Title
Investigating the OGT-TET interaction in vitro and in mouse embryonic stem cells

Author
Hrit, Joel

Publication Date
2018

Peer reviewed|Thesis/dissertation

Investigating the OGT-TET interaction in vitro and in mouse embryonic stem cells

by

Joel Avery Hrit

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

Acknowledgments

I want to thank all of my science teachers over the years, especially Donald Bomeli and Yasmeen Youngs, who instilled in me a love and passion for science. Thank you to my undergraduate research advisor, Aaron Goldstrohm, who welcomed me into his lab and introduced me to scientific research. I'm deeply indebted to his patient mentorship and training.

Thank you to all of my friends in my graduate school class, especially Johnny Rodriguez, Kelly Crotty, Fernando Meza-Gutierrez, and Andre Lazar. You've been dear friends ever since I moved alone to San Francisco. I'm so grateful for all the fun times we've had over the last six years. Thank you to the current and former members of the Panning lab for friendship, mentorship, and support. I especially want to thank my friends Leeanne Goodrich, Betsy Martin, and Luke Strauskulage for assistance in many forms, from performing and troubleshooting experiments to drinking beers to forget about experiments. Thank you also to Sailaja Peddada, whose expert advice helped me find my footing in the first couple years of graduate school.

Thank you to my parents, Tom and Lisa Hrit, who raised me to understand the value of education, hard work, and critical thinking and supported my pursuits from day one (literally).

I can't give enough thanks to my beautiful wife and best friend, Gabriella Hrit. She has been my biggest encouragement and support though the all the trials and triumphs of graduate school. Her contribution to the work described in this thesis can't be overstated.

Finally, thank you to my incredible thesis advisor Barbara Panning. Her great depth of scientific knowledge and skill are only one facet among many that make her such a good mentor. She has given me extensive guidance and mentorship along with extensive freedom to explore in my graduate school research. Her constant patience and calm demeanor have kept me grounded through difficulty and frustration. I'm so glad I was able to perform my graduate work in her lab.

The text of chapter 2 of this thesis, in revised form, is under revision and review for publication at eLife. It includes the following contributions from collaborators:

- The zebrafish rescue experiments in Figure 2.6 were performed by Cheng Li in the lab of Mary Goll.
- The analysis of RNA-seq data in Figure 2.7D was performed by Leeanne Goodrich.
- The mass spectrometry in Figure 2.7F was performed by Jenna Fernandez in the lab of Natalia Tretyakova.
- The whole genome bisulfite sequencing in Figure 2.8 was performed by Bang-An Wang, with the assistance of Joseph R. Neay and Rosa Castanon, in the lab of Joseph R. Ecker

The mass spectrometry in chapters 3 and 4 was performed by Jason Maynard in the lab of Alma Burlingame.

Investigating the OGT-TET interaction *in vitro* and in mouse embryonic stem cells

By Joel Avery Hrit

Abstract

Proper spatial and temporal control of gene expression is necessary for cellular survival and proper function. Addition of a methyl group to the 5' carbon of cytosine in DNA (5mC) is a major mechanism used to modulate gene expression. The Ten-Eleven Translocation (TET) family of enzymes iteratively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). These modifications function both as stable epigenetic marks and transient intermediates in the demethylation of DNA. Improper placement of these epigenetic marks often causes death or disease. Among the proteins that interact with TET enzymes is *O*-linked N-acetylglucosamine (*O*-GlcNAc) Transferase (OGT). OGT is the sole enzyme responsible for attaching a GlcNAc sugar to serine, threonine, and cysteine residues of over 1,000 nuclear, cytoplasmic, and mitochondrial proteins. OGT has been termed a "nutrient sensor" because its activity requires the sugar donor UDP-GlcNAc, whose abundance is dependent upon the levels of various cellular metabolites. Thus the reversible *O*-GlcNAc modification dynamically regulates the functions of OGT's targets in response to nutrient status. OGT stably interacts with and modifies TET proteins and its genome-wide distribution overlaps significantly with TETs. However, the significance of the OGT-TET interactions are poorly understood. In this work, we explore the consequences of the OGT-TET interactions *in vitro* and in mouse embryonic stem cells (mESCs). We show that OGT directly binds and modifies TET1 *in vitro*, and the *O*-GlcNAc modification enhances TET1 activity. We

identify a point mutation in TET1 that disrupts its interaction with OGT and use this to interrogate the effects on TET activity, gene expression, and epigenetic patterning of disrupting the OGT-TET1 interaction in mESCs. To assess the importance of the OGT-TET interaction for OGT function, we use quantitative SILAC mass spectrometry to examine proteome-wide changes in *O*-GlcNAcylation in mESCs when Tets are deleted. We also identify sites of *O*-GlcNAcylation on TET1 and TET2, further analyze the interactions between OGT and all three TETs, and examine the effect of *O*-GlcNAcylation on TET2 and TET3 activity *in vitro*. Our results link metabolism and epigenetic control, which may be relevant to the developmental and disease processes regulated by these two enzymes.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Appendix

Chapter 1

Introduction

The importance of epigenetics

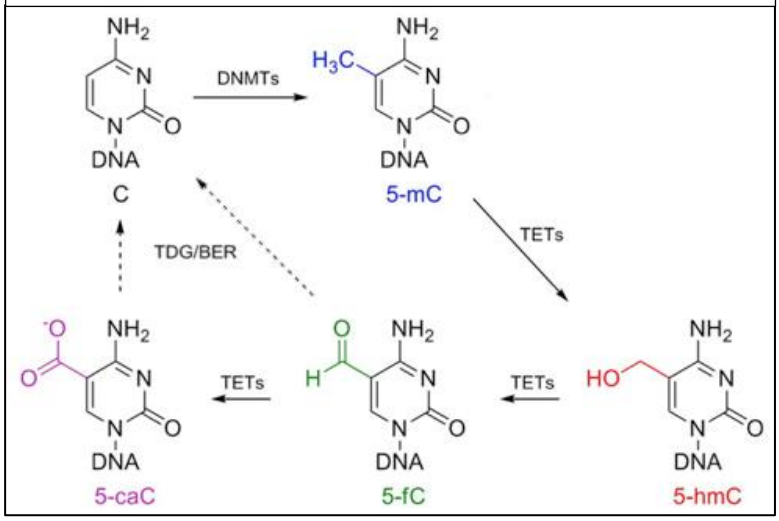
An organism's unique DNA sequence, their genetic code, accounts for many of the observable differences between individuals. However, the variety encoded in a cell's genetic material goes far beyond differences in DNA sequence. Every cell comprising a multicellular organism contains the same DNA molecule, yet different cell types take on vastly different characteristics to perform the variety of functions necessary in a complex organism. Thus epigenetics, chemical modifications to DNA itself and its associated histone proteins that regulate gene expression, has emerged as an important field of study. Many developmental disorders and disease states can be traced back to aberrant epigenetic patterning resulting in improper temporal and spatial expression of genes. Thus, studying the effects of individual epigenetic marks on gene expression and the enzymes that write, read, and erase those marks is important both for a fuller understanding of basic biology and for the development of treatments for diseases and disorders caused by epigenetic misregulation.

DNA methylation and TET enzymes

For decades, the only characterized epigenetic modification on the DNA molecule itself was 5-methylcytosine (5mC). This mark has been studied extensively and is generally associated with silencing of its associated genes, although this depends in part on its location in the genome and the histone modifications with which it overlaps(Guibert & Weber, 2013; Smith & Meissner, 2013). In the past decade, however, the epigenome has expanded with the discovery and characterization of the Ten-Eleven Translocation (TET) family of enzymes, which oxidize 5mC in three sequential reactions to form 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (He et al., 2011; S. Ito et al., 2011; Kriaucionis & Heintz, 2009;

Tahiliani et al., 2009) (Figure 1.1). These three DNA modifications are thought to serve at least two purposes in epigenetic regulation. First, like 5mC, its three derivatives 5hmC, 5fC, and 5caC can recruit specific factors to DNA while excluding others(Frauer, Hoffmann, et al., 2011a; Spruijt et al., 2013), and 5hmC and 5fC are stably present at certain genomic regions(Bachman et al., 2015; 2014). This suggests that these modifications may function as stable marks that influence the expression of their associated genes. Second, the oxidation catalyzed by TETs is part of an active DNA demethylation pathway, since 5fC and 5caC can be recognized and excised by DNA glycosylases and replaced with unmodified cytosine(Cortellino et al., 2011; He et al., 2011; Müller, Bauer, Siegl, Rottach, & Leonhardt, 2014; A. R. Weber et al., 2016). This is the only known process by which cells can actively remove 5mC.

Figure 1.1 – DNA methylation cycle
Cytosine in DNA can be methylated by DNA methyltransferases (DNMTs) to form 5mC. TET enzymes sequentially oxidize 5mC to 5hmC, 5fC, and 5caC. 5fC and 5caC can be recognized and excised by the glycosylase TDG and replaced with C by the base excision repair (BER) machinery.



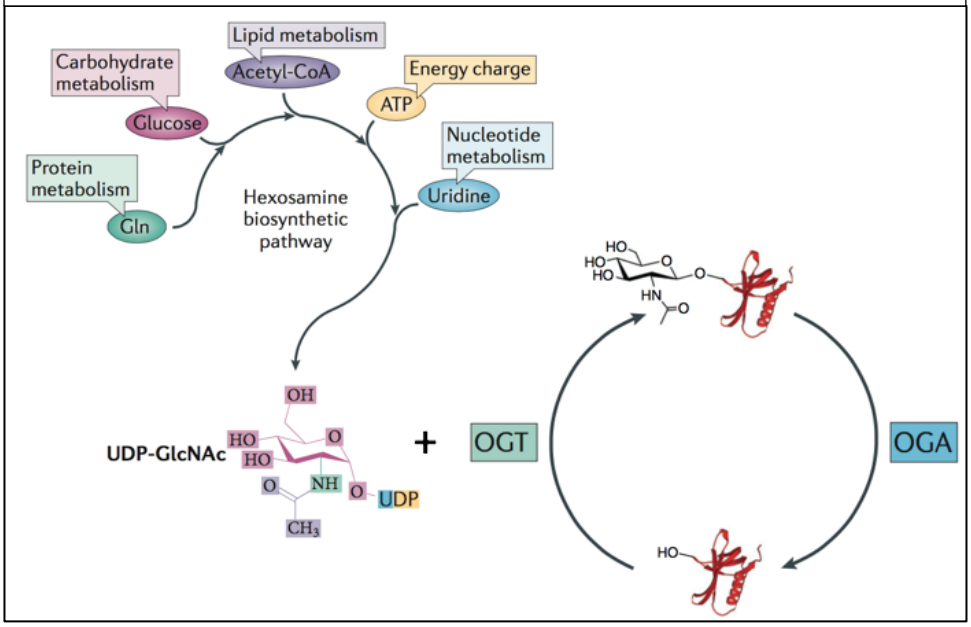
Despite significant effort to understand TETs and the DNA modifications they catalyze, our understanding is poor due in part to the complexity of the problem. The three TET proteins (TET1, TET2, and TET3) all catalyze the same chemical reactions but are differentially expressed in different tissues and stages of development and target different regions of the genome for oxidation(Fidalgo et al., 2016; Globisch et al., 2010; Huang et al., 2014; S. Ito et al., 2010; Koh et al., 2011). The presence of all three oxidized bases in the genome means that TETs perform anywhere from one to three oxidation reactions on a given cytosine before dissociating. And the specific effects on gene expression of 5hmC, 5fC, and 5caC are not straightforward, being dependent on where in the genome they are found and what other epigenetic marks are present(Delatte, Deplus, & Fuks, 2014). What is clear is the importance of TETs in human health and disease: Tet loss of function mutations and global loss of 5hmC are common features in various cancers(Baylin & Jones, 2011; Ko et al., 2010; Lian et al., 2012). This aberrant epigenetic state is thought to contribute to the misregulation of gene expression necessary for cancer. Thus it is important to investigate the numerous open questions about how TETs and their oxidized cytosine products are regulated in space and time. This work focuses on one promising candidate regulator of TET enzyme function: *O*-GlcNAc transferase (OGT), discussed next.

Protein glycosylation by OGT

Reversible post-translational modifications (PTMs) on proteins regulate their function by altering their activity, localization, protein-protein interactions, and stability. Many proteins depend on PTMs to function properly under various conditions and stresses both inside and outside the cell. One important PTM is the addition of *O*-linked N-acetylglucosamine (*O*-

GlcNAc) to serine, threonine and cysteine residues on over 1,000 intracellular proteins by the enzyme O-linked N-acetylglucosamine transferase (*O*-GlcNAc transferase, or OGT) (Haltiwanger, Holt, & Hart, 1990; Maynard, Burlingame, & Medzihradszky, 2016). The sugar donor used by OGT, UDP-GlcNAc, is synthesized by the hexosamine biosynthetic pathway, which takes inputs from the metabolism of sugars, amino acids, lipids, nucleotides, and other metabolites (Figure 1.2). Since OGT's overall activity and specific targets are dependent on the amount of UDP-GlcNAc(Shen, Gloster, Yuzwa, & Vocadlo, 2012; Vocadlo, 2012), OGT is thought to function as a "nutrient sensor," dynamically regulating the functions of its targets in response to nutrient availability(Levine & Walker, 2016; Ruan, Singh, Li, Wu, & Yang, 2013). OGT modifies proteins involved in virtually every known cellular process, which makes it absolutely essential for survival. Ogt deletion causes early embryonic lethality, and mutation or misregulation of Ogt is tied to many diseases, most notably diabetes and various types of cancer(Hart, Slawson, Ramirez-Correa, & Lagerlof, 2011; Levine & Walker, 2016).

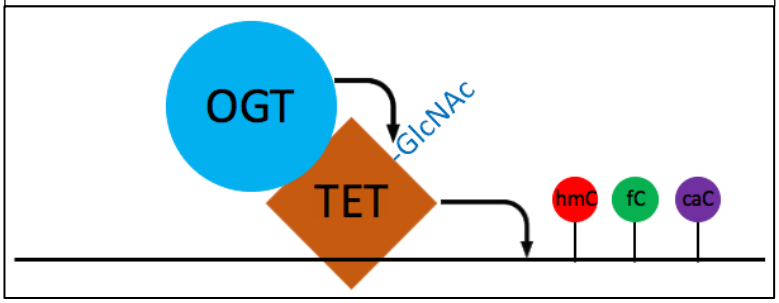
Figure 1.2 – Protein *O*-GlcNAcylation by OGT
OGT adds *O*-linked N-acetylglucosamine (*O*-GlcNAc) to serine, threonine, and cysteine residues of target proteins using the UDP-GlcNAc cofactor. UDP-GlcNAc is synthesized from numerous cellular metabolites via the hexosamine biosynthetic pathway, causing OGT's activity to vary in response to nutrient status.



The TET-OGT interaction

In 2013, three separate studies demonstrated an interaction between OGT and TET proteins(Q. Chen, Chen, Bian, Fujiki, & Yu, 2013; Deplus et al., 2013; Vella et al., 2013). In summary, it was found that OGT directly interacts with and *O*-GlcNAcylates all three TETs and OGT's genome-wide distribution overlaps significantly with TETs, especially TET1 (Figure 1.3). One group also showed that knockdown of Tet1 significantly depleted OGT from chromatin, suggesting that OGT may depend on TETs for recruitment to the genome(Vella et al., 2013).

Figure 1.3 – Diagram of the TET-OGT protein-protein interaction
OGT directly interacts with and modifies all three TET proteins and co-localizes with OGT on DNA.



The *O*-GlcNAc modification regulates numerous aspects of its target proteins, including subcellular localization, protein-protein interactions, catalytic activity, stability, and others(Levine & Walker, 2016). OGT could be exerting any or none of these regulatory effects on TET proteins. Thus one important question is, are OGT and *O*-GlcNAc regulating TET enzymes and if so, how? Regulation of TET enzymes by OGT would suggest a tempting model linking metabolic sensing to epigenetic control: a cell could potentially modulate its transcriptional program in response to changing nutrient availability through differential activity of OGT toward TETs (Figure 1.4).

On the other hand, one could ask the question, are TET proteins regulating OGT? Among OGT's cellular targets are many chromatin-associated proteins, including transcription factors and histones(Myers, Panning, & Burlingame, 2011a; Ruan et al., 2013; Sakabe, Wang, & Hart, 2010). However, although some OGT is bound to chromatin in cells, OGT is not known to bind DNA or chromatin by itself *in vitro*. The overlap of OGT and TETs on chromatin and the depletion of OGT from chromatin upon knockdown of Tet1(Vella et al., 2013) hint that perhaps TETs are important for bringing OGT to chromatin. Recruitment of OGT to specific genomic

sites by TETs could be necessary for OGT to modify its chromatin-bound targets (Figure 1.5). Thus, both possibilities of reciprocal co-regulation between OGT and TET proteins should be investigated to understand the biological significance of the TET-OGT interaction.

Figure 1.4 – Model for regulation of TET enzymes by OGT
In this model OGT, in response to the cell's nutrient status, modifies TET enzymes to regulate their DNA modification activity to achieve a desired gene expression pattern.

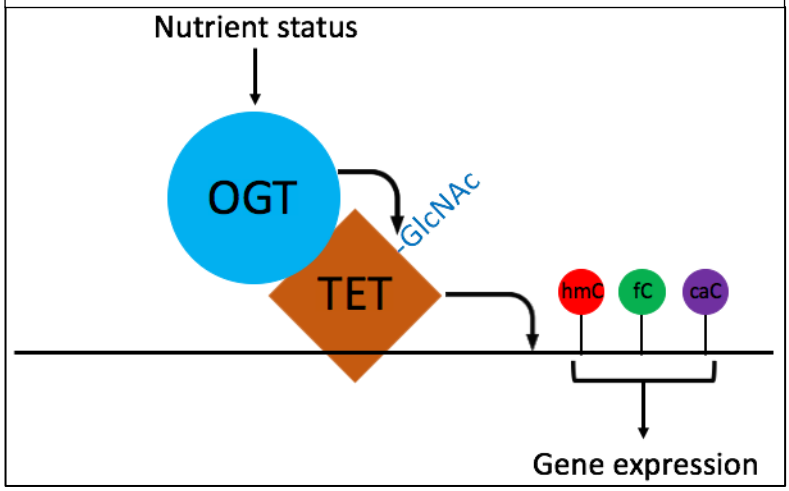
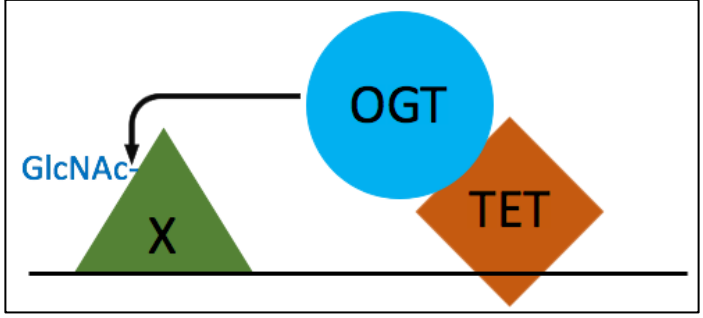


Figure 1.5 – Model for regulation of OGT by TETs
In this model TET enzymes recruit OGT to DNA, allowing OGT to modify its chromatin-associated targets.



Mouse embryonic stem cells as a model system to study the TET-OGT interaction

Mouse embryonic stem cells (mESCs) are derived from the inner cell mass of the blastocyst at embryonic day 4.5. In culture these cells exhibit both self-renewal (the ability to divide indefinitely to produce more stem cells) and pluripotency (the ability to become any mouse somatic cell type). mESCs in culture are a widely accepted model system to study mammalian stem cell biology and early development.

Both OGT and TET proteins are important in stem cells and early development. Knockout of *Ogt* causes early embryonic lethality(Shafi et al., 2000), and depletion of OGT from mESCs leads to spontaneous differentiation and cell death. This is consistent with OGT's critical role in regulating a plethora of biological processes, including the pluripotency transcription factor network(Levine & Walker, 2016; Myers et al., 2016). In contrast, all three Tet genes can be disrupted in mESCs without loss of viability, but *Tet1/2/3* triple knockout (TKO) mESCs exhibit severe differentiation defects in culture(Dawlaty et al., 2014). Consistent with this, *Tet1/2/3* TKO mice are inviable(Dawlaty et al., 2014). Ogt, Tet1, and Tet2 are all highly expressed in male mESCs(Q. Chen et al., 2013; Vella et al., 2013), and female mESCs express Tet3 as well (unpublished data). Therefore, mESCs are a suitable model system to study the importance of the TET-OGT interaction for the functions of OGT and all the TETs in embryonic stem cell biology and early mammalian development.

Summary

The TET family of epigenetic regulators and the nutrient-sensing post-translational modifier OGT are important in early mammalian development. OGT interacts with and modifies all three TET proteins, hinting that OGT may regulate TET enzyme activity. This would suggest

a model that ties nutrient sensing to epigenetics, in which a cell could regulate its gene expression in response to changing availability of metabolites through OGT-mediated control of TET activity (Figure 1.4). Conversely, since OGT modifies numerous DNA-bound proteins in mESCs but appears to be dependent on TETs for chromatin association, TETs may regulate the activity of OGT by recruiting it to specific sites on chromatin where it can *O*-GlcNAcylate target proteins (Figure 1.5). The first model, that OGT regulates the activity of TETs, is explored in chapter 2 using mutational analysis, *in vitro* biochemical assays, and cell biological and genome-wide profiling experiments in mESCs. The second model, that TETs recruit OGT to chromatin to modify target proteins, is explored in chapter 3 using a quantitative mass spectrometry screen. The remaining chapters detail further analysis into the binding of OGT to the three TETs and its effect on TET protein activity *in vitro*.

Chapter 2

OGT binds a conserved C-terminal domain of TET1 to regulate TET1 activity and function in development

Introduction

Methylation at the 5' position of cytosine in DNA is a widespread epigenetic regulator of gene expression. Proper deposition and removal of this mark is indispensable for normal vertebrate development, and misregulation of DNA methylation is a common feature in many diseases(Guibert & Weber, 2013; Smith & Meissner, 2013). The discovery of the Ten-Eleven Translocation (TET) family of enzymes, which iteratively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), has expanded the epigenome(He et al., 2011; S. Ito et al., 2010; 2011; Kriaucionis & Heintz, 2009; Tahiliani et al., 2009). These modified cytosines have multiple roles, functioning both as transient intermediates in an active DNA demethylation pathway(Cortellino et al., 2011; Gao et al., 2013; Guo, Su, Zhong, Ming, & Song, 2011; He et al., 2011; A. R. Weber et al., 2016) and as stable epigenetic marks(Bachman et al., 2014; 2015) that may recruit specific readers(Spruijt et al., 2013).

One interesting interaction partner of TET proteins is *O*-linked N-acetylglucosamine (*O*-GlcNAc) Transferase (OGT). OGT is the sole enzyme responsible for attaching a GlcNAc sugar to serine, threonine, and cysteine residues of over 1,000 nuclear, cytoplasmic, and mitochondrial proteins(Haltiwanger et al., 1990; Hanover, Krause, & Love, 2012; Maynard et al., 2016). Like phosphorylation, *O*-GlcNAcylation is a reversible modification that affects the function of target proteins. OGT's targets regulate gene expression(Hardivillé & Hart, 2016; Lewis & Hanover, 2014), metabolism(Bullen et al., 2014; Hanover et al., 2012; Ruan et al., 2013), and signaling(Durning, Flanagan-Steet, Prasad, & Wells, 2016; Hanover et al., 2005), consistent with OGT's role in development and disease(Hart et al., 2011; Levine & Walker, 2016).

OGT stably interacts with and modifies all three TET proteins and its genome-wide distribution overlaps significantly with TETs(Q. Chen et al., 2013; Deplus et al., 2013; Vella et al., 2013). Two studies in mouse embryonic stem cells (mESCs) have suggested that TET1 and OGT may be intimately linked in regulation of gene expression, as depleting either enzyme reduced the chromatin association of the other and affected expression of its target genes(Shi et al., 2013; Vella et al., 2013). However, it is unclear to what extent these genome-wide changes are direct effects of perturbing the TET1-OGT interaction. Further work is necessary to uncover the biological importance of the partnership between TET1 and OGT.

In this work, we map the interaction between TET1 and OGT to a small C-terminal region of TET1, which is both necessary and sufficient to bind OGT. We show for the first time that OGT modifies the catalytic domain of TET1 *in vitro* and enhances its catalytic activity. We also use mutant TET1 to show that the TET1-OGT interaction promotes TET1 function in the developing zebrafish embryo. Finally, we show that in mESCs a mutation in TET1 that impairs its interaction with OGT results in alterations in gene expression and in abundance of 5mC and TET2. Together these results suggest that OGT regulates TET1 activity, indicating that the TET1-OGT interaction may be two-fold in function – allowing TET1 to recruit OGT to specific genomic loci and allowing OGT to modulate TET1 activity.

Materials and Methods

Cell Culture

The mESC line LF2 and its derivatives were routinely passaged by standard methods in KO-DMEM, 10% FBS, 2 mM glutamine, 1X non-essential amino acids, 0.1 mM b-

mercaptoethanol and recombinant leukemia inhibitory factor. HEK293T cells were cultured in DMEM, 10% FBS, and 2 mM glutamine.

Recombinant protein purification

Full-length human OGT in the pBJG vector was transformed into BL-21 DE3 E. coli. A liquid culture was grown in LB + 50ug/mL kanamycin at 37C until OD_{600} reached 1.0. IPTG was added to 1mM final and the culture was induced at 16C overnight. Cells were pelleted by centrifugation and resuspended in 5mL BugBuster (Novagen) + protease inhibitors (Sigma Aldrich) per gram of cell pellet. Cells were lysed on an orbital shaker for 20 minutes at room temperature. The lysate was clarified by centrifugation at 30,000g for 30 minutes at 4C. Clarified lysate was bound to Ni-NTA resin (Qiagen) at 4C and then poured over a disposable column. The column was washed with 6 column volumes of wash buffer 1 (20mM Tris pH 8, 1mM CHAPS, 10% glycerol, 5mM BME, 10mM imidazole, 250mM NaCl) followed by 6 column volumes of wash buffer 2 (wash buffer 1 with 50mM imidazole). The protein was eluted in 4 column volumes of elution buffer (20mM Tris pH 8, 1mM CHAPS, 5mM BME, 250mM imidazole, 250mM NaCl). Positive fractions were pooled and dialyzed into storage buffer (20mM Tris pH 8, 1mM CHAPS, 0.5mM THP, 10% glycerol, 150mM NaCl, 1mM EDTA), flash frozen in liquid nitrogen and stored at -80C in small aliquots.

Mouse TET1 catalytic domain (aa1367-2039) was expressed in sf9 insect cells according to the Bac-to-Bac Baculovirus Expression System. Constructs were cloned into the pFastBac HTA vector and transformed in DH10Bac E. coli for recombination into a bacmid. Bacmid containing the insert was isolated and used to transfect adherent sf9 cells for 6 days at 25C. Cell media (P1 virus) was isolated and used to infect 20mL of sf9 cells in suspension for 3 days. Cell

media (P2 virus) was isolated and used to infect a larger sf9 suspension culture for 3 days. Cells were pelleted by centrifugation, resuspended in lysis buffer (20mM Tris pH 8, 1% Triton, 10% glycerol, 20mM imidazole, 50mM NaCl, 1mM $MgCl_2$, 0.5mM TCEP, protease inhibitors, 2.5U/mL benzonase), and lysed by douncing and agitation at 4C for 1 hour. The lysate was clarified by centrifugation at 48,000g for 30 minutes at 4C and bound to Ni-NTA resin (Qiagen) at 4C, then poured over a disposable column. The column was washed with 5 column volumes of wash buffer (20mM Tris pH 8, 0.3% Triton, 10% glycerol, 20mM imidazole, 250mM NaCl, 0.5mM TCEP, protease inhibitors). The protein was eluted in 5 column volumes of elution buffer (20mM Tris pH 8, 250mM imidazole, 250mM NaCl, 0.5mM TCEP, protease inhibitors). Positive fractions were pooled and dialyzed overnight into storage buffer (20mM Tris pH 8, 150mM NaCl, 0.5mM TCEP). Dialyzed protein was purified by size exclusion chromatography on a 120mL Superdex 200 column (GE Healthcare). Positive fractions were pooled, concentrated, flash frozen in liquid nitrogen and stored at -80C in small aliquots.

Overexpression in HEK293T cells and immunoprecipitation

Mouse Tet1 catalytic domain (aa1367-2039) and truncations and mutations thereof were cloned into the pcDNA3b vector. GFP fusion constructs were cloned into the pcDNA3.1 vector. Human OGT constructs were cloned into the pcDNA4 vector. Plasmids were transiently transfected into adherent HEK293T cells at 70-90% confluency using the Lipofectamine 2000 transfection reagent (ThermoFisher) for 1-3 days.

