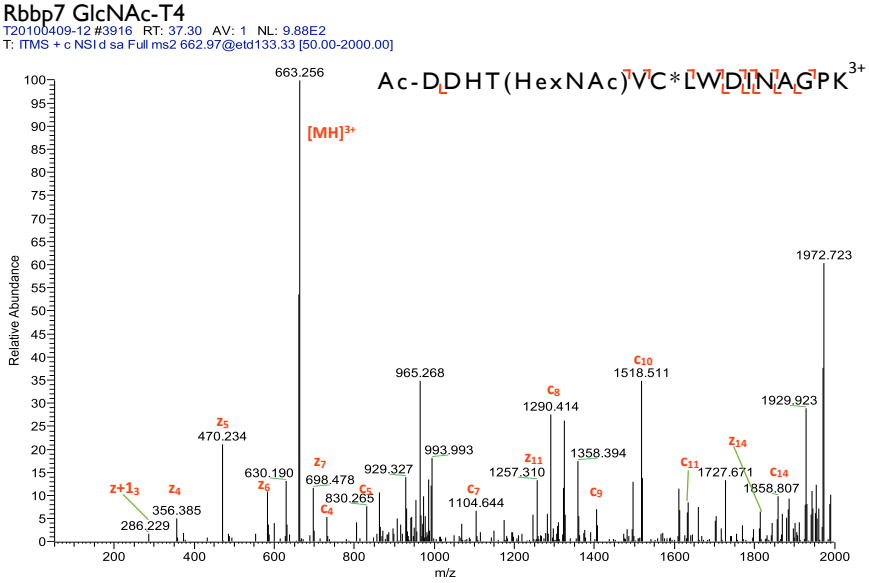
RanBP2 GlcNAc-T1138

T20100409-16 #2994 RT: 29.50 AV: 1 NL: 2.33E3

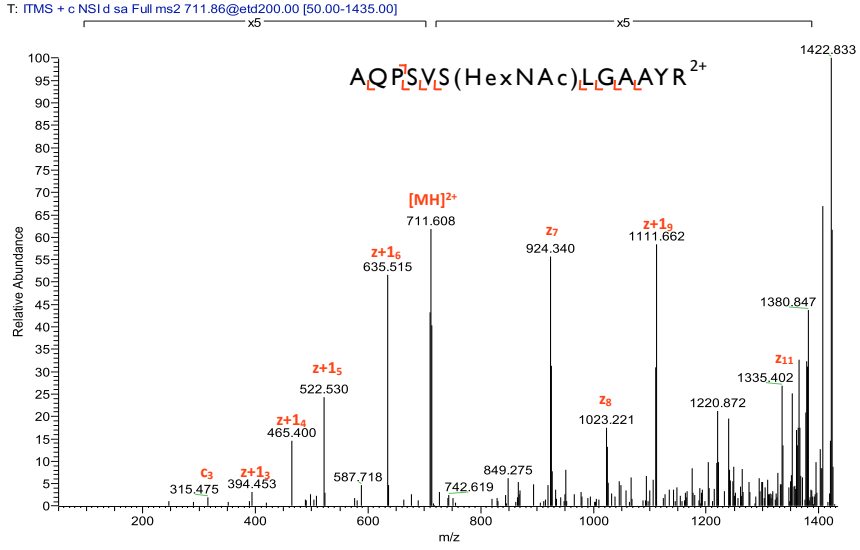
T: ITMS + c NSI d sa Full ms2 771.39@etd200.00 [50.00-1555.00]



Rbbp7 GlcNAc-T4

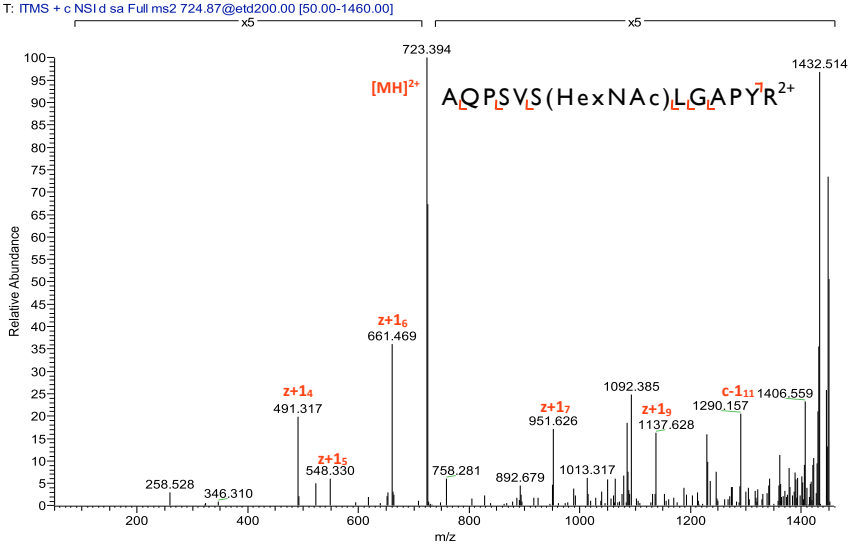


Rbm14 GlcNAc-S244



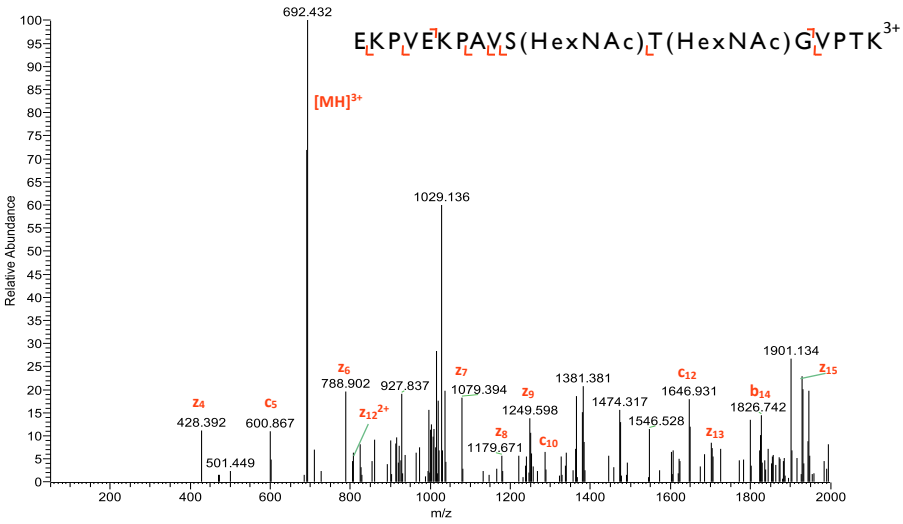
Rbm14 GlcNAc-S280

T20100409-20 #2462 RT: 25.55 AV: 1 NL: 2.77E3
T: ITMS + c NSI d sa Full ms2 724.87@etd200.00 [50.00-1460.00]



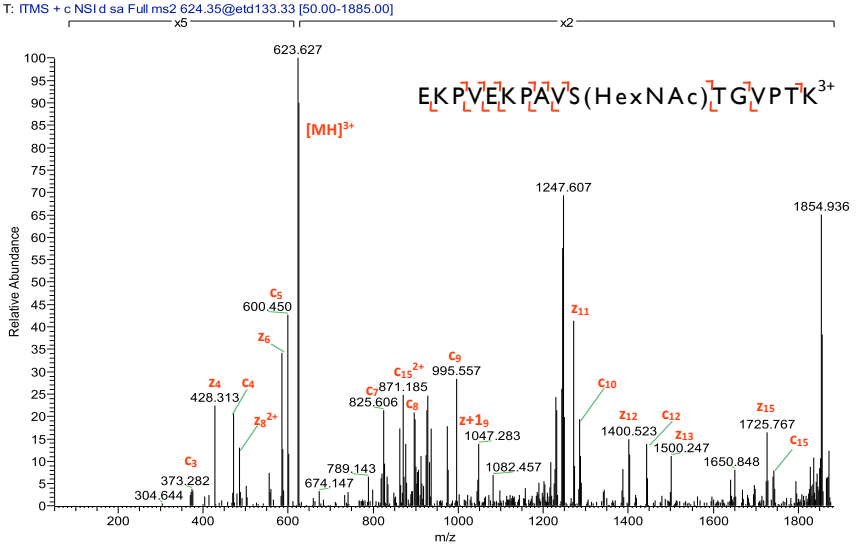
Rprd2 GlcNAc-S103 & T104

T9081311 #1258 RT: 18.66 AV: 1 NL: 1.94E2
T: ITMS + c NSI d sa Full ms2 692.04@etd200.00 [50.00-2000.00]



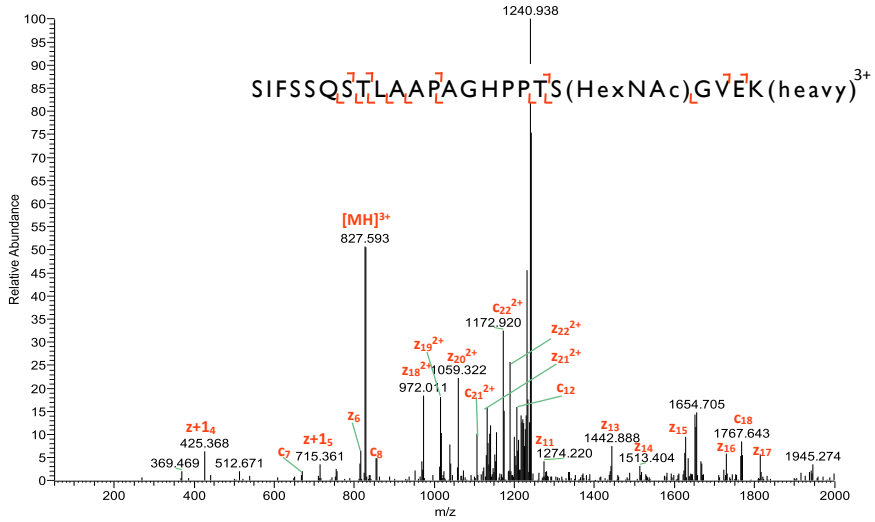
Rprd2 GlcNAc-S103

T20100409-15 #1471 RT: 16.39 AV: 1 NL: 2.97E3
T: ITMS + c NSI d sa Full ms2 624.35@etd133.33 [50.00-1885.00]



