1-1-2016

Epigenetic Regulation of Embryonic and Intestinal Stem Cells by DNA Hydroxymethylation

Rinho Kim
University of Pennsylvania, rinho.kim1@gmail.com

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Bioinformatics Commons, Developmental Biology Commons, and the Genetics Commons

Recommended Citation
Kim, Rinho, "Epigenetic Regulation of Embryonic and Intestinal Stem Cells by DNA Hydroxymethylation" (2016). Publicly Accessible Penn Dissertations. 1814.
http://repository.upenn.edu/edissertations/1814

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/1814
For more information, please contact libraryrepository@pobox.upenn.edu.
Epigenetic Regulation of Embryonic and Intestinal Stem Cells by DNA Hydroxymethylation

Abstract
DNA methylation at the 5 position of cytosine is a well-characterized epigenetic modification that is important for essential cellular processes including stem cell proliferation and differentiation. 5-methylcytosine (5mC) can be oxidized by the Ten Eleven Translocation (TET) family of enzymes to 5-hydroxymethylcytosine (5hmC), which has been suggested as either an intermediate in DNA demethylation or as an independent epigenetic modification that directly regulates gene expression. Gene silencing and activation are important processes in embryonic stem cells (ESCs) and intestinal stem cells (ISCs) for their self-renewal and differentiation, but our understanding of the potential role of 5hmC and TET hydroxylases in these processes is still limited.

Here, I generate genome-wide maps of the 5hmC mark in ISCs and their differentiated progeny and utilize Tet1-/- mice model to investigate the role of TET1 in the intestinal epithelium. Genes with high levels of hydroxymethylation in ISCs are strongly enriched for developmental regulation functions. The Tet1-deficient postnatal intestine shows significantly reduced numbers of proliferative cells and decreased Wnt target genes expression, which correlates with lower 5hmC levels at their promoters. These data demonstrate that Tet1-mediated DNA hydroxymethylation is important for the self-renewal of the intestinal epithelium.

To determine a novel role of 5hmC, we perform integrative analyses of various epigenomic sequencing datasets, and identify a group of distal transcription factor binding sites as putative silenced enhancers, which are lacking active histone marks and nascent transcription, but are nevertheless highly enriched for the 5hmC modification in ESCs. During lineage specification of ESCs, these silenced elements lose 5hmC and acquire the H3K4me1/2 histone marks and become active. I demonstrate that these elements function as enhancers when the 5hmC mark is removed by a reporter assays. These data suggest that 5hmC suppresses the activity of a specific subset of enhancers in ESCs.

In summary, DNA hydroxymethylation contributes to both transcriptional activation and repression for epigenetic gene regulation in stem cells.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Cell & Molecular Biology

First Advisor
Klaus H. Kaestner

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1814
Keywords
5-hydroxymethylcytosine (5hmC), Epigenetics and DNA hydroxymethylation, Intestinal stem cell specification and development, Next generation sequencing, Stem cell gene regulation, Ten-eleven translocation 1 (Tet1)

Subject Categories
Bioinformatics | Developmental Biology | Genetics

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/1814
EPIGENETIC REGULATION OF EMBRYONIC AND INTESTINAL STEM CELLS
BY DNA HYDROXYMETHYLATION

Rinho Kim

A DISSERTATION
in
Cell and Molecular Biology
Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2016

Supervisor of Dissertation
____________________
Klaus H. Kaestner, Ph.D.
Thomas and Evelyn Suor Butterworth Professor in Genetics

Graduate Group Chairperson
____________________
Daniel S. Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee
Zhaolan (Joe) Zhou, Ph.D. Associate Professor of Genetics
Christopher J. Lengner, Ph.D. Assistant Professor of Biomedical Sciences
Kenneth S. Zaret, Ph.D. Joseph Leidy Professor of Cell and Developmental Biology
John P. Lynch, M.D., Ph.D. Associate Professor of Medicine
Dedicated to my grandfather, Mr. Dae-hyun Kim.
ABSTRACT

Epigenetic Regulation of Embryonic and Intestinal Stem Cells

by DNA hydroxymethylation

Rinho Kim
Dr. Klaus H. Kaestner

DNA methylation at the 5 position of cytosine is a well-characterized epigenetic modification that is important for essential cellular processes including stem cell proliferation and differentiation. 5-methylcytosine (5mC) can be oxidized by the Ten Eleven Translocation (TET) family of enzymes to 5-hydroxymethylcytosine (5hmC), which has been suggested as either an intermediate in DNA demethylation or as an independent epigenetic modification that directly regulates gene expression. Gene silencing and activation are important processes in embryonic stem cells (ESCs) and intestinal stem cells (ISCs) for their self-renewal and differentiation, but our understanding of the potential role of 5hmC and TET hydroxylases in these processes is still limited.

Here, I generate genome-wide maps of the 5hmC mark in ISCs and their differentiated progeny and utilize Tet1−/− mice model to investigate the role of TET1 in the intestinal epithelium. Genes with high levels of hydroxymethylation in ISCs are strongly enriched for developmental regulation functions. The Tet1-deficient postnatal intestine shows significantly reduced numbers of proliferative cells and decreased Wnt
target genes expression, which correlates with lower 5hmC levels at their promoters. These data demonstrate that Tet1-mediated DNA hydroxymethylation is important for the self-renewal of the intestinal epithelium.

To determine a novel role of 5hmC, we perform integrative analyses of various epigenomic sequencing datasets, and identify a group of distal transcription factor binding sites as putative silenced enhancers, which are lacking active histone marks and nascent transcription, but are nevertheless highly enriched for the 5hmC modification in ESCs. During lineage specification of ESCs, these silenced elements lose 5hmC and acquire the H3K4me1/2 histone marks and become active. I demonstrate that these elements function as enhancers when the 5hmC mark is removed by a reporter assays. These data suggest that 5hmC suppresses the activity of a specific subset of enhancers in ESCs.