Transiently transfected HEK293T cells were harvested, pelleted, and lysed in IP lysis buffer (50mM Tris pH 8, 200mM NaCl, 1% NP40, 1x HALT protease/phosphatase inhibitors). For pulldown of FLAG-tagged constructs, cell lysate was bound to anti-FLAG M2 magnetic

beads (Sigma Aldrich) at 4C. For pulldown of GFP constructs, cell lysate was bound to magnetic protein G dynabeads (ThermoFisher) conjugated to the JL8 GFP monoclonal antibody (Clontech) at 4C. Beads were washed 3 times with IP wash buffer (50mM Tris pH 8, 200mM NaCl, 0.2% NP40, 1x HALT protease/phosphatase inhibitors). Bound proteins were eluted by boiling in SDS sample buffer.

In vitro transcription/translation and immunoprecipitation

GFP fused to TET C-terminus peptides were cloned into the pcDNA3.1 vector and transcribed and translated *in vitro* using the TNT Quick Coupled Transcription/Translation System (Promega).

For immunoprecipitation, recombinant His-tagged OGT was coupled to His-Tag isolation dynabeads (ThermoFisher). Beads were bound to *in vitro* translation extract diluted 1:1 in binding buffer (40mM Tris pH 8, 200mM NaCl, 40mM imidazole, 0.1% NP40) at 4C. Beads were washed 3 times with wash buffer (20mM Tris pH 8, 150mM NaCl, 20mM imidazole, 0.1% NP40). Bound proteins were eluted by boiling in SDS sample buffer.

Recombinant protein binding assay

20uL reactions containing 2.5uM rOGT and 2.5uM rTET1 CD wt or D2018A were assembled in binding buffer (50mM Tris pH 7.5, 100mM NaCl, 0.02% Tween-20) and pre-incubated at room temperature for 15 minutes. TET1 antibody (Millipore 09-872) was bound to magnetic Protein G Dynabeads (Invitrogen), and beads added to reactions following pre-incubation. Reactions were bound to beads for 10 minutes at room temperature. Beads were

washed 3 times with 100uL binding buffer, and bound proteins were recovered by boiling in SDS sample buffer and analyzed by SDS-PAGE and coomassie stain.

Western blots

For western blot, proteins were separated on a denaturing SDS-PAGE gel and transferred to PVDF membrane. Membranes were blocked in PBST + 5% nonfat dry milk at room temp for >10 minutes or at 4C overnight. Primary antibodies used for western blot were: FLAG M2 monoclonal antibody (Sigma Aldrich F1804), TET2 monoclonal antibody (Millipore MABE462), TET3 polyclonal antibody (Millipore ABE383), OGT polyclonal antibody (Santa Cruz sc32921), OGT monoclonal antibody (Cell Signaling D1D8Q), His6 monoclonal antibody (Thermo MA1-21315), JL8 GFP monoclonal antibody (Clontech), and *O*-GlcNAc RL2 monoclonal antibody (Abcam). Secondary antibodies used were goat anti-mouse HRP and goat anti-rabbit HRP from BioRad. Blots were incubated with Pico Chemiluminescent Substrate (ThermoFisher) and exposed to film in a dark room.

Slot blot

DNA samples were denatured in 400mM NaOH + 10mM EDTA by heating to 95C for 10 minutes. Samples were placed on ice and neutralized by addition of 1 volume of cold NH_4OAc pH 7.2. DNA was loaded onto a Hybond N+ nylon membrane (GE) by vacuum using a slot blot apparatus. The membrane was dried at 37C and DNA was covalently linked to the membrane by UV crosslinking (700uJ/cm^2 for 3 minutes). Antibody binding and signal detection were performed as outlined for western blotting using 5hmC monoclonal antibody (Active Motif 39791).

For the loading control, membranes were analyzed using the Biotin Chromogenic Detection Kit (Thermo Scientific) according to the protocol. Briefly, membranes were blocked, probed with streptavidin conjugated to alkaline phosphatase (AP), and incubated in the AP substrate BCIP-T (5-bromo-4-chloro-3-indolyl phosphate, p-toluidine salt). Cleavage of BCIP-T causes formation of a blue precipitate.

For quantification of slot blots, at least 3 biological replicates were used. Signal was normalized to the loading control and significance was determined using the unpaired t test.

Preparation of lambda DNA substrate

Linear genomic DNA from phage lambda (dam-, dcm-) containing 12bp 5' overhangs was purchased from Thermo Scientific. Biotinylation was performed by annealing and ligating a complementary biotinylated DNA oligo. Reactions containing 175ng/uL lambda DNA, 2uM biotinylated oligo, and 10mM ATP were assembled in 1x T4 DNA ligase buffer, heated to 65C, and cooled slowly to room temperature to anneal. 10uL T4 DNA ligase was added and ligation was performed overnight at room temperature. Biotinylated lambda DNA was purified by PEG precipitation. To a 500uL ligation reaction, 250uL of PEG8000 + 10mM $MgCl_2$ was added and reaction was incubated at 4C overnight with rotation. The next day DNA was pelleted by centrifugation at 14,000g at 4C for 5 minutes. Pellet was washed with 1mL of 75% ethanol and resuspended in TE.

Biotinylated lambda DNA was methylated using M.SssI CpG methyltransferase from NEB. 20uL reactions containing 500ng lambda DNA, 640uM S-adenosylmethionine, and 4 units methyltransferase were assembled in 1x NEBuffer 2 supplemented with 20mM Tris pH 8 and

incubated at 37C for 1 hour. Complete methylation was confirmed by digestion with the methylation-sensitive restriction enzyme BstUI from NEB.

In vitro TET1 CD *O*-GlcNAcylation

In vitro modification of rTET1 CD with rOGT was performed as follows: 10uL reactions containing 1uM rTET1 CD, 1-5uM rOGT, and 1mM UDP-GlcNAc were assembled in reaction buffer (50mM HEPES pH 6.8, 150mM NaCl, 10% glycerol, 0.5mM TCEP) and incubated at 37C for 30-60 minutes or at 4C for 18-24 hours.

In vitro TET1 CD activity assays

20uL reactions containing 100ng biotinylated, methylated lambda DNA, rTET1 CD (from frozen aliquots or from *in vitro O*-GlcNAcylation reactions), and TET cofactors (1mM alpha-ketoglutarate, 2mM ascorbic acid, 100uM ferrous ammonium sulfate) were assembled in reaction buffer (50mM HEPES pH 6.8, 100mM NaCl) and incubated at 37C for 10-60 minutes. Reactions were stopped by addition of 1 volume of 2M NaOH + 50mM EDTA and DNA was analyzed by slot blot.

Generation of mouse embryonic stem cell lines

mESC lines (Figure S2.1A, B) were derived using CRISPR-Cas9 genome editing. A guide RNA to the Tet1 3'UTR was cloned into the px459-Cas9-2A-Puro plasmid using published protocols(Ran et al., 2013) with minor modifications. Templates for homology directed repair were amplified from Gene Blocks (IDT) (Tables S1 and S2). Plasmid and template were co-transfected into LF2 mESCs using FuGENE HD (Promega) according to

manufacturer protocol. After two days cells were selected with puromycin for 48 hours, then allowed to grow in antibiotic-free media. Cells were monitored for green or red fluorescence (indicating homology directed repair) and fluorescent cells were isolated by FACS 1-2 weeks after transfection. All cell lines were propagated from single cells and correct insertion was confirmed by PCR genotyping (Figure S2.2B, C).

Genome-wide profiling of 5mC revealed a 25kb deletion in WT but not D2018A cells (Figure S2.2A). Analysis of cells with wild-type TET1 and one intact copy of this region shows that the differences in gene expression and 5mC levels are caused by the TET1 D2018A mutation rather than the 25kb deletion (Figure S2.2B, C).

Mass spectrometry

Genomic DNA (3ug) was subjected to hydrolysis with PDE I (3.6 U), PDE II (3.2 U), DNase I (50U), and alkaline phosphatase (10 U) in 10 mM Tris HCl/15 mM $MgCl_2$ buffer (pH 7) at 37°C overnight. The hydrolysates were spiked with $^{13}C_{10}{}^{15}N_2$-5-methyl-2′-deoxycytidine (1 pmol) and 5-hydroxymethyl-d_2-2′-deoxycytidine-6-d_1 (500 fmol) (internal standards for mass spectrometry) and filtered through Nanosep 10K Omega filters (Pall Corporation, Port Washington, NY).

Quantitation of mC and hmC was performed using a Dionex Ultimate 3000UHPLC (Thermo Fisher, Waltham MA) interfaced with a Thermo TSQ Vantage mass spectrometer (Thermo Fisher). Chromatographic separation was achieved on a Luna Omega Polar C18 column (150 x 1.0mm, 1.6µm, Phenomenex, Torrance CA) heated to 50℃ and eluted at a flow rate of 50 µL/min with a gradient of 0.1% acetic acid in H_2O (A) and acetonitrile (B). A linear gradient of 1% to 5% B in 5.7 min was used, followed by an increase to 20% B over 1.1 min and a further

increase to 50% B in 1.1 min. Solvent composition was a returned to initial conditions (1% B) and the column was re-equilibrated for 7 min. Under these conditions, mC and $^{13}C_{10}{}^{15}N_2$-MeC eluted at 3.7 min, both hmC and the internal standard D_3-hmC eluted at 2.9 min. Quantitation was achieved by monitoring the transitions *m/z* 258.2 [M + H^+] → *m/z* 142.1 [M – deoxyribose + H^+] for hmC, *m/z* 261.2 [M + H^+] → *m/z* 145.1 [M – deoxyribose + H^+] for D_3-hmC, *m/z* 242.1 [M + H^+] → *m/z* 126.1 [M + H^+] for mC, *m/z* 254.2 [M + H^+] → *m/z* 133.1 [M + H^+] for $^{13}C_{10}{}^{15}N_2$-mC. Optimal mass spectrometry conditions were determined by infusion of authentic standards. Typical settings on the mass spectrometer were: a spray voltage of 3500 V, a sheath gas of 12 units, the declustering voltage was 5 V, the RF lens was 55 V, the vaporizer temperature was 75 ℃, and the ion transfer tube was maintained at 350 ℃. The full-width at half-maximum (FWHM) was maintained at 0.7 for both Q1 and Q3. Fragmentation was induced using a collision gas of 1.5 mTorr and a collision energy of 10.3 V for mC and 10.6 V for hmC.

RNA-seq

RNA was extracted from mESCs (two biological replicates each for wild type and Tet1 D2018A, and two technical replicates for each biological replicate) using Zymo DirectZol RNA miniprep kit. 200ng of RNA per sample was used for Lexogen Ribocop rRNA depletion. Libraries were prepared from 8uL of Ribocop-treated RNA using Lexogen SENSE Total RNA-seq Library Prep Kit.

Libraries were sequenced on an Illumina Hiseq 4000 with single-end 50 base reads. Reads were aligned to the mouse genome (Ensembl build GRCm38.p6) and gene counts were created using STAR_2.5.3a. Normalization and differential expression analysis was performed with DESeq2 v1.20.0. Data was visualized using Matplotlib.

Methyl-seq library preparation and whole genome bisulfite sequencing

Genomic DNA was extracted from 1 million mESCs (two biological replicates each for wild type and Tet1 D2018A) using the DNeasy kit from Qiagen. Methyl-Seq libraries were prepared using Accel-NGS® Methyl-Seq DNA Library Kit with 30 ng gDNA for each sample and were sequenced on Illumina HiSeq 4000 with 150 bases pair-ended reads. For data processing, the raw reads were first trimmed to remove Illumina adapters and PCR duplicates and then mapped to mm10 mouse reference genome using Bismark (https://www.bioinformatics.babraham.ac.uk/projects/bismark/). The alignment files (BAM file) generated were analyzed by MethylPy (https://github.com/yupenghe/methylpy)(Lister et al., 2013) to get the methylation level at individual cytosines. Generally, the methylation level is defined as the ratio of the sum of methylated basecall counts over the total basecall counts at each individual pairwise cytosine on both strands. The significantly methylated cytosine sites (DMSs) were identified using a binomial test for each CpG context with FDR < 0.01 as described previously(Ma et al., 2014). Differentially methylated regions (DMRs) were defined by the DMRfind function in MethylPy by joining at least two DMSs within 250bp.

RT-qPCR

Total RNA was isolated from mESCs using Direct-zol RNA miniprep kit from Zymo. 1ug of RNA was used for cDNA synthesis using the iScript Reverse Transcription kit from BioRad. cDNA was used for qPCR using the SensiFAST SYBR Lo-Rox kit from Bioline. Relative gene expression levels were calculated using the $\Delta\Delta C_t$ method. See Table S3.

Zebrafish mRNA rescue experiments

Zebrafish husbandry was conducted under full animal use and care guidelines with approval by the Memorial Sloan-Kettering animal care and use committee. For mRNA rescue experiments, mTET1D2018A and mTET1wt plasmids were linearized by NotI digestion. Capped RNA was synthesized using mMessage mMachine (Ambion) with T7 RNA polymerase. RNA was injected into one-cell-stage embryos derived from $tet2^{mk17/mk17}$, $tet3^{mk18/+}$ intercrosses at the concentration of 100pg/embryo(C. Li et al., 2015). Injected embryos were raised under standard conditions at 28.5°C until 30 hours post-fertilization (hpf) at which point they were fixed for *in situ* hybridization using an antisense probe for *runx1*. The *runx1* probe is described in(Kalev-Zylinska et al., 2002)*; in situ* hybridization was performed using standard methods, and runx1 levels were scored across samples without knowledge of the associated experimental conditions(C. Thisse & Thisse, 2008). *tet2/3* double mutants were identified based on morphological criteria and mutants were confirmed by PCR genotyping after in situ hybridization using previously described primers(C. Li et al., 2015).

For sample size estimation for rescue experiments, we assume a background mean of 20% positive animals in control groups. We anticipate a significant change would result in at least a 30% difference between the experimental and control means with a standard deviation of no more than 10. Using the 1-Sample Z-test method, for a specified power of 95% the minimum sample size is 4. Typically, zebrafish crosses generate far more embryos than required. Experiments are conducted using all available embryos. The experiment is discarded if numbers for any sample are below this minimum threshold when embryos are genotyped at the end of the experimental period. Injections were separately performed on clutches from five independent

crosses; p values are based on these replicates and were derived from the unpaired two-tailed *t* test.

For the dot blot, genomic DNA was isolated from larvae at 30hpf by phenol-chloroform extraction and ethanol precipitation. Following RNase treatment and denaturation, 2-fold serially diluted DNA was spotted onto nitrocellulose membranes. Cross-linked membranes were incubated with 0.02% methylene blue to validate uniform DNA loading. Membranes were blocked with 5% BSA and incubated with anti-5hmC antibody (1:10,000; Active Motif) followed by a horseradish peroxidase-conjugated antibody (1:15,000; Active Motif). Signal was detected using the ECL Prime Detection Kit (GE). The results of three independent experiments were quantified using ImageJ at the lowest dilution and exposure where signal was observed in Tet1 injected embryos. To normalize across blots, all values are presented as the ratio of 5hmC signal in experimental animals divided by wildtype control signal from the same blot.

Results

A short C-terminal region of TET1 is necessary for binding to OGT

TET1 and OGT interact with each other and are mutually dependent for their localization to chromatin(Vella et al., 2013). To understand the role of this association, it is necessary to specifically disrupt the TET1-OGT interaction. All three TETs interact with OGT via their catalytic domains(Q. Chen et al., 2013; Deplus et al., 2013; R. Ito et al., 2014). We sought to identify the region within the TET1 catalytic domain (TET1 CD) responsible for binding to OGT. The TET1 CD consists of a cysteine-rich N-terminal region necessary for co-factor and substrate binding, a catalytic fold consisting of two lobes separated by a spacer of unknown

function, and a short C-terminal region also of unknown function (Figure 2.1A). We transiently transfected HEK293T cells with FLAG-tagged mouse TET1 CD constructs bearing deletions of each of these regions, some of which failed to express (Figure 2.1B). Because HEK293T cells have low levels of endogenous OGT, we also co-expressed His-tagged human OGT (identical to mouse at 1042 of 1046 residues). TET1 constructs were immunoprecipitated (IPed) using a FLAG antibody and analyzed for interaction with OGT. We found that deletion of only the 45 residue C-terminus of TET1 (hereafter C45) prevented detectable interaction with OGT (Figure 2.1B, TET1 CD del. 4). To exclude the possibility that this result is an artifact of OGT overexpression, we repeated the experiment overexpressing only TET1. TET1 CD, but not TET1 CD ΔC45, interacted with endogenous OGT, confirming that the C45 is necessary for this interaction (Figure 2.2).

OGT has two major domains: the N-terminus consists of 13.5 tetratricopeptide repeat (TPR) protein-protein interaction domains, and the C-terminus contains the bilobed catalytic domain (Figure 2.1C). We made internal deletions of several sets of TPRs to ask which are responsible for binding to the TET1 CD. We co-transfected HEK293T cells with FLAG-TET1 CD and His6-tagged OGT constructs and performed FLAG IP and western blot as above. We found that all the TPR deletions tested impaired the interaction with TET1 CD, with deletion of TPRs 7-9, 10-12, or 13-13.5 being most severe (Figure 2.1C). This result suggests that all of OGT's TPRs may be involved in binding to the TET1 CD, or that deletion of a set of TPRs disrupts the overall structure of the repeats in a way that disfavors binding.

Figure 2.1 - The short TET1 C-terminus is required for interaction with OGT
A) Domain architecture of TET1. B) Diagram of FLAG-tagged TET1 CD constructs expressed in HEK293T cells (upper). FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing FLAG-TET1 CD truncations and His-OGT (lower). C) Diagram of His-tagged OGT constructs expressed in HEK293T cells (upper). FLAG and His western blot of input and FLAG IPs from HEK293T cells transiently expressing FLAG-TET1 CD and His-OGT TPR deletions (lower).

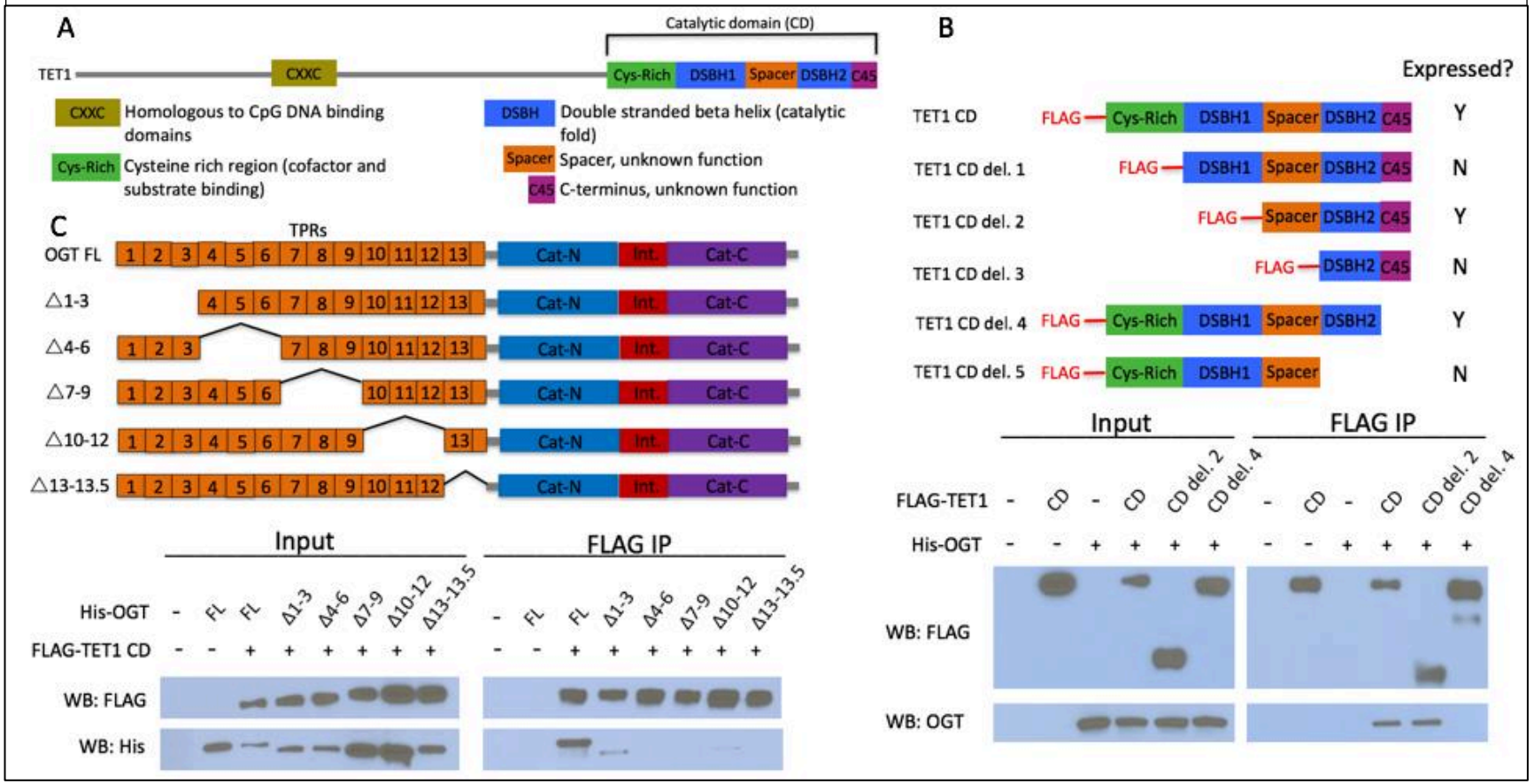
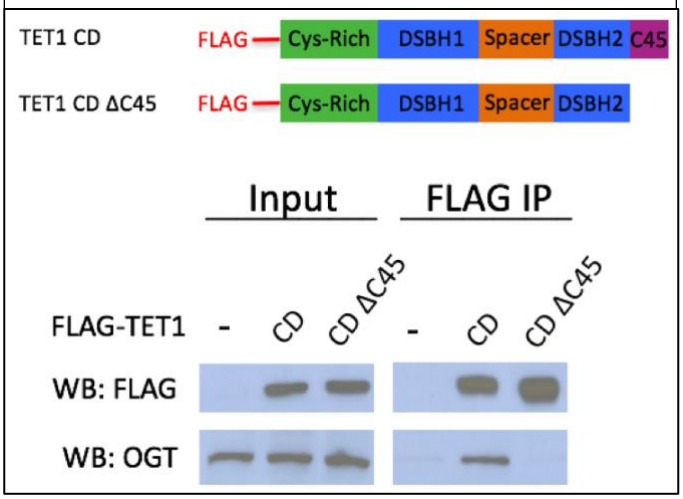


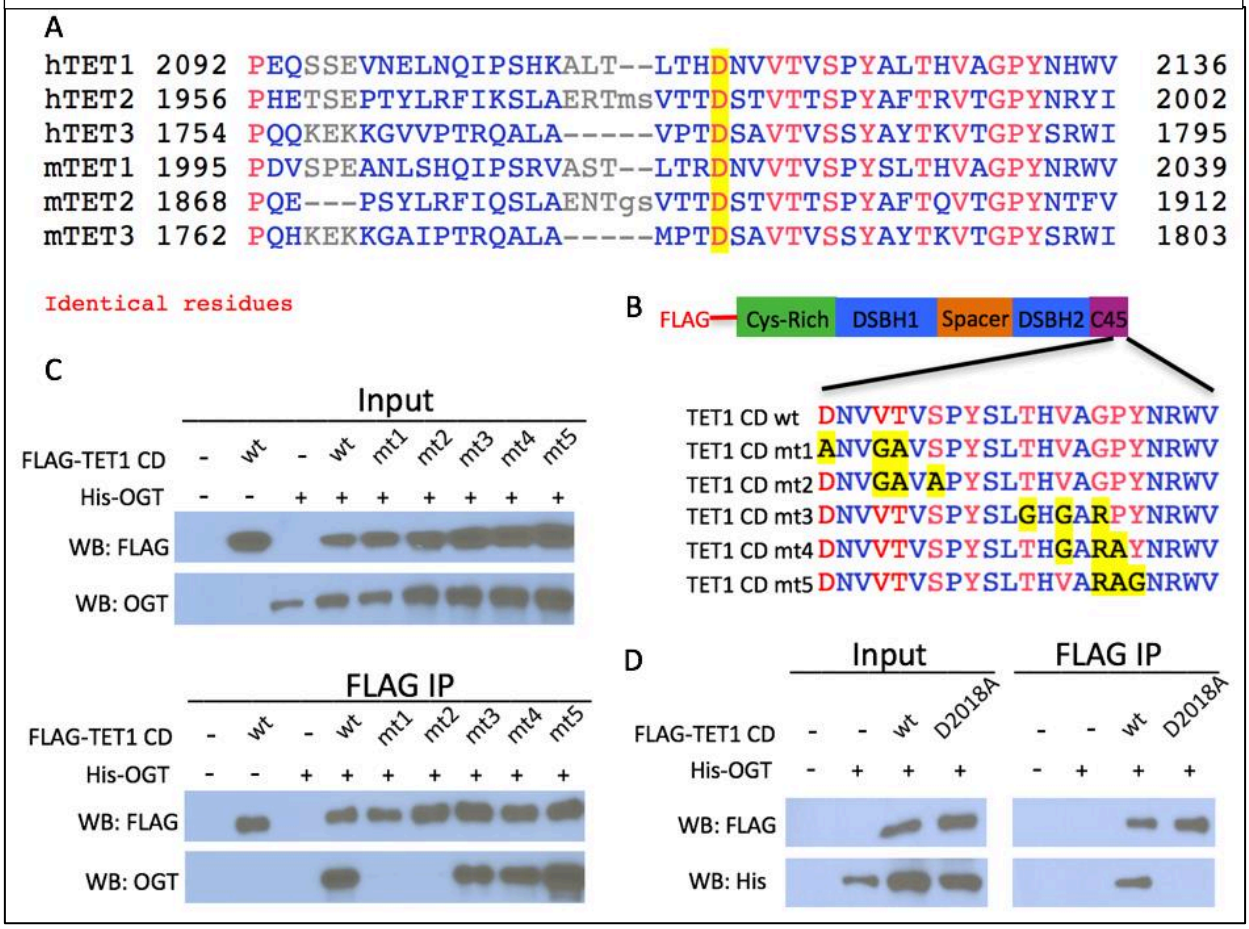
Figure 2.2 - TET1 C45 is necessary for interaction with endogenous OGT
FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing FLAG-TET1 CD or FLAG-TET1 CD ΔC45.



Conserved residues in the TET1 C45 are necessary for the TET1-OGT interaction

An alignment of the TET1 C45 region with the C-termini of TET2 and TET3 revealed several conserved residues (Figure 2.3A). We mutated clusters of three conserved residues in the TET1 C45 of FLAG-tagged TET1 CD (Figure 2.3B) and co-expressed these constructs with His-OGT in HEK293T cells. FLAG pulldowns revealed that two sets of point mutations disrupted the interaction with OGT: mutation of D2018, V2021, and T2022, or mutation of V2021, T2022, and S2024 (Figure 2.3C, mt1 and mt2). These results suggested that the residues between D2018 and S2024 are crucial for the interaction between TET1 and OGT. Further mutational analysis revealed that altering D2018 to A (D2018A) eliminated detectable interaction between FLAG-tagged TET1 CD and His-OGT (Figure 2.3D).

Figure 2.3 - Conserved residues in the TET1 C45 are necessary for the TET1-OGT interaction
A) Alignment of the C-termini of human (h) and mouse (m) TETs 1, 2, and 3. A conserved aspartate residue mutated in D is highlighted. B) Diagram of FLAG-tagged TET1 CD constructs expressed in HEK293T cells. C) FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing FLAG-TET1 CD triple point mutants and His-OGT. D) FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing His-OGT and FLAG-TET1 CD or FLAG-TET1 CD D2018A.