Rprd2 GlcNAc-S1110

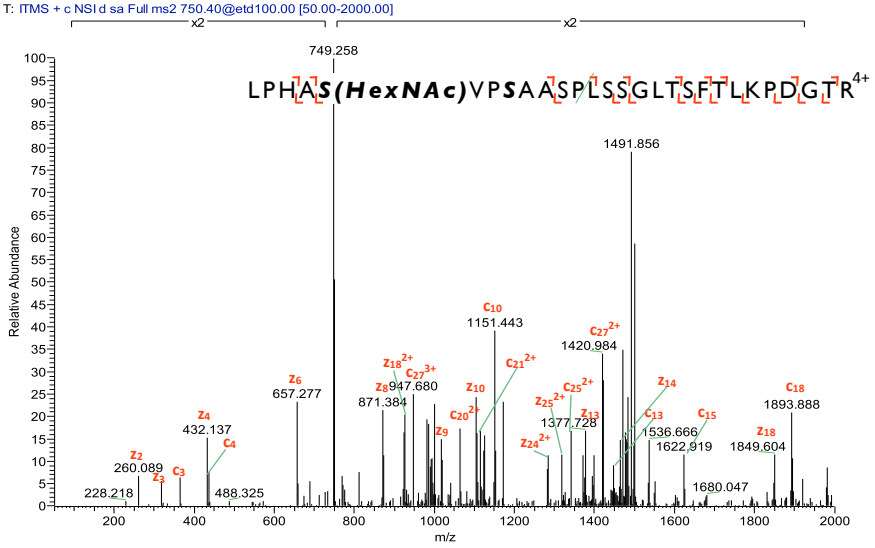
T20100618-17 #2803 RT: 28.98 AV: 1 NL: 1.76E3
T: ITMS + c NSI d sa Full ms2 827.75@etd133.33 [50.00-2000.00]



Sal4 GlcNAc-S299 ***or*** S302

T20100616-06 #3975 RT: 39.47 AV: 1 NL: 3.60E3

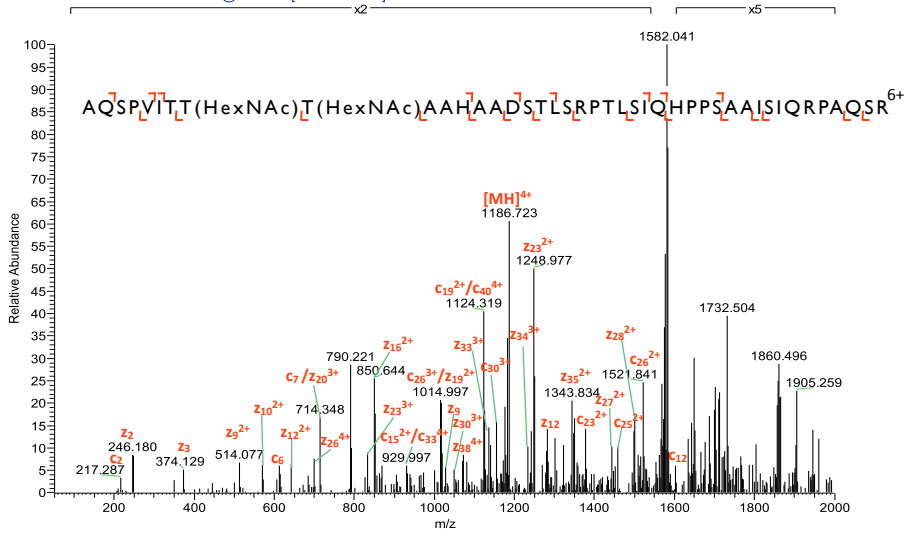
T: ITMS + c NSI d sa Full ms2 750.40@etd100.00 [50.00-2000.00]



Sap130 GlcNAc-T284 & T285

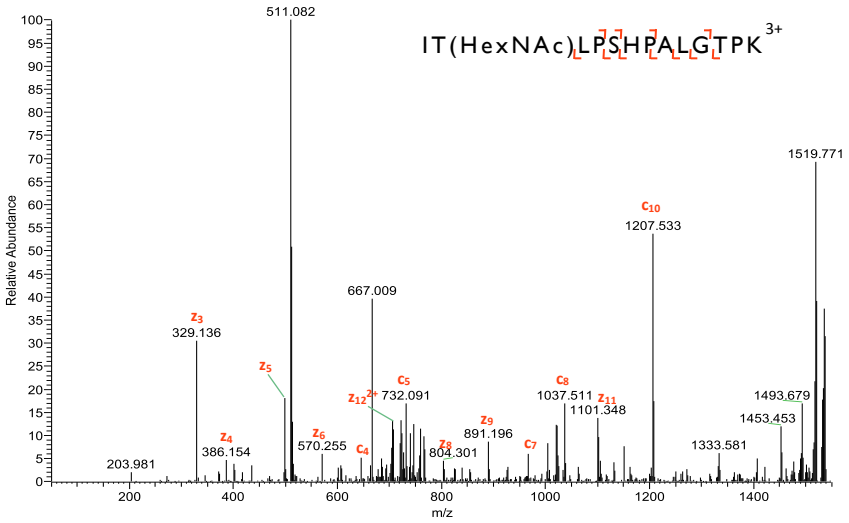
T20100618-24 #2936 RT: 32.70 AV: 1 NL: 4.10E3

T: ITMS + c NSI d sa Full ms2 791.42@etd66.67 [50.00-2000.00]



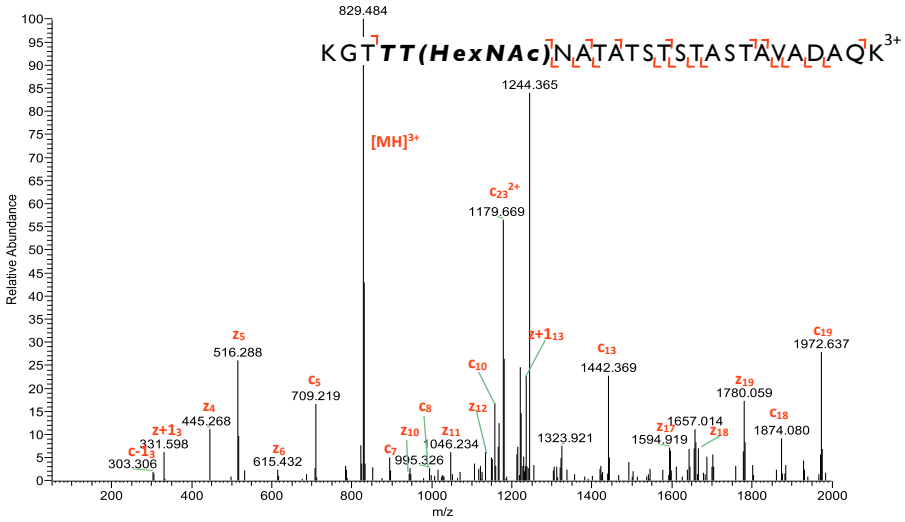
Sap130 GlcNAc-T320

T20100618-20 #2556 RT: 27.12 AV: 1 NL: 4.61E3
T: ITMS + c NSI d sa Full ms2 512.29@etd133.33 [50.00-1550.00]

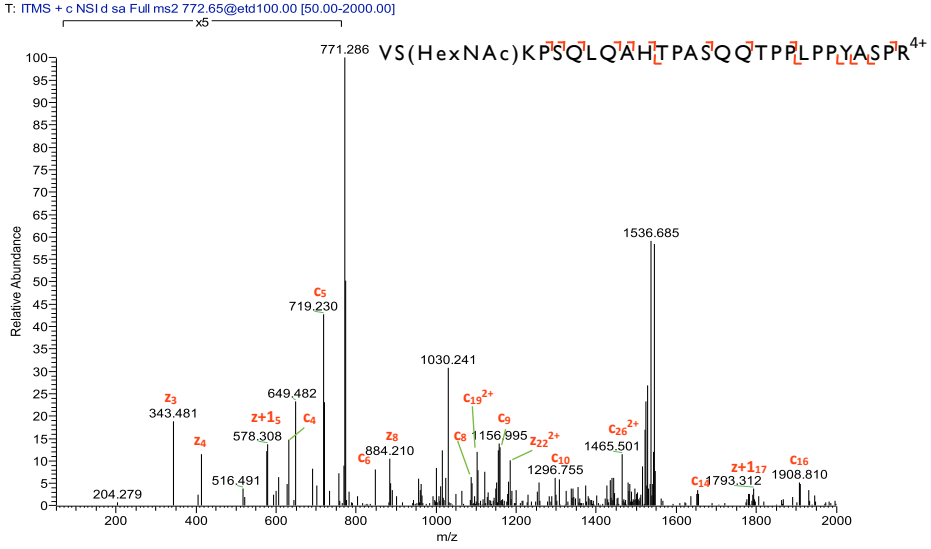


Sap30bp GlcNAc-T232 *or* T233

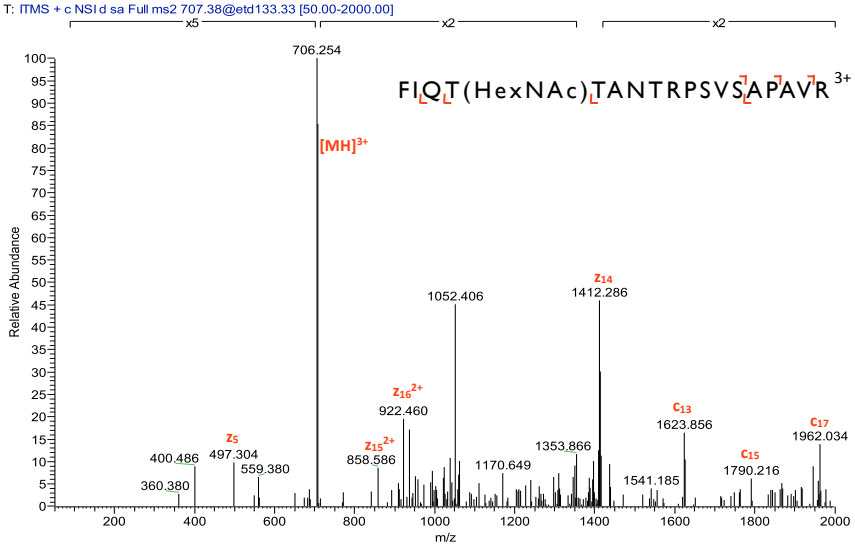
T20100618-12 #1481 RT: 17.65 AV: 1 NL: 6.65E2
T: ITMS + c NSI d sa Full ms2 830.08@etd133.33 [50.00-2000.00]