In summary, DNA hydroxymethylation contributes to both transcriptional activation and repression for epigenetic gene regulation in stem cells.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................ iii

LIST OF TABLES ............................................................................................................................... vi

LIST OF ILLUSTRATIONS .............................................................................................................. vii

CHAPTER 1 ....................................................................................................................................... 1

Introduction ...................................................................................................................................... 1

1.1 Development and structure of the intestinal epithelium .......................................................... 2

1.2 Wnt signaling in intestinal development and homeostasis ..................................................... 5

1.3 Intestinal stem cells and organoids .......................................................................................... 9

1.4 Epigenetic regulations in the intestinal epithelium ................................................................. 14

1.5 TET-mediated DNA hydroxymethylation in DNA demethylation pathway ......................... 20

1.6 An Independent Role of 5hmC in Gene Regulation ............................................................... 23

1.7 Reference. ................................................................................................................................ 25

CHAPTER 2 ....................................................................................................................................... 37

Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation ................................................................................................................................. 37

2.1 Summary ................................................................................................................................... 38

2.2 Background ............................................................................................................................... 39

2.3 Methods ..................................................................................................................................... 41

2.4 Results ....................................................................................................................................... 47

2.5 Discussion ................................................................................................................................. 63

2.6 Reference ................................................................................................................................... 65

CHAPTER 3 ....................................................................................................................................... 70

5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation ........................................................................... 70

3.1 Summary ................................................................................................................................... 71

3.2 Background ............................................................................................................................... 72

3.3 Methods ..................................................................................................................................... 74

3.4 Results ....................................................................................................................................... 77

3.5 Discussion ................................................................................................................................. 95

3.6 Reference ................................................................................................................................... 99

CHAPTER 4 ....................................................................................................................................... 105

Conclusions and Future direction .................................................................................................. 105

4.1 TET1-mediated DNA hydroxymethylation is required for the regulation of gene expression in intestinal stem cells. .............................................................. 106

4.2 5hmC can function as a repressive mark at distal transcription factor binding sites 109

4.3 The dual role of 5hmC in gene activation and repression ...................................................... 111

4.4 Reference ................................................................................................................................... 112
LIST OF TABLES

| Table 2.1 | RT-qPCR primers | 45 |
| Table 2.2 | hMeDIP-qPCR primers | 46 |
| Table 2.3 | Previous and current studies of Tet1 null mice | 56 |
| Table 3.1 | Datasets. | 75 |
| Table 3.2 | The frequency of transcription factor occupancy in cluster 2. | 85 |
| Figure 1.1 | Development of the intestine | 2 |
| Figure 1.2 | Villus and crypt formation | 3 |
| Figure 1.3 | Wnt signaling pathway | 6 |
| Figure 1.4 | Wnt signaling in intestinal homeostasis | 8 |
| Figure 1.5 | Intestinal stem cells: +4 and CBC stem cells. | 10 |
| Figure 1.6 | Intestinal stem cell hierarchy. | 11 |
| Figure 1.7 | Intestinal organoid structure. | 12 |
| Figure 1.8 | Intestinal organoids and spheroids. | 13 |
| Figure 1.9 | Two main components of epigenetic modification. | 14 |
| Figure 1.10 | Histone modifications in gene regulation. | 16 |
| Figure 1.11 | De novo and maintenance DNA methylation. | 17 |
| Figure 1.12 | DNA methylation changes in the intestinal epithelium. | 18 |
| Figure 1.13 | DNA methylation is essential for the postnatal intestinal maturation. | 19 |
| Figure 1.14 | 5mC oxidation and demethylation process. | 21 |
| Figure 1.15 | The dual role of 5hmC and TET in gene regulation. | 23 |
| Figure 2.1 | Differential enrichment of DNA methylation and hydroxymethylation in the intestinal epithelium. | 48 |
| Figure 2.2 | Isolation of Lgr5⁺ intestinal stem and villus differentiated cells. | 49 |
| Figure 2.3 | Differential 5hmC distributions in intestinal stem and differentiated cells. | 50 |
| Figure 2.4 | Marker gene expression level in Lgr5⁺ stem and differentiated cells. | 51 |
| Figure 2.5 | Overall 5hmC levels in Lgr5+ stem cells and differentiated cells. | 53 |
| Figure 2.6 | Dynamics of 5hmC and gene expression changes during intestinal stem cell differentiation. | 53 |
| Figure 2.7 | Differential expression of Tet1-3 in Lgr5⁺ stem and differentiated cells. | 55 |
| Figure 2.8 | Postnatal growth retardation and lethality of *Tet1* null mice. | 56 |
Figure 2.9  Developmental defects during postnatal intestinal maturation.  

Figure 2.10  Decreased level of Wnt target expression in the postnatal Tet1 null mice intestine.  

Figure 2.11  Reduction of Lgr5 expression in Tet1 null mice intestine.  

Figure 2.12  Mature organoid-forming potential is reduced in the absence of Tet1.  

Figure 2.13  Differential 5hmC level at the promoter of Wnt target genes.  

Figure 3.1  5hmC profile at promoters and enhancers.  

Figure 3.2  5hmC and other epigenetic modifications in ESCs.  

Figure 3.3  Comparison of the characteristics of each cluster.  

Figure 3.4  Comparison of the 5hmC patterns for each cluster.  

Figure 3.5  The 5hmC profile of cluster 2 using TAB-Seq.  

Figure 3.6  The 5hmC clusters in hESCs.  

Figure 3.7  The 5hmC clusters in mature adipocytes.  

Figure 3.8  The average profiles of TFs at cluster 2.  

Figure 3.9  Functional analysis for TFBSs with 5hmC.  

Figure 3.10  Chromatin organization at the Sorcs2 gene in mESCs.  

Figure 3.11  Lineage specific activation of distal TFBSs in cluster 2.  

Figure 3.12  The gene expression change for the target genes for each cluster.  

Figure 3.13  Enhancer activity of distal TFBSs lacking 5hmC in mESCs.  

Figure 3.14  The gene expression changes of the target genes after Tet1 knockdown for each cluster.  

Figure 3.15  The 5hmC in mESC and NPC at the TFBSs in mESCs.  

Figure 3.16  5hmC at CTCF binding sites in cluster 2.
CHAPTER 1

Introduction
1.1 Development and structure of the intestinal epithelium

The intestine is a simple tubular organ that connects the stomach to the anus, and is essential for the absorption of nutrients and water. In mice, the gastrointestinal tract originates from the endoderm at embryonic day (E) 7.5, and is regionalized along the anterior-posterior axis into foregut, midgut and hindgut, guided by epithelial-mesenchymal interactions at E 8.5 (Figure 1.1). The small intestine and colon are developed from the midgut and hindgut, respectively, while the foregut gives rise to the esophagus and stomach, as well as the gut-associated organs thyroid, ling, liver, and pancreas (Zorn and Wells, 2009).