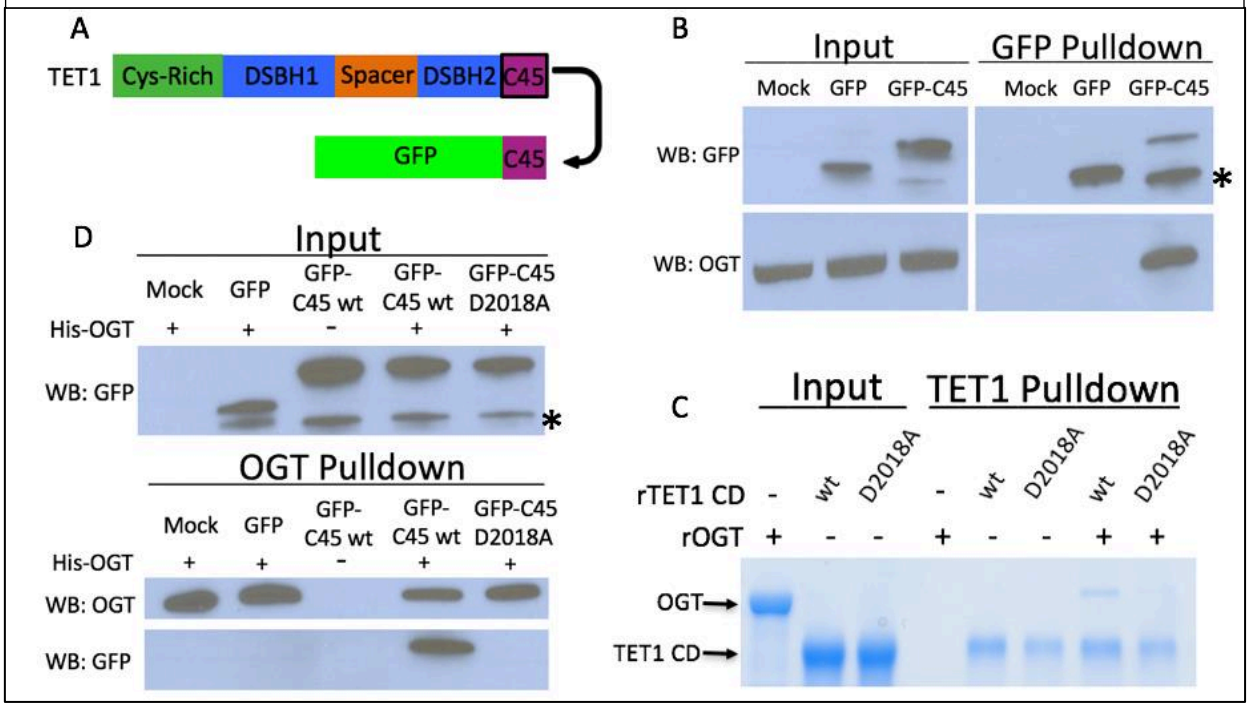


The TET1 C-terminus is sufficient for binding to OGT

Having shown that the TET1 C45 is necessary for the interaction with OGT, we next examined if it is also sufficient to bind OGT. We fused the TET1 C45 to the C-terminus of GFP (Figure 2.4A) and investigated its interaction with OGT. We transiently transfected GFP or GFP-C45 into HEK293T cells and pulled down with a GFP antibody. We found that GFP-C45, but not GFP alone, bound OGT (Figure 2.4B), indicating that the TET1 C45 is sufficient for interaction with OGT.

To determine if the interaction between TET1 CD and OGT is direct, we employed recombinant proteins in pulldown assays using beads conjugated to a TET1 antibody. We used recombinant human OGT (rOGT) isolated from *E. coli* and recombinant mouse TET1 catalytic domain (aa1367-2039), either wild type (rTET1 wt) or D2018A (rD2018A) purified from sf9 cells. rTET1 wt, but not beads alone, pulled down rOGT, indicating a direct interaction between these proteins (Figure 2.4C). rD2018A did not pull down rOGT, consistent with our mutational analysis in cells. Then we used an *in vitro* transcription/translation extract to produce GFP and GFP-C45, incubated each with rOGT, and found that the TET1 C45 is sufficient to confer binding to rOGT (Figure 2.4D). The D2018A mutation in the GFP-C45 was also sufficient to prevent rOGT binding (Figure 2.4D), consistent with the behavior of TET1 CD D2018A in cells. Together these results indicate that the TET1-OGT interaction is direct and mediated by the TET1 C45.

Figure 2.4 - The TET1 C45 is sufficient for interaction with OGT in cells and *in vitro*
A) Schematic of the TET1 C45 fusion to the C-terminus of GFP. B) GFP and OGT western blot of inputs and GFP IPs from HEK293T cells transiently expressing GFP or GFP-TET1 C45. *Truncated GFP. C) Coomassie stained protein gel of inputs and TET1 IPs from *in vitro* binding reactions containing rOGT and rTET1 CD wild type or D2018A. No UDP-GlcNAc was included in these reactions. D) GFP and OGT western blot of inputs and OGT IPs from *in vitro* binding reactions containing rOGT and *in vitro* translated GFP constructs. *Truncated GFP.

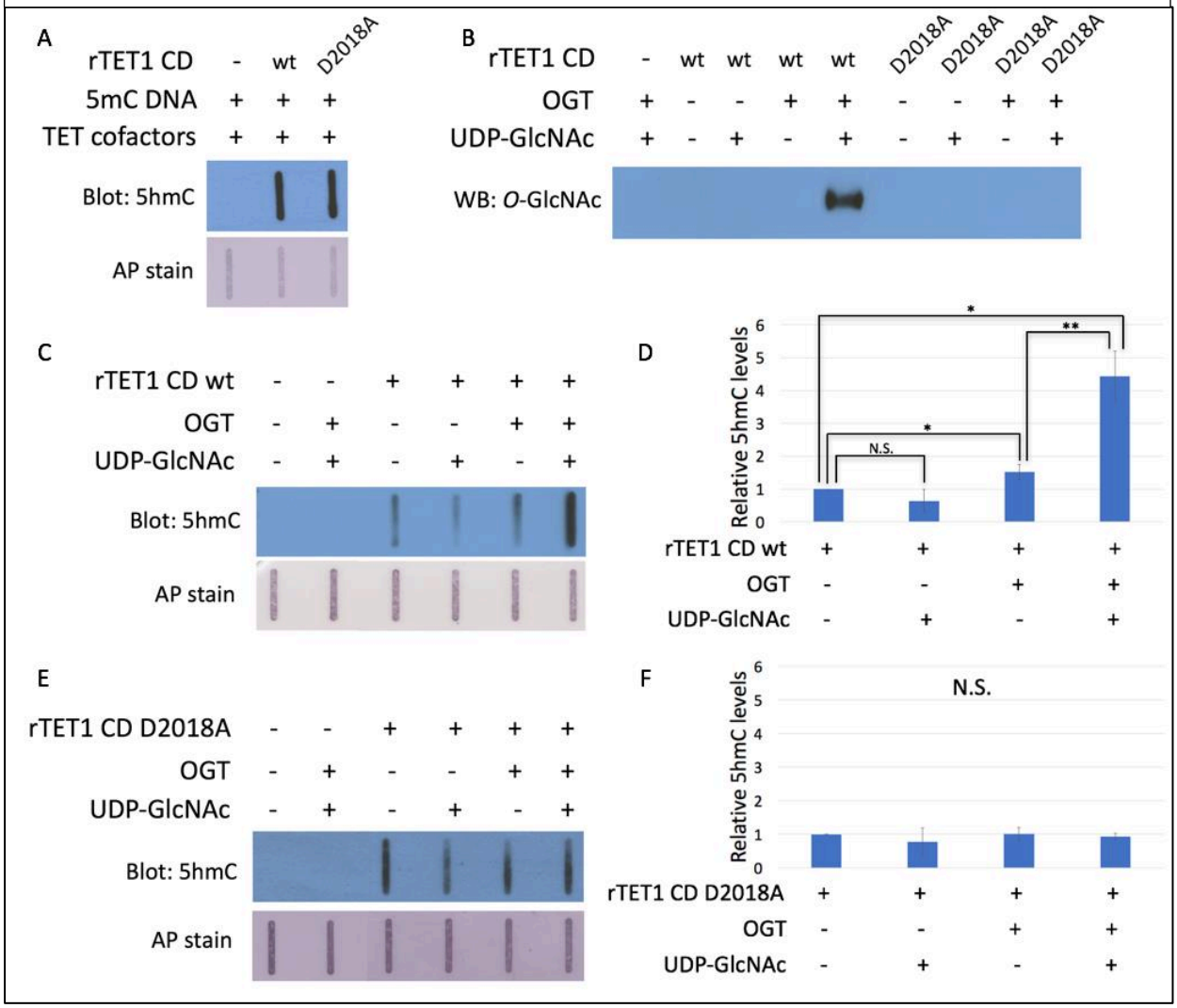


The D2018A mutation impairs TET1 CD stimulation by OGT

We employed the D2018A mutation to investigate the effects of perturbing the TET1-OGT interaction on rTET1 activity. rTET1 wt and rD2018A catalyzed formation of 5hmC on an *in vitro* methylated lambda DNA substrate (Figure 2.5A). Incubation with rOGT and OGT's cofactor UDP-GlcNAc resulted in *O*-GlcNAcylation of rTET1 wt but not rD2018A (Figure 2.5B).

To explore whether *O*-GlcNAcylation affects TET1 CD activity, we incubated rTET1 wt and rD2018A with UDP-GlcNAc and rOGT individually or together and assessed 5hmC production. Addition of UDP-GlcNAc did not significantly affect activity of rTET1 wt or rD2018A. Incubation with rOGT alone slightly enhanced 5hmC synthesis by rTET1 wt (1.3. - 1.7-fold), but not rD2018A. We observed robust stimulation of TET activity (4-5-fold) when rTET1 wt but not rD2018A was incubated with rOGT and UDP-GlcNAc (Figures 2.5C-F). These results suggest that while the TET1-OGT protein-protein interaction may slightly enhance TET1's activity, the *O*-GlcNAc modification is responsible for the majority of the observed stimulation.

Figure 2.5 - The D2018A mutation impairs TET1 CD stimulation by OGT
A) 5hmC slot blot of biotinylated 5mC containing lambda DNA from rTET1 CD activity assays. Alkaline phosphatase staining was used to detect biotin as a loading control. B) Western blot for *O*-GlcNAc in *in vitro O*-GlcNAcylation reactions. C) 5hmC slot blot of biotinylated 5mC containing lambda DNA from rTET1 wt activity assays. Alkaline phosphatase staining was used to detect biotin as a loading control. D) Quantification of 5hmC levels from rTET1 wt activity assays. Results are from 3-5 slot blots and normalized to rTET1 wt alone. E) 5hmC slot blot of biotinylated 5mC containing lambda DNA from rD2018A activity assays. Alkaline phosphatase staining was used to detect biotin as a loading control. F) Quantification of 5hmC levels from rD2018A activity assays. Results are from 3-5 slot blots and normalized to rD2018A alone. Error bars denote s.d. *P<0.01, **P<0.01, N.S. – not significant.



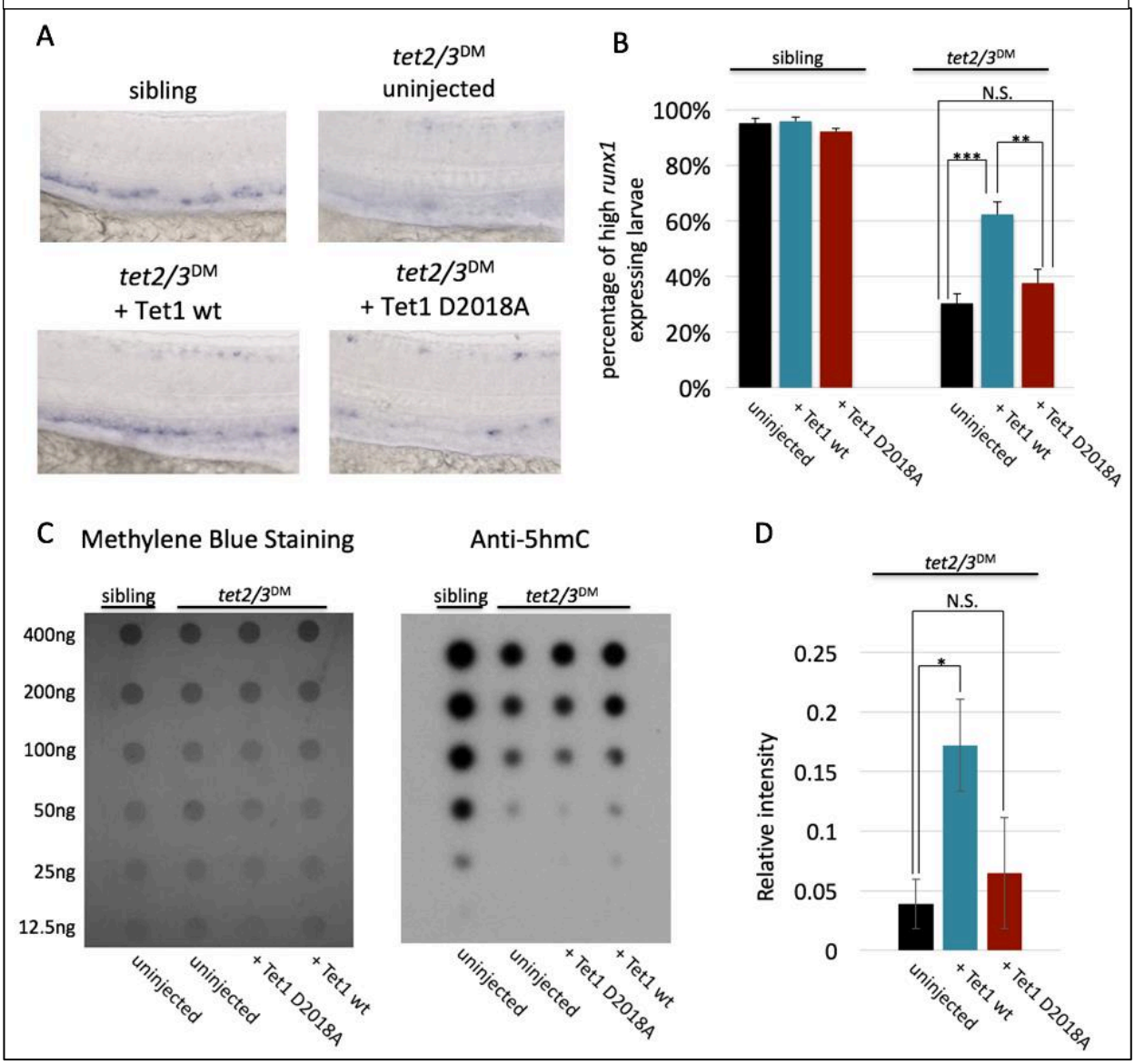
The TET-OGT interaction promotes TET1 function in the zebrafish embryo

We used zebrafish as a model system to ask whether the D2018A mutation affects TET function during development. Deletion analysis of *tet*s in zebrafish showed that Tet2 and Tet3 are the most important in development, while Tet1 contribution is relatively limited(C. Li et al., 2015). Deletion of both *tet2* and *tet3* (*tet2/3*^{DM}) causes a severe decrease in 5hmC levels accompanied by larval lethality owing to abnormalities including defects in hematopoietic stem cell (HSC) production. Reduced HSC production is visualized by reductions in the transcription factor *runx1*, which marks HSCs in the dorsal aorta of wild-type embryos, but is largely absent from this region in *tet2/3*^{DM} embryos. 5hmC levels and *runx1* expression are rescued by injection of human TET2 or TET3 mRNA into one-cell-stage embryos(C. Li et al., 2015).

Given strong sequence conservation among vertebrate TET/Tet proteins, we asked if over expression of mouse Tet1 mRNA could also rescue HSC production in *tet2/3*^{DM} zebrafish embryos and if this rescue is OGT interaction-dependent. To this end, *tet2/3*^{DM} embryos were injected with wild type or D2018A mutant encoding mouse Tet1 mRNA at the one cell stage. At 30 hours post fertilization (hpf) embryos were fixed and the presence of *runx1* positive HSCs in the dorsal aorta was assessed by *in situ* hybridization (Figure 2.6A). Tet1 wild type mRNA significantly increased the percentage of embryos with strong *runx1* labeling in the dorsal aorta (high *runx1*), while Tet1 D2018A mRNA failed to rescue *runx1* positive cells (Figures 2.6A-B). We also performed dot blots with genomic DNA from these embryos to measure levels of 5hmC (Figure 2.6C). On average, embryos injected with wild type Tet1 mRNA showed a modest but significant increase in 5hmC relative to uninjected *tet2/3*^{DM} embryos, while injection of TET1 D2018A mRNA did not show a significant increase (Figure 2.6D). These results suggest that the

TET1-OGT interaction promotes both TET1's catalytic activity and its ability to rescue *runx1* expression in this system.

Figure 2.6 - The TET1-OGT interaction promotes TET1 function in the zebrafish embryo
A) Representative images of *runx1* labeling in the dorsal aorta of wild type or *tet2/3*^{DM} zebrafish embryos, uninjected or injected with mRNA encoding mouse Tet1 wild type or D2018A. B) Percentage of embryos with high *runx1* expression along the dorsal aorta (*P<0.05, **P<0.01, ***P<0.001, N.S. – not significant). C) 5hmC dot blot of genomic DNA from wild type or *tet2/3*^{DM} zebrafish embryos injected with Tet1 wild type or D2018A mRNA. Methylene blue was used as a loading control. D) Quantification of 5hmC levels from 3 dot blots, normalized to methylene blue staining (*P<0.05, **P<0.01, ***P<0.001, N.S. – not significant).



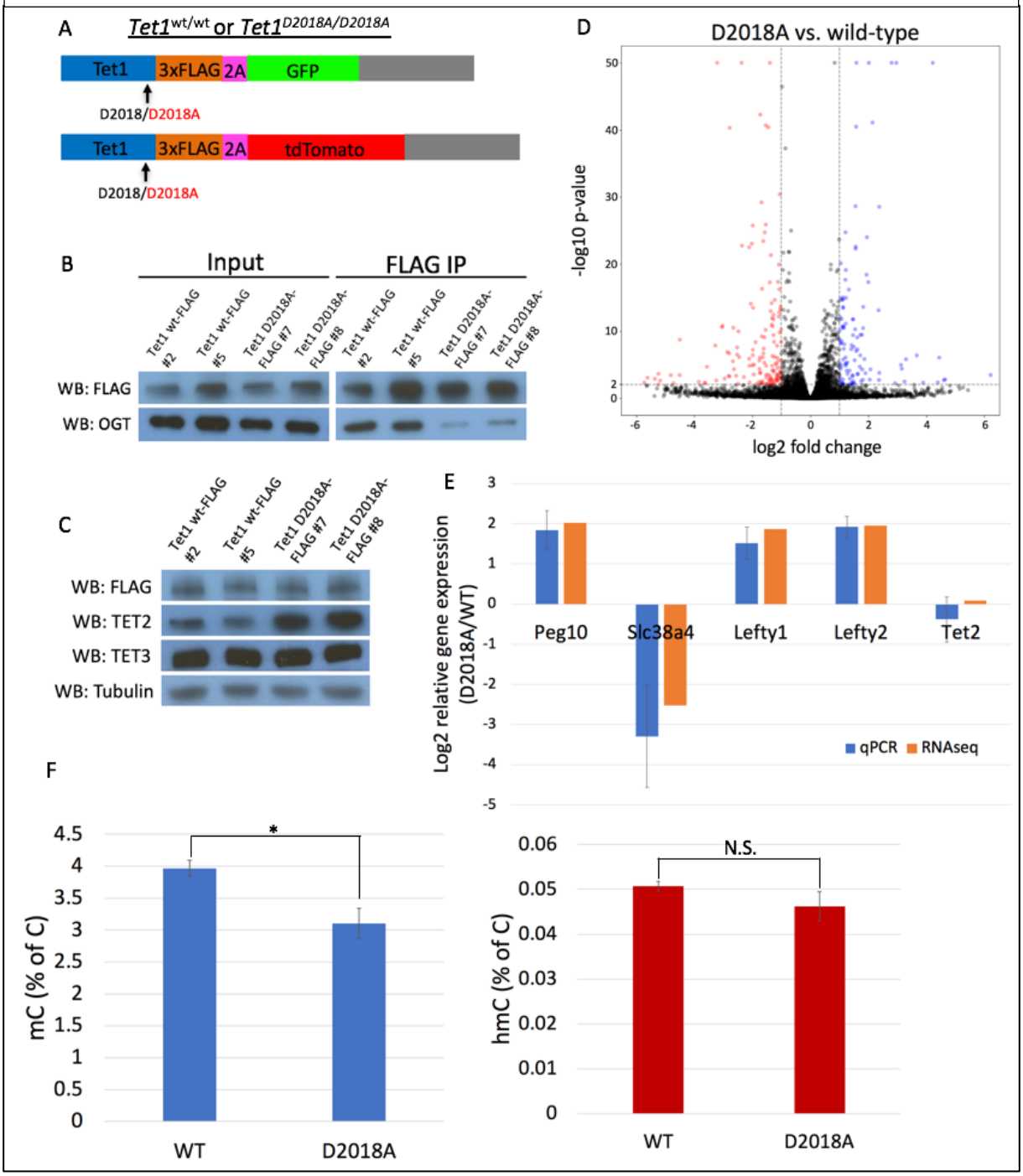
The D2018A mutation alters gene expression and 5mC levels in mESCs

Given the defect of TET1 D2018A in the zebrafish system, we decided to explore the effect of this mutation in mammalian cells. To this end, we generated a D2018A mutation in both copies of the *Tet1* gene in mESCs (Figure 2.7A and Figure S2.1). A FLAG tag was also introduced onto the C-terminus of wild type (WT) or D2018A mutant (D2018A) TET1. We first tested whether D2018 was necessary for the TET1-OGT interaction in the context of endogenous full length TET1 in these cells. FLAG pulldowns revealed that the D2018A mutation reduced, but did not eliminate, co-IP of OGT with TET1 (Figure 2.7B). Levels of TET2 protein were significantly increased in D2018A cells compared to WT (Figure 2.7C), suggesting the cells may be compensating for impaired TET1 function by producing more TET2.

To determine whether the TET1 D2018A mutation affected gene expression, we compared WT and D2018A mESCs using RNA-seq. We identified 378 genes whose expression changed by 2-fold or more in D2018A cells compared to WT (157 upregulated and 221 downregulated)(Figure 2.7D). In spite of the increased TET2 protein levels in D2018A cells, we did not observe increased abundance of *Tet2* transcripts (Figure 2.7E), nor was the stability of TET2 protein altered in D2018A cells compared to WT (Figure S2.3).

To examine how the TET1 D2018A mutation affected DNA modifications, we used LC-MS/MS to measure levels of 5mC and 5hmC in WT and D2018A mESCs. The total amount of 5mC was about 25% lower in D2018A cells compared to WT, while levels of 5hmC were not significantly different (Figure 2.7F).

Figure 2.7 – The D2018A mutation alters gene expression and 5mC levels in mESCs
A) Schematic of WT-FLAG and D2018A-FLAG mESC lines. B) FLAG and OGT western blot of inputs and FLAG IPs from WT-FLAG and D2018A-FLAG mESCs. C) Western blots for FLAG, TET2, and TET3 of protein extracts from WT-FLAG and D2018A-FLAG mESCs. D) Volcano plot showing differentially expressed genes in D2018A vs. WT mESCs. Red: decreased expression (log2 fold change > -1, Benjamini-Hochberg adjusted p-value < 0.01). Blue: increased expression (log2 fold change > 1, adjusted p-value < 0.01) E) qPCR analysis of selected differentially expressed genes. F) Mass spec quantification of mC and hmC levels in WT and D2018A cells (*P<0.05, N.S. – not significant).

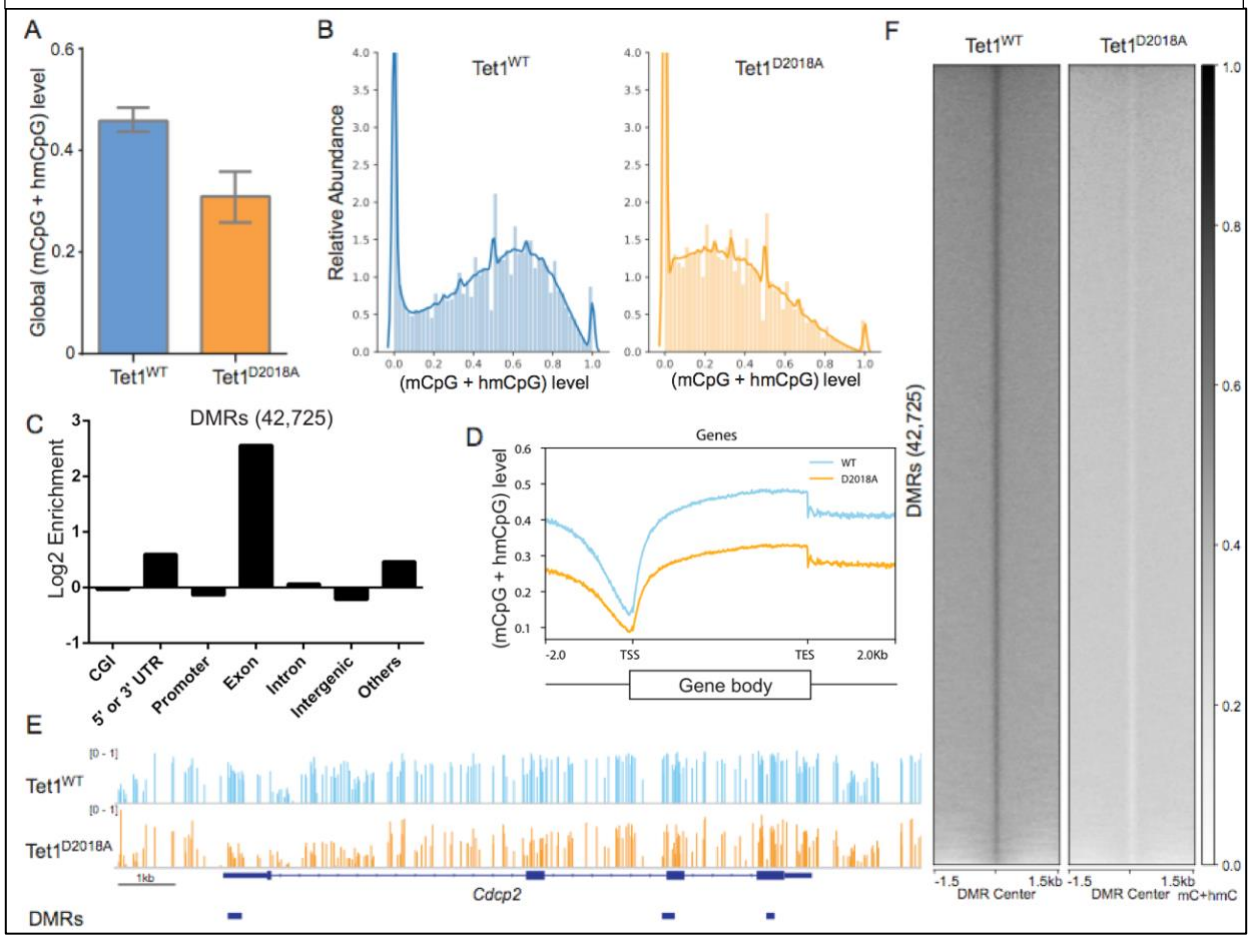


The D2018A mutation reduces 5mC levels without altering distribution

To determine if perturbing the TET1-OGT interaction affected distribution of CpG modifications, we performed whole genome bisulfite sequencing (WGBS-seq) and hmC-Seal on WT and D2018A mESCs. Consistent with our MS results, we observed reduced levels of 5mC + 5hmC in D2018A cells compared to WT (Figure 2.8A, B).

We identified 42,725 differentially methylated regions (DMRs), which were significantly enriched in exons over other genomic regions (Figure 2.8C). We found no evidence of a change in the distribution of CpG modification; rather, 5mC + 5hmC was reduced genome-wide in D2018A cells, with all DMRs showing loss of CpG modification in D2018A cells compared to WT (Figure 2.8D-F). We found no correlation between average methylation of cytosines at promoters (+/- 2kb window) and gene expression, suggesting that changes in CpG modifications do not directly underlie the altered gene expression caused by perturbing the TET1-OGT interaction.

Figure 2.8 – Whole genome bisulfite sequencing of WT and D2018A mESCs
A) Genome-wide levels of mCpG + hmCpG. B) Distribution of mC + hmC levels for individual CG sites fitting with kernel density estimate (KDE). C) Genomic annotations of hypo CG-DMRs in D2018A mESCs. CGI: CpG island. D) Distribution of averaged mCpG + hmCpG level at all genes. E) An example of hypo CG-DRMs in exons of *Cdcp2* gene in D2018A mESCs. F) Average mCpG + hmCpG level ((h)mCG/CG) of individual ranked DMRs and flanking regions (+/- 1.5kb).



Discussion

A unique OGT interaction domain?

We identified a 45-amino acid domain of TET1 that is both necessary and sufficient for binding of OGT. To our knowledge, this is the first time that a small protein domain has been identified that confers stable binding to OGT. The vast majority of OGT targets do not bind to OGT tightly enough to be detected in co-IP experiments, suggesting that OGT's interaction with TET proteins is unusually strong. For determination of the crystal structure of the human TET2 catalytic domain in complex with DNA, the corresponding C-terminal region was deleted(Hu et al., 2013), suggesting that it may be unstructured. When bound to OGT this domain may become structured, and structural studies of OGT bound to C45 could shed light on what features make this domain uniquely able to interact stably with OGT and how OGT may stimulate TET1 activity.

An alternative or additional role for the stable TET-OGT interaction may be recruitment of OGT to chromatin by TET proteins. Loss of TET1 causes loss of OGT from chromatin(Vella et al., 2013) and induces similar changes in transcription in both wild-type mESCs and mESCs lacking DNA methylation(Williams et al., 2011). This raises the possibility that TET proteins may recruit OGT to chromatin to regulate gene expression independent of 5mC oxidation. Consistent with this possibility, OGT modifies many transcription factors and chromatin regulators in mESCs(Myers, Panning, & Burlingame, 2011a)(Figure 2.9). Thus it may be that the stable TET1-OGT interaction promotes both regulation of TET1 activity by *O*-GlcNAcylation as well as recruitment of OGT to chromatin. Notably, our results show that TET1 D2018A does not rescue 5hmC levels in *tet2/3*^{DM} zebrafish embryos to the same extent as the wild type protein,

suggesting that at least part of the role of the TET1-OGT interaction *in vivo* is regulation of TET1 activity.