Sin3a GlcNAc-S251

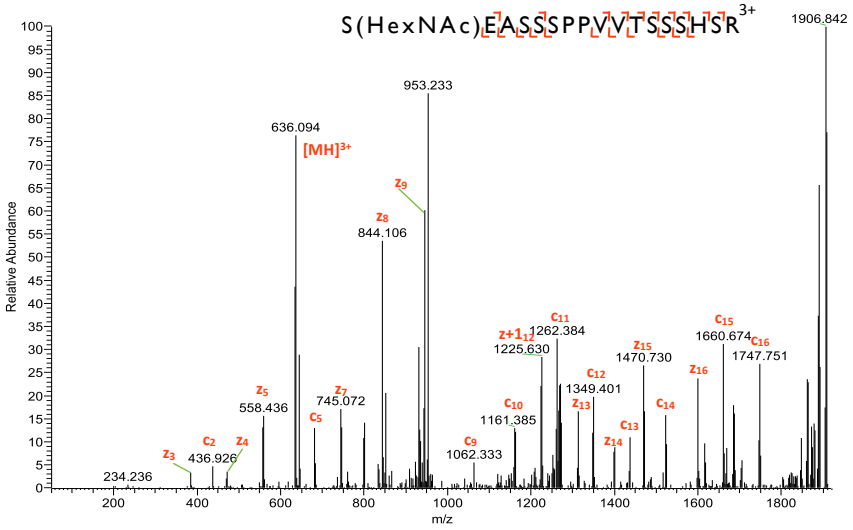


Snbo1 GlcNAc-T124



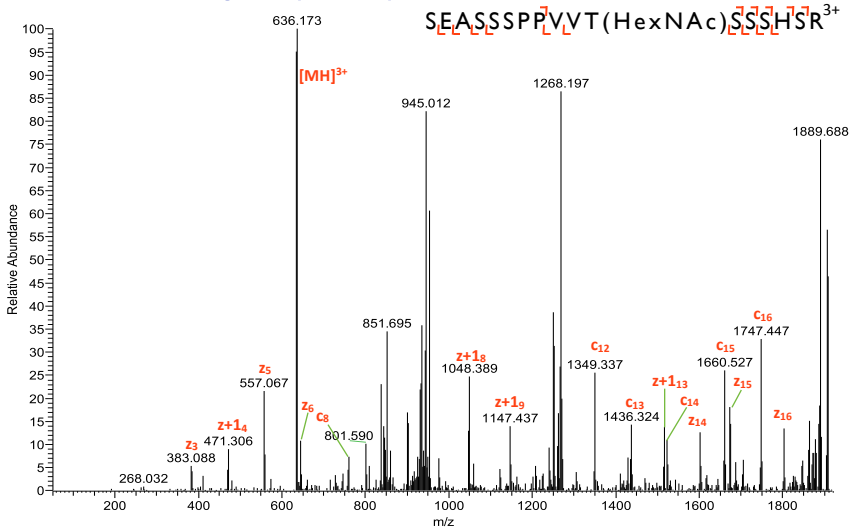
Sox2 GlcNAc-S248

T20100409-11 #1061 RT: 12.52 AV: 1 NL: 1.80E3
T: ITMS + c NSI d sa Full ms2 635.97@etd133.33 [50.00-1920.00]



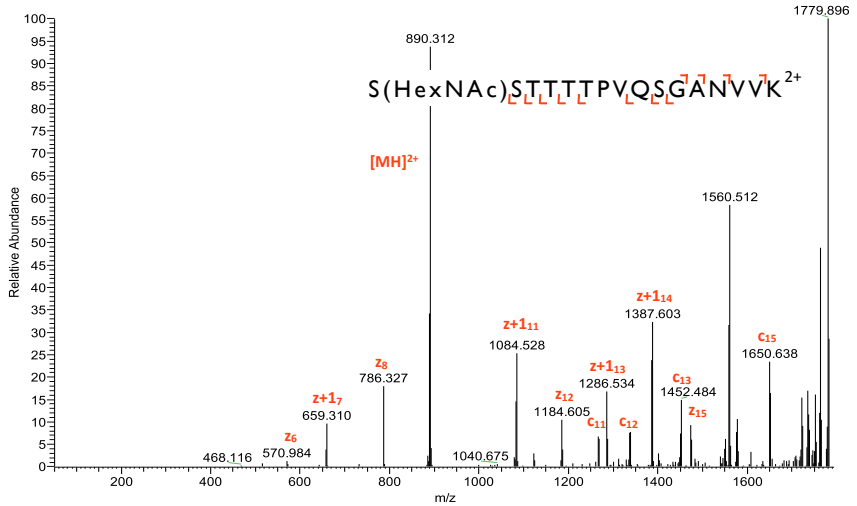
Sox2 GlcNAc-T258

T9100410 #1032 RT: 13.80 AV: 1 NL: 1.99E3
T: ITMS + c NSI d sa Full ms2 635.63@etd133.33 [50.00-1920.00]



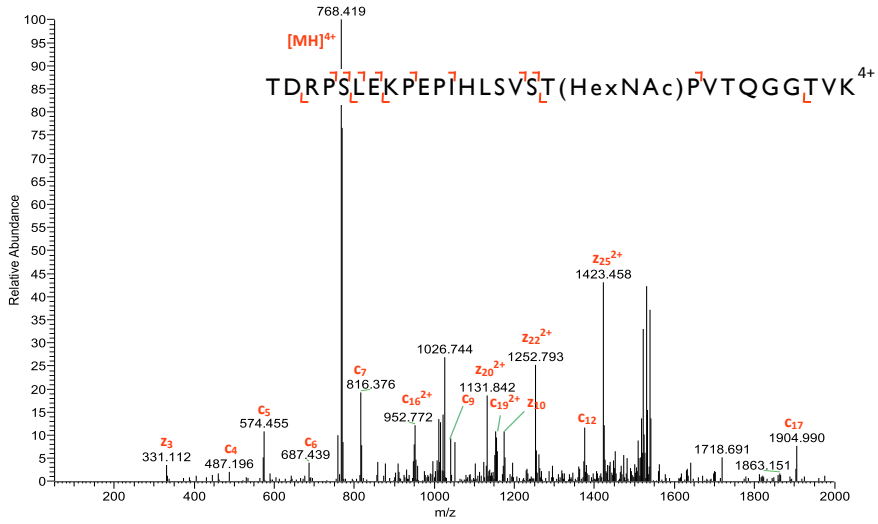
Sp2 GlcNAc-S186

T20100409-15 #1800 RT: 19.19 AV: 1 NL: 1.62E3
T: ITMS + c NSI d sa Full ms2 890.96@etd200.00 [50.00-1795.00]



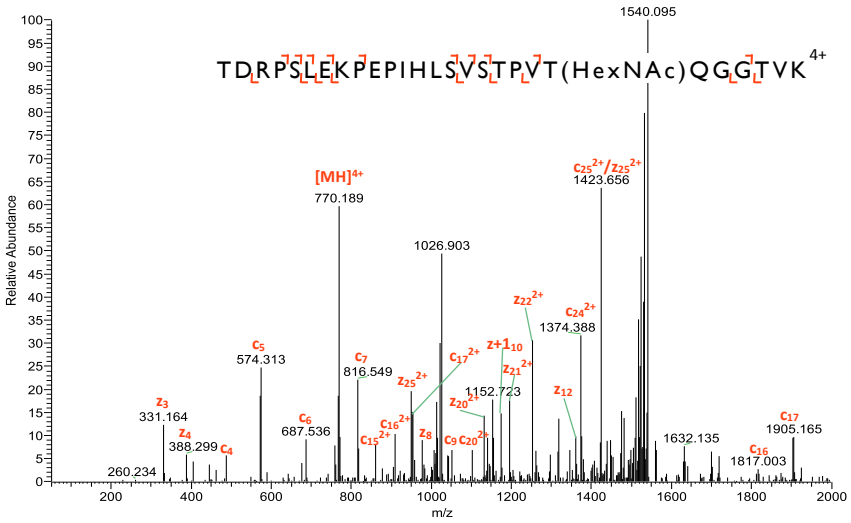
Spen GlcNAc-T2896

T20100409-19 #3046 RT: 30.59 AV: 1 NL: 2.19E3
T: ITMS + c NSI d sa Full ms2 770.17@etd100.00 [50.00-2000.00]



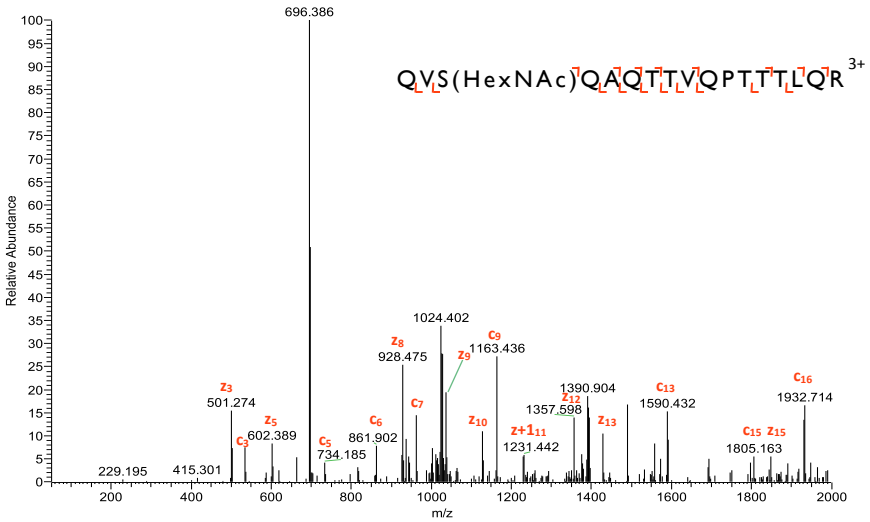
Spen GlcNAc-T2899

T20100618-17 #2924 RT: 29.94 AV: 1 NL: 2.15E3
T: ITMS + c NSI d sa Full ms2 770.16@etd100.00 [50.00-2000.00]



Taf4a GlcNAc-S123

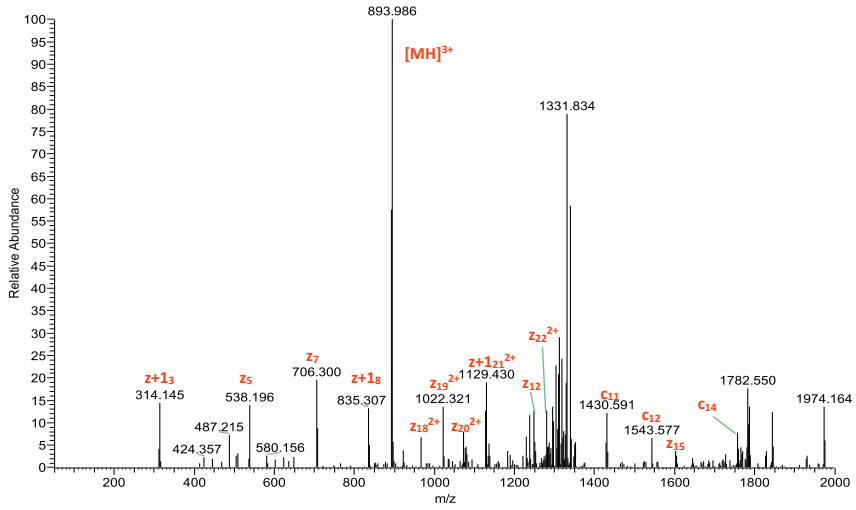
T20100618-14 #1860 RT: 20.81 AV: 1 NL: 1.09E3
T: ITMS + c NSI d sa Full ms2 697.70@etd133.33 [50.00-2000.00]



Taf6 GlcNAc-T49

T20101031-20 #3211 RT: 31.52 AV: 1 NL: 6.80E3
T: ITMS + c NSI d sa Full ms2 893.49@etd133.33 [50.00-2000.