![Figure 1.1 Development of the intestine.](image)

(Left) The major events during endoderm organ formation between E 7.5 to 9.5. (Right) Endoderm cell lineages of the gastrointestinal tract. Adapted from (Zorn and Wells, 2009).
The intestinal epithelium is structured as a pseudostratified epithelium until E 13.5 (Figure 1.2). After this point, mesenchymal clusters appear around epithelial cells, which secrete hedgehog (Hh) and platelet-derived growth factor A (Pdgfa), and drive villus morphogenesis at E 14.5 (Figure 1.2; (Karlsson et al., 2000; Madison et al., 2005; Walton et al., 2012)). During villus morphogenesis, the pseudostratified epithelium is transformed into a columnar epithelium with stereotypical villi that project into the lumen, and proliferative intervillus regions. The intervillus epithelium is reshaped to develop crypts in the first week after birth in mice. The maturation of villus and crypt continues until the weaning age, when mice are approximately 3-4 weeks old (Wells and Spence, 2014). During human development, villus and crypt development occur prenatally.

**Figure 1.2 Villus and crypt formation.** At E 14.5, mesenchymal clusters (orange) is required for villus morphogenesis. As the epithelium remolds between E 14.5 to 16.5, the fetal intestinal epithelium consist of villus (green) and proliferative intervillus (yellow) regions. The mature adult crypt contains the intestinal stem cells (yellow) and secretory Paneth cells (red), as well as several additional cell types, such as enterocytes (green), goblet cells (blue) and enteroendocrine cells (purple). Adapted from (Wells and Spence, 2014).
The mature small intestinal epithelium is composed of a single layer of cells that contains stem, transit-amplifying and four different lineages of differentiated cells, such as enterocytes, goblet cells, enteroendocrine cells and Paneth cells (Figure 1.2; (Cheng and Leblond, 1974b)). Enterocytes are absorptive cells, which have a brush border on their apical surface and also secrete hydrolytic enzymes essential for the digestion of food. Goblet cells secrete mucus, which lubricates and protects the intestinal wall from luminal contents. Enteroendocrine cells are rare and secrete various hormones such as peptides and catecholamines. Paneth cells reside at the bottom of the crypt and secrete antibacterial proteins such as lyzosomes and defensins in the small intestine, but not the colon.

The intestinal epithelium is the most rapidly renewing tissue in human body. This renewal process takes only five days to complete and is fueled by intestinal stem cells (Heath, 1996; Potten and Loeffler, 1990). Intestinal stem cells reside at the crypt base and give rise to rapid-cycling transit-amplifying cells that differentiate into mature epithelial cells migrating up along the crypt-villus axis, and that are shed at the top of the villi. Only Paneth cells move downward to the bottom of the crypt and have a longer life span of 6-8 weeks. Rapid renewal and distinctive morphology make the intestinal epithelium an excellent model to study the mechanisms underlying cell proliferation and differentiation in adult tissues. To maintain normal function of the intestinal epithelium, precise regulation of cell proliferation and differentiation is essential (Crosnier et al., 2006).
1.2 Wnt signaling in intestinal development and homeostasis

Wnt signaling is involved in many aspects of embryonic development and adult tissue homeostasis, and the signaling molecules are highly conserved across metazoans (van Amerongen and Nusse, 2009). The canonical Wnt pathway relies on β-catenin as its key effector. In the absence of a Wnt signal, cytosolic β-catenin (β-cat) is maintained at a minimum by the 'destruction complex composed of ademomatous polyposis coli (APC), Axin, glycogen synthase kinase 3 β (GSK3β) and Disheveled (Dvl)' (Figure 1.3; (Logan and Nusse, 2004)). Wnt signaling is activated by binding of secreted Wnt ligands to the transmembrane Frizzled receptor and the low-density lipoprotein receptor protein (LRP) that mediate downstream Wnt signaling (Figure 1.3; (Bhanot et al., 1996; Pinson et al., 2000)). Binding of R-spondins (R-spo) to leucine-rich repeat-containing G-protein coupled receptor (Lgr) 4/5/6 receptor stabilizes Frizzled to potentiate Wnt signaling (Figure 1.3; (Carmon et al., 2012; de Lau et al., 2011)). Activation of the Wnt pathway leads to destabilization of the destruction complex and translocation of β-catenin. Nuclear β-catenin interacts with DNA-binding proteins termed somewhat misleadingly 'T-cell factor/lymphoid enhancer-binding factor (TCF/LEF)' to regulate transcription of context-specific Wnt target genes (Figure 1.3; (Clevers and van de Wetering, 1997; Logan and Nusse, 2004)).
Figure 1.3 Wnt signaling pathway. (Left) In the absence of Wnt ligand, the destruction complex containing APC, Axin, GSK3β, and Dvl phosphorylates β-catenin and leads its degradation in the cytoplasm. In the nucleus, TCF/LEF is inactive and Wnt target genes are “OFF”. (Right) Upon binding of Wnt ligand to the Frizzled and the LRP with binding of R-spo to Lgr for enhancing signals, the destruction complex is disintegrated. Accumulated β-catenin could translocate into the nucleus and interact with TCF/LEF transcription factor to upregulate Wnt target genes “ON”. Adapted from (Jansson et al., 2015).

Wnt signaling regulates multiple aspects of embryonic intestinal development and adult intestinal homeostasis. During embryonic development, β-catenin is known as an essential factor for endoderm formation (Imai et al., 2000; Lickert et al., 2002) In mice, Wnt signaling is active at E 7.5 in the posterior endoderm and transiently induces caudal-related homeobox 2 (Cdx2), the master regulator of intestinal development (Gao et al., 2009; Sherwood et al., 2011). Until E 14.5, the Tcf4-deficient intestine shows no noticeable difference in the tissue organization and cell proliferation. However, at E 16.5 and afterward, a reduced number of villi and loss of proliferating cells in intervillus
regions are observed in the small intestine of the mutant mice, leading to neonatal lethality (Korinek et al., 1998). In addition, canonical Wnt signaling is active in intervillus regions after birth, suggesting that Wnt signaling may play an important role in de novo crypt formation and crypt fission in mice (Kim et al., 2007).