OGT stimulation of TET activity

Our results show for the first time that OGT can modify a TET protein *in vitro*, and that *O*-GlcNAcylation stimulates the activity of a TET protein *in vitro*. We have identified 8 sites of *O*-GlcNAcylation within the TET1 CD (data not shown), which precludes a simple analysis of which sites are important for stimulation. Detailed studies of individual sites of modification will be required to resolve this question.

Our data are consistent with a role for OGT in TET1 regulation in cells and *in vivo*. OGT also directly interacts with TET2 and TET3, suggesting that it may regulate all three TET proteins. Notably, although all three TETs catalyze the same reaction, they show a number of differences that are likely to determine their biological role. Different TET proteins are expressed in different cell types and at different stages of development(Dawlaty et al., 2011; Koh et al., 2011; Z. Li et al., 2011; Zhao et al., 2015). TET1 and TET2 appear to target different genomic regions(Huang et al., 2014) and to promote different pluripotent states in mESCs(Fidalgo et al., 2016). The mechanisms responsible for these differences are not well understood. We suggest that OGT is a strong candidate for regulation of TET enzymes.

Regulation of TETs by OGT in development

Our result that wild type TET1 mRNA, but not TET1 mRNA carrying a mutation that can impair interaction with OGT, can rescue *tet2/3*^{DM} zebrafish suggests that OGT regulation of TET enzymes may play a role in development. The importance of both TET proteins and OGT in

development has been thoroughly established. Zebrafish lacking *tet2* and *tet3* die as larvae(C. Li et al., 2015), and knockout of *Tet* genes in mice yields developmental phenotypes of varying severities, with knockout of all three *Tet*s together being embryonic lethal(Dawlaty et al., 2011; 2014; 2013; Z. Li et al., 2011). Similarly, OGT is absolutely essential for development in mice(Shafi et al., 2000) and zebrafish(Webster et al., 2009), though its vast number of targets have made it difficult to narrow down more specifically why OGT is necessary. Our results suggest that TETs are important OGT targets in development.

The TET1-OGT interaction regulates gene expression and DNA methylation in mESCs

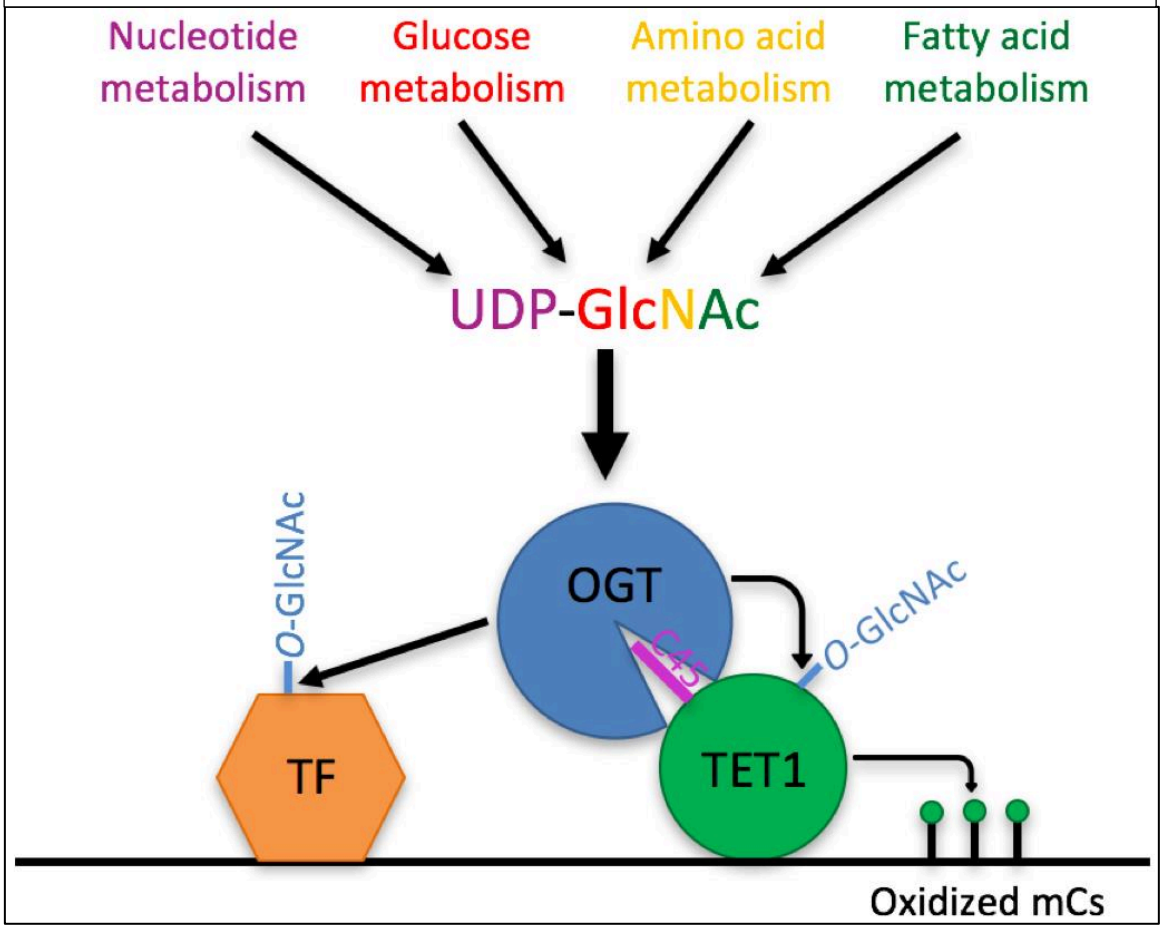
The D2018A mutation reduced the TET1-OGT interaction in mESCs and altered gene expression and reduced bulk 5mC levels with no significant changes in 5hmC levels. Bisulfite sequencing showed a genome-wide reduction in 5mC+5hmC, while mass spectrometry showed no significant change in 5hmC. Together these results suggest that the bisulfite sequencing largely reflects changes in 5mC. We did not find any correlations between differentially expressed genes and DMRs, which may be attributable to indirect effects of disrupting the TET1-OGT interaction. In support of this, the D2018A mESCs expressed more TET2 than WT TET1 mESCs, suggesting that TET1 regulates the other TETs in mESCs. TET1 and TET2 regulate different genomic regions in mESCs(Huang et al., 2014), so increased TET2 activity could result in epigenetic changes at regions not regulated by TET1. In support of this, an enrichment of DMRs was observed in gene bodies, which are targets of TET2, not TET1(Huang et al., 2014). Increased levels of TET2 could increase 5mC oxidation at gene bodies leading to demethylation, which would explain the loss of genic 5mC in D2018A cells compared to WT.

Compensation by TET2 for loss of TET1 activity could also explain why bulk levels of 5hmC were unchanged by the D2018A mutation.

A connection between metabolism and the epigenome

OGT has been proposed to act as a metabolic sensor because its cofactor, UDP-GlcNAc, is synthesized via the hexosamine biosynthetic pathway (HBP), which is fed by pathways metabolizing glucose, amino acids, fatty acids, and nucleotides(Hart et al., 2011). UDP-GlcNAc levels change in response to flux through these pathways(Marshall, Nadeau, & Yamasaki, 2004; McClain, 2002; Weigert et al., 2003), leading to the hypothesis that OGT activity may vary in response to the nutrient status of the cell. Thus the enhancement of TET1 activity by OGT and the significant overlap of the two enzymes on chromatin(Vella et al., 2013) suggest a model in which OGT may regulate the epigenome in response to nutrient status by controlling TET1 activity (Figure 2.9).

Figure 2.9: Model
Model showing two roles of the TET1-OGT interaction in regulation of gene expression. OGT's activity is regulated by the abundance of its cofactor UDP-GlcNAc, whose synthesis has inputs from nucleotide, glucose, amino acid, and fatty acid metabolism. OGT (blue circle) binds to TET1 (large green circle) via the TET1 C45 (purple line). OGT modifies TET1 and regulates its catalytic activity (small green circles representing modified cytosines). At the same time, TET1 binding to DNA brings OGT into proximity of other DNA-bound transcription factors (orange hexagon), which OGT also modifies and regulates.



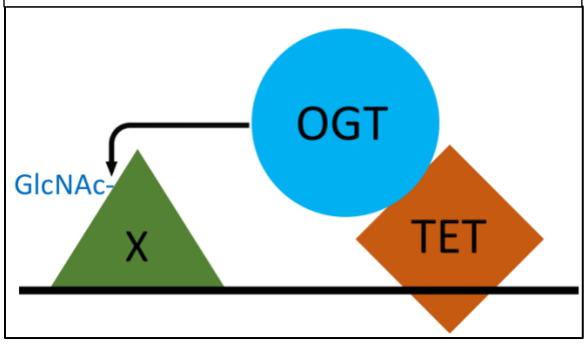
Chapter 3

The contribution of TET1 and TET2 to nuclear protein *O*-GlcNAcylation in mESCs

Introduction

Collectively, the stable interaction between OGT and TET proteins, the modification of TET proteins by OGT, and the overlap of OGT and TET proteins on chromatin suggest at least two types of models: 1. TET proteins may affect the activity of OGT, and/or 2. OGT may affect the activity of the TET proteins. This chapter details experiments conducted to test a model of the first type. OGT is known to modify many chromatin-associated proteins, including numerous transcription factors(Myers, Panning, & Burlingame, 2011b; Ruan et al., 2013; Sakabe et al., 2010), but OGT alone has not been observed to bind DNA. Thus, one model is that the TET proteins, which do stably bind DNA, recruit OGT to chromatin so that it may modify chromatin-associated targets (Figure 3.1).

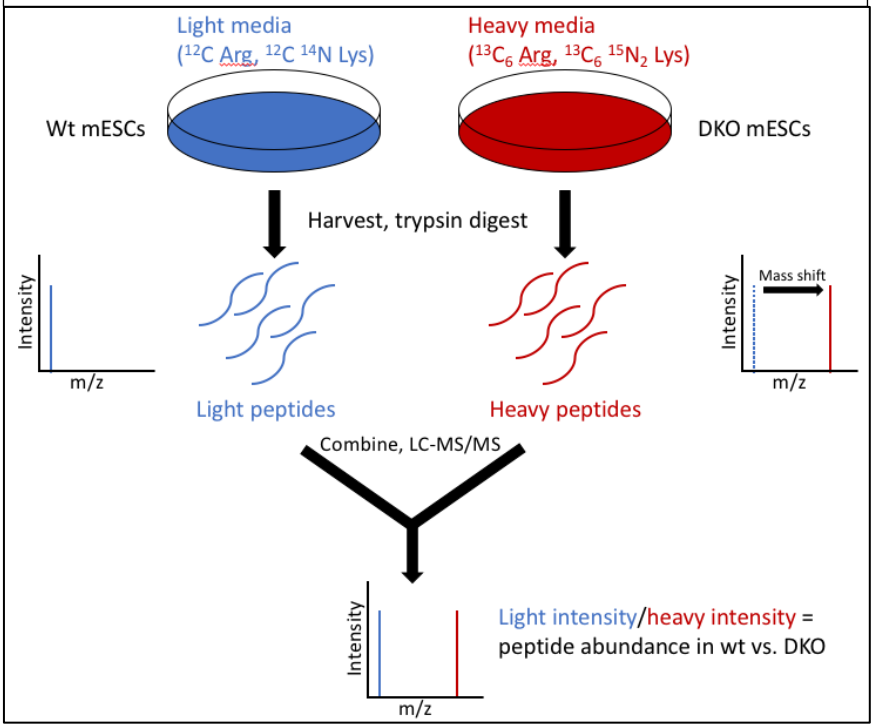
Figure 3.1 – Model for regulation of OGT by TETs
In this model TET proteins recruit OGT to DNA, permitting OGT to modify its numerous chromatin-associated targets.



One prediction of this model is that if the interaction between OGT and TETs is broken, patterns of *O*-GlcNAcylation of OGT's chromatin targets would be significantly altered. To test this hypothesis, we obtained male mESCs in which *Tet1* and *Tet2* are both knocked out (DKO cells), along with a matching wild-type cell line (wt cells){Dawlaty:2013em}. A *Tet1/Tet2/Tet3*

triple knockout cell line was not available when these experiments began, and Tet3 expression is negligible in male mESCs. Thus, in the DKO cells the recruitment of OGT to chromatin by TETs should be severely impaired. We used stable isotope labeling by amino acids in culture (SILAC){Ong:2012hf} to compare patterns of nuclear protein *O*-GlcNAcylation in wt cells versus DKO cells (Figure 3.2). In this method, one cell line is grown in normal media (light) while the second is grown in media containing heavy isotopes of lysine and arginine (heavy). After sufficient passaging proteins in heavy media are >99% labeled with heavy lysine and arginine. Equal amounts of nuclear protein from each cell line are mixed, digested with trypsin, enriched for *O*-GlcNAc modified peptides, and analyzed by mass spectrometry. A given peptide labeled with heavy lysine/arginine will have a higher m/z ratio than the same unlabeled peptide. The ratio between the intensities of the light peptide and the heavy peptide gives the ratio of the abundances of that peptide in the two cell types.

Figure 3.2 – Overview of SILAC method



Materials and Methods

Cell culture

Wt and DKO mESCs derived by the Jaenisch lab{Dawlaty:2013em} were adapted to feeder-free conditions and cultured in 2i media (DMEM/F12, 0.5x N2 supplement, 0.5x B27 supplement, 0.5mg/mL BSA, 1uM MEK inhibitor (Stemgent 04-0006), 3uM GSK3B inhibitor (1-Azakenpaullone), 0.15mM monothioglycerol, 2mM glutamine, and human leukemia inhibitory factor (LIF)). For SILAC, cells were cultured in 2i media prepared with SILAC DMEM/F12 (lacking lysine and arginine) and supplemented with 28ug/mL light or heavy arginine (Fisher 89989 and 88210, respectively), 49ug/mL light or heavy lysine (Fisher 89987 and 88209, respectively), and 200ug/mL light proline. After >5 passages in heavy media, heavy isotope incorporation was confirmed and cells were expanded for mass spectrometry.

Nuclei preparation

For isolation of nuclei, mESCs were lysed in buffer 1 (10mM Tris pH 8, 320mM sucrose, 3mM $CaCl_2$, 2mM MgOAc, 0.1mM EDTA, 0.1% Triton X-100, 1mM DTT, and protease inhibitors). Lysed cells were diluted in two volumes of buffer 2 (10mM Tris pH 8, 2M sucrose, 5mM MgOAc, 0.1mM EDTA, 5mM DTT, and protease inhibitors), then layered over buffer 2 in a centrifuge tube. Nuclei were pelleted by centrifugation at 37,000rpm at 4C for 45 minutes and recovered.

Lectin weak affinity chromatography

Nuclear pellets were solubilized in 8M urea, 50mM ammonium bicarbonate pH 8.0, 4x Phosphatase Inhibitor Cocktails I and III (Sigma), and 20μM PUGNAc. The mixture was

reduced for 30 min at 57°C with 5mM dithiothreitol and subsequently carbamidomethylated using 10mM iodoacetamide for 30 min at room temperature in the dark. Lysates were diluted to 2M urea with 50mM ammonium bicarbonate, pH 8.0, and digested overnight at 37°C with sequencing-grade trypsin (Promega) at an enzyme-to-substrate ratio of 1:50 (w/w). After digestion, samples were acidified with formic acid (Sigma-Aldrich) and desalted using a 360mg C18 Sep-Pak SPE cartridge (Waters). Desalted samples were dried to completeness using a SpeedVac concentrator (Thermo Electron). Desalted tryptic peptides were resuspended in 500 µl of LWAC buffer (100mM Tris, pH 7.5, 150mM NaCl, 10mM $MgCl_2$, 10mM $CaCl_2$, and 5% acetonitrile). Glycopeptides were enriched with a POROS-WGA column, collected, and desalted inline using a Luna 10µ C18 column (Phenomenex). A total of three rounds of LWAC enrichment were performed. LWAC-enriched glycopeptides were subsequently fractionated by high-pH reverse-phase liquid chromatography.

Mass spectrometry analysis

All samples were analyzed on an Orbitrap Velos (Thermo Fisher Scientific) equipped with a nano-Acquity UPLC (Waters). Peptides were fractionated on a 15cm × 75µm ID 3µm C18 EASY-Spray column using a linear gradient from 2 to 35% solvent B over 60 min. Survey mass measurements were performed using the Orbitrap, scanning from *m*/*z* 350–2000. The three most abundant multiply charged ions were computer selected for higher energy collisional dissociation (HCD) and electron transfer dissociation (ETD) analysis. The trigger intensity was set to 2000. Supplemental activation was enabled. The ETD fragments were measured in the linear trap, whereas HCD fragments were measured in the Orbitrap. Each sample was injected twice; the first analysis selected only 2+ precursor ions, and in the second analysis, 2+ precursor ions were

excluded. Peaklists were extracted using Proteome Discoverer 1.4. ETD data were searched twice against the UniProt Mus musculus database (downloaded June 17, 2013) (and concatenated with a randomized sequence for each entry) using Protein Prospector (version 5.10.15). Cleavage specificity was set as tryptic, allowing for two missed cleavages. Carbamidomethylation of Cys was set as a constant modification. The required mass accuracy was 20 ppm for precursor ions and 0.8 Da for ETD fragments. Both searches included $^{13}C(6)$ labelled Arg and $^{13}C(6)^{15}N(2)$ labelled Lys; acetylation of protein N termini; oxidation of Met; cyclization of N-terminal Gln; and HexNAc modification of Ser, Thr, and Asn, as variable modifications. The first search also allowed for the following extended N-linked glycans on Asn: HexNAc2, HexNAc2Hex2, HexNAc2Hex2Fuc, HexNAc2Hex2Fuc, HexNAc2Hex3, HexNAc2Hex3Fuc, HexNAc2Hex4, HexNAc2Hex4Fuc, HexNAc2Hex5, HexNAc2Hex5Fuc, HexNAc2Hex6, HexNAc2Hex7, HexNAc2Hex8, HexNAc2Hex9, HexNAc3Hex3Fuc, HexNAc3Hex5Fuc, HexNAc3Hex5FucSA, HexNAc4Hex3Fuc, HexNAc4Hex4Fuc2, HexNAc4Hex4SA, HexNAc4Hex5Fuc, HexNAc4Hex5Fuc2, HexNAc4Hex6Fuc, HexNAc5Hex3Fuc, HexNAc5Hex4Fuc, HexNAc5Hex4Fuc2, HexNAc5Hex4SA, HexNAcFuc. The second search allowed for the following extended O-linked glycans on Ser and Thr: HexNAc2, HexNAc2Hex2, HexNAcFuc, HexNAcHex, HexNAcHexSA, HexNAcHexSA2, HexNAcSA, as well as HexNAc and HexNAc2 on Asn. Three modifications per peptide were permitted. HCD data were searched with the same parameters except that fragment ion mass accuracy was 30 ppm and the only glycans set as variable modifications were HexNAc on Asn, Ser, and Thr and as a neutral loss. Modified peptides were identified with a protein and peptide false discovery rate of 1%. The $\log_2$ fold change between WT and KO O-GlcNAc peptide MS1

peak areas were median normalized and used to detect site specific O-GlcNAc differences between the samples.

Results/Discussion

928 unique *O*-GlcNAcylated peptides were identified, representing 307 unique sites of modification on 286 proteins. Applying a 2-fold change cutoff, 91 *O*-GlcNAcylated peptides were changed in abundance in wt cells compared to DKO, representing 53 proteins. 42 of these proteins had only 1 changed *O*-GlcNAc peptide, while the remaining 11 proteins had 2-7 changed *O*-GlcNAc peptides each (Table 3.1).

Table 3.1 – proteins with more than one altered *O*-GlcNAc peptide	
Protein	**Number of unique *O*-GlcNAc peptides changed in wt vs. DKO**
SIN3A	7
HCFC1	7
MLLT10	7
PROSER1	7
QSER1	5
SAP130	4
KLF4	3
VIM	3
EP400	2
NUP214	2
YEATS2	2

Peptides less *O*-GlcNAcylated in DKO
Peptides more *O*-GlcNAcylated in DKO
Both less and more *O*-GlcNAcylated peptides

We focused analysis on the top few proteins in Table 3.1, which represent targets whose *O*-GlcNAcylation is most drastically affected by the loss of *Tet1* and *Tet2*.

SIN3A, SAP130, HCF1

SIN3A, a known interaction partner of OGT, is a scaffold protein that takes part in numerous multi-protein complexes that both activate and repress transcription. SAP130 (SIN3-Associated Polypeptide, 130kDa) is a component of one such repressive complex containing HDAC1, a histone deacetylase. HCF1 is another transcriptional coregulator that interacts with both OGT and SIN3A and regulates cell cycle progression. *O*-GlcNAcylation of HCF1 by OGT is necessary for proteolytic maturation of HCF1, a process essential for its function{Capotosti:2011kd}. Both SIN3A and HCF1 can also co-IP TET1. Numerous co-IP experiments and size exclusion chromatography suggest that all these proteins may exist together in large (>670kDa) multiprotein complexes{Vella:2013ez}. The exact composition and architecture of these complexes is unknown, but if TET1 is central to their formation then it is reasonable that its absence would alter OGT's ability to modify other components like SIN3A, SAP130, and HCF1. However, this model is complicated by the fact that loss of TETs has different effects on different proteins (Table 3.1) – SIN3A is less *O*-GlcNAcylated in the absence of TETs, HCF1 is more *O*-GlcNAcylated, and SAP130 has some more and some less modified peptides. The complex crosstalk between TETs, OGT, and these OGT targets will require further experiments to clarify.

PROSER1 and QSER1

Little is known about PROSER1 (proline/serine rich 1) or QSER1 (glutamine/serine rich 1). Neither has any functional characterization or predicted protein domains. However, the few protein-protein interactions that have been described for PROSER and QSER make them interesting in the context of OGT and TETs.

PROSER1 is an interaction partner of KDM6A/UTX, an OGT target and H3K27me2/3 demethylase implicated in the activation of HOX gene transcription during development{Agger:2007hi}. KDM6A/UTX also interacts with the protein complex responsible for H3K4 methylation at these HOX genes.

QSER1 was identified as an interaction partner of multiple components of a recently characterized EMSY/KDM5A/SIN3B complex that also regulates gene expression{Varier:2016je}. This complex is also predicted to contain SIN3A and HDACs, potentially linking it to the SIN3A/OGT/TET1 complex(es) discussed above.

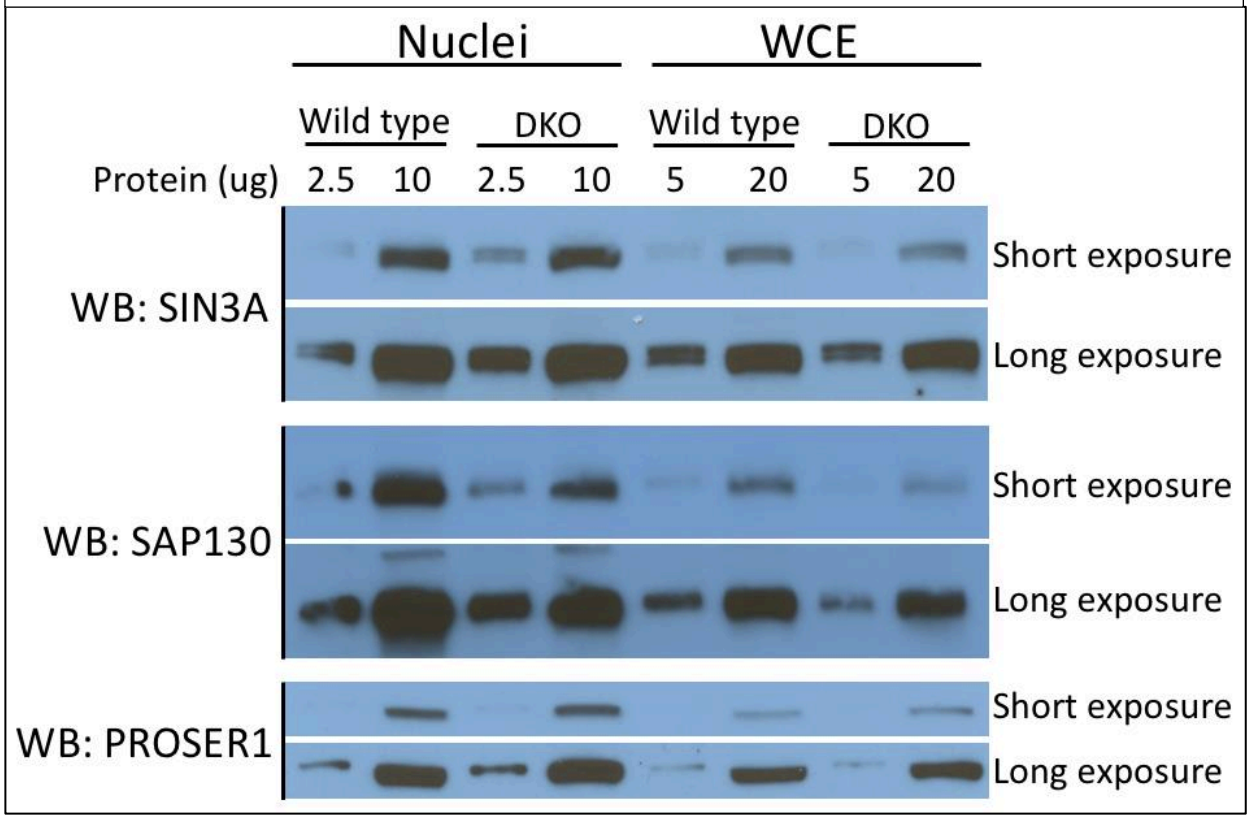
Thus, although nothing definitive can be said about the biological significance of *O*-GlcNAcylation of PROSER and QSER, their interactions with transcriptional regulatory protein complexes, which contain other OGT targets and OGT/TET1 interacting proteins, suggests that they may have significance in the context of gene expression regulation by OGT and TETs.

Followup

A difference in abundance of an *O*-GlcNAcylated peptide between two cell types could have two non-mutually exclusive causes: 1. a difference in the activity of OGT toward the protein and/or 2. a difference in total expression of the protein. Thus, the levels of proteins of interest in wt and DKO cells need to be compared to assess the relative contributions of these two factors to the differences in *O*-GlcNAcylation. To begin this process, the levels of SIN3A, SAP130, and PROSER1 in wt and DKO nuclei and whole cell extracts (WCE) were compared by western blotting (Figure 3.3). The data show that both SIN3A and PROSER1 are expressed comparably in the two cell types, both in nuclei and WCE, suggesting that the loss of *O*-GlcNAcylation of these proteins in DKO cells compared to wt can be attributed to reduced OGT

activity. SAP130, in contrast, appears to be at least 2-fold less expressed in DKO cells than wt. Of the four *O*-GlcNAcylated SAP130 peptides identified, three are 2-3-fold less abundant in DKO cells, while one is 2-fold more abundant. Thus the three less abundant *O*-GlcNAcylated peptides may simply reflect the decrease in total protein, while the peptide with more *O*-GlcNAcylation in DKO cells suggests that OGT's activity toward this region of SAP130 could be increased in the absence of TETs.

Figure 3.3 – Western blot analysis of proteins of interest in wt and DKO cells



Chapter 4

Identification of *O*-GlcNAc sites on TET1 and TET2 *in vitro* and in mESCs

Introduction

The result that *O*-GlcNAcylation of mTET1 CD stimulates its activity *in vitro* raises the question of where the site(s) of modification are and, if there are multiple sites, what the effects are of individual sites of modification. Previous experiments have identified numerous *O*-GlcNAcylation sites on all three human TET proteins in cells(Bauer et al., 2015), while only one site has been identified on mouse TET1 from mESCs(Myers, Panning, & Burlingame, 2011a). The SILAC experiment described in chapter 3 identified numerous sites of *O*-GlcNAcylation, as well as *O*-GlcNAcylated peptides with ambiguous site assignments, on both TET1 and TET2 from mESCs. As a complementary approach, we analyzed *in vitro O*-GlcNAcylated mTET1 CD and mTET2 CD by ETD-LC-MS/MS to map sites of modification by OGT. In addition to endpoint experiments in which TET1 CD and TET2 CD were modified by OGT overnight, we conducted a timecourse experiment on TET1 CD to examine the order in which the sites were modified. These experiments showed almost complete overlap between the sites identified in mESCs and on *in vitro* modified material, and suggest a potential model for the mechanism of TET1 stimulation by OGT which requires further investigation.

Materials and Methods

In vitro TET *O*-GlcNAcylation

O-GlcNAcylation of recombinant TET1 and TET2 was performed according to the methods in chapter 2.