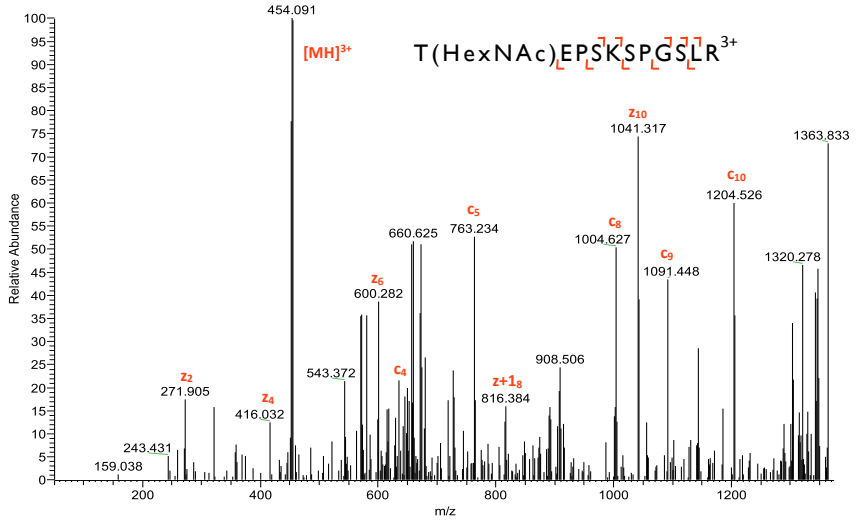
TT(HexNAc)LTITQPRPTLTLSQAPQPGPR^{+3}



Tcf20 GlcNAc-T725

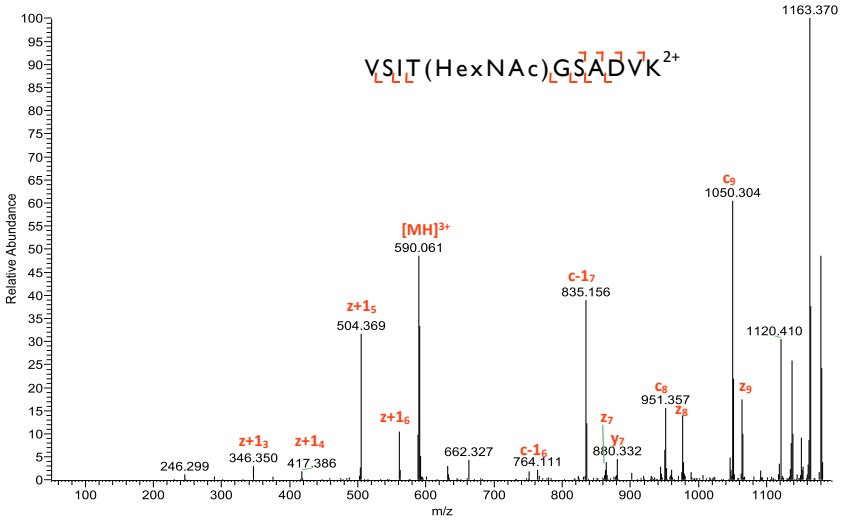
T20100409-12 #822 RT: 10.17 AV: 1 NL: 3.11E2
T: ITMS + c NSI d sa Full ms2 454.57@etd133.33 [50.00-1375.00]

T(HexNAc)EPSKSPGSLR^{3+}



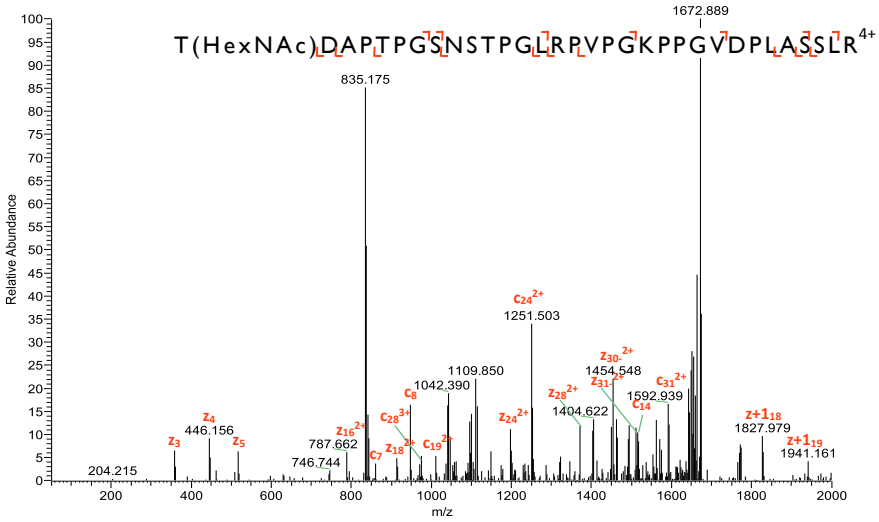
Tet1 GlcNAc-T535

T20100618-11 #1779 RT: 19.62 AV: 1 NL: 2.56E3
T: ITMS + c NSI d sa Full ms2 590.31@etd200.00 [50.00-1195.00]



Tle4 GlcNAc-T340

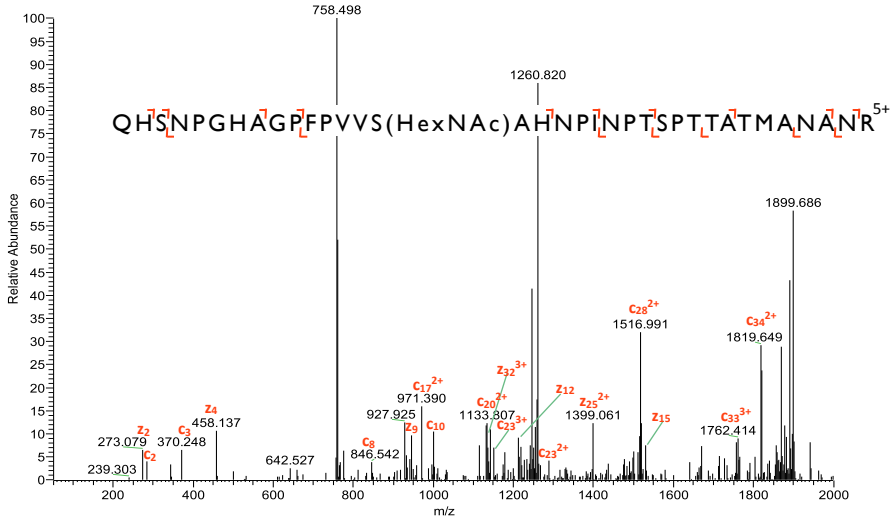
T20100618-19 #3344 RT: 33.43 AV: 1 NL: 3.18E3
T: ITMS + c NSI d sa Full ms2 836.44@etd100.00 [50.00-2000.00]



Trim33 GlcNAc-S350

T20100618-23 #2704 RT: 29.76 AV: 1 NL: 1.39E3

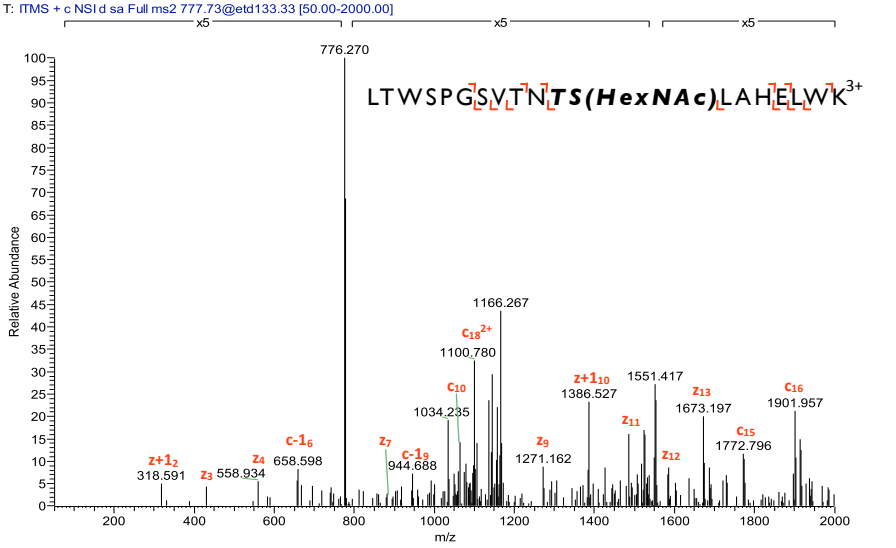
T: ITMS + c NSI d sa Full ms2 759.97@etd80.00 [50.00-2000.00]



Tnrc6a GlcNAc-T1646 ***or*** S1647

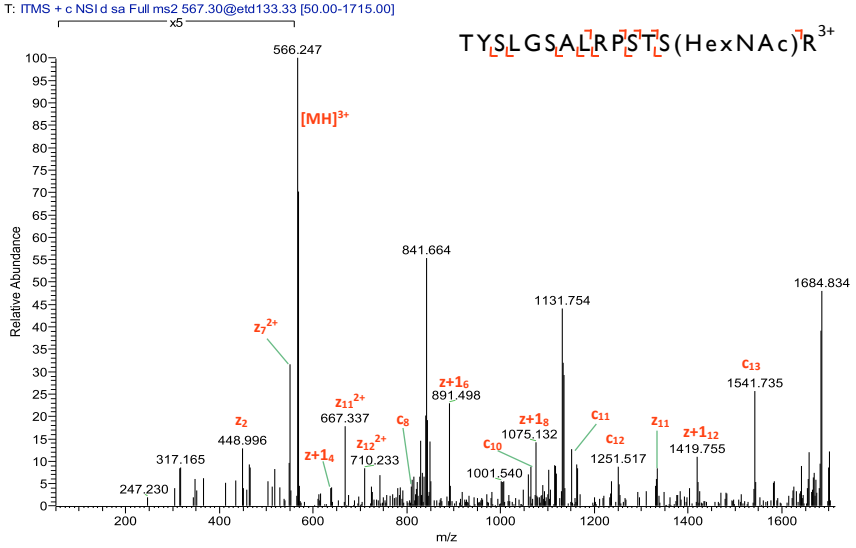
T20100409-24 #4539 RT: 44.59 AV: 1 NL: 1.82E3

T: ITMS + c NSI d sa Full ms2 777.73@etd133.33 [50.00-2000.00]



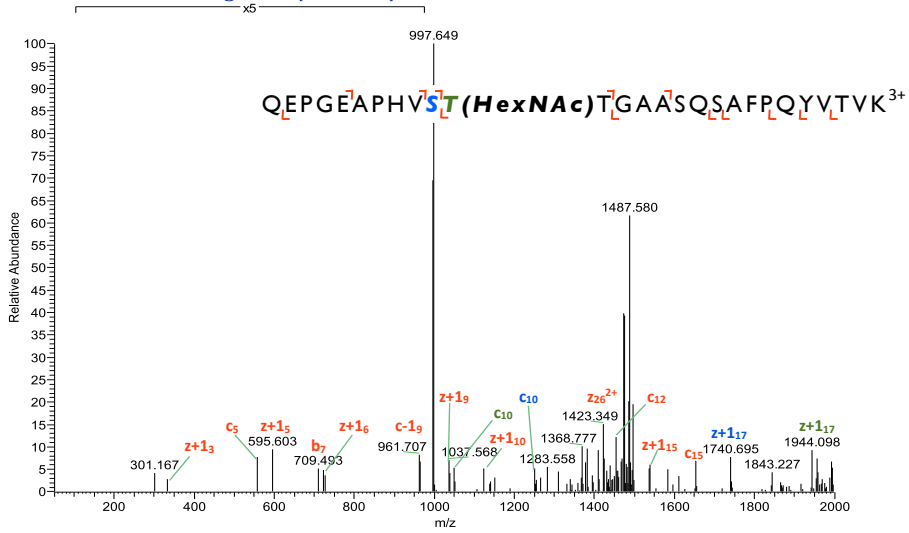
Vim GlcNAc-S49

T20100409-25 #2365 RT: 25.85 AV: 1 NL: 8.43E2
T: ITMS + c NSI d sa Full ms2 567.30@etd133.33 [50.00-1715.00]