The Wnt pathway is the master regulator of intestinal epithelial regeneration in the adult. Active Wnt signaling in intestinal crypt is a major driving force for intestinal stem cell renewal and proliferation of transit-amplifying (TA) cells. Tight regulation of Wnt signaling is critical for the intestinal epithelial homeostasis (Gregorieff and Clevers, 2005). Reduced Wnt signaling results in loss of crypt regenerative capacity. Overexpression of the Wnt antagonist Dickkopf1 in adult transgenic mice or intestinal epithelial-specific deletion of β-catenin leads to reduced proliferation, loss of crypts and reduced numbers of secretory cells (Ireland et al., 2004; Pinto et al., 2003). Recently, ablation of Wnt-secreting Forkhead box L1+ (Foxl1+) intestinal mesenchymal cells or deletion of the Wntless gene, which is required for the secretion of Wnts, in adult mice, was shown to cause loss of intestinal stem cells and impairment of intestinal homeostasis (Figure 1.4; (Aoki et al., 2016; Valenta et al., 2016)). In contrast, aberrant activation of Wnt signaling leads to intestinal neoplasia. For instance, deletion of Apc in Lgr5+ intestinal stem cells results in constitutively active Wnt signaling and leads to widespread transformation (Barker et al., 2009).
Figure 1.4 Wnt signaling in intestinal homeostasis. (Left) Intestinal homeostasis is maintained in normal condition. Wnts are secreted from Paneth cells and mesenchymal cells. (Middle and Right) Wnts secretion is globally blocked by deletion of Wntless gene. (Middle) Loss of intestinal stem and TA cells without Wnt signaling. (Right) Intestinal stem and TA cells are survived with exogenous reconstitution of Wnt3a or Wnt2b. Adapted from (Valenta et al., 2016).
1.3 Intestinal stem cells and organoids

Two intestinal stem cell compartments have been reported in the literature: the ‘+4 cell’ model and the 'crypt base columnar cell' model. +4 cells reside immediately above the Paneth cells in the ‘+4’ position when counted from the crypt base. This was a population characterized by slow cycling and long term retention of a label incorporated into DNA during S-phase, and proposed to be the intestinal stem cell (Potten et al., 1974). Many studies identify various +4 stem cell markers including Bmi1, homeodomain-only (Hopx), telomerase reverse transcriptase (Tert) and Leu-rich repeats and immunoglobulin-like domains 1 (Lrig1), which were confirmed by genetic lineage tracing to contribute to all differentiated epithelial cell lineages. (Figure 1.5; (Montgomery et al., 2011; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011)). The situation might be even more complex, as it was suggested recently that +4 cells marked by Hopx/Bmi1-CreER are non-overlapping with the +4 label retaining cells (Li et al., 2016).

Crypt base columnar (CBC) cells are located at the crypt base and intermingled with Paneth cells (Cheng and Leblond, 1974a). Recently, several markers of CBC cells have been identified, such as Lgr5, Musashi homologue 1 (Msi1), olfactomedin 4 (Olfm4) and Achaete–Scute homologue 2 (Ascl2). These cells were confirmed by genetic lineage tracing and in vitro culture as actively dividing stem cells (Figure 1.5; (Barker et al., 2007; He et al., 2007; Sato et al., 2009; van der Flier et al., 2009a; van der Flier et al., 2009b)).
Figure 1.5 Intestinal stem cells: +4 and CBC stem cells. Summary of intestinal stem cell pools with their markers. +4 stem cells are located the fourth position from the crypt bottom. They are relatively quiescent, resistant to acute damage and retain DNA labels. CBC stem cells are intermingled with Paneth cells. They are highly proliferative and sensitive to damage. Adapted from (Barker, 2014).

Interestingly, proliferative CBC stem cells are dispensable for intestinal epithelial homeostasis, as quiescent +4 stem cells are activated when CBCs are ablated. Thus upon deletion of Lgr5+ stem cells by genetic ablation or radiation, Bmi1+ stem cells actively give rise to Lgr5+ stem cells to rapidly recover the entire intestinal epithelial cell sheet (Tian et al., 2011; Yan et al., 2012). Conversely, Lgr5+ stem cells can also give rise to Bmi1+ stem cells in intestinal organoids in vitro (Takeda et al., 2011). Precursor/progenitor cells of the intestinal epithelium also show their cellular dedifferentiation capacity (Figure 1.6). Absorptive progenitor cells expressing alkaline phosphatase around position +6 upward and secretory progenitor cells expressing the Notch ligandDll1 in the +4 position can dedifferentiate into stem cells in vivo to
regenerate entire crypt-villus units upon stem cell loss (Tetteh et al., 2016; van Es et al., 2012b). A quiescent label-retaining cell (LRC) population around the +4 position serves as Paneth and enteroendocrine cells precursor, and does not contribute to the stem cell pool during homeostasis. However, a subset of this population expresses Lgr5 and +4 cell markers and gives rise to proliferative intestinal stem cells under conditions of cytotoxic damage (Buczacki et al., 2013).

Figure 1.6 Intestinal stem cell hierarchy. Quiescent (Bmi1\textsuperscript{+} +4 cells) and active cycling (Lgr5\textsuperscript{+} CBC) intestinal stem cells produces absorptive progenitor, which give rise to enterocyte, and secretory precursor cell that located around the +4 position and generates cycling secretory progenitor cells, which produce goblet, Paneth and enteroendocrine cells. Both precursor and progenitor cell can repopulate Lgr5\textsuperscript{+} CBC cycling stem cell after loss of Lgr5\textsuperscript{+} CBC cell by injury. Adapted from (Visvader and Clevers, 2016).
Intestinal stem cells and crypts from the adult intestine grow into three-dimensional intestinal organoids in culture with Matrigel matrix (Jung et al., 2011; Sato et al., 2009). These organoids consist of crypt and villus domains, and contain stem cell and all the differentiated cell types, which mimic the *in vivo* structure of the intestinal epithelium (Figure 1.7). Intestinal organoid growth requires epidermal growth factor and R-spondin1, which enhances Wnt signaling via Lgr receptors (Figure 1.3; (Carmon et al., 2012; de Lau et al., 2011)).

![Figure 1.7 Intestinal organoid structure.](image)

However, intestinal progenitor cells from fetal and neonatal intestine grow into undifferentiated hollow spheroids in the same culture condition (Figure 1.8; (Fordham et al., 2013; Mustata et al., 2013)). These spheroids are morphologically similar to APC null adult organoids, but express low level of Wnt target genes including *Lgr5* and *Axin2*.
Figure 1.8 Intestinal organoids and spheroids. Intestinal spheroid consists of highly proliferative cells without terminally differentiated cells. There is no crypt-villus domain in intestinal spheroid. Adapted from (Fordham et al., 2013)
1.4 Epigenetic regulations in the intestinal epithelium

While the signaling pathways and transcriptional regulators involved in intestinal development and homeostasis have been investigated extensively, the role of epigenetic factors in these processes is largely unknown. Epigenetics is defined as heritable phenotypic traits, which affect gene function but are independent of alteration to the DNA sequence itself (Jaenisch and Bird, 2003). Epigenetic control is important for the establishment of cell-specific gene expression profiles and thus for self-renewal and differentiation of stem cells (Spivakov and Fisher, 2007). There are two main components of epigenetic modification - histone modifications and DNA methylation (Figure 1.9).