Sample digestion and preparation

In vitro TET *O*-GlcNAcylation reactions were run on an SDS-PAGE gel and stained with Coomassie blue. TET protein bands were excised and cut up in cubic pieces 1mm long, washed twice with 50% acetonitrile in 25mM ammonium bicarbonate (NH_4HCO_3) and vacuum dried. The gel samples were reduced with DTT (10mM in 25mM NH_4HCO_3, 56°C for 1h), alkylated with iodoacetamide (55mM in 25mM NH_4HCO_3, room temperature for 1 h), and vacuum dried again. Samples were rehydrated in at least 32 µl of digestion buffer (10 ng/µl trypsin or chymotrypsin in 25mM NH_4HCO_3) (actual volume was based on the original size of the gel slice), and covered with the minimum volume of NH_4HCO_3. After an overnight digestion at 37°C, peptides were extracted twice with a solution containing 50% acetonitrile and 5% formic acid. The extracted digests were vacuum-dried and resuspended in 20 µl of 1% formic acid in water.

Mass spectrometry analysis

For overnight *O*-GlcNAcylation of TETs, samples were analyzed as described in chapter 3. Data was analyzed as described in chapter 3 with the exception that no labeling was considered and the only Glycan modifications considered were HexNAc on N, S, or T. All *O*-GlcNAc positive spectra were manually verified.

For timecourse analysis of mTET1 CD *O*-GlcNAcylation, peptides were analyzed on an Orbitrap Fusion Lumos (Thermo) using EThcD fragmentation with an HCD collision energy of 25%. Peaklists were generated with PAVA and searched using Protein Prospector. MS1 peak areas were quantified using Skyline(Schilling et al., 2012). *O*-GlcNAc modified peptide peak areas were normalized against unmodified peptides from TET1 and OGT.

Results

The combined results of the SILAC experiment in mESCs and the overnight *O*-GlcNAcylation of TETs *in vitro* are summarized in Tables 4.1-4.2 and Figures 4.1-4.2 below. These results demonstrate that there are numerous sites of *O*-GlcNAcylation (>10) on both mTET1 and mTET2. The modified residues observed *in vitro* match very well with residues and peptides that are *O*-GlcNAcylated in mESCs, arguing for the physiological relevance of the *in vitro* assay.

Table 4.1 – *O*-GlcNAcylation sites on mTET1
Sites and peptides are color coded to indicate protein domain (see Figure 4.1).

mTET1				
***In vitro* modified CD**		**mESCs (SILAC experiments)**		
Sites	**Ambiguous**	**Sites**	**Ambiguous**	**Peptides**
S1567 (DSBH1)	S1753 or S1755 (spacer)	T327	T550 or S553 (CXXC)	aa269-290
S1777 (spacer)	S1768 or T1769 (spacer)	S533 (CXXC)	S790 or T792 or T793 or S794	aa315-332
T1781 (spacer)	S2004 or S2009 (C45)	T535 (CXXC)	S1768 or T1769 or S1777 or S1778 or T1781 or S1782 (spacer)	aa532-541 (CXXC)
S1998 (C45)	S2024 or S2027 (C45)	T550 (CXXC)	S2004 or S2009 (C45)	aa542-568 (CXXC)
S2009 (C45)		S790		aa778-805
T2022 (C45)		T792		aa787-805
S2027 (C45)		S794		aa1052-1063
		S1055		aa1127-1137
		S2009 (C45)		aa1734-1756 (spacer)
				aa1766-1788 (spacer)
				aa1993-2010 (C45)

Figure 4.1 – Domain architecture of mTET1

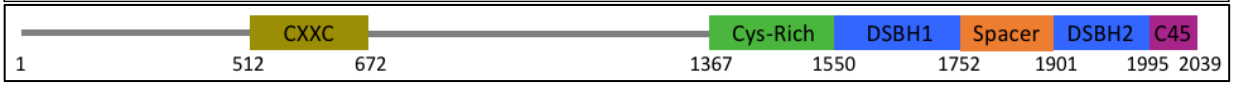
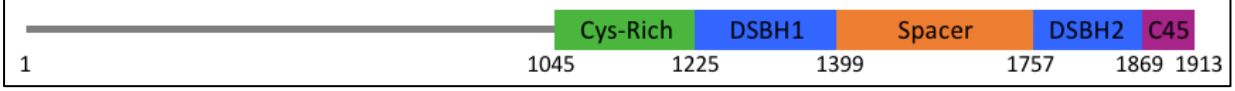


Table 4.2 – *O*-GlcNAcylation sites on mTET2
Sites and peptides are color coded to indicate protein domain (see Figure 4.2).

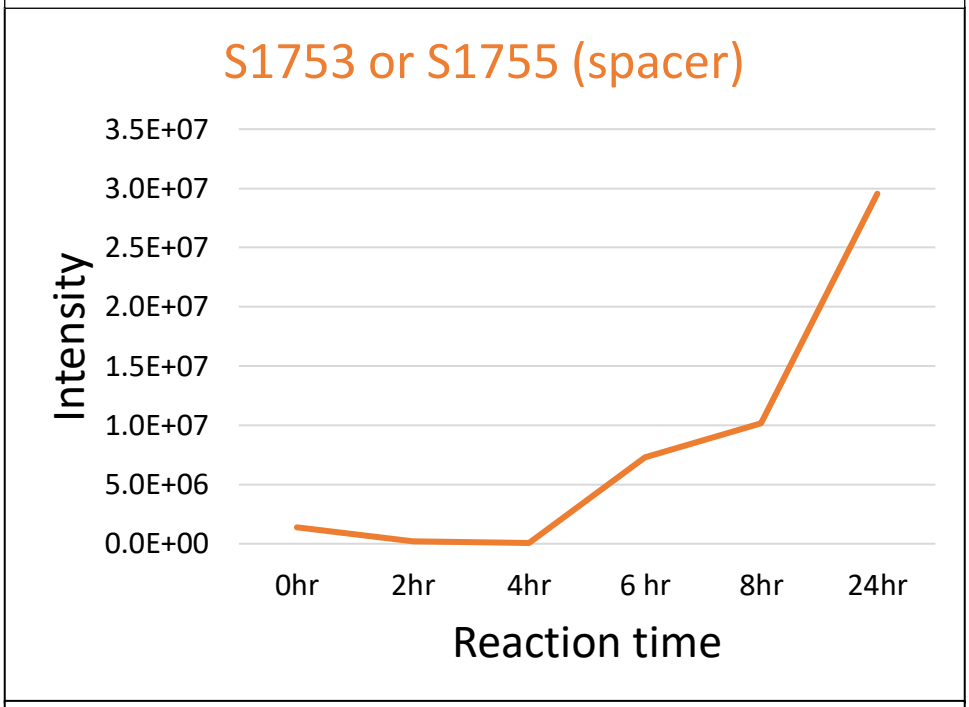
mTET2				
***In vitro* modified CD**		**mESCs (SILAC experiments)**		
Sites	**Ambiguous**	**Sites**	**Ambiguous**	**Peptides**
S1619 (spacer)	S1694 + T1676 or S1677 (spacer)	T95	T95 or S97	aa373-391
T1645 (spacer)	T1595 or T1604 (spacer)	S744	T153 or T154	aa435-469
S1663 (spacer)	T1631 or T1632 (spacer)	T1631 (spacer)	T379 or T381 or S384	aa616-646
S1665 (spacer)	T1685 or S1688 (spacer)	T1672 (spacer)	T625 or S628	aa625-646
S1668 (spacer)			S779 or S780	aa665-702
			T1631 or T1632 or T1635 (spacer)	aa1588-1616 (spacer)
			T1676 or S1677 (spacer)	aa1628-1653 (spacer)
				aa1665-1675 (spacer)
				aa1676-1705 (spacer)

Figure 4.2 – Domain architecture of mTET2



In addition to the experiments above, the *in vitro O*-GlcNAcylation of mTET1 CD over a timecourse from 0-24 hours was examined. Six modifications were found and tracked (three in the spacer, three in the C45). Results are shown in Figures 4.3 and 4.4 below.

Figure 4.3 – Timecourse of mTET1 CD *O*-GlcNAcylation



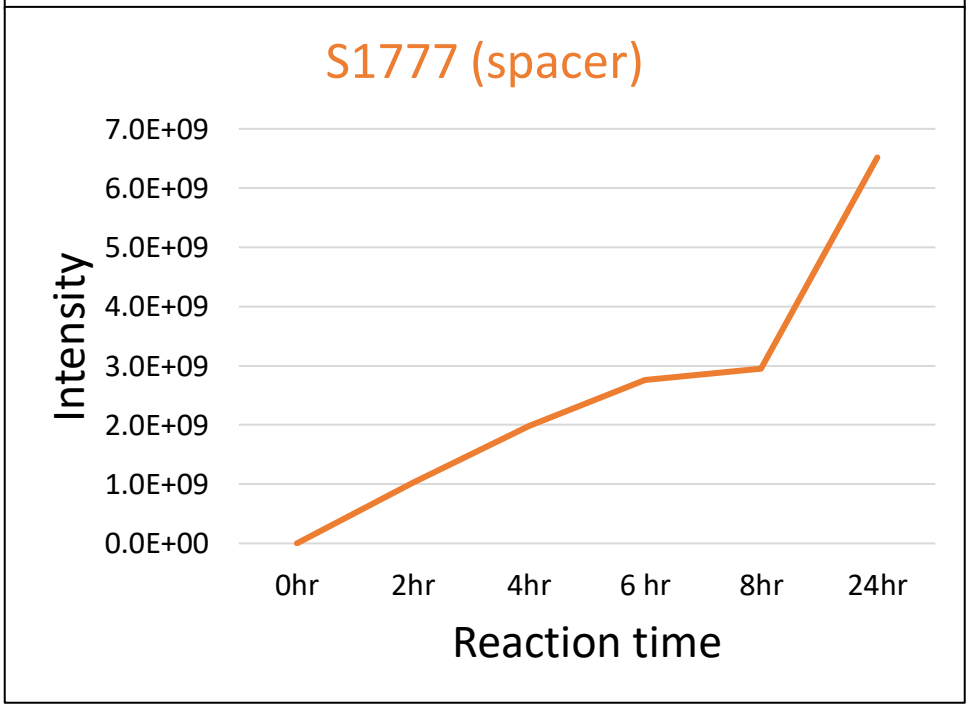


Figure 4.3 – Timecourse of mTET1 CD *O*-GlcNAcylation

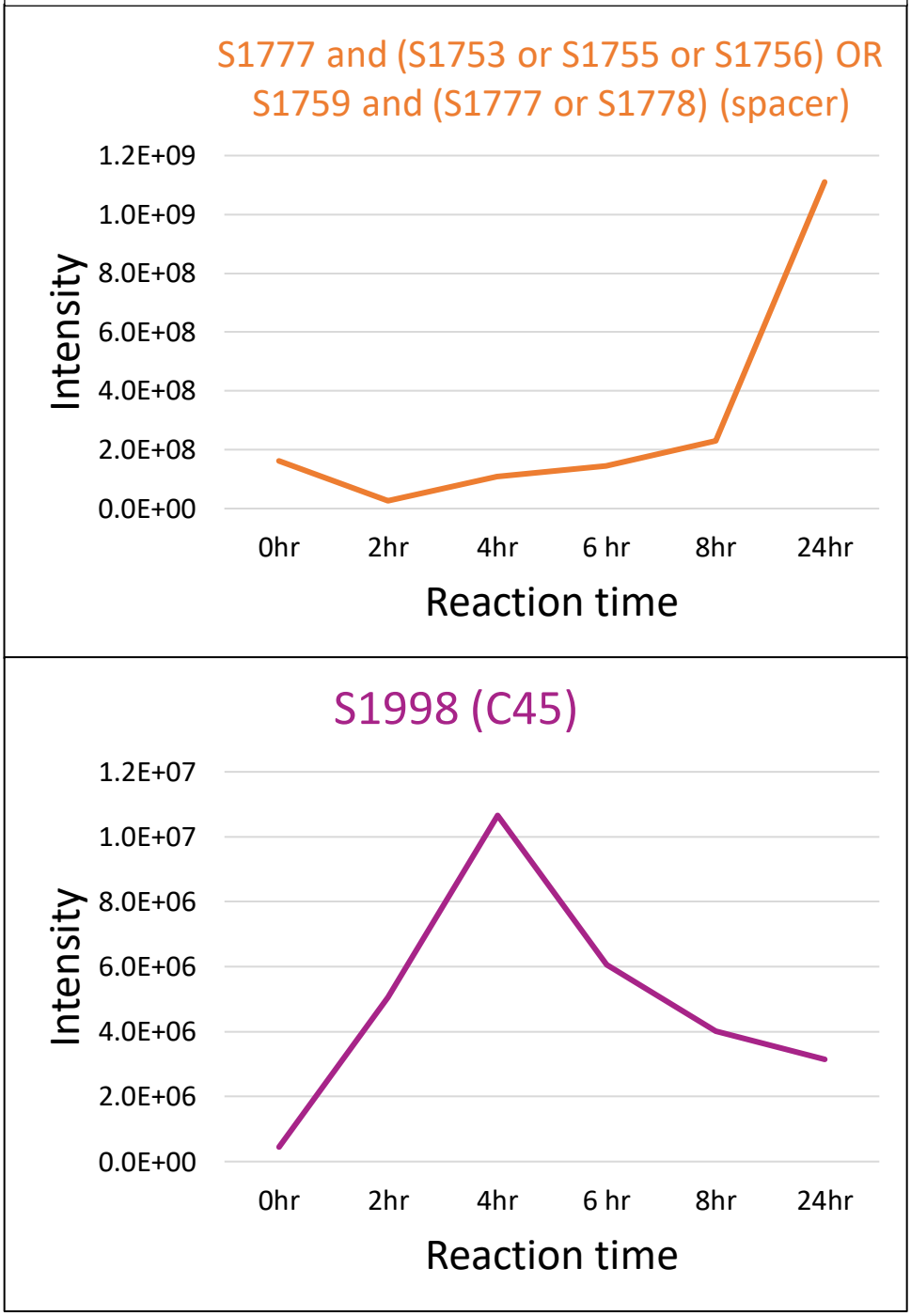
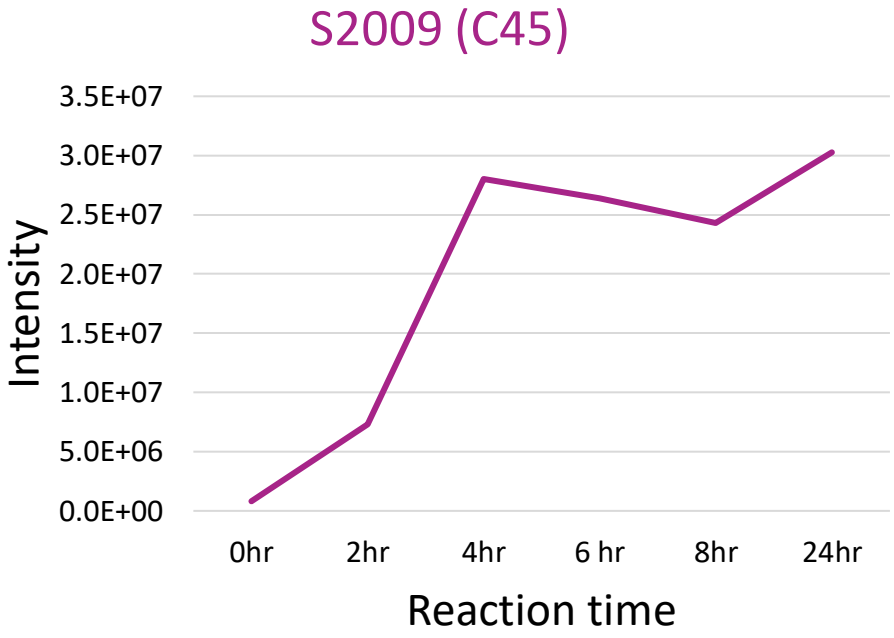


Figure 4.3 – Timecourse of mTET1 CD *O*-GlcNAcylation



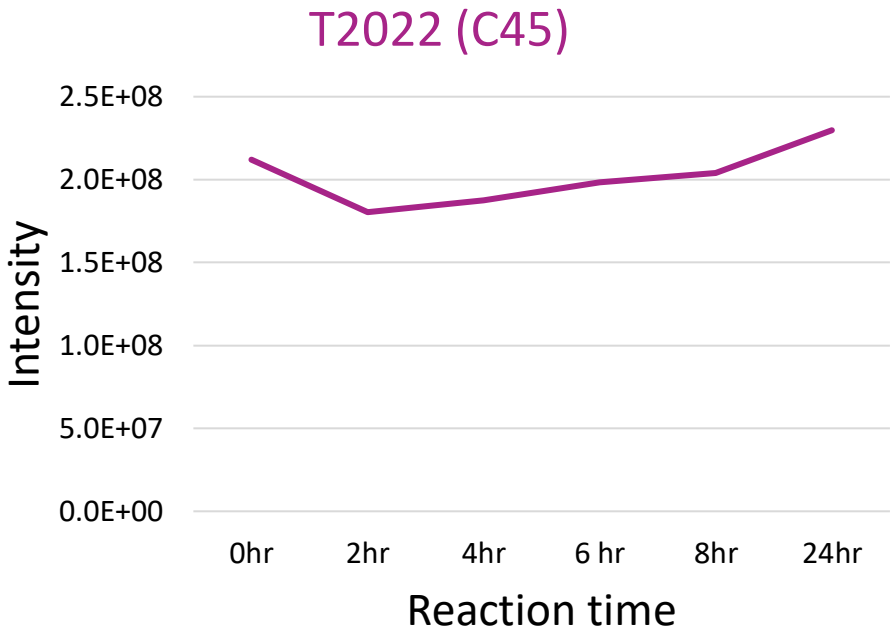
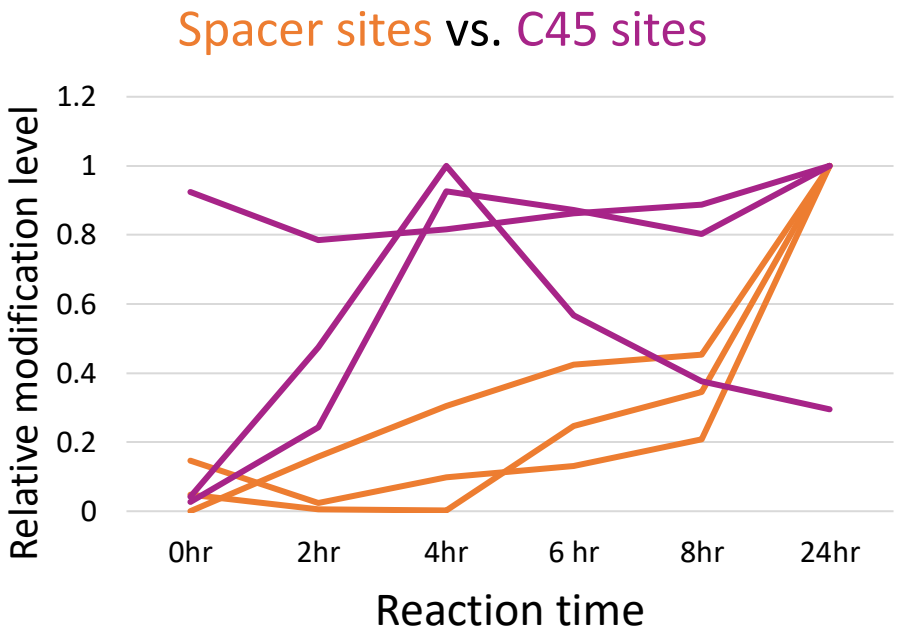


Figure 4.4 – Timecourse of mTET1 CD *O*-GlcNAcylation: overlay of all sites



Discussion/Analysis

Overall there is excellent agreement between the sites identified in mESCs and on *in vitro* modified material, with almost every *O*-GlcNAcylated residue *in vitro* corresponding to an *O*-GlcNAcylated residue or peptide in cells (Tables 4.1-4.2).

Sites in the TET1 CXXC domain

In addition to *O*-GlcNAcylated T535 of TET1, which was identified previously(Myers, Panning, & Burlingame, 2011a), two new sites of modification in the CXXC domain, T533 and T550, are identified, as well as a possible modification on S553. This domain was named for its homology to CXXC DNA binding domains found in other proteins, but studies on the DNA

binding capacity of the TET1 CXXC domain are in conflict(Frauer, Rottach, et al., 2011b; Xu et al., 2011). If this domain does bind DNA, it is possible that *O*-GlcNAcylation regulates its DNA binding activity.

Site in the TET1 DSBH domain

Interestingly, the single *O*-GlcNAc site found in the mTET1 DSBH domain (S1567) occurs at a residue within a 32 amino acid stretch that is not conserved in any other mouse or human TET protein (Figure 4.5). The significance of this particular *O*-GlcNAc site is unknown. This modification has only been observed on *in vitro* modified mTET1 CD and not on material from mESCs.

Figure 4.5 – *O*-GlcNAc site in the mTET1 DSBH domain
Alignment of a short region of mTET1 with mTET2-3 and hTET1-3. S1567, a site of *O*-GlcNAcylation observed on *in vitro* modified mTET1 CD, is highlighted.

```
hTET1   1603   PSSPLH--------------------------------EKNLEDNLQSLA   1620
hTET2   1313   GDDPKE-------------------------------EEKLESHLQNLS   1330
hTET3   1008   GDNPKE-------------------------------EEVLRKSFQDLA   1025
mTET1   1551   PNYPLHnyykritgmssegsdvktgwiipdrktlisreEKQLEKNLQELA   1600
mTET2   1226   GAEPKE-------------------------------EERLGSHLQNLA   1243
mTET3   1016   GDNPKE-------------------------------EEVLRNSFQDLA   1033

Identical residues
```

Sites in the spacer and C45 domains

It is interesting that the vast majority of sites mapped in the catalytic domains of TET1 and TET2, both in cells and *in vitro*, are found in either the spacer or C45 domains. TET1 has at least four *O*-GlcNAc sites in the spacer and at least four in the C45, and all of the sites in the TET2 catalytic domain (there are at least eleven) are in the spacer.

From a biophysical standpoint, this result makes sense. OGT is known to preferentially modify unstructured peptides(Lazarus, Nam, Jiang, Sliz, & Walker, 2011), and both the spacer and C45 domains are presumed to be unstructured. In addition, OGT binds to the C45 domains of both TET1 and TET2, so its active site is presumably also near or bound to this domain.

Interestingly, TETs are catalytically active *in vitro* when the spacer domain, the C45, or both, are deleted(Hu et al., 2013). How then might *O*-GlcNAcylation of these non-essential domains stimulate the activity of TETs? A broader related question is, how can addition of numerous uncharged, bulky sugar groups to an enzyme stimulate its activity? Further in-depth studies are required to answer this question, but one possible model is that the spacer domain of TETs is an autoinhibitory region and that *O*-GlcNAcylation relieves inhibition, possibly because the bulky sugar groups force it into a different conformation (Figure 4.6). The spacer domain could interfere with substrate binding, with catalytic chemistry, or both. This model is highly speculative, but preliminary evidence for it will be discussed in the next chapter.

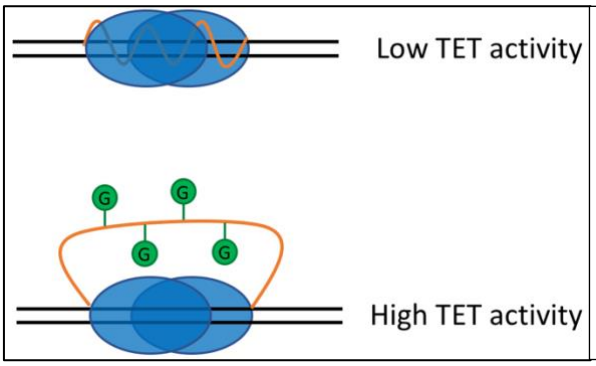


Figure 4.6 – Model of *O*-GlcNAc-mediated stimulation of TET1 CD activity
In this model the bilobed catalytic fold (blue ovals) is unable to efficiently modify DNA (black lines) because of interference by the spacer (orange line). Placement of *O*-GlcNAc on the spacer (green lollipops) forces it out of the inhibitory conformation, stimulating catalytic activity.

Temporal order of modifications

The timecourse experiment tracked six sites or sets of sites that are modified over time (Figure 4.3). Small apparent decreases in modification at a given site with time may be attributed to error within the method. However, it is difficult to explain why *O*-GlcNAcylation of S2009 appears to peak at 4 hours and decrease significantly by 24 hours. It also appears that T2022 is *O*-GlcNAcylated even at time zero. Since mTET1 CD is purified from sf9 insect cells, which appear to have an OGT homolog, it is possible that TET1 CD is *O*-GlcNAcylated at some level in these cells before purification.

When all six plots are overlaid on one another, an interesting pattern emerges (Figure 4.4). It appears that the sites within the C45 domain are modified first, followed by the sites in the spacer. It may be that OGT modifies the C45 most efficiently because it is bound there, and only afterward does it modify the spacer which is further away in space. This experiment needs to be repeated to determine if these trends are reproducible.

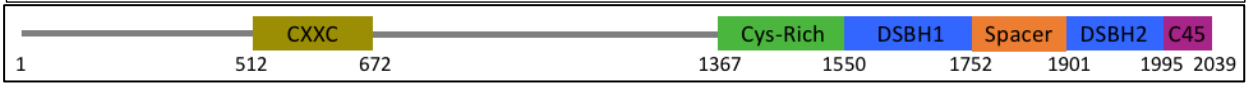
Chapter 5

Autoinhibition of TET by the spacer domain: a possible model for stimulation of TET activity by OGT

Introduction

The previous chapter discussed sites of *O*-GlcNAcylation on mTET1. All but one of the sites found in the mTET1 catalytic domain (CD) are either in the C45 domain, which binds OGT, or the spacer region (Figure 5.1). Both of these domains are predicted to be unstructured and are dispensable for *in vitro* catalytic activity of the purified enzyme. These data raise the question of how *O*-GlcNAcylation of unstructured, non-essential domain(s) stimulates the activity of mTET1 CD. One possibility is the "spacer inhibition model," in which the naked spacer region acts to inhibit mTET1's catalytic activity and spacer *O*-GlcNAcylation relieves this inhibition, perhaps because the numerous large sugar groups preclude it from interacting with the catalytic fold of TET (Figure 5.2). As a first test of this speculative model, I purified recombinant mTET1 spacer from *E. coli* and added it to *in vitro* rTET1 CD activity assays.

Figure 5.1 – Domain architecture of mTET1



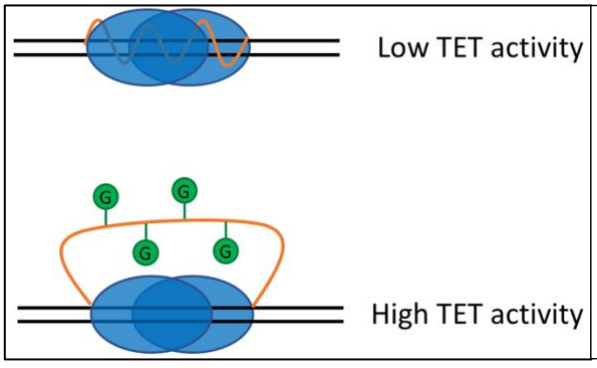


Figure 5.2 – Model of *O*-GlcNAc-mediated stimulation of TET1 CD activity
In this model the bilobed catalytic fold (blue ovals) is unable to efficiently modify DNA (black lines) because of interference by the spacer (orange line). Placement of *O*-GlcNAc on the spacer (green lollipops) forces it out of the inhibitory conformation, stimulating catalytic activity.

Materials and Methods

Protein purification

mTET1 spacer domain bearing a C-terminal 8xHis tag was cloned into the pBJG plasmid and expressed in BL21 DE3 cells. A 150mL culture of LB + 50ug/mL kanamycin was grown to OD_{600} = 0.7, then induced by addition 50uM IPTG and grown at 37C for 3 hours. The cell pellet was lysed in 5mL lysis buffer (BugBuster + 20mM imidazole, 0.5mM TCEP, complete EDTA-free protease inhibitors) with agitation at 4C for 30 minutes. The lysate was clarified by centrifugation at 15,000rpm at 4C for 10 minutes and bound to 500uL Ni-NTA resin with mixing at 4C for >1 hour. The lysate was poured over a 10mL disposable BioRad column in the cold room, washed with 10 column volumes (5mL) of wash buffer (20mM Tris pH 8, 1mM CHAPS, 50mM imidazole, 250mM NaCl, 0.5mM TCEP, complete EDTA-free protease inhibitors) and eluted with 6 column volumes (3mL) of elution buffer (20mM Tris pH 8, 250mM imidazole, 250mM NaCl, 0.5mM TCEP, complete EDTA-free protease inhibitors) in 500uL fractions. Fractions were analyzed by SDS-PAGE, and positive fractions were pooled and dialyzed into storage buffer (20mM Tris pH 8, 150mM NaCl, 10% glycerol, 0.5mM TCEP). Dialyzed protein was flash frozen in small aliquots and stored at -80C.

mTET1 catalytic domain was expressed and purified from sf9 insect cells according to the methods in chapter 2.