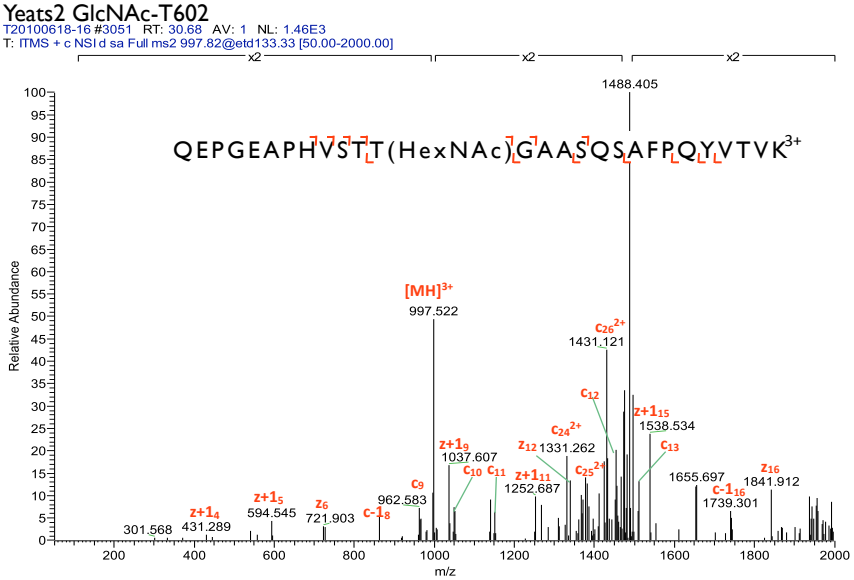


Yeats2 GlcNAc-Mixture of S600 and T601

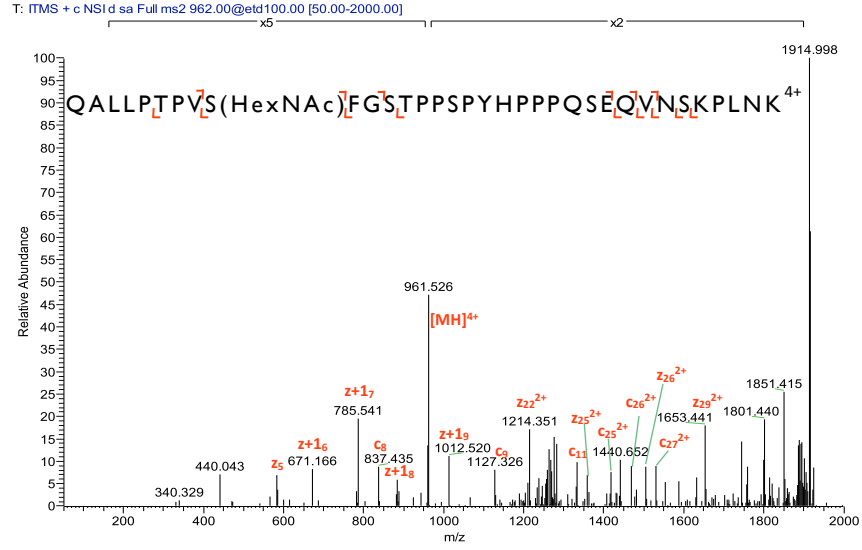
T20100409-18 #3179 RT: 31.61 AV: 1 NL: 6.43E2
T: ITMS + c NSI d sa Full ms2 997.82@etd133.33 [50.00-2000.00]



Yeats2 GlcNAc-T602

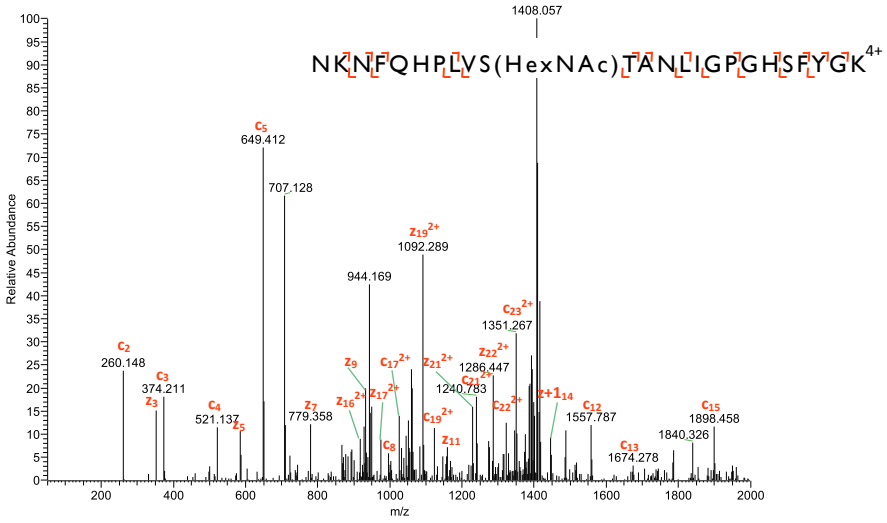


Ylpm1 GlcNAc-S636



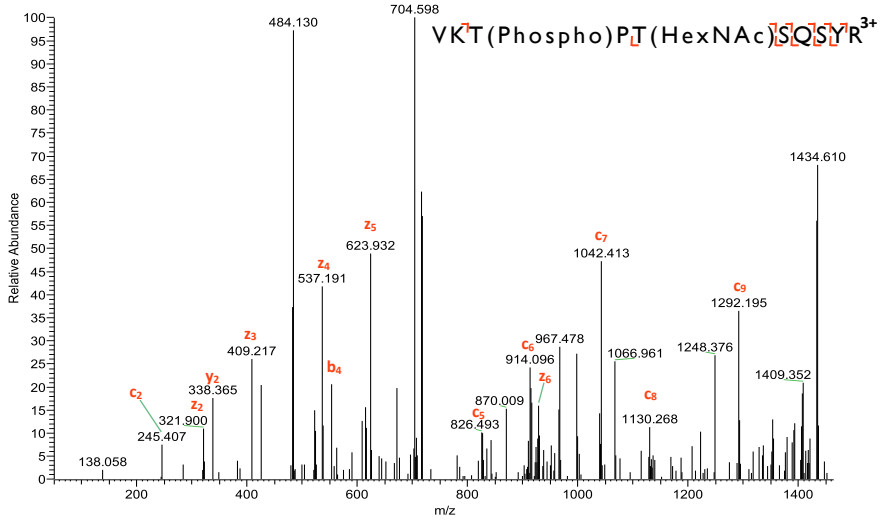
Zfhx3 GlcNAc-S352

T20100618-26 #2999 RT: 32.40 AV: 1 NL: 1.07E3
T: ITMS + c NSI d sa Full ms2 708.37@etd100.00 [50.00-2000.00]

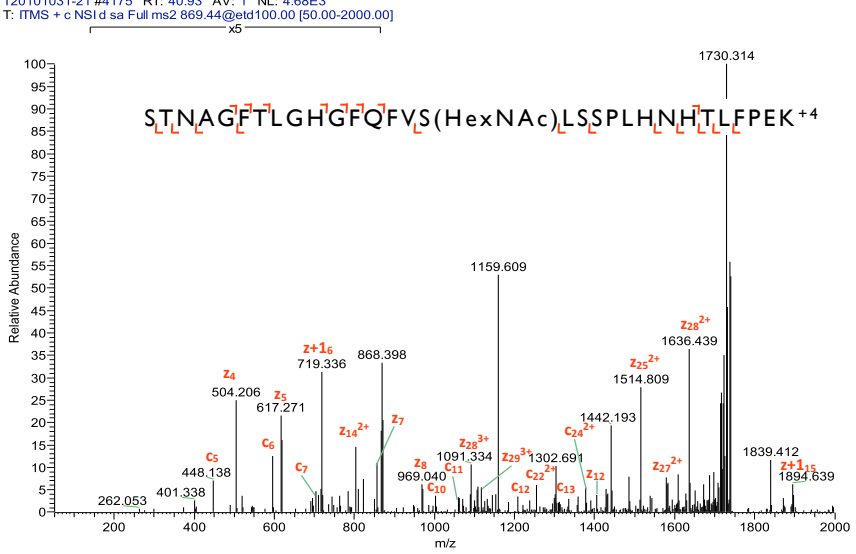


Zfp281 GlcNAc-T888

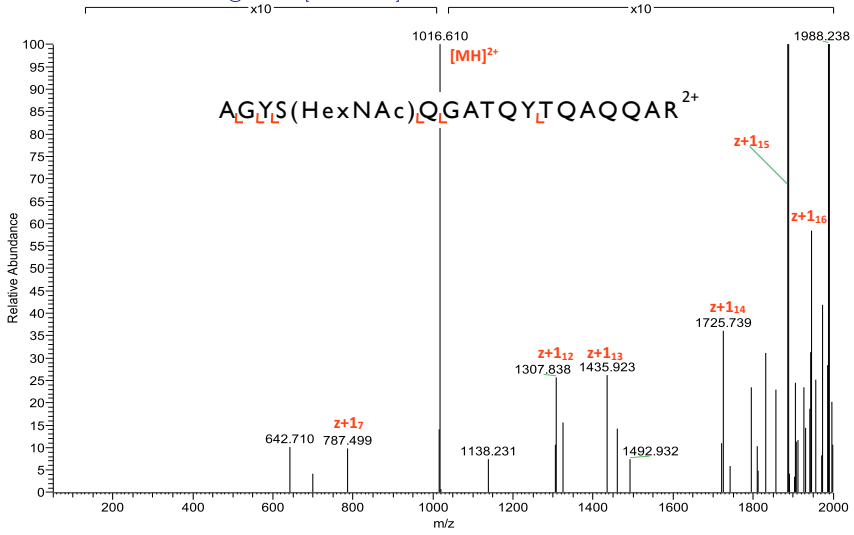
T9100412 #809 RT: 12.50 AV: 1 NL: 3.24E2
T: ITMS + c NSI d sa Full ms2 483.89@etd133.33 [50.00-1465.00]