Figure 1.9 Two main components of epigenetic modification. DNA methylation is a process by which a methyl group is added to cytosine, mostly on cytosine-guanine
dinucleotide. Histone modification is a covalent post-translation modification including methylation and acetylation in globular domain or tail. Adapted from (Qiu, 2006).

DNA is wrapped around the histone octamer, which contains two copies each of the histone proteins H2A, H2B, H3, and H4, and is packed into chromatin. The basic unit of chromatin is the nucleosome that consists of 147 base pair of DNA and a histone octamer (Luger et al., 1997). Post-translational modifications on histone proteins alter DNA accessibility and transcription during cell fate determination by themselves, or by recruiting effector proteins (Jenuwein and Allis, 2001; Tropberger and Schneider, 2013). Methylation (Me) and acetylation (Ac) on lysine (K) residue of the histone 3 (H3) tail have been the best characterized histone modifications to date. Although H3K4me3 and H3K27me3 mark active and inactive genes, respectively, they are co-localized as bivalent modifications at developmental genes poised for rapid activation following a differentiation signal (Bernstein et al., 2006; Ringrose and Paro, 2004). With regards to promoter-distal regulatory elements, H3K4me2 and H3K27Ac are enriched at active enhancers, which are also frequently occupied by the histone acetyltransferase p300. Poised enhancers have both H3K4me1 and H3K27me3, and are also often bound by the Polycomb Repressive Complex 2 (PRC2) (Figure 1.10; (Zhou et al., 2011)).
Figure 1.10 Histone modifications in gene regulation. H3K4me2 and H3K27Ac mark active enhancers with the histone acetyltransferase p300, lineage-specific transcription factors and RNA polymerase II, producing enhancer RNAs (eRNAs). Poised enhancers have H3K4me1 and H3K27me3 with the Polycomb Repressive Complex 2 (PRC2), and do not give rise to eRNAs. Repressed enhancers are enriched in H3K27me3 and PRC2, and are characterized by dense nucleosome packing. Adapted from (Nguyen et al., 2015).

Genome-wide H3K4me2 profiles in Lgr5+ intestinal stem cells, secretory progenitors, and enterocytes are relatively similar (Kim et al., 2014). This finding suggests that enhancer chromatin is ‘open’ in any cell type of the intestinal epithelium, but their activities might be regulated by other epigenetic modifications or transcription factor bindings for proper differentiation.

Another major epigenetic modification is methylation at the 5-carbon of the cytosine resulting in 5-methylcytosine (5mC). 5mC plays critical roles in many biological processes such as gene silencing, genomic imprinting, and genomic stability by
suppressing transposable and repetitive elements (Miranda and Jones, 2007). There are three DNA methyltransferases (Dnmt), termed ‘Dnmt1, -3a and -3b’ in mammals. Dnmt3a and -3b are classified as \textit{de novo} methyltransferases and Dnmt1 is considered as maintenance methyltransferase to copy methylation pattern to newly synthesized daughter strands during replication, although this strict biochemically based classification has been challenged recently by in vivo gene ablation studies (Elliott et al., 2016). Mice with ablation of the Dnmt genes exhibit global hypomethylation, and embryonic lethality, suggesting an essential role for methylation during development (Figure 1.11; (Bestor, 2000)).

\textbf{Figure 1.11 De novo and maintenance DNA methylation.} Initial methylation patterns are established by the \textit{de novo} DNA methyltransferases DNMT3A and DNMT3B during early development, When DNA replication and cell division occur, the maintenance
methyltransferase, DNMT1 maintains established methylation patterns in daughter cells. Adapted from (Wu and Zhang, 2010).

Alterations of the DNA methylome during the process of differentiation of Lgr5\(^+\) intestinal stem into differentiated villus cells in adult mice have been reported recently (Kaaij et al., 2013; Sheaffer et al., 2014). Genes highly expressed in Lgr5\(^+\) intestinal stem cells, such as Olfm4 and ‘Hes family basic helix-loop-helix transcription factor 1’ (Hes1), gain methylation while their expression is repressed in differentiated cells. Intestinal enterocyte marker genes, such as alkaline phosphatase, intestinal (Alpi) and lactase (Lct), lose DNA methylation during the differentiation process, which correlates with increased binding of Cdx2 and its partner protein Hepatocyte Nuclear Factor 4\(\alpha\) (Hnf4\(\alpha\)), and activation of gene expression in differentiated cells (Figure 1.12; (Kaaij et al., 2013; Sheaffer et al., 2014)). Conditional deletion of Dnmt1 in the entire intestinal epithelium or in Lgr5\(^+\) stem cells of adult mice leads to crypt expansion or altered stem cell morphologies, respectively (Figure 1.12; (Sheaffer et al., 2014; Yu et al., 2015)).

![Figure 1.12 DNA methylation changes in the intestinal epithelium. Lgr5\(^+\) intestinal](image)
stem cells (LGR5+ ISC), differentiated intestinal epithelial cells (DIFF IEC), and Dnmt1-deleted intestinal epithelium (ΔDNMT1). Adapted from (Sheaffer et al., 2014)

Loss of Tcf4 correlates with increased methylation levels at the differentially methylated regions, which are located near Tcf4 binding sites in intestinal crypt (Kaaij et al., 2013; Schuijers et al., 2014; van Es et al., 2012a). These findings suggest that the dynamics of DNA methylation plays an important role in enhancer activities, and is associated with Wnt target genes in intestinal epithelial differentiation.

Loss of $Dnmt1$ in the developing intestine results in severe intestinal abnormalities and partial lethality before weaning (Figure 1.13; (Elliott et al., 2015; Yu et al., 2015)). These results demonstrate that DNA methylation plays an important role in intestinal maturation and homeostasis.