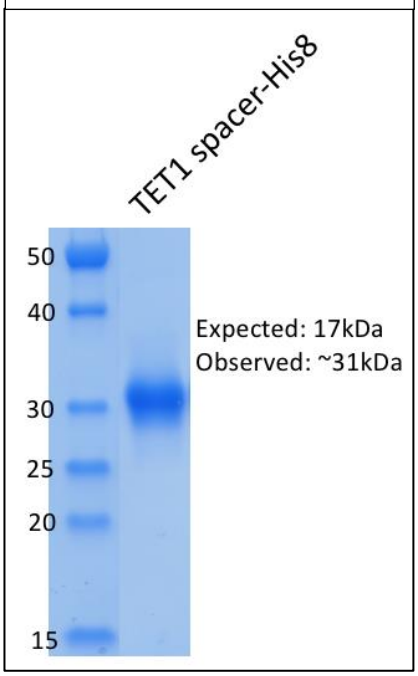
TET activity assays

TET activity assays were performed as indicated in chapter 2, with the addition of varying concentrations of recombinant TET1 spacer domain.

Results

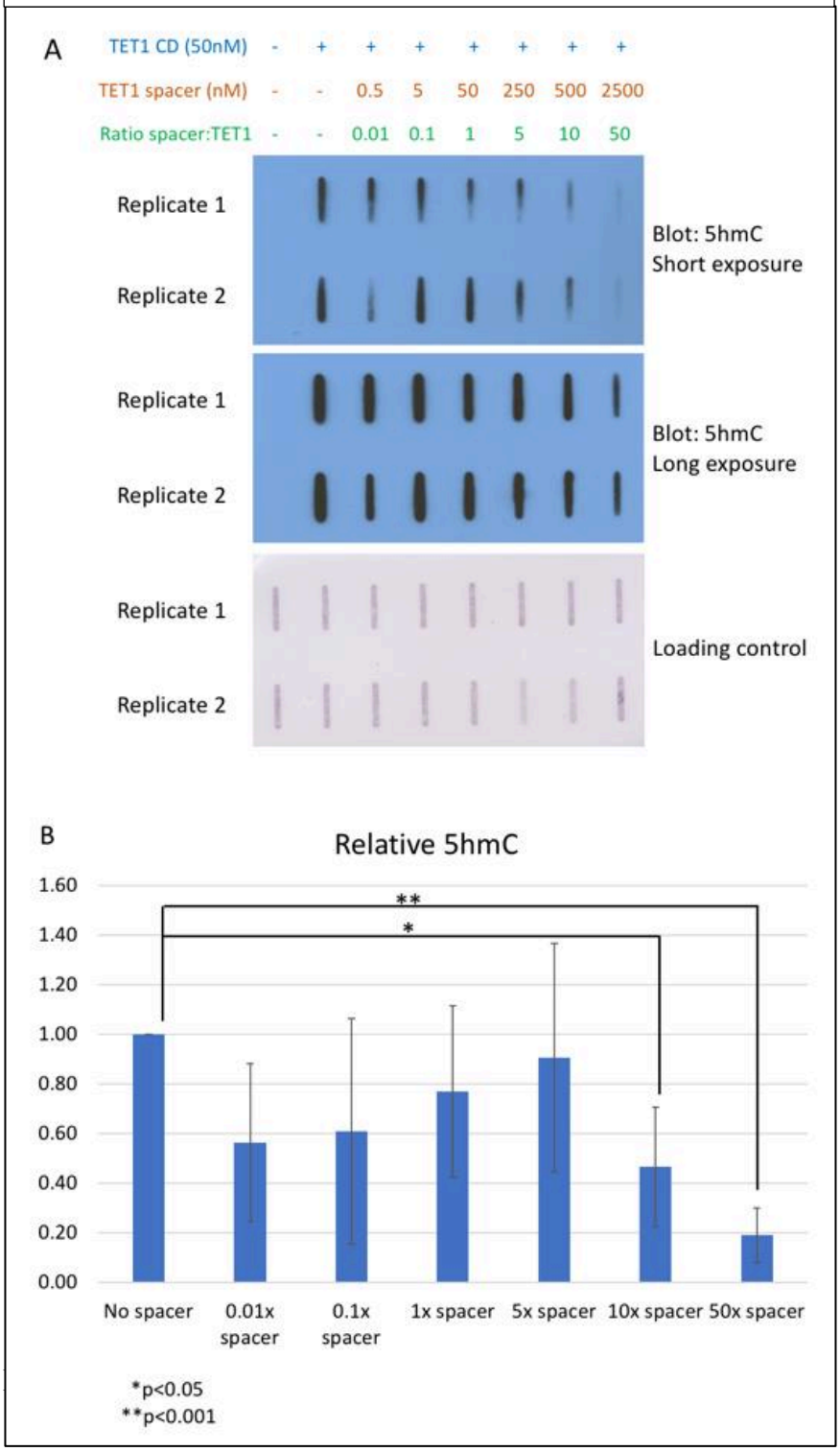
Highly pure mTET1 spacer was obtained from *E. coli* (Figure 5.3). The peptide did not migrate as expected on a denaturing SDS-PAGE gel, running with an apparent molecular weight of ~31kDa as opposed to its actual mass of 17kDa. The reason for this discrepancy is unknown; it could reflect unusual biophysical properties of the spacer which may or may not have biological significance.

Figure 5.3 – mTET1 spacer domain purified from *E. coli*.



Next, the purified spacer was added in increasing amounts to *in vitro* rTET1 CD activity assays. Figure 5.4 shows slot blots from two representative experiments, as well as quantification from three experiments. Despite the large variability reflected in the error bars in figure 5.4B, there is a reproducible and statistically significant decrease in TET activity at 10x and 50x molar excess of spacer over TET1 CD.

Figure 5.4 – rTET1 CD activity assays with spacer domain added
A) Slot blots from two representative TET activity assays. B) Quantification of 5hmC signal from three independent experiments.



Discussion

The decrease in TET activity upon addition of 10-50x molar excess of spacer is interesting but by no means conclusive. It is unknown whether this effect is a result of specific enzyme inhibition by the spacer domain, or simply a nonspecific effect of the presence of a large peptide. There are several straightforward experiments that could answer this question and either support or disprove the model in figure 5.2. For example, if inhibition of TET is a specific effect of the spacer domain then addition of an equivalent amount of a peptide of similar size should have less or no effect on TET1 CD activity. Another prediction of the model in figure 5.2 is that if the spacer is *O*-GlcNAcylated prior to being added to TET activity assays its effect on TET activity should be reduced or eliminated. Finally, if the spacer does have inhibitory activity then rTET1 CD with the spacer deleted may be more active than the full catalytic domain, as well as more sensitive to inhibition by spacer added in trans.

Chapter 6

Mutational analysis of the interactions between OGT and TET1, TET2, and TET3

Introduction

OGT directly interacts with the catalytic domains of all three TET proteins(Q. Chen et al., 2013; Deplus et al., 2013; Vella et al., 2013). Mutational analysis of the interaction between OGT and mTET1 CD is presented in chapter 2, showing that the C-terminal 45 amino acids (C45) of mTET1 are necessary and sufficient for interaction with OGT and that the mTET1 D2018A mutation disrupts the OGT-mTET1 CD interaction. This chapter presents further analysis of the interactions between OGT and TET proteins, including experiments with more mutations in the mTET1 C45 as well as analysis of the interactions between OGT and mTET2, mTET3, and hTET3.

Materials and Methods

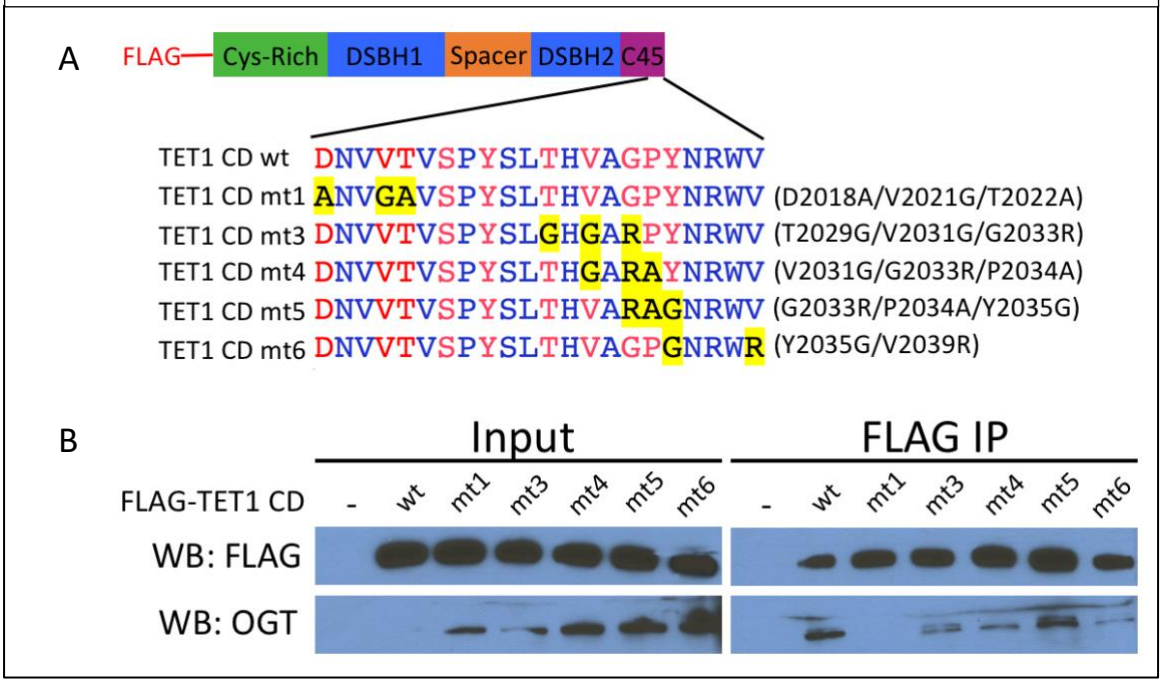
All experiments were performed according to the methods detailed in chapter 2.

Results

Multiple residues between mTET1 D2018 and V2039 are involved in binding OGT

The experiments in chapter 2 showed that mutation of D2018/V2021/T2022 together or V2021/T2022/S2024 together has by far the largest effect on binding of OGT to mTET1 CD (Figure 2.3). Further analysis shows that amino acids after S2024 also have a smaller effect on binding of OGT to mTET1 CD (Figure 6.1).

Figure 6.1 – Residues beyond T2022 are important for the OGT-mTET1 CD interaction
A) Diagram of FLAG-tagged mTET1 CD constructs expressed in HEK293T cells. B) FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing FLAG-TET1 CD constructs in A.

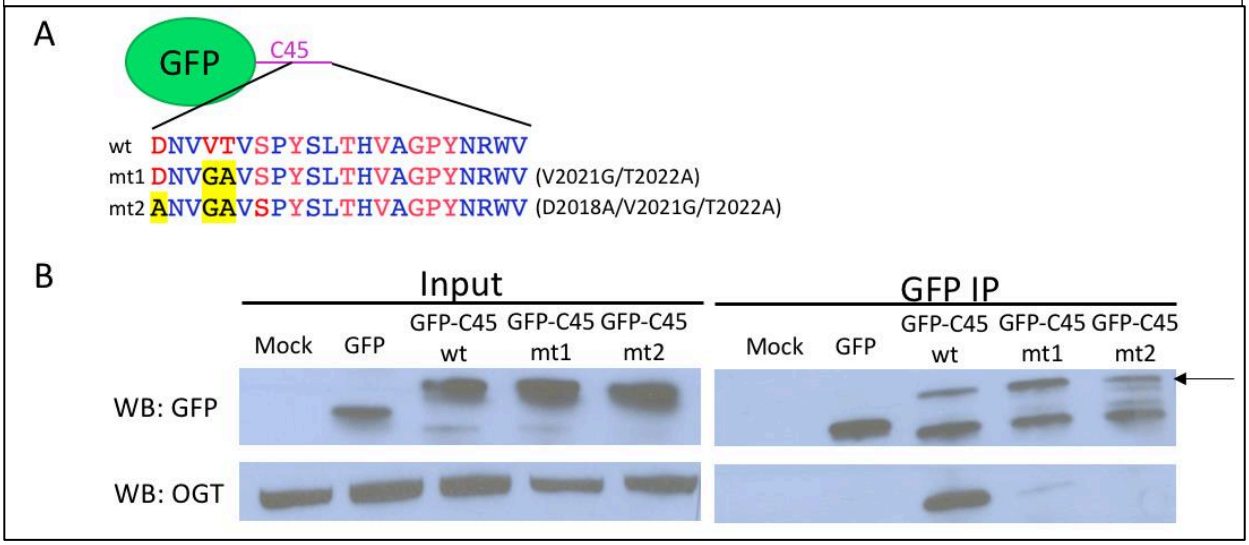


In Figure 6.1 mutants 3, 4, and 6 all modestly reduce the interaction between OGT and TET1 CD. Since mutant 5 appears to have little or no effect, residues G2033, P2034, and Y2035 are probably not important for OGT binding. This suggests that residues T2029, V2031, and V2039 are involved in binding OGT but are not as important as D2018, V2021, and T2022. Note that the absence of OGT in the first two input lanes in Figure 6.1 likely reflects a western blotting problem. This analysis is based on comparison to the amount of OGT pulled down by wild-type TET1 CD, which assumes that in reality the amount of OGT in each input sample is approximately equal.

V2021 and T2022 are involved in binding of OGT to the mTET1 C45

The mTET1 C45 is sufficient for interaction with OGT, and the single D2018A point mutation completely disrupts this interaction (Figure 2.4). Prior to this experiment, a V2021G/T2022A double mutant and D2018A/V2021G/T2022A triple mutant were tested in this context. The GFP-mTET1 C45 double mutant retains only the smallest hint of binding to OGT, while the triple mutant has no detectable binding affinity whatsoever for OGT (Figure 6.2). This indicates that in addition to D2018, V2021 and T2022 are also important for the OGT-mTET1 C45 interaction.

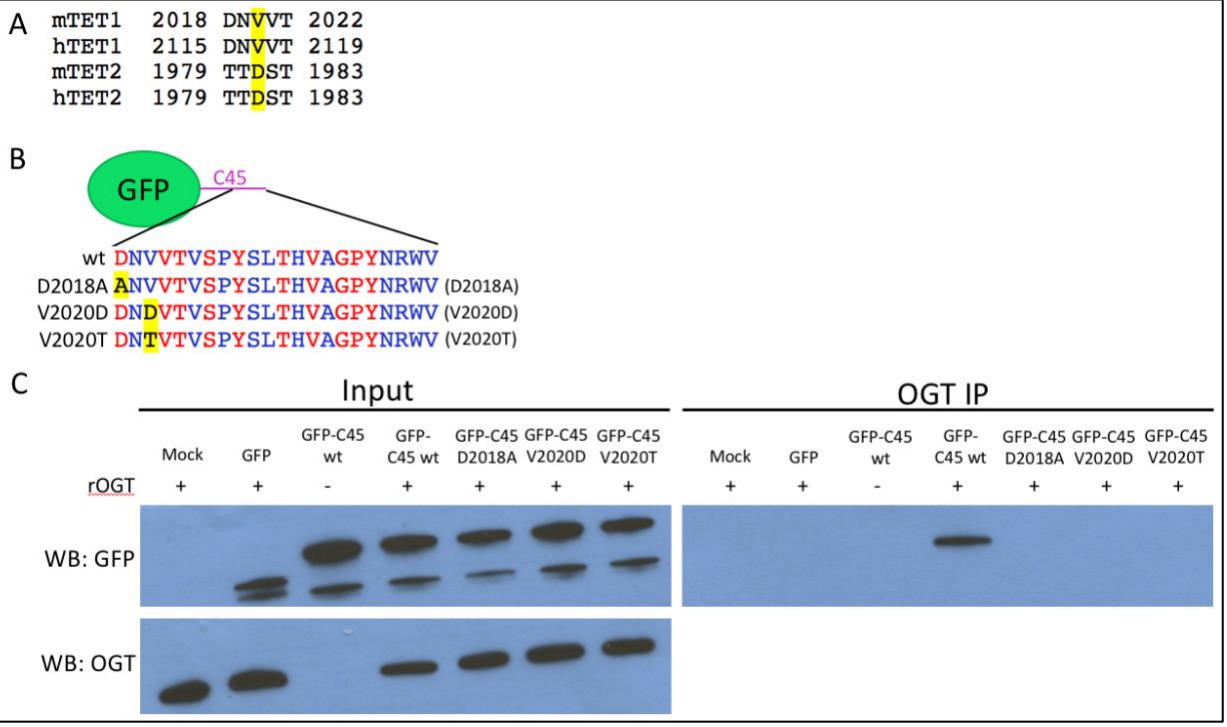
Figure 6.2 – V2021 and T2022 are important for the OGT-mTET1 C45 interaction
A) Diagram of GFP-mTET1 C45 constructs expressed in HEK293T cells. B) GFP and OGT western blot of inputs and GFP IPs from HEK293T cells transiently expressing constructs in A. Arrow indicates the correct band.



V2020 is important for OGT binding to mTET1 C45

There is evidence that OGT binds more strongly to TET2 and TET3 than TET1 (both the mouse and human proteins) (Q. Chen et al., 2013; Deplus et al., 2013) (Figure 6.6). I spent some time speculating how the three TET C-termini might bind to OGT, based on the structure of OGT bound to a peptide from HCF(Lazarus et al., 2013) combined with my data above about the residues in the mTET1 C45 important for binding OGT. I created an edited TET protein alignment that I predicted would align residues of the C-termini that bound to the same region of OGT. Among the differences between the C-termini of TET1 and TET2 in this alignment, I noticed that residue V2020 of mTET1 is conserved as valine in hTET1, but in both mouse and human TET2 it is aspartate (Figure 6.3A). My speculation with the structure of OGT lead me to predict that an aspartate or threonine might be favorable here while a valine might be unfavorable. To test this idea, I mutated V2020 to D or T in the GFP-mTET1 C45 construct and tested binding to OGT (Figure 6.3B, C). Contrary to my expectation, both mutations of V2020 completely ablated binding to OGT. Nevertheless, this experiment does demonstrate that residue V2020 of mTET1 is important for binding of OGT to the mTET1 C45.

Figure 6.3 – V2020 is important for the OGT-mTET1 C45 interaction
A) Alignment of a region of the mouse and human TET1 and TET2 C-termini, based on speculation about which residues might bind the same region of OGT. B) Diagram of GFP-mTET1 C45 constructs. C) GFP and OGT western blot of inputs and OGT IPs from *in vitro* binding reactions containing rOGT and *in vitro* translated GFP constructs.

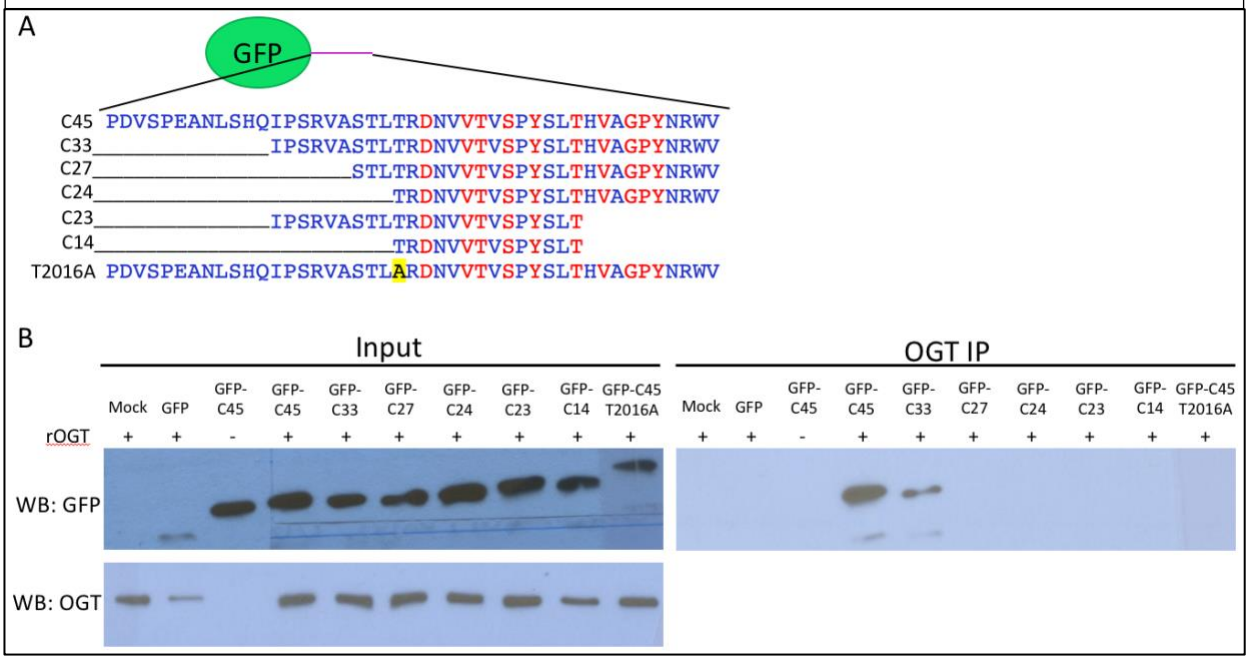


Residues prior to D2018 are important for binding of OGT to the mTET1 C45

Thus far the mutational analyses have focused on the C-terminal half of the mTET1 C45 because all of the identically conserved residues are found in this region. However, experiments with truncations of the C45 fused to GFP suggest that less conserved residues in the C45 N-terminus are also important for binding to OGT (Figure 6.4). Deleting the first 12 amino acids of the C45 significantly impairs binding to OGT, and deleting the first 18 amino acids completely ablates binding. In addition, mutation of T2016 (which is also threonine in mTET2 but proline in

mTET3) to alanine completely prevents binding of OGT, implicating this amino acid in the OGT-mTET1 interaction.

Figure 6.4 – Residues prior to D2018 are important for the OGT-mTET1 C45 interaction
A) Diagram of GFP-mTET1 C45 constructs. B) GFP and OGT western blot of inputs and OGT IPs from *in vitro* binding reactions containing rOGT and *in vitro* translated GFP constructs.

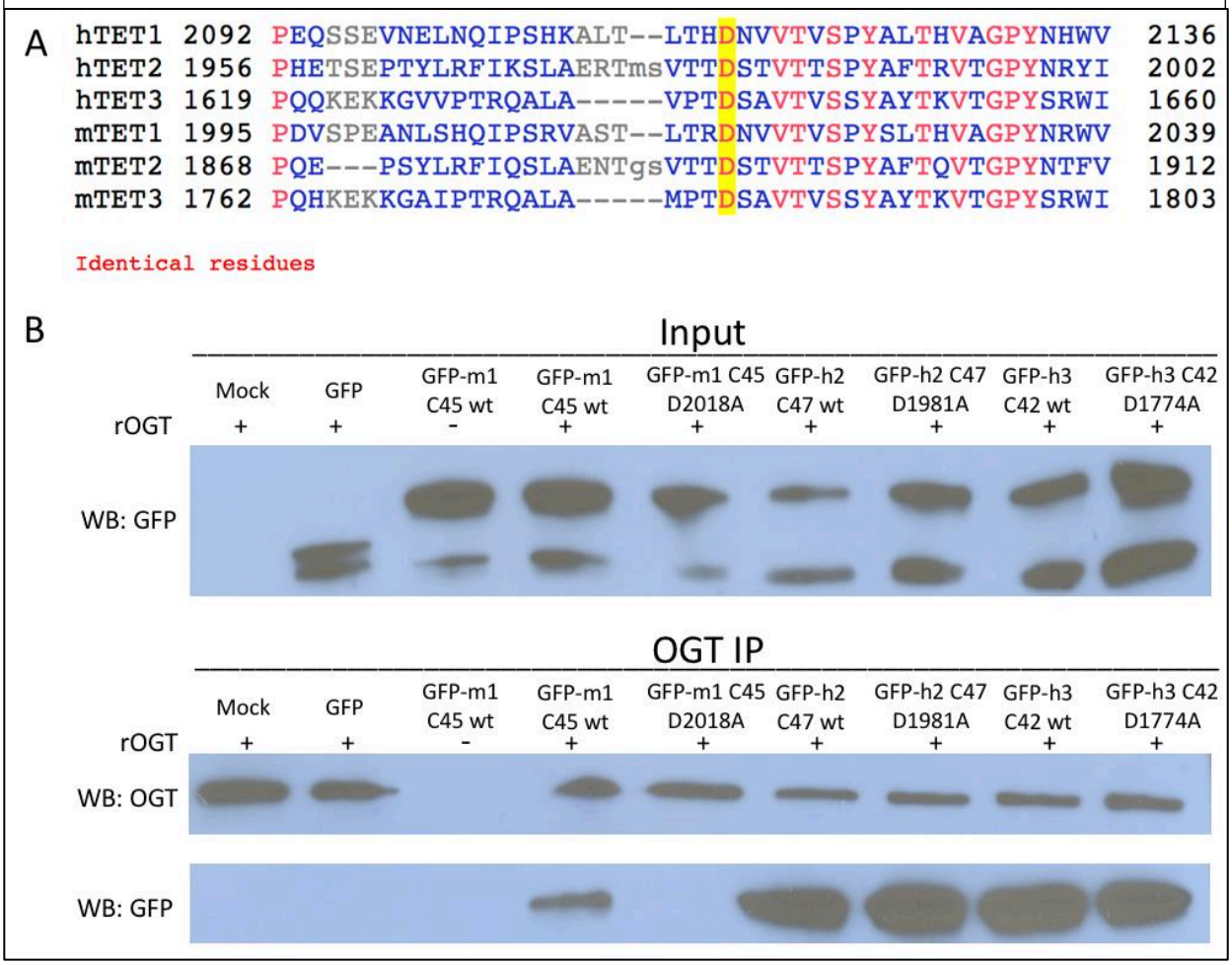


Binding of the C-termini of hTET2 and hTET3 to OGT

The mTET1 C45 region is highly conserved throughout all 3 mouse and human TET proteins, including identical conservation of D2018 and other residues important for the OGT-mTET1 interaction (Figure 6.5A). In this experiment I tested whether the C-termini of hTET2 and hTET3 are also sufficient to bind rOGT *in vitro* and what the effect is of mutating the aspartates corresponding to D2018 of mTET1 (Figure 6.5B). rOGT binds to the hTET2 C47 and hTET3 C42 more tightly than to mTET1 C45, and in both GFP-hTET C-termini fusions the

aspartate-to-alanine mutation has no detectable effect on the interaction. This suggests that OGT may bind to the hTET2 and hTET3 C-termini differently than it binds the mTET1 C45. Alternatively, the aspartate residue may be involved in binding but the increased strength of the interaction means that the single aspartate-to-alanine change is not sufficient to perturb binding (however, see next section).

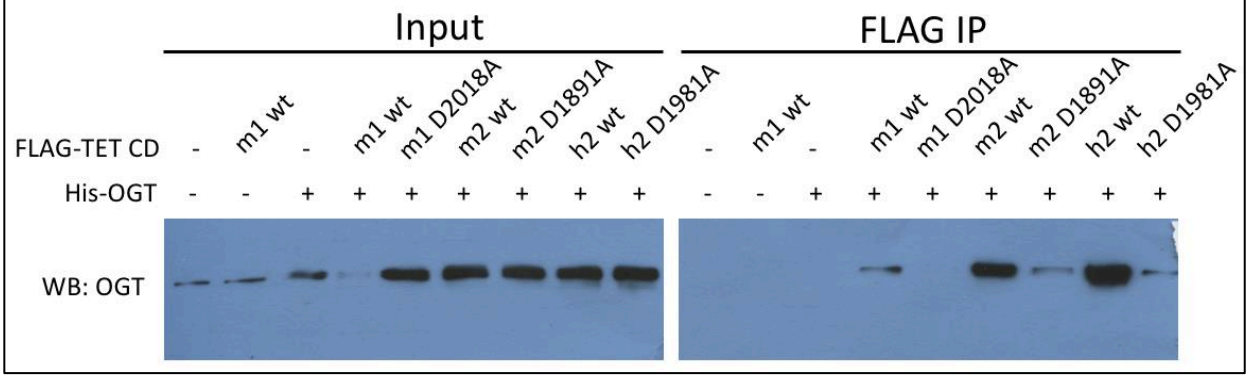
Figure 6.5 – hTET2 and hTET3 C-termini are sufficient to bind OGT
A) Alignment of the C-termini of the 3 mouse and human TET proteins. mTET1 D2018 and the corresponding aspartates in the other TETs are highlighted. B) GFP and OGT western blot of inputs and OGT IPs from *in vitro* binding reactions containing rOGT and *in vitro* translated GFP constructs.