Zfp281 GlcNAc-S691

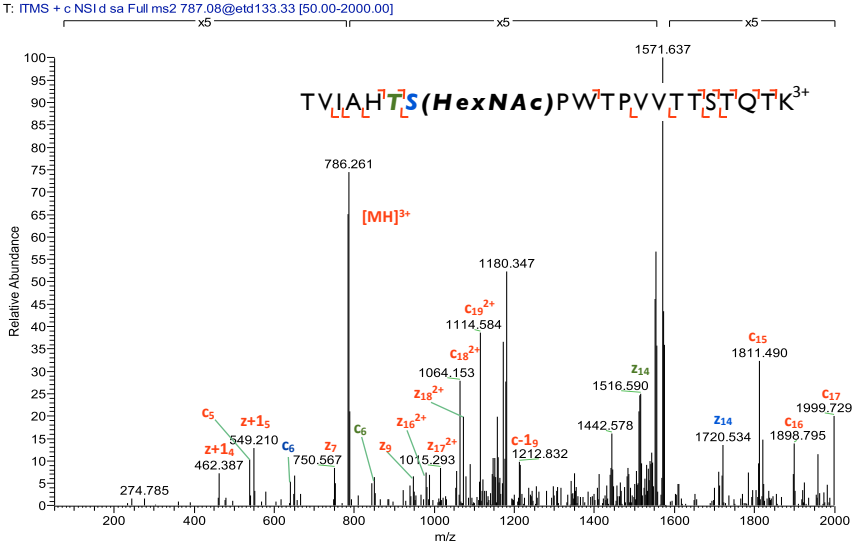


Zfr GlcNAc-S195



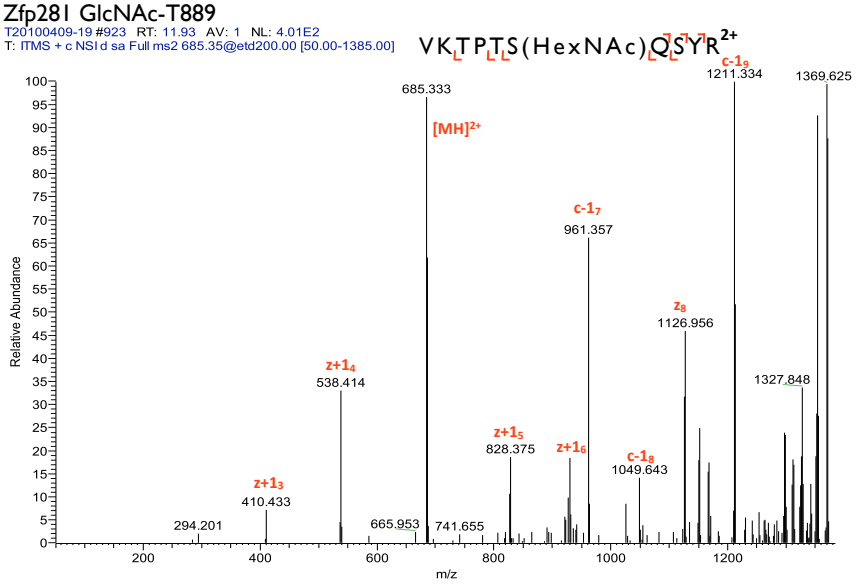
Znf318 Mixture of GlcNAc-T1185 and S1186

T20100409-23 #2953 RT: 30.08 AV: 1 NL: 3.27E3
T: ITMS + c NSI d sa Full ms2 787.08@etd133.33 [50.00-2000.00]



Zfp281 GlcNAc-T889

T20100409-19 #923 RT: 11.93 AV: 1 NL: 4.01E2
T: ITMS + c NSI d sa Full ms2 685.35@etd200.00 [50.00-1385.00]

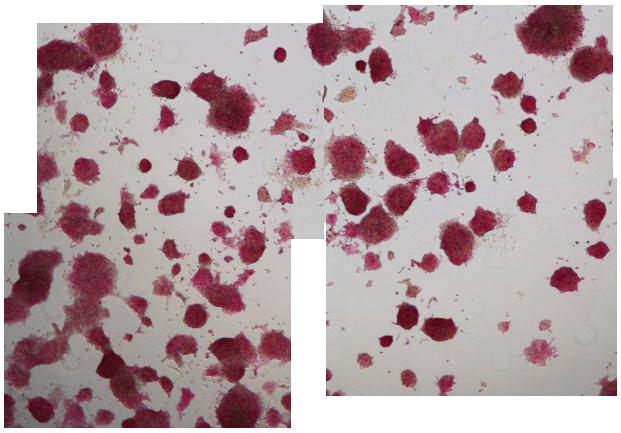


Appendix II Alkaline phosphatase staining and colony morphology of fSOX2-Tg and fS248A-Tg cells

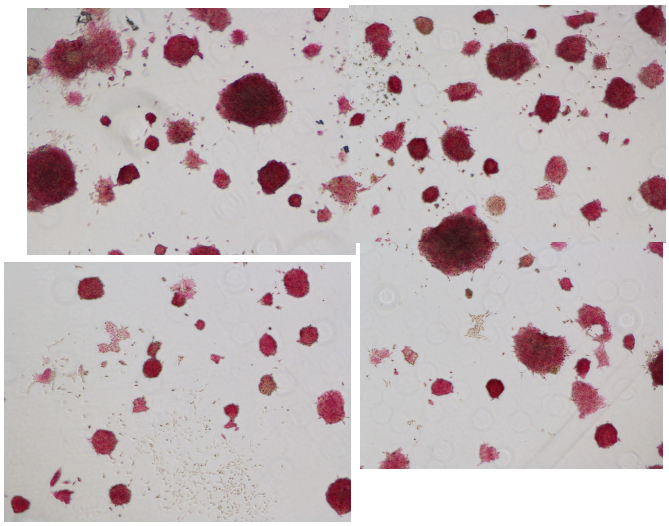
Alkaline Phosphatase Images of Differential SOX2 Interactor Knockdowns

Images of alkaline phosphatase staining for fSOX2-Tg or fS248A-Tg cells with the knockdowns of *Gfp* or genes that were identified to differentially interact with 3xF-SOX2^{WT} or 3xF-SOX2^{S248A}. See Chapter 5 Figure9C for quantiation. Each figure is four concatenated fields of view at 4x.