**Figure 1.13 DNA methylation is essential for the postnatal intestinal maturation.** Intestinal epithelium specific deletion of Dnmt1 in the developing intestine. Loss of $Dnmt1$ shows villus atrophy and vacuolated cells in the intestinal epithelium of postnatal day 7 mice. Adapted from (Yu et al., 2015)
1.5 TET-mediated DNA hydroxymethylation in DNA demethylation pathway

DNA methylation is relatively stable epigenetic modification, but its levels are tightly regulated during development (Reik et al., 2001). In primordial germ cells and zygotes of mammals, for example, global 5mC decreases rapidly, suggesting a precise mechanism of DNA demethylation. Ten-eleven translocation (TET) enzymes are identified as dioxygenases that can convert 5mC to 5-hydroxymethylcytosine (5hmC) in various mammalian cells such as embryonic stem cells, brain and bone marrow (Kriaucionis and Heintz, 2009; Li and Liu, 2011; Tahiliani et al., 2009).

5hmC and TET enzymes are involved in both passive and active DNA demethylation in mammals (Wu and Zhang, 2010). 5hmC can facilitate passive demethylation during replication due to poor recognition of the 5hmC-modified DNA segments by the maintenance DNA methyltransferase DNMT1 (Inoue and Zhang, 2011; Valinluck and Sowers, 2007). In terms of active demethylation, TET enzymes are able to further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC), which are excised by thymidine DNA glycosylase (TDG), followed by replacement of abasic sites with unmethylated cytosines by the base excision repair pathway (Figure 1.14; (He et al., 2011; Maiti and Drohat, 2011)). These findings suggest that 5hmC is the first step of DNA demethylation and may be involved in transcriptional activation by removing repressive 5mC (Ficz et al., 2011; Nestor et al., 2012).
**Figure 1.14 5mC oxidation and demethylation process.** 5mC is generated by DNMT, which can be oxidized by TET enzymes to generate 5hmC, 5fC and 5caC. All oxidative derivatives can be diluted out during replication, due to lack of recognition of DNMT1 (Passive demethylation). 5fC and 5caC can be excised by TDG and repaired by the base excision repair mechanism to C (Active demethylation). Adapted from (Ficz, 2015)

The mammalian Tet gene family has three members: Tet1, Tet2 and Tet3 (Ito et al., 2010). All of Tet proteins contain a C-terminal catalytic domain that exhibits Fe^{2+} and 2-oxoglutarate-dependent dioxygenase activity and oxidize 5mC to 5hmC/5fC/5caC (Ito et al., 2011). This catalytic domain is used for a targeted DNA demethylation at endogenous loci by fusing to engineered transcription activator–like effector (TALE) repeat arrays. TALE-TET catalytic domain fusion protein can lead to demethylation at the targeted promoters in human cells and upregulation of endogenous human gene expressions (Maeder et al., 2013).

Tet1 is highly expressed in ES cells and is required for normal ES cell differentiation (Dawlaty et al., 2011; Ito et al., 2010). Tet1^{−/−} mice on a mixed background
are grossly normal, but have a smaller body size than litter mate controls during the postnatal period and exhibit defects in memory function (Dawlaty et al., 2011; Rudenko et al., 2013; Xin et al., 2015; Zhang et al., 2013). In contrast, a different allele of Tet1−/− mice on a pure C57BL/6 background showed partial embryonic lethality, smaller body size, and defects in fertility (Kang et al., 2015; Yamaguchi et al., 2012). Tet2−/− mice develop abnormally high numbers of myeloid cells and blood cancer in adulthood. Tet2 loss in hematopoietic stem cells leads to increased stem cell self-renewal, DNA hypermethylation at active enhancers and leukemogenesis (Ko et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011; Rasmussen et al., 2015; Shide et al., 2012). Tet3−/− mice show embryonic and neonatal lethality (Gu et al., 2011). Tet3 followed by Tet2 are the most highly expressed Tet genes in the embryonic mouse brain, and their expression levels are increased during neuronal differentiation along with the increase of 5hmC levels. Loss of Tet3 and/or Tet2 causes a defect in neural progenitor cell maintenance and differentiation during brain development (Hahn et al., 2013; Li et al., 2015; Lister et al., 2013).

In chapter 2, I will present a novel role for TET1-mediated DNA hydroxymethylation in Wnt target gene expression in intestinal maturation and homeostasis.
1.6 An Independent Role of 5hmC in Gene Regulation

The levels of 5hmC in various tissues is much greater than would be expected if it were just an intermediate product in the DNA demethylation pathways, and are also dynamically changed during differentiation and aging (Bachman et al., 2014; Szulwach et al., 2011b; Tahiliani et al., 2009). Thus, 5hmC may be a new semi-stable epigenetic mark, recruiting unique chromatin or transcriptional modifiers, and providing another level of gene expression regulation (Frauer et al., 2011; Mellén et al., 2012; Spruijt et al., 2013; Yildirim et al., 2011).

Although the enrichment 5hmC in promoters and gene bodies is correlated with active gene expression (Figure 1.15; (Ficz et al., 2011; Song et al., 2010)), 5hmC is also found at bivalent promoters of poised developmental genes (Figure 1.15; (Pastor et al., 2011; Wu et al., 2011)). This indicates 5hmC may have a role in pausing or silencing gene expression (Robertson et al., 2011; Szulwach et al., 2011a). Dual functions of 5hmC in gene regulation could be explained by its key role as a central node in epigenomic network in ES cells (Juan et al., 2016).
Figure 1.15 The dual role of 5hmC and TET in gene regulation. 5hmC and TET proteins may block DNMTs or Methyl-binding proteins to maintain unmethylated status in active gene promoter. However, 5hmC and TET enzymes are also found in regulatory regions of transcriptionally poised gene, interacting with the SIN3A co-repressor complex and PRC. Adapted from (Branco et al., 2012).