Binding of mTET2 CD and hTET2 CD to OGT

Finally, I tested the effect of the aspartate-to-alanine mutation on the interaction between OGT and both human and mouse TET2 catalytic domains by IP-western blot from HEK293T cells (Figure 6.5A and Figure 6.6). Pulldown of OGT was much more robust with both TET2 constructs than with mTET1 CD, suggesting that mouse and human TET2 may interact with OGT more strongly than mouse TET1, consistent with published data(Q. Chen et al., 2013; Deplus et al., 2013). However, the input signal for OGT is very low for the mTET1 CD wild-type reaction, which could reflect a technical problem or failure to overexpress OGT in this sample. The former would also explain the poor pulldown of OGT by mTET1 CD wild-type. In addition, the aspartate-to-alanine mutation has a severe effect on pulldown of OGT by both human and mouse TET2 CDs, although a small amount of binding to OGT is evident unlike in the case of mTET1 CD D2018A. This is unexpected in light of the data in the previous section, in which GFP fused to hTET2 C47 wild-type and D1981A both robustly pulled down OGT (Figure 6.5). This may suggest that OGT's interaction with the TET C-termini alone is different than its interaction with the entire TET catalytic domains. A large caveat to this experiment, however, is that a good FLAG western blot was not obtained, so it is uncertain if expression of the FLAG-TET constructs was uniform across samples.

Figure 6.6 – Human and mouse TET2 CDs require the conserved aspartate for robust interaction with OGT
FLAG and OGT western blot of inputs and FLAG IPs from HEK293T cells transiently expressing FLAG-tagged mTET1 CD, mTET2 CD, and hTET2 CD.



Discussion

A summary of the TET mutagenesis experiments is provided in table 6.1 below. Overall, the results show that both the first and second halves of the mTET1 C45 are important for the interaction with OGT, and identify specific identically conserved residues in the second half that are critical. Mutation of the identically conserved mTET1 D2018 to A disrupts the interaction of mTET1 CD with OGT, and the corresponding aspartate residue in the mouse and human TET2 proteins appears to be important for their binding to OGT as well.

Table 6.1 – Summary of mutagenesis experiments characterizing the TET-OGT protein:protein interactions

mTET1 CD in HEK293T cells		
Mutation	**Binding to OGT**	**Reference**
D2018A/V2021G/T2022A	Eliminated	Figure 2.3
V2021G/T2022A/S2024A	Eliminated	Figure 2.3
T2029G/V2031G/G2033R	Probably impaired	Figure 6.1; general notebook #3 pgs 475-476 Caveat: OGT input blot
V2031G/G2033R/P2034A	Probably impaired	Figure 6.1; general notebook #3 pgs 475-476 Caveat: OGT input blot
G2033R/P2034A/Y2035G	Probably unaffected	Figure 6.1; general notebook #3 pgs 475-476 Caveat: OGT input blot
Y2035G/V2039R	Probably impaired	Figure 6.1; general notebook #3 pgs 475-476 Caveat: OGT input blot
D2018A	Eliminated	Figure 2.3
GFP-mTET1 C45 in HEK293T cells		
Mutation	**Binding to OGT**	**Reference**
D2018A/V2021G/T2022A	Eliminated	Figure 6.2; general notebook #2 pgs 391-392
V2021G/T2022A	Almost eliminated	Figure 6.2; general notebook #2 pgs 391-392
GFP-mTET1 C45 in IVTT system		
Mutation	**Binding to OGT**	**Reference**
C33 (Δ1-12)	Impaired	Figure 6.4; general notebook #3 pgs 410-412
C27 (Δ1-18)	Eliminated	Figure 6.4; general notebook #3 pgs 410-412
C24 (Δ1-21)	Eliminated	Figure 6.4; general notebook #3 pgs 410-412
C23 (Δ1-12, Δ36-45)	Eliminated	Figure 6.4; general notebook #3 pgs 410-412
C14 (Δ1-21, Δ36-45)	Eliminated	Figure 6.4; general notebook #3 pgs 410-412
T2016A	Eliminated	Figure 6.4; general notebook #3 pgs 410-412
D2018A	Eliminated	Figure 6.3; general notebook #6 pgs 953-955
V2020D	Eliminated	Figure 6.3; general notebook #6 pgs 953-955
V2020T	Eliminated	Figure 6.3; general notebook #6 pgs 953-955
GFP-hTET2 C47/GFP-hTET3 C42 in IVTT system		
Mutation	**Binding to OGT**	**Reference**
hTET2 D1981A/hTET3 D1774A (analogous to mTET1 D2018A)	Unaffected	Figure 6.5; general notebook #6 pgs 936-940
mTET2 CD/hTET2 CD in HEK293T cells		
Mutation	**Binding to OGT**	**Reference**
mTET2 D1891A/hTET2 D1981A (analogous to mTET1 D2018A)	Probably almost eliminated	Figure 6.6; general notebook #9 pgs 1403-1404 Caveat: no FLAG blot

Chapter 7

The effect of OGT on TET2 and TET3 activity *in vitro*

Introduction

OGT interacts directly with and modifies all three TET proteins(Q. Chen et al., 2013; Deplus et al., 2013; Vella et al., 2013). Since OGT stimulates mTET1 CD activity *in vitro* via *O*-GlcNAcylation, the question arises whether OGT can similarly stimulate the other TET proteins. This chapter presents preliminary data investigating the effect of recombinant OGT (rOGT) on recombinant mouse TET2 catalytic domain (mTET2 CD) and recombinant human TET3 catalytic domain (hTET3 CD).

Materials and Methods

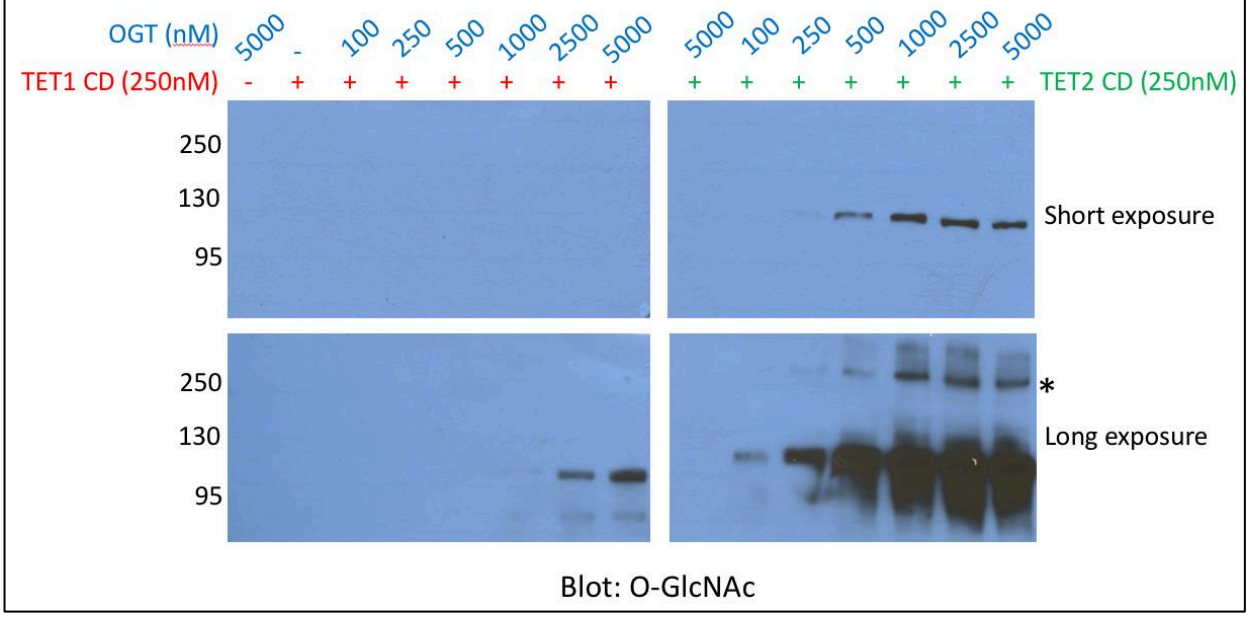
All experiments were performed according to the methods detailed in chapter 2.

Results

OGT modifies mTET2 CD *in vitro*

To test if OGT can modify mTET2 CD *in vitro* and compare the efficiencies of modification of TET1 vs. TET2 by OGT, mTET1 CD and mTET2 CD were incubated with varying amounts of OGT in the presence of UDP-GlcNAc. Reactions were analyzed by western blot with the RL2 *O*-GlcNAc antibody (Figure 7.1). OGT modified TET2 far more efficiently than TET1 – *O*-GlcNAcylation of TET2 was achieved by substoichiometric amounts of OGT (250nM mTET2 CD + 100nM OGT), while TET1 *O*-GlcNAcylation is undetectable until a 4-10-fold molar excess of OGT is supplied (250nM mTET1 CD + 1000-2500nM OGT). At each concentration of OGT tested the level of *O*-GlcNAcylation of TET2 is vastly higher than of TET1. This result makes sense in light of the evidence for tighter binding of OGT to TET2 than to TET1(Q. Chen et al., 2013; Deplus et al., 2013) (chapter 6).

Figure 7.1 – *O*-GlcNAcylation of mTET1 CD and mTET2 CD by OGT *in vitro*
O-GlcNAc western blots of *in vitro O*-GlcNAcylation reactions containing mTET1 CD or mTET2 CD, UDP-GlcNAc, and varying amounts of OGT.



OGT stimulates mTET2 CD activity *in vitro*

To test the effect of OGT and *O*-GlcNAcylation on TET2 activity, activity assays were performed using mTET2 CD, with or without preincubation with OGT and UDP-GlcNAc, and analyzed by slot blot (Figure 7.2). The data show that mTET2 CD activity is stimulated by OGT, and in this experiment stimulation of TET2 by OGT was more robust than stimulation of TET1. Replicates of this experiment are needed to verify this result, and further experiments should be performed to determine whether *O*-GlcNAcylation of mTET2 CD is necessary for OGT-mediated stimulation of TET2 activity. It should be noted that *O*-GlcNAcylation of the TETs in this experiment was never verified, so it is unknown if TET2 was successfully modified by OGT.

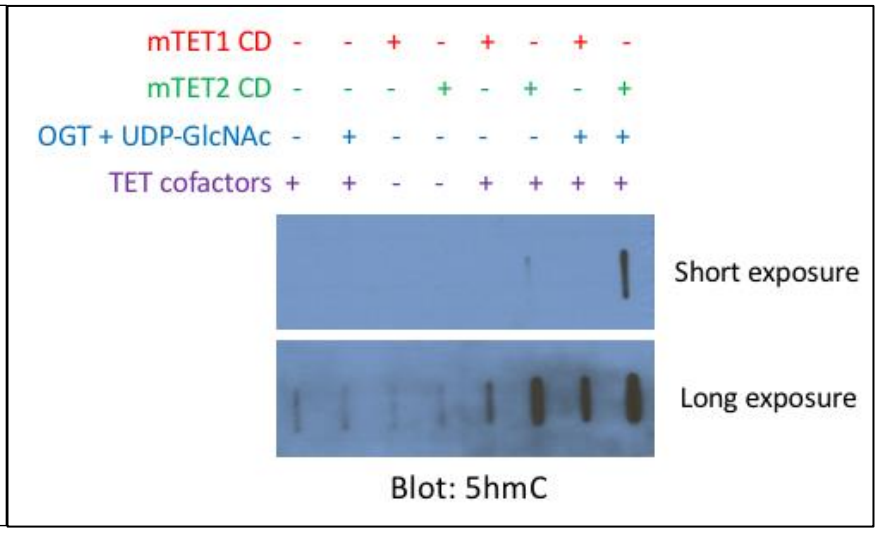


Figure 7.2 – Stimulation of mTET2 CD activity by OGT. 5hmC slot blot of *in vitro* TET activity assays with mTET1 CD or mTET2 CD, with or without preincubation with OGT and UDP-GlcNAc

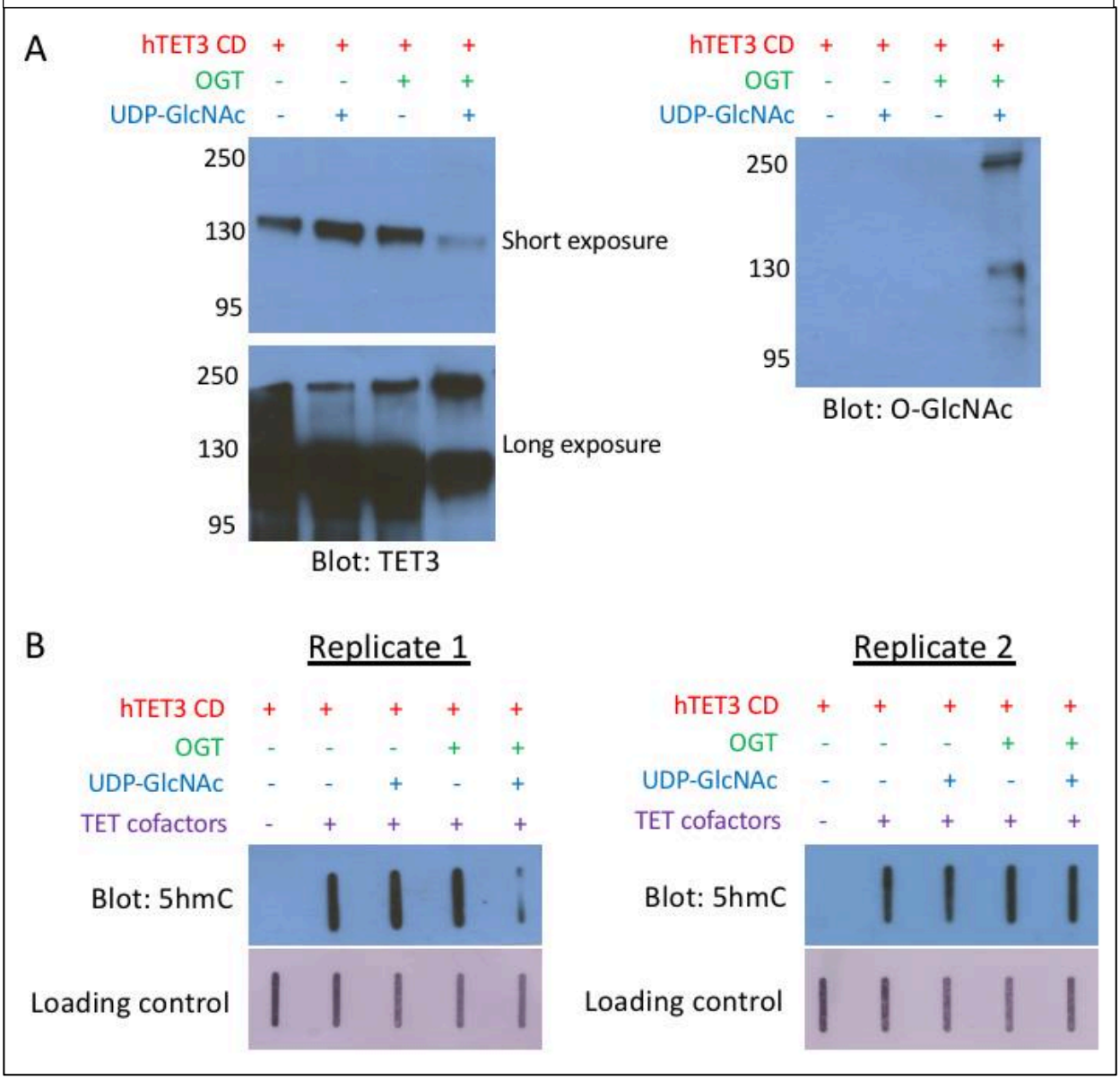
OGT modifies and might inhibit hTET3 CD *in vitro*

Analogous TET activity assays were also performed to test the effect of OGT and *O*-GlcNAcylation on hTET3 CD *in vitro* (Figure 7.3). When analyzed by western blot, two bands were recognized by the TET3 antibody: by far the most abundant species runs at ~130kDa, as expected for hTET3 CD, but a much fainter band is also picked up around 250kDa (Figure 7.3A). The *O*-GlcNAc western blot also recognized these same two bands, but in this case their intensities were approximately equal. The identity of the slower migrating band is unknown; it is recognized by the TET3 antibody and modified by OGT, so it may be an alternative form of TET3. However, TET3 is not expected to form any higher order complexes in the denaturing, reducing environment of the gel. Regardless, these data show that recombinant OGT can modify recombinant hTET3 CD *in vitro*.

Experiments testing the effect of *O*-GlcNAcylation on hTET3 CD activity have so far been inconsistent (Figure 7.3B). In some cases *O*-GlcNAcylated TET3 appears to have less activity than the unmodified enzyme (in contrast to mTET1 and mTET2), while other

experiments show no apparent change in activity when TET3 is modified by OGT. More experiments will be required to confidently assess the effect of OGT and *O*-GlcNAcylation on hTET3 CD activity *in vitro*.

Figure 7.3 – *O*-GlcNAcylation and activity of hTET3 CD *in vitro*
A) Western blots of *in vitro O*-GlcNAcylation reactions containing hTET3 CD, UDP-GlcNAc, and OGT. B) Slot blots of *in vitro* hTET3 CD activity assays



Discussion

These results show that recombinant OGT can modify multiple recombinant TET proteins *in vitro* (mTET2 CD and hTET3 CD) and suggest that *O*-GlcNAcylation may regulate the activity of these TETs as well. More experiments are necessary to confirm and/or clarify the effect of OGT on mTET2/hTET3 activity and to distinguish between the effects of the TET-OGT protein-protein interaction versus TET *O*-GlcNAcylation.

Chapter 8

Miscellaneous observations on the *in vitro* activities of TET1 CD and OGT

Introduction

Significant optimization was required to get the *in vitro* assays presented in this work up and running (for *O*-GlcNAcylating recombinant TET proteins with recombinant OGT, and measuring the activity of recombinant TETs). Through this process a few important parameters for OGT and TET activity were identified, which shall be discussed here. Specifically, the activities of both OGT and TETs are significantly pH-dependent in a physiologically relevant range, and *O*-GlcNAcylation of recombinant mTET1 CD by recombinant OGT is nonlinear with time.

Materials and Methods

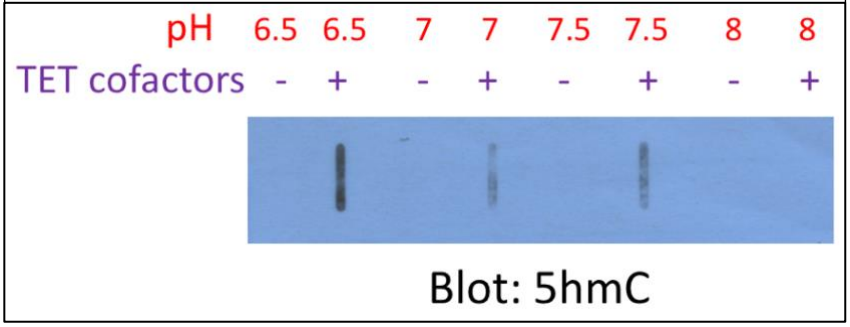
All experiments were performed according to the methods detailed in chapter 2.

Results

pH dependence of mTET1 CD activity

The activity of mTET1 CD on a methylated lambda DNA substrate was assayed by slot blot over a pH range from 6.5 to 8.0 (Figure 8.1). The experiment indicates that mTET1 CD is most active at pH 6.5, with moderately reduced activity at pH 7.0 and 7.5 and a complete loss of activity at pH 8.0. It is not known how the activity of mTET1 CD is affected by pH below 6.5.

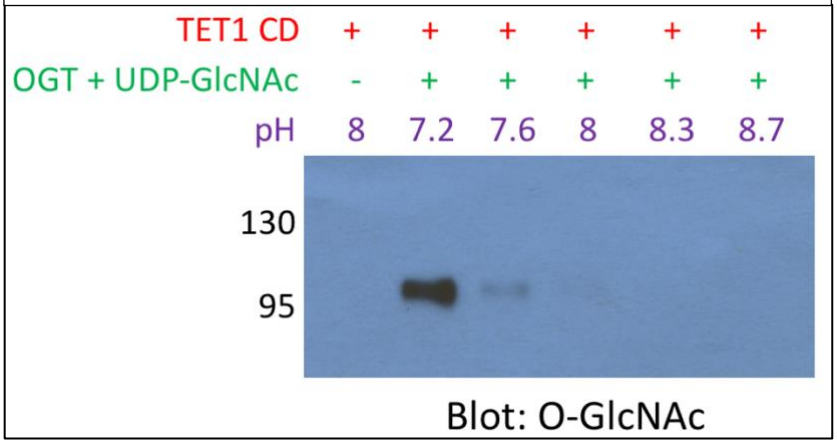
Figure 8.1 – mTET1 CD activity varies between pH 6.5 and 8.0
5hmC slot blot of *in vitro* reactions containing methylated lambda DNA, mTET1 CD, and TET cofactors at varying pH.



pH dependence of OGT activity

The activity of recombinant OGT toward recombinant mTET1 CD over a pH range from 7.2 to 8.7 was interrogated by *O*-GlcNAc western blot (Figure 8.2). OGT was by far the most active at pH 7.2, with a severe drop in activity at pH 7.6 and no detectable activity at pH 8.0-8.7. OGT's activity at pH less than 7.2 has not been explored.

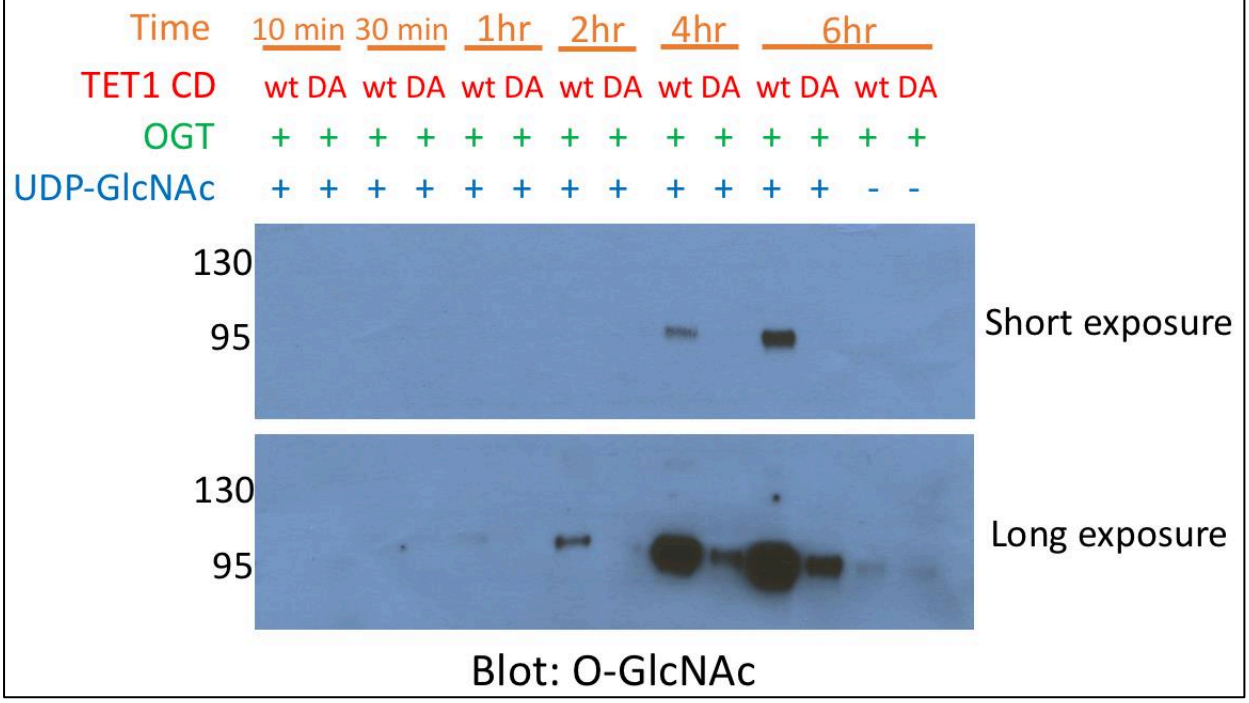
Figure 8.2 – OGT activity varies between pH 7.2 and 8.7
O-GlcNAc western blot of *in vitro* reactions containing mTET1 CD, OGT, and UDP-GlcNAc at varying pH.



Nonlinear time dependence of OGT activity

In order to determine the incubation time necessary for robust and specific *in vitro O*-GlcNAcylation of mTET1 CD, the activity of recombinant OGT toward recombinant mTET1 CD wild-type and D2018A was measured over a timecourse from 10 minutes to 6 hours (Figure 8.3). At all timepoints *O*-GlcNAcylation of wild-type TET was far more efficient than D2018A, consistent with the disrupted interaction between OGT and TET1 D2018A. However, unexpectedly, the activity of OGT toward both wild-type and D2018A TET1 was nonlinear with time. TET1 wild-type, for example, was not detectably modified at all until 1 hour, with *O*-GlcNAcylation increasing substantially between 1 and 2 hours, followed by a massive increase between 2 and 4 hours, and a further increase of at least 5-fold between 4 and 6 hours. There are multiple feasible models that could explain this behavior. Because there are at least 10 sites of *O*-GlcNAcylation on mTET1 CD (chapter 4), the dynamics of OGT's activity toward TET1 are likely to be complex. One possibility is that TET1 CD *O*-GlcNAcylation could initially be very slow, but once modification at one or a few sites has occurred, further *O*-GlcNAcylation of other sites or other TET molecules may be much more favorable. Another factor is that different *O*-GlcNAc sites may be unequally accessible to antibody binding on a membrane, such that western blot signal may not linearly scale with total TET1 *O*-GlcNAcylation. Exploring the biochemistry responsible for this result would require significant experimental effort and more sophisticated methods for measuring TET1 *O*-GlcNAcylation at particular sites.

Figure 8.3 – *In vitro O*-GlcNAcylation of mTET1 CD varies non-linearly with time
O-GlcNAc western blot of *in vitro* reactions containing mTET1 CD wild-type or D2018A, OGT, and UDP-GlcNAc over a timecourse from 10 minutes to 6 hours.



Discussion

These data explore the effects on the *in vitro* activities of TET1 CD and OGT of pH and time. This analysis should be useful for the design of future experiments investigating TET1 and OGT enzymology.

Appendix

Supplemental figures and tables associated with chapter 2

Figure S2.1 – Generation of mESC lines
A) Schematic of mESC lines. DNA encoding a 3xFLAG tag was added to the 3' end of both alleles of *Tet1*, followed by a 2A sequence and a fluorescent protein (GFP or tdTomato). The 2A sequence causes ribosome skipping, resulting in separate translation of TET1-3xFLAG and 2A-GFP or 2A-tdTomato. Purple line: template used for homology-directed repair (HDR). Horizontal arrows: primers used for PCR genotyping. Vertical arrows: D2018 residue. B) PCR genotyping of independently derived, clonal, targeted mESC lines using primers indicated in A.

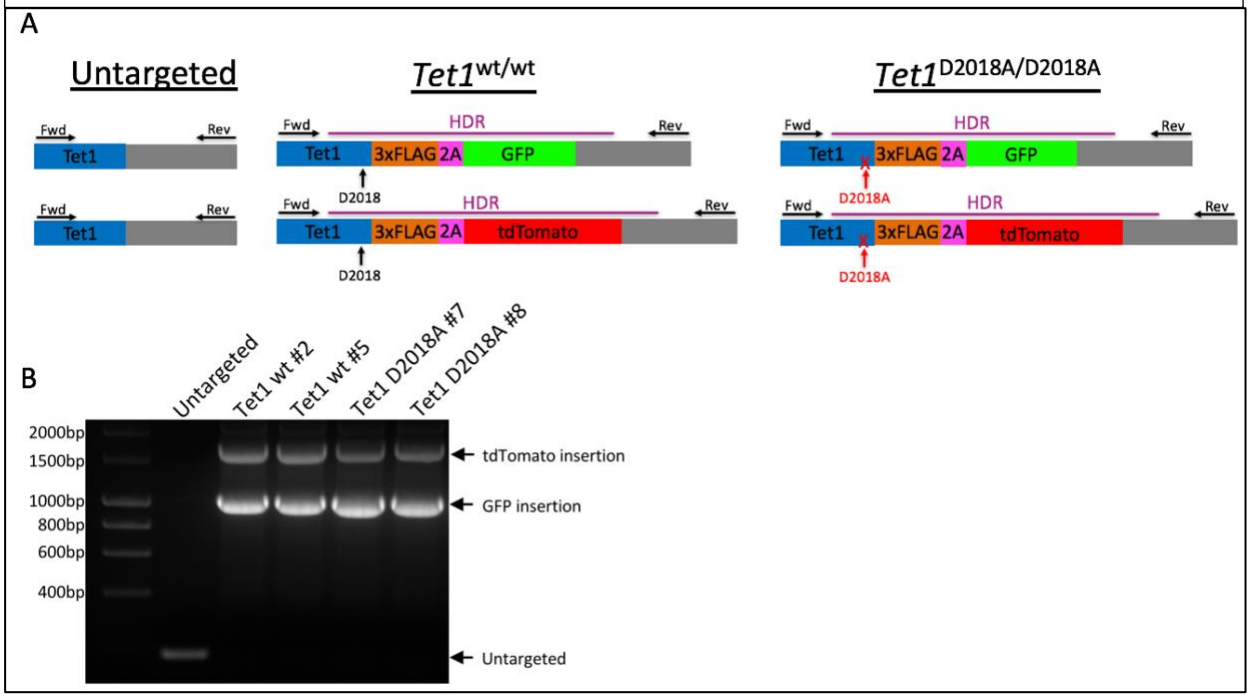


Figure S2.2 – Analysis of 25kb deletion in WT cells
A) Coverage from WGBS-seq showing ~25kb deleted in WT cells. B) qPCR comparing gene expression in WT, D2018A, and WT single targeted (het-1 and het-2) cell lines. Het-2 cells have wild-type TET1 and one intact copy of the 25kb region. C) Mass spec comparing 5mC levels in WT, D2018A, and WT single targeted (het-1 and het-2) cells (*P<0.05).