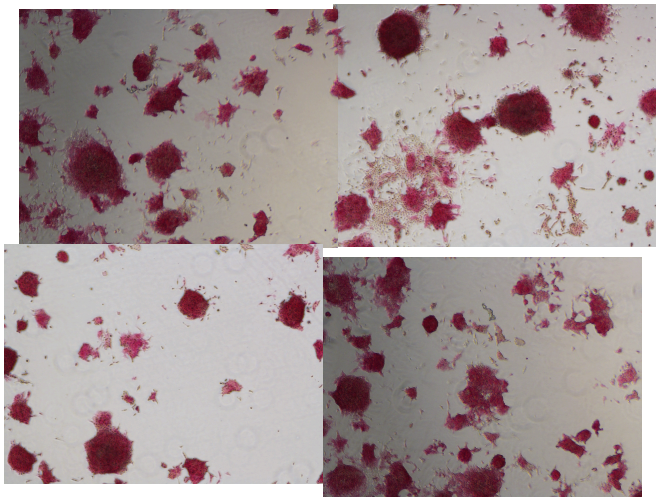
fSOX2-Tg mock



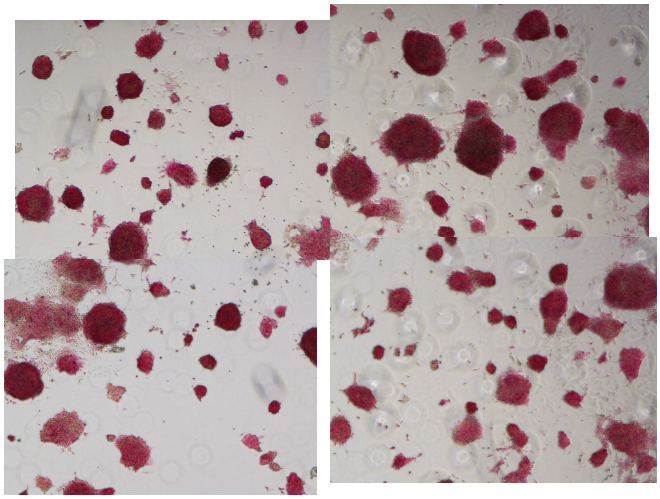
fSOX2-Tg Dnmt1



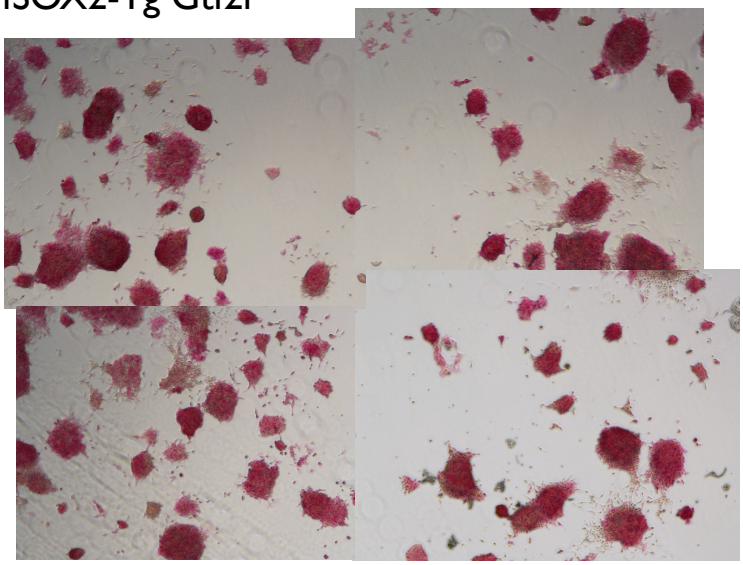
fSOX2-Tg Brg1



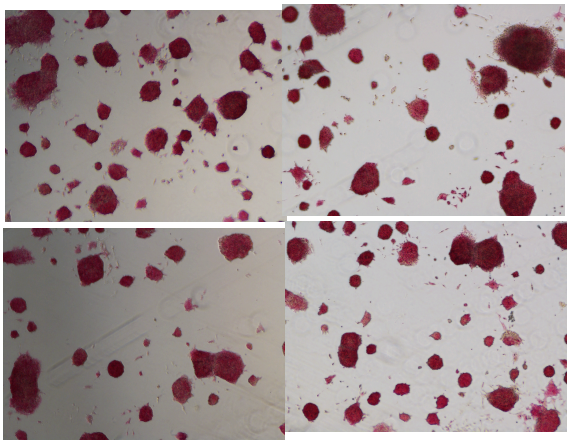
fSOX2-Tg GataD2beta



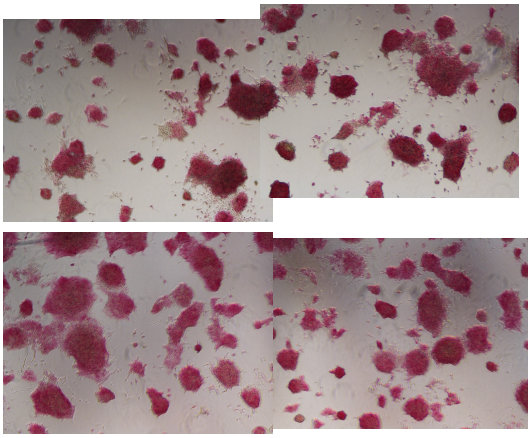
fSOX2-Tg Gtf2i



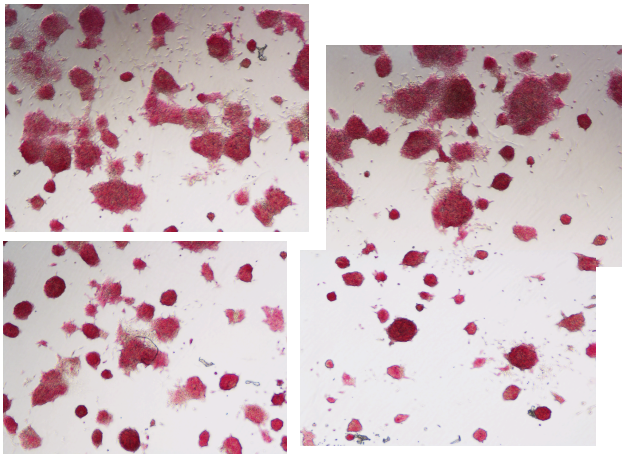
fSOX2-Tg Mbd3



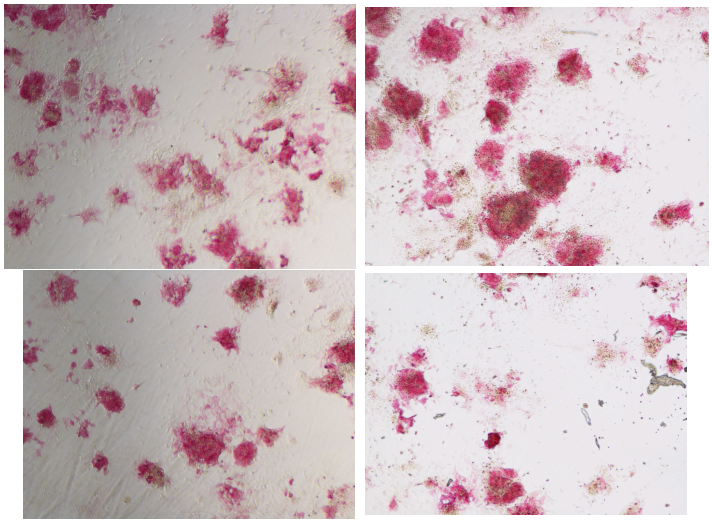
fSOX2-Tg Msh2



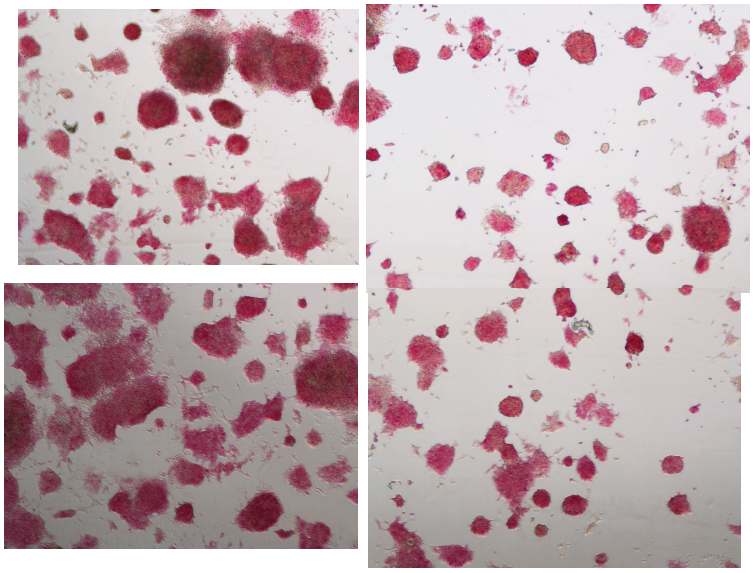
fSOX2-Tg Mta2



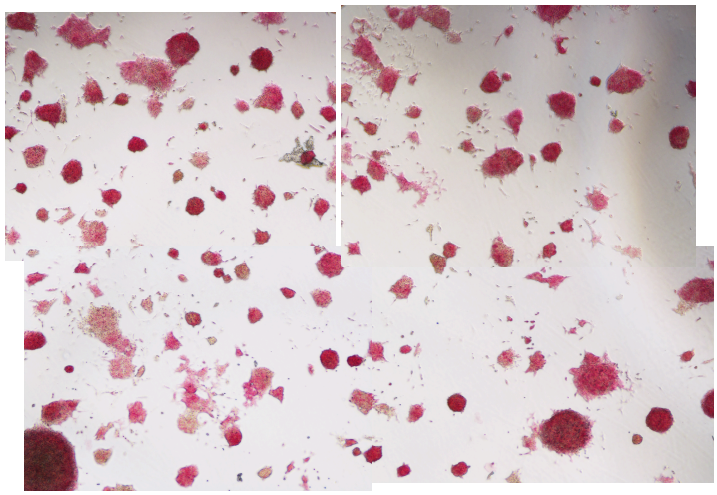
fSOX2-Tg Oct4



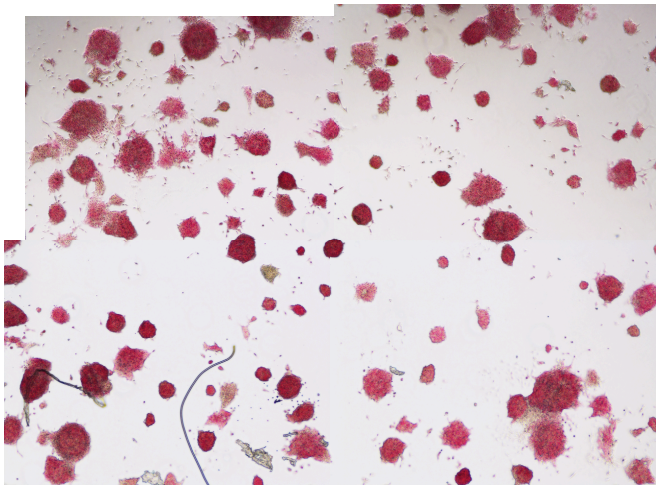
fSOX2-Tg Parp1



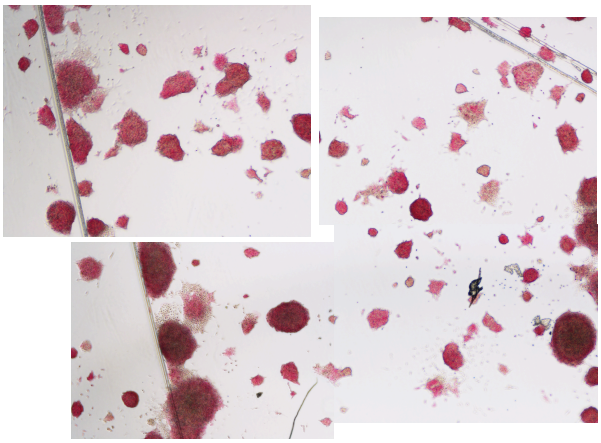
fSOX2-Tg Stk38



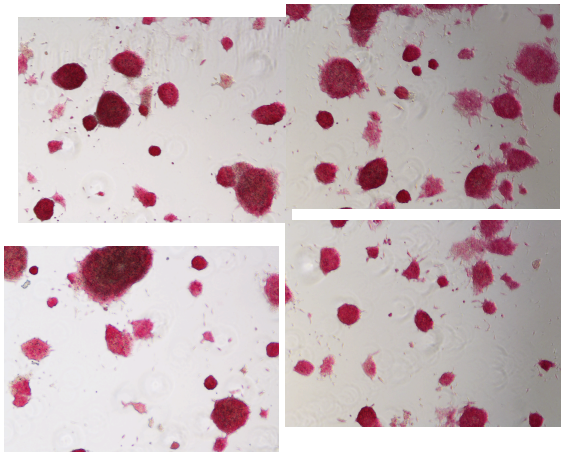
fSOX2-Tg Wdr76



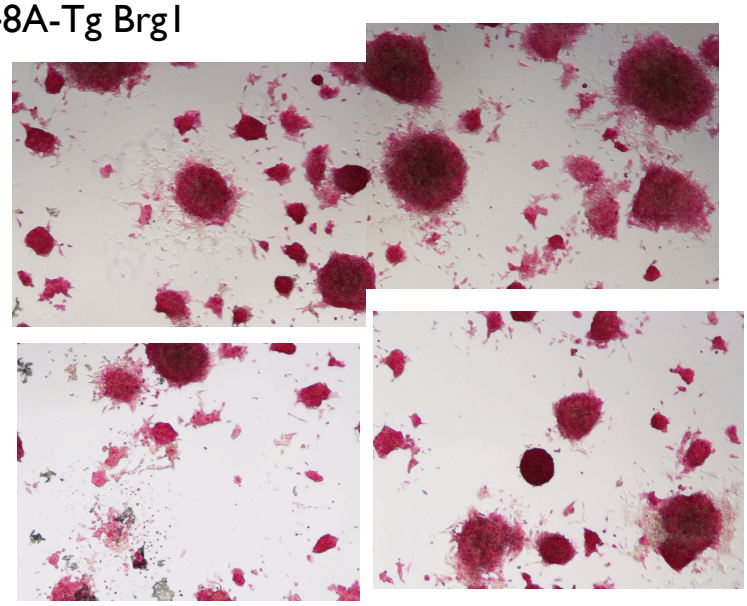
fSOX2-Tg Xrcc5