In chapter 3, I will present a repressive role for 5hmC at putative distal enhancers, which is characterized by transcription factor bindings and absence of histone enhancer marks, and suggests that 5hmC is a new epigenetic mark for silenced enhancers.
1.7 Reference


CHAPTER 2

Epigenetic regulation of intestinal stem cells by Tet1-mediated DNA hydroxymethylation


*Submitted*
2.1 Summary

Methylated cytosines are associated with gene silencing. The Ten-eleven Translocation (TET) hydroxylases, which oxidize methylated cytosines to 5-hydroxymethylcytosine (5hmC), are essential for cytosine demethylation. Gene silencing and activation are important processes in intestinal stem cell (ISC) differentiation and maintenance, but the potential role of TET hydroxylases in these processes has not yet been examined. Here, we generate genome-wide maps of the 5hmC mark in ISCs and their differentiated progeny. Genes with high levels of hydroxymethylation in ISCs are strongly associated with Wnt signaling and developmental processes. We thus investigated Tet1-deficient mice and found that they are growth-retarded, exhibit partial postnatal lethality, and have significantly reduced numbers of proliferative cells in the postnatal intestine. In the Tet1-deficient crypt, decreased expression of Wnt target genes such as Axin2 and Lgr5 correlates with lower 5hmC levels at the promoter of Wnt target genes. These data demonstrate that Tet1-mediated DNA hydroxymethylation plays a critical role in the epigenetic regulation of the Wnt pathway in intestinal stem and progenitor cells, and consequently in the self-renewal of the intestinal epithelium.
2.2 Background

DNA methylation is a widespread epigenetic modification that regulates gene expression in development and cellular differentiation (Reik et al., 2001; Sheaffer et al., 2014; Smith and Meissner, 2013). DNA methylation is dynamic, yet tightly regulated to prevent aberrant patterns of 5-methylcytosine (5mC), which are common in cancer (Baylin et al., 2001; Esteller, 2007). 5mC is established and maintained by DNA methyltransferase (DNMT) enzymes (Li et al., 1992; Okano et al., 1999). Recently, oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) by the Ten-eleven translocation (TET) gene family was proposed as a novel mechanism for removal of 5mC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This can occur by at least two non-mutually exclusive mechanisms. First, 5hmC can be further oxidized to 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC), again by the Tet enzymes, though at much lower efficacy than oxidation from 5mC to 5hmC (Ito et al., 2011). Both 5fC and 5caC can be glycosylated by thymine DNA glycosylase and replaced by unmodified cytosine via base excision repair (Maiti and Drohat, 2011; Weber et al., 2016). Second, because hemimethylated 5hmC nucleotides are not recognized by the maintenance DNA methyltransferases following DNA replication during S-phase, hydroxymethylation of specific sites will lead to their targeted, passive demethylation in replicating cells. Thus, hydroxymethylation via the Tet enzymes offers an elegant pathway for even lineage-committed cells to change their DNA methylation status.

Global methylome analyses in recent years have shown that the DNA methylation status is dynamic with aging even within the same cell type (Avrahami et al., 2015).
Likewise, we and others have demonstrated previously that DNA methylation at specific loci changes during differentiation of intestinal stem cells (ISCs) to enterocytes in adult mice (Kaaij et al., 2013; Sheaffer et al., 2014) and during ISC maturation in the postnatal mice (Yu et al., 2015); however, how these elements are targeted for demethylation is currently unknown.

The epithelium of the small intestine is the most rapidly self-renewing tissue in mammals, and even in humans the epithelium is replaced every three to five days. Active Wnt signaling in the intestinal crypt is a major driving force for intestinal stem cell renewal and proliferation of transit-amplifying cells. Tight regulation of Wnt signaling is critical for the intestinal epithelial homeostasis (Gregorieff and Clevers, 2005). Rapidly cycling Lgr5+ stem cells at the bottom of the intestinal crypt give rise to transit-amplifying cells that divide up to six times further (Rao and Wang, 2010) before differentiating into post-mitotic, functional cells as they exit the crypt zone and migrate up the crypt-villus axis. This process is accompanied by dramatic changes in the transcriptional program, which are mediated, at least in part, by epigenetic changes such as the aforementioned realignment of DNA methylation status (Kaaij et al., 2013; Sheaffer et al., 2014). In fact, in the absence of Dnmt1 alone or both Dnmt1 and Dnmt3b, the intestinal epithelium is not viable in the postnatal period (Yu et al., 2015) or in adult (Elliott et al., 2016), respectively. Here, we investigate to what extent targeted oxidation of 5mC to 5hmC contributes to intestinal epithelial health.
2.3 Methods

**Mice.** All procedures involving mice were conducted under a protocol approved by the Institutional Animal Care and Use Committee (IACUC) of the University of Pennsylvania. Eight to twelve week-old C57BL/6J mice were used for all experiments unless noted otherwise. For sorting of Lgr5+ stem cell, we employed Lgr5-EGFP-Ires-CreERT2 mice (Barker et al., 2007). Tet1-/+ mice (Dawlaty et al., 2011) were backcrossed to C57BL/6J mice for five generations. Eight week-old Tet1+/- mice were then intercrossed to generate Tet1-/- mice. Pups were monitored for litter size and body weight gain. Survival rate calculation and statistical analyses were performed using Prism Graphpad software. Tet1+/- mice were crossed to Lgr5-EGFP-Ires-CreERT2 mice and Tet1+/-;Lgr5-EGFP-Ires-CreERT2 mice were intercrossed to generated Tet1+/-;Lgr5-EGFP-Ires-CreERT2 mice.

**Dot Blot.** For the analysis of hydroxymethylation levels, DNA samples were denatured at 95 °C for 10 min and spotted onto Hybond-N+ nitrocellulose membranes (GE Healthcare). The membrane was blocked with CAS-Block (Thermo Fisher Scientific), followed by overnight incubation with an anti-5hmC antibody (Active Motif) at 4°C. Membranes were incubated with horseradish peroxidase (HRP)-conjugated goat anti-mouse antibodies (GE Healthcare) for 30 min at room temperature and developed using the ECL+ prime blotting detection system (GE Healthcare).
**Immunohistochemistry.** Dissected tissues were washed in cold PBS and fixed overnight in 4% paraformaldehyde at 4°C before paraffin embedding. Antigen retrieval was performed using Retriever in buffer A (Electron Microscopy Sciences), and tissue sections were incubated overnight at 4°C with anti-5hmC (Active Motif), anti-Ki67 (BD Pharmingen), anti-Axin2 (Abcam), anti-Sox9 (Millipore), anti-CyclinD1 (Biocare Medical) and anti-GFP antibodies (Aves Labs). After incubation with secondary antibodies (Vector Lab) for 2hrs at room temperature, samples were mounted in fluorescent mounting medium (Dako) or developed using the Vectastain Elite ABC kit (Vector Lab). Images were acquired using a Nikon Eclipse 80i fluorescence microscope and a Leica SP8 confocal microscope.