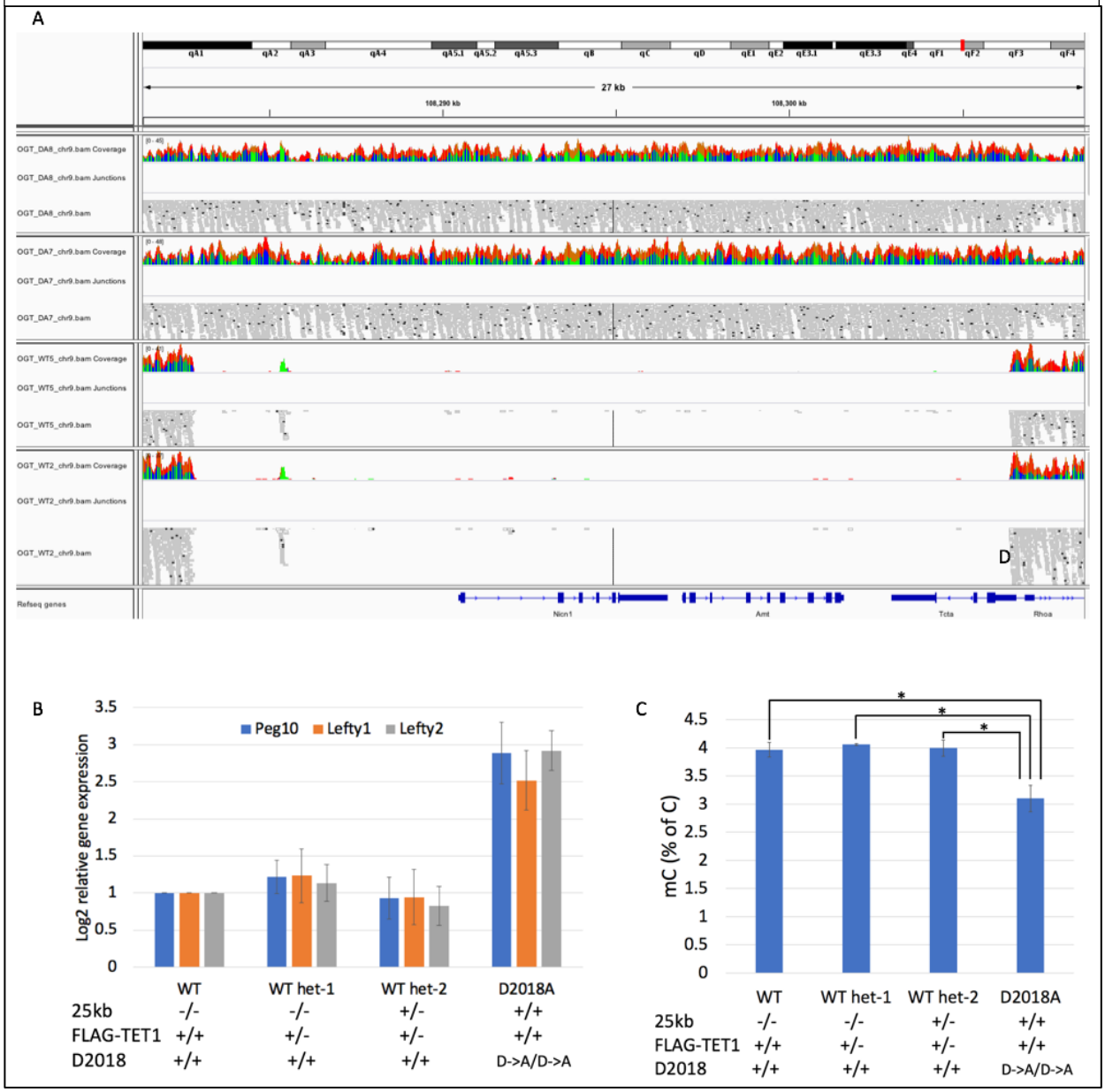


Figure S2.3 – Analysis of TET2 protein stability
TET2 western blots of protein lysates from WT and D2018A mESCs treated with cycloheximide (50ug/mL) at indicated timepoints.

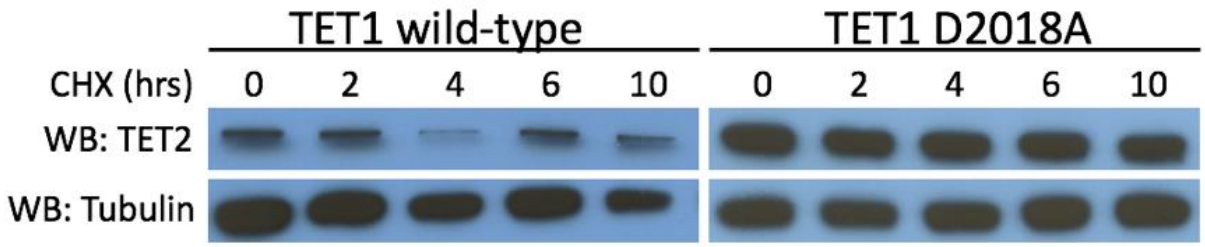


Table S1:
Primers used for creating and genotyping mESC lines

Name	Purpose	Sequence
WtAmpFwd	Forward primer for amplifying Tet1 wt Gene Blocks to make HDR template	atcaaccttaacccgagaca
MutAmpFwd	Forward primer for amplifying Tet1 D2018A Gene Blocks to make HDR template	tcaaccttaacccgagcc
AmpRev	Reverse primer for amplifying Tet1 wt and D2018A Gene Blocks to make HDR template	ctttttaacagcaccggaaa
GenotypeFwd	Forward primer for genotyping Tet1 allele	tgatgtatcccccgaagc
GenotypeRev	Reverse primer for genotyping Tet1 allele	cccactacaccacattagca

Table S2:
Gene blocks amplified to make HDR templates

Name	Sequence
Tet1 wild type-3xF-T2A-GFP	gcagaccgggagtgtcctgatgtatcccccgaagccaatttatcacaccaaattccttctcgagttgcatcaa ccttaacccgagacaatgttgttaccgtgtccccatactctctcactcatgttgcgggaccatacaatcgttgg gtcgactacaaagaccatgacggtgattataaagatcatgatatcgattacaaggatgacgatgacaaggg aagcggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggacctgtgagca agggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccac aagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaaatttatttgca cgacagggaagctgcccgtgccctggcccaccctcgttacgaccctaacatatggcgtgcagtgcttcag ccgctacccggatcatatgaagcaacacgacttctttaagtcagccatgcccgaaggctacgtccaggagc gcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctgg agtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaactt caagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccca tcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagacc ccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatgga cgagctgtacaagtaaaagcttctctcatgtaatgcatttgctaatgtggtgtagtgggtatttttgtttgtttgttt gttttcttttgtttttttgtttttttccggtgctgttaaaaagaaagtcattctgttgtttactgtagctttgtttcgcccat ttc
Tet1 D2018A-3xF-T2A-GFP	gcagaccgggagtgtcctgatgtatcccccgaagccaatttatcacaccaaattccttctcgagttgcatcaa ccttaacccgagccaatgttgttaccgtgtccccatactctctcactcatgttgcgggaccatacaatcgttgg gtcgactacaaagaccatgacggtgattataaagatcatgatatcgattacaaggatgacgatgacaaggg aagcggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggacctgtgagca agggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccac aagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaaatttatttgca cgacagggaagctgcccgtgccctggcccaccctcgttacgaccctaacatatggcgtgcagtgcttcag ccgctacccggatcatatgaagcaacacgacttctttaagtcagccatgcccgaaggctacgtccaggagc gcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccc tggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctgg agtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaactt caagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacaccccca tcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagacc ccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatgga cgagctgtacaagtaaaagcttctctcatgtaatgcatttgctaatgtggtgtagtgggtatttttgtttgtttgttt gttttcttttgtttttttgtttttttccggtgctgttaaaaagaaagtcattctgttgtttactgtagctttgtttcgcccat ttc

Name	Sequence
Tet1 wild type-3xF-T2A-tdTomato	gcagaccgggagtgtcctgatgtatcccccgaagccaatttatcacaccaaattccttctcgagttgcatcaa ccttaacccgagacaatgttgttaccgtgtccccatactctctcactcatgttgcgggaccatacaatcgttgg gtcgactacaaagaccatgacggtgattataaagatcatgatatcgattacaaggatgacgatgacaaggg aagcggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggacctgtttccaa aggggaggaagtcattaaggaatttatgaggttcaaagtgcgcatggagggatctatgaacggccacgaa tttgagatagaaggcgaaggcgagggaaggccctacgagggcactcagactgctaagctgaaagtaact aagggtggtcctctgcctttcgcctgggatatcctgtcaccccagtttatgtacggtagtaaagcttatgtgaa gcatcccgctgatatacctgactataaaaaactgtccttcccagagggcttcaagtgggagcgagtaatgaa ctttgaagatggtggactggttaccgttacccaagattcatctttgcaggacggaacattgatctacaaggtc aagatgcggggcactaacttcccacccgacgggccagtcatgcagaagaagactatgggctgggaagct agtactgagcgactctaccctagagatggtgtcttgaaaggggagattcatcaagcactgaaattgaaaga cggcggtcattacctcgtcgaattcaaaaccatatacatggccaaaaagcctgtgcaactgccagggtatta ttatgtcgacacaaaactcgatataaccagccataatgaagattataccatagtcgaacaatatgaacgctct gaaggacgacatcatttgttcttgggacatgggactggatccacaggatccggttcctctggaacagcatcc tccgaagacaataatatggccgtaataaaagaattcatgcgattcaaagtgagaatggaaggaagtatgaat ggtcacgagtttgaaatcgagggagaaggagagggtcggccctatgagggtacacagacagctaagttg aaggttactaagggcggccctcttcccttgcttgggatattctctccccacaattcatgtacgggtccaaggc ttacgtaaaacatcccgctgatatacccgattacaaaaaactgtccttccccgaaggctttaaatgggaaagg gtgatgaatttcgaggacgggggattggtaactgtcacacaggattcctctcttcaagatggaacactgattt acaaggtaaaaatgagagggaccaactttccccctgatgggcccgtgatgcaaaagaaaaccatgggctg ggaagcatctaccgagagactttatcccagggacggcgttcttaagggagagattcaccaagctttgaaac ttaaggatggaggtcactacctcgtggagtttaagacaatatatatggcaaaaaaaccagtccaactccccg gatactattacgttgataccaaactggacataacttctcataacgaggactacactatagtggaacaatatga acgctctgagggtcgacaccacctttttcctgtatggaatggatgaactgtataagtagtaaaagcttctctcat gtaatgcatttgctaatgtggtgtagtgggtatttttgtttgtttgtttgttttctttttgttttttttgttttttccggtgctgt taaaaagaaagtcattctgttgtttactgtagctttgtttcgcccatttc

Name	Sequence
Tet1 D2018A-3xF-T2A-tdTomato	gcagaccgggagtgtcctgatgtatcccccgaagccaatttatcacaccaaattccttctcgagttgcatcaa ccttaacccgagccaatgttgttaccgtgtccccatactctctcactcatgttgcgggaccatacaatcgttgg gtcgactacaaagaccatgacggtgattataaagatcatgatatcgattacaaggatgacgatgacaaggg aagcggagagggcagaggaagtctgctaacatgcggtgacgtcgaggagaatcctggacctgtttccaa aggggaggaagtcattaaggaatttatgaggttcaaagtgcgcatggagggatctatgaacggccacgaa tttgagatagaaggcgaaggcgagggaaggccctacgagggcactcagactgctaagctgaaagtaact aagggtggtcctctgcctttcgcctgggatatcctgtcaccccagtttatgtacggtagtaaagcttatgtgaa gcatcccgctgatatacctgactataaaaaactgtccttcccagagggcttcaagtgggagcgagtaatgaa ctttgaagatggtggactggttaccgttacccaagattcatctttgcaggacggaacattgatctacaaggtc aagatgcggggcactaacttcccacccgacgggccagtcatgcagaagaagactatgggctgggaagct agtactgagcgactctaccctagagatggtgtcttgaaaggggagattcatcaagcactgaaattgaaaga cggcggtcattacctcgtcgaattcaaaaccatatacatggccaaaaagcctgtgcaactgccagggtatta ttatgtcgacacaaaactcgatataaccagccataatgaagattataccatagtcgaacaatatgaacgctct gaaggacgacatcatttgttcttgggacatgggactggatccacaggatccggttcctctggaacagcatcc tccgaagacaataatatggccgtaataaaagaattcatgcgattcaaagtgagaatggaaggaagtatgaat ggtcacgagtttgaaatcgagggagaaggagagggtcggccctatgagggtacacagacagctaagttg aaggttactaagggcggccctcttcccttgcttgggatattctctccccacaattcatgtacgggtccaaggc ttacgtaaaacatcccgctgatatacccgattacaaaaaactgtccttccccgaaggctttaaatgggaaagg gtgatgaatttcgaggacgggggattggtaactgtcacacaggattcctctcttcaagatggaacactgattt acaaggtaaaaatgagagggaccaactttccccctgatgggcccgtgatgcaaaagaaaaccatgggctg ggaagcatctaccgagagactttatcccagggacggcgttcttaagggagagattcaccaagctttgaaac ttaaggatggaggtcactacctcgtggagtttaagacaatatatatggcaaaaaaaccagtccaactccccg gatactattacgttgataccaaactggacataacttctcataacgaggactacactatagtggaacaatatga acgctctgagggtcgacaccacctttttcctgtatggaatggatgaactgtataagtagtaaaagcttctctcat gtaatgcatttgctaatgtggtgtagtgggtatttttgtttgtttgtttgttttctttttgttttttttgttttttccggtgctgt taaaaagaaagtcattctgttgtttactgtagctttgtttcgcccatttc

Table S3:
Primers used for qPCR

Gene	Primers
Peg10	Fwd: gaatcctcgtgtggaacag
	Rev: cagttggaggaaccaccc
Slc38a4	Fwd: gccaaggaaggagggtctc
	Rev: ggctccaatgttctgcattg
Lefty1	Fwd: ctatggagctcaaggcaatt
	Rev: gttctaggatccagttctcg
Lefty2	Fwd: ggttcctgacgtatgaatgt
	Rev: ctccttcacactgacaatca
Tet2	Fwd: gtcaacaggacatgatccaggag
	Rev: cctgttccatcaggcttgct

References

Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H. E., Iurlaro, M., Reik, W., et al. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nature Publishing Group*, *11*(8), 555–557. http://doi.org/10.1038/nchembio.1848

Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., & Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nature Chemistry*, *6*(12), 1049–1055. http://doi.org/10.1038/nchem.2064

Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Müller, U., et al. (2015). Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *The Journal of Biological Chemistry*, *290*(8), 4801–4812. http://doi.org/10.1074/jbc.M114.605881

Baylin, S. B., & Jones, P. A. (2011). A decade of exploring the cancer epigenome — biological and translational implications, 1–9. http://doi.org/10.1038/nrc3130

Bullen, J. W., Balsbaugh, J. L., Chanda, D., Shabanowitz, J., Hunt, D. F., Neumann, D., & Hart, G. W. (2014). Cross-talk between two essential nutrient-sensitive enzymes: O-GlcNAc transferase (OGT) and AMP-activated protein kinase (AMPK). *The Journal of Biological Chemistry*, *289*(15), 10592–10606. http://doi.org/10.1074/jbc.M113.523068

Chen, Q., Chen, Y., Bian, C., Fujiki, R., & Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, *493*(7433), 561–564. http://doi.org/10.1038/nature11742

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, *146*(1), 67–79. http://doi.org/10.1016/j.cell.2011.06.020

Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental Cell*, *29*(1), 102–111. http://doi.org/10.1016/j.devcel.2014.03.003

Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Developmental Cell*, *24*(3), 310–323. http://doi.org/10.1016/j.devcel.2012.12.015

Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y.-C., Markoulaki, S., Cheng, A. W., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, *9*(2), 166–175. http://doi.org/10.1016/j.stem.2011.07.010

Delatte, B., Deplus, R., & Fuks, F. (2014). Playing TETris with DNA modifications. *The EMBO Journal*, *33*(11), 1198–1211. http://doi.org/10.15252/embj.201488290

Deplus, R., Delatte, B., Schwinn, M. K., Defrance, M., Méndez, J., Murphy, N., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO Journal*, *32*(5), 645–655. http://doi.org/10.1038/emboj.2012.357

Durning, S. P., Flanagan-Steet, H., Prasad, N., & Wells, L. (2016). O-Linked β-N-acetylglucosamine (O-GlcNAc) Acts as a Glucose Sensor to Epigenetically Regulate the Insulin Gene in Pancreatic Beta Cells. *The Journal of Biological Chemistry*, *291*(5), 2107–2118. http://doi.org/10.1074/jbc.M115.693580

Fidalgo, M., Huang, X., Guallar, D., Sanchez-Priego, C., Valdes, V. J., Saunders, A., et al. (2016). Zfp281 Coordinates Opposing Functions of Tet1 and Tet2 in Pluripotent States. *Cell Stem Cell*, *19*(3), 355–369. http://doi.org/10.1016/j.stem.2016.05.025

Frauer, C., Hoffmann, T., Hoffmann, T., Bultmann, S., Casa, V., Casa, V., et al. (2011a). Recognition of 5-Hydroxymethylcytosine by the Uhrf1 SRA Domain. *PLoS ONE*, *6*(6), e21306. http://doi.org/10.1371/journal.pone.0021306.s010

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., et al. (2011b). Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1. *PLoS ONE*, *6*(2), e16627. http://doi.org/10.1371/journal.pone.0016627.s001

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., et al. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell*, *12*(4), 453–469. http://doi.org/10.1016/j.stem.2013.02.005

Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., et al. (2010). Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE*, *5*(12), e15367. http://doi.org/10.1371/journal.pone.0015367.s005

Guibert, S., & Weber, M. (2013). Functions of DNA methylation and hydroxymethylation in mammalian development. *Current Topics in Developmental Biology*, *104*, 47–83. http://doi.org/10.1016/B978-0-12-416027-9.00002-4

Guo, J. U., Su, Y., Zhong, C., Ming, G.-L., & Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, *145*(3), 423–434. http://doi.org/10.1016/j.cell.2011.03.022

Haltiwanger, R. S., Holt, G. D., & Hart, G. W. (1990). Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *Journal of Biological Chemistry*, *265*(5), 2563–2568.

Hanover, J. A., Forsythe, M. E., Hennessey, P. T., Brodigan, T. M., Love, D. C., Ashwell, G., & Krause, M. (2005). A Caenorhabditis elegans model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proceedings of the National Academy of Sciences*, *102*(32), 11266–11271. http://doi.org/10.1073/pnas.0408771102

Hanover, J. A., Krause, M. W., & Love, D. C. (2012). Bittersweet memories: linking metabolism to epigenetics through O-GlcNAcylation. *Nature Reviews. Molecular Cell Biology*, *13*(5), 312–321. http://doi.org/10.1038/nrm3334

Hardivillé, S., & Hart, G. W. (2016). Nutrient regulation of gene expression by O-GlcNAcylation of chromatin. *Current Opinion in Chemical Biology*, *33*, 88–94. http://doi.org/10.1016/j.cbpa.2016.06.005

Hart, G. W., Slawson, C., Ramirez-Correa, G., & Lagerlof, O. (2011). Cross talk between O-GlcNAcylation and phosphorylation: roles in signaling, transcription, and chronic disease. *Annual Review of Biochemistry*, *80*, 825–858. http://doi.org/10.1146/annurev-biochem-060608-102511

He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, *333*(6047), 1303–1307. http://doi.org/10.1126/science.1210944

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., et al. (2013). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell*, *155*(7), 1545–1555. http://doi.org/10.1016/j.cell.2013.11.020

Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences*, *111*(4), 1361–1366. http://doi.org/10.1073/pnas.1322921111

Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., et al. (2014). TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes to Cells*, *19*(1), 52–65. http://doi.org/10.1111/gtc.12107

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129–1133. http://doi.org/10.1038/nature09303

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300–1303. http://doi.org/10.1126/science.1210597

Kalev-Zylinska, M. L., Horsfield, J. A., Flores, M. V. C., Postlethwait, J. H., Vitas, M. R., Baas, A. M., et al. (2002). Runx1 is required for zebrafish blood and vessel development and expression of a human RUNX1-CBF2T1 transgene advances a model for studies of leukemogenesis. *Development*, *129*(8), 2015–2030.

Ko, M., Huang, Y., Jankowska, A. M., Pape, U. J., Tahiliani, M., Bandukwala, H. S., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, *468*(7325), 839–843. http://doi.org/10.1038/nature09586

Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, *8*(2), 200–213. http://doi.org/10.1016/j.stem.2011.01.008

Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, *324*(5929), 929–930. http://doi.org/10.1126/science.1169786

Lazarus, M. B., Jiang, J., Kapuria, V., Bhuiyan, T., Janetzko, J., Zandberg, W. F., et al. (2013). HCF-1 is cleaved in the active site of O-GlcNAc transferase. *Science*, *342*(6163), 1235–1239. http://doi.org/10.1126/science.1243990

Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P., & Walker, S. (2011). Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature*, *469*(7331), 564–567. http://doi.org/10.1038/nature09638

Levine, Z. G., & Walker, S. (2016). The Biochemistry of O-GlcNAc Transferase: Which Functions Make It Essential in Mammalian Cells? *Annual Review of Biochemistry*, *85*, 631–657. http://doi.org/10.1146/annurev-biochem-060713-035344

Lewis, B. A., & Hanover, J. A. (2014). O-GlcNAc and the epigenetic regulation of gene expression. *The Journal of Biological Chemistry*, *289*(50), 34440–34448. http://doi.org/10.1074/jbc.R114.595439

Li, C., Lan, Y., Schwartz-Orbach, L., Korol, E., Tahiliani, M., Evans, T., & Goll, M. G. (2015). Overlapping Requirements for Tet2 and Tet3 in Normal Development and Hematopoietic Stem Cell Emergence. *CellReports*, *12*(7), 1133–1143. http://doi.org/10.1016/j.celrep.2015.07.025

Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B. E., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*, *118*(17), 4509–4518. http://doi.org/10.1182/blood-2010-12-325241

Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., et al. (2012). Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell*, *150*(6), 1135–1146. http://doi.org/10.1016/j.cell.2012.07.033

Lister, R., Mukamel, E. A., Nery, J. R., Urich, M., Puddifoot, C. A., Johnson, N. D., et al. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science*, *341*(6146), 1237905–1237905. http://doi.org/10.1126/science.1237905

Ma, H., Morey, R., O'Neil, R. C., He, Y., Daughtry, B., Schultz, M. D., et al. (2014). Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature*, *511*(7508), 177–183. http://doi.org/10.1038/nature13551

Marshall, S., Nadeau, O., & Yamasaki, K. (2004). Dynamic Actions of Glucose and Glucosamine on Hexosamine Biosynthesis in Isolated Adipocytes: DIFFERENTIAL EFFECTS ON GLUCOSAMINE 6-PHOSPHATE, UDP-N-ACETYLGLUCOSAMINE, AND ATP LEVELS. *Journal of Biological Chemistry*, *279*(34), 35313–35319. http://doi.org/10.1074/jbc.M404133200

Maynard, J. C., Burlingame, A. L., & Medzihradszky, K. F. (2016). Cysteine S-linked N-acetylglucosamine (S-GlcNAcylation), A New Post-translational Modification in Mammals. *Molecular & Cellular Proteomics*, *15*(11), 3405–3411. http://doi.org/10.1074/mcp.M116.061549

McClain, D. A. (2002). Hexosamines as mediators of nutrient sensing and regulation in diabetes. *Journal of Diabetes and Its Complications*, *16*, 72–80.

Müller, U., Bauer, C., Siegl, M., Rottach, A., & Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Research*, *42*(13), 8592–8604. http://doi.org/10.1093/nar/gku552

Myers, S. A., Panning, B., & Burlingame, A. L. (2011a). Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences*, *108*(23), 9490–9495. http://doi.org/10.1073/pnas.1019289108

Myers, S. A., Panning, B., & Burlingame, A. L. (2011b). Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(23), 9490–9495. http://doi.org/10.1073/pnas.1019289108

Myers, S. A., Peddada, S., Chatterjee, N., Friedrich, T., Tomoda, K., Krings, G., et al. (2016). SOX2 O-GlcNAcylation alters its protein- protein interactions and genomic occupancy to modulate gene expression in pluripotent cells, 1–20. http://doi.org/10.7554/eLife.10647.001

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308. http://doi.org/10.1038/nprot.2013.143

Ruan, H.-B., Singh, J. P., Li, M.-D., Wu, J., & Yang, X. (2013). Cracking the O-GlcNAc code in metabolism. *Trends in Endocrinology and Metabolism: TEM*, *24*(6), 301–309. http://doi.org/10.1016/j.tem.2013.02.002

Sakabe, K., Wang, Z., & Hart, G. W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(46), 19915–19920. http://doi.org/10.1073/pnas.1009023107

Schilling, B., Rardin, M. J., MacLean, B. X., Zawadzka, A. M., Frewen, B. E., Cusack, M. P., et al. (2012). Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline. *Molecular & Cellular Proteomics*, *11*(5), 202–214. http://doi.org/10.1074/mcp.M112.017707

Shafi, R., Iyer, S. P. N., Ellies, L. G., O'Donnell, N., Marek, K., Chiu, D., et al. (2000). The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonicstem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences*, *97*(11), 5735–5739.

Shen, D. L., Gloster, T. M., Yuzwa, S. A., & Vocadlo, D. J. (2012). Insights into O-linked N-acetylglucosamine ([0-9]O-GlcNAc) processing and dynamics through kinetic analysis of O-GlcNAc transferase and O-GlcNAcase activity on protein substrates. *The Journal of Biological Chemistry*, *287*(19), 15395–15408. http://doi.org/10.1074/jbc.M111.310664

Shi, F.-T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M. A., et al. (2013). Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *The Journal of Biological Chemistry*, *288*(29), 20776–20784. http://doi.org/10.1074/jbc.M113.460386

Smith, Z. D., & Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220. http://doi.org/10.1038/nrg3354

Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell*, *152*(5), 1146–1159. http://doi.org/10.1016/j.cell.2013.02.004

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, *324*(5929), 930–935. http://doi.org/10.1126/science.1170116

Thisse, C., & Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nature Protocols*, *3*(1), 59–69. http://doi.org/10.1038/nprot.2007.514

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., et al. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Molecular Cell*, *49*(4), 645–656. http://doi.org/10.1016/j.molcel.2012.12.019

Vocadlo, D. J. (2012). O-GlcNAc processing enzymes: catalytic mechanisms, substrate specificity, and enzyme regulation. *Current Opinion in Chemical Biology*, *16*(5-6), 488–497. http://doi.org/10.1016/j.cbpa.2012.10.021

Weber, A. R., Krawczyk, C., Robertson, A. B., Kuśnierczyk, A., Vågbø, C. B., Schuermann, D., et al. (2016). Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nature Communications*, *7*, 10806. http://doi.org/10.1038/ncomms10806

Webster, D. M., Webster, D. M., Teo, C., Teo, C., Sun, Y., Sun, Y., et al. (2009). O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. *BMC Developmental Biology*, *9*(1), 28. http://doi.org/10.1186/1471-213X-9-28

Weigert, C., Klopfer, K., Kausch, C., Brodbeck, K., Stumvoll, M., Haring, H., & Schleicher, E. (2003). Palmitate-Induced Activation of the Hexosamine Pathway in Human Myotubes Increased Expression of Glutamine:Fructose-6-Phosphate Aminotransferase. *Diabetes*, *52*, 1–7.

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A. C., Rappsilber, J., & Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, *473*(7347), 343–348. http://doi.org/10.1038/nature10066

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Molecular Cell*, *42*(4), 451–464. http://doi.org/10.1016/j.molcel.2011.04.005

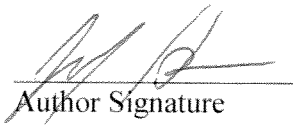
Zhao, Z., Chen, L., Dawlaty, M. M., Pan, F., Weeks, O., Zhou, Y., et al. (2015). Combined Loss of Tet1 and Tet2 Promotes B Cell, but Not Myeloid Malignancies, in Mice. *CellReports*, *13*(8), 1692–1704. http://doi.org/10.1016/j.celrep.2015.10.037

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

9/5/18

Date