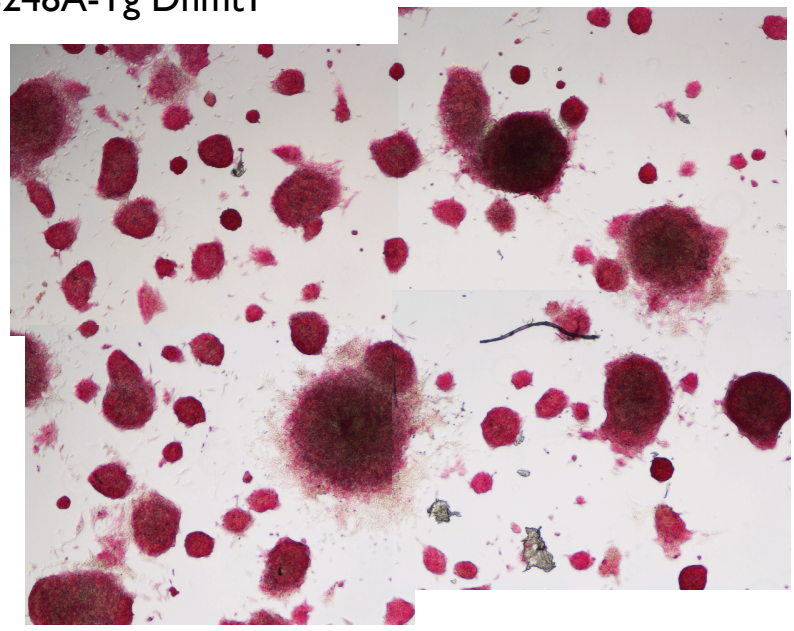
fS248A-Tg mock



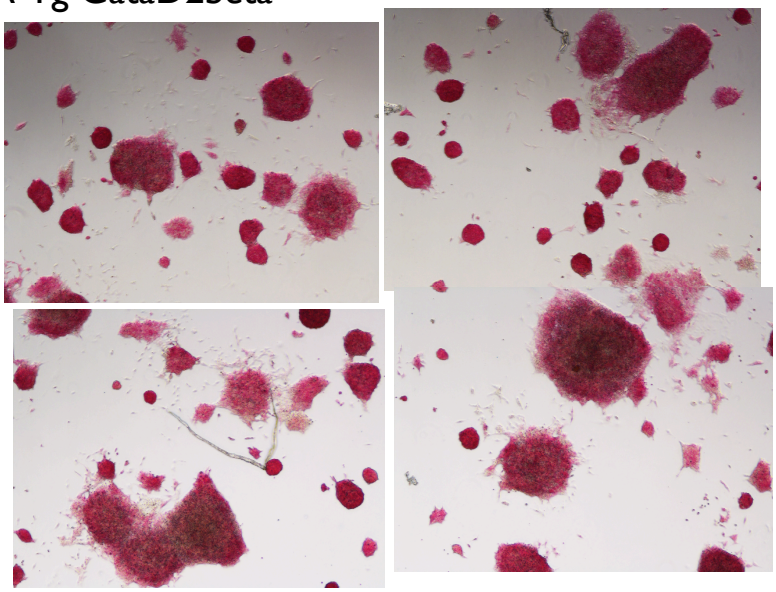
fS248A-Tg Brg1



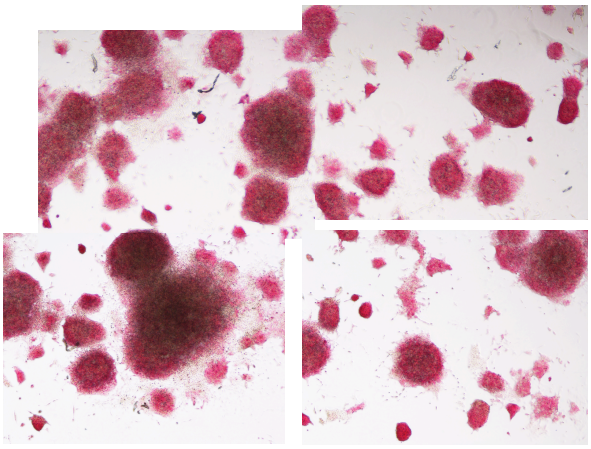
fS248A-Tg Dnmt1



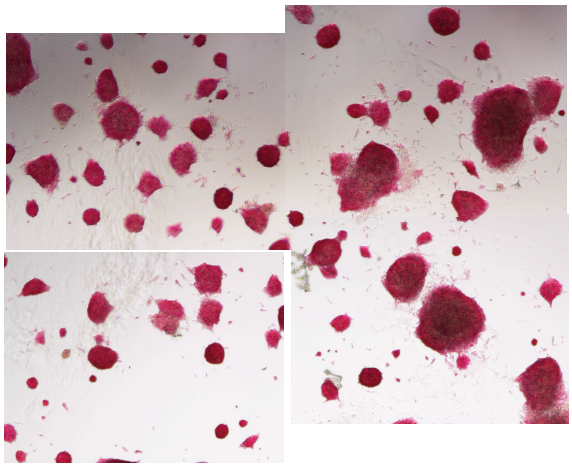
fS248A-Tg GataD2beta



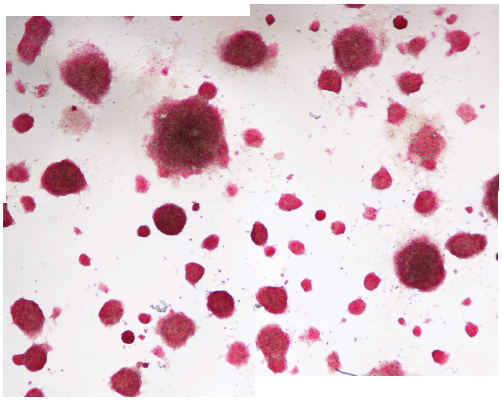
fS248A-Tg Gtf2i



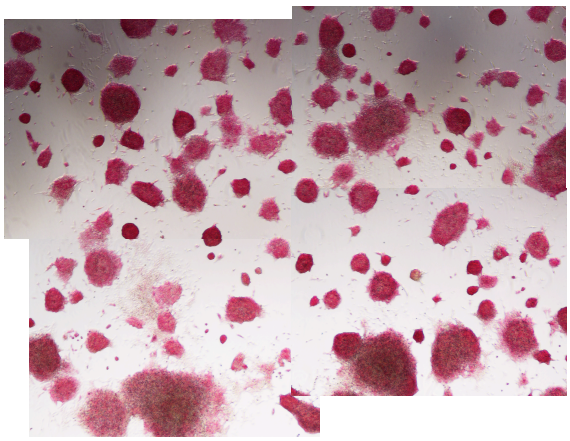
fS248A-Tg Mbd3



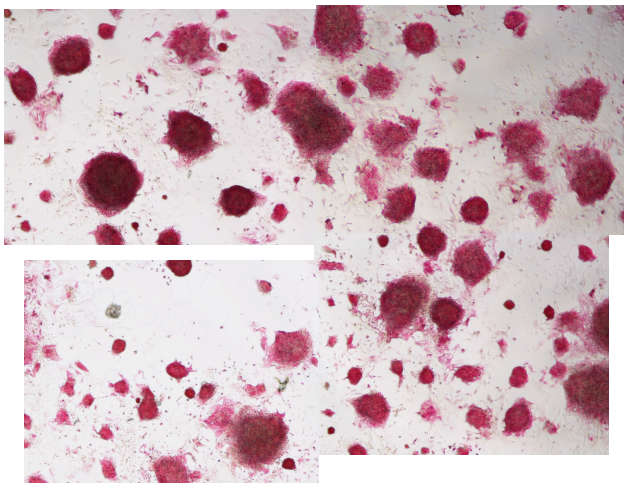
fS248A-Tg Msh2



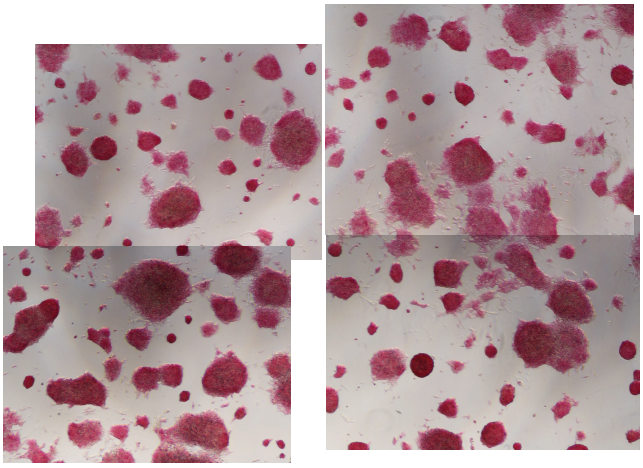
fS248A-Tg Mta2



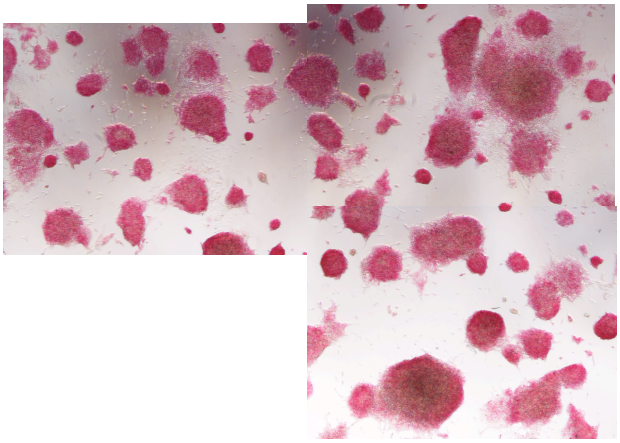
fS248A-Tg Oct4



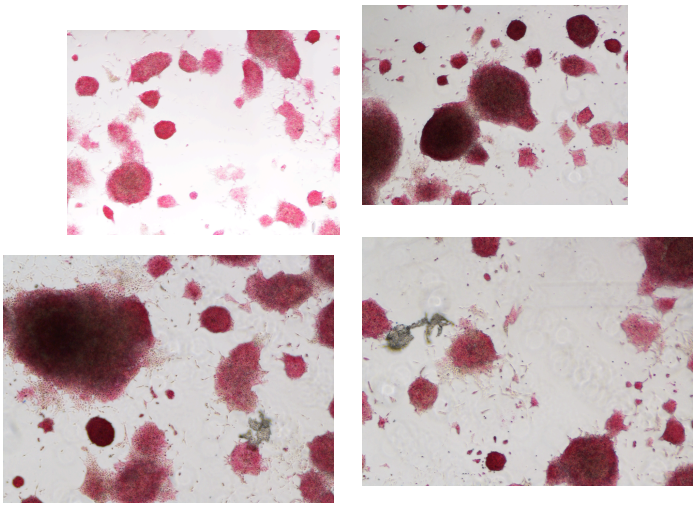
fS248A-Tg Parp1



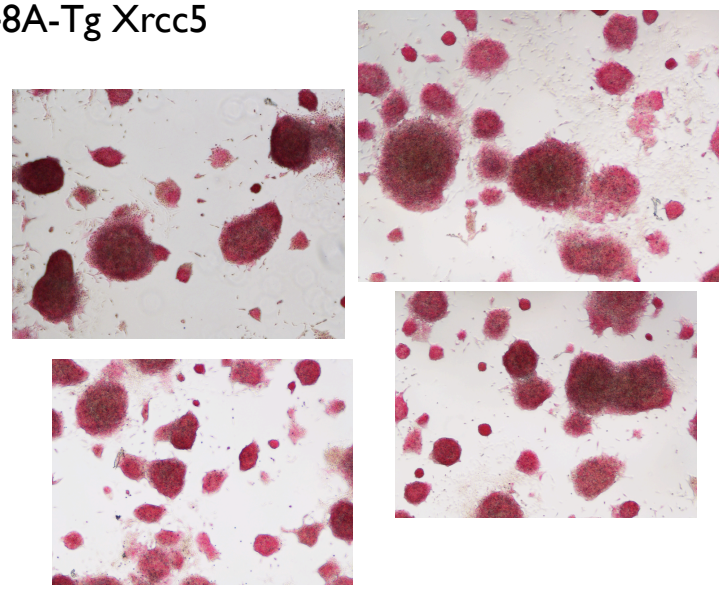
fS248A-Tg Stk38



fS248A-Tg Wdr76



fS248A-Tg Xrcc5



Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

4 July 2014

Date