**Lgr5⁺ cell, crypt, and villus isolation.** Small intestines were dissected and opened longitudinally. The tissues were washed in cold PBS and scraped gently on the luminal side with a glass slide to obtain villus cells. The remaining tissue was incubated with rotation in 5mM EDTA/HBSS for 30 min at 4°C. After EDTA incubation, the tissue was vigorously shaken for 15 seconds. The first fraction was villus-rich and thus discarded. After further shaking, the supernatant enriched for intestinal crypts was passed through a 70 µm cell strainer to remove residual villus material and was centrifuged at 600 rpm for 2 min to collect crypts. For Lgr5⁺ cell sorting, isolated crypts were dissociated with TrypLe express (Thermo Fisher Scientific) for 20 min at 37°C and shaken every 5 min to prevent cell clumping. Dissociated cells were passed through a 40 µm cell strainer and washed with HBSS. GFP⁺ cells were separated on a FACS Diva (BD Bioscience).
**DNA hydroxymethylation analysis.** Genomic DNA was isolated from sorted Lgr5<sup>+</sup> cells, crypt and villus cells using the AllPrep kit (Qiagen) and sonicated to an average size of 150-300 bp (Covaris). DNA fragments were denatured (10 min at 95 °C) and immunoprecipitated using 2µl of anti-5hmC antibody (Active Motif) and 10µl Protein G beads (Thermo Fisher Scientific) in IP buffer (10 mM sodium phosphate, pH 7.0; 140 mM NaCl; 0.05% Triton X-100). For hMeDIP-qPCR, hydroxymethylated DNA and input DNA were quantified on an Mx3005P qPCR System (Applied Biosystems) using the Brilliant II SYBR Green QPCR master mix (Agilent). For hMeDIP-seq, DNA fragments were ligated with Illumina adaptors using the NEBNext® DNA Library Prep Master Mix Set (NEB). Following adaptor ligation, DNA fragments were denatured and immunoprecipitated using an anti-5hmC antibody (Active Motif). Hydroxymethylated DNA was amplified with adapter-specific primers (12 cycles). Amplified fragments ranging from 150 to 200 bp were size-selected followed by sequencing on a HiSeq2000 (Illumina).

**Mapping of sequencing data and bioinformatics.** Reads were aligned to the mouse reference genome (NCBI build 37, mm9) using Bowtie (Langmead et al., 2008). Only unique reads were used for peak calling and annotation by HOMER (Heinz et al., 2010). BedGraph files were generated and viewed on the UCSC Genome Browser. Differentially hydroxymethylated Regions (DhMRs) were determined by comparing 5hmC peak levels in one sample to the other in each direction (fold change cutoff = 5, P
value cutoff = 1.00 x 10^-4). Gene ontology analysis was performed with DAVID (Huang et al., 2009).

**RNA isolation, qPCR, and mRNA-seq analysis.** Total RNA and complementary DNA (cDNA) were prepared using AllPrep kit (Qiagen,) and Superscript First-Strand synthesis system (Thermo Fisher Scientific). RT-qPCR reactions were performed on an Mx3005P qPCR System (Applied Biosystems) using Brilliant II SYBR Green QPCR master mix (Agilent). Relative expression levels were determined using comparative Ct values after normalizing to *Tbp*. RNA-seq method was previously detailed (Sheaffer et al., 2014).

**Intestine organoid culture.** Isolated intestinal crypts were resuspended in Matrigel (Corning) and covered with standard EGF/Noggin/R-spondin medium (Sato et al., 2009). Culture medium was changed every other day.
### Table 2.1 RT-qPCR primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Strand</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1</td>
<td>F</td>
<td>GAGCCTGTTTCCCTCGATGTGG</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CAAACCCACCTGAGGCTGTT</td>
</tr>
<tr>
<td>Tet2</td>
<td>F</td>
<td>AACCTGGCTACTGTCATTGCTCCA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ATGTTCCTGCTGGCTCTGTGGGAA</td>
</tr>
<tr>
<td>Tet3</td>
<td>F</td>
<td>CCGGATTGAGGATTCATCTCAC</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AAGATAACAATCACGCGTTCT</td>
</tr>
<tr>
<td>Lgr5</td>
<td>F</td>
<td>TAAAGACGAGGCAACAGTG</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GCCTTCAGGTCTCTCTCAAA</td>
</tr>
<tr>
<td>Creb3l3</td>
<td>F</td>
<td>GAACAGGAAACCAGGCCCTTT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CTCTGCATCTCTCTGGTGTC</td>
</tr>
<tr>
<td>Vimentin</td>
<td>F</td>
<td>CGGAAAGTGGAATCCTTGCA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CACATCGATCTGGACATGCTGT</td>
</tr>
<tr>
<td>Hemoglobin β</td>
<td>F</td>
<td>TGATGCTGAGAAGGGCTGCTGCTCTGT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GTGCCTTTGAGGCTGCTCAAGTG</td>
</tr>
<tr>
<td>Name</td>
<td>Genomic location</td>
<td>Strand</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Lgr5 promoter</td>
<td>chr10:115,026,111-115,026,229</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
</tr>
<tr>
<td>Axin2 promoter</td>
<td>chr11:108,780,916-108,781,024</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
</tr>
<tr>
<td>Sox9 promoter</td>
<td>chr11:112,641,826-112,641,940</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
</tr>
</tbody>
</table>
2.4 Results

5hmC level is high in intestinal villus epithelial cells

To determine global 5hmC levels in ISCs, we performed 5hmC dot blot analysis of genomic DNA from isolated epithelial cells as well as brain and embryonic stem cells (ESCs) for comparison. We found that 5hmC levels in the gut are almost as high as those present in the brain and ESCs (Figure 2.1A). Hydroxymethylation of cytosines has been suggested to play an important role in neurons due to the abundance of this mark in the brain (Hahn et al., 2013; Rudenko et al., 2013). Therefore, we hypothesized that this epigenetic mark might also play an essential role in intestinal epithelial cells. We examined the distribution of global 5hmC compared to 5mC in the intestinal epithelium by immunofluorescence detection of each mark with specific antibodies. 5mC was present in all epithelial cells, with highest levels in the crypt where stem and transit-amplifying cells reside (Figure 2.1B). We identified proliferating stem and transit-amplifying cells by co-staining with Ki67, a marker of proliferation. Interestingly, we found that the strongest 5hmC signal was detected in the nuclei of differentiated villus cells, while the signal decreased in the crypt region (Figure 2.1C). This finding is consistent with previous reports that 5hmC is depleted in proliferating cancer cells compared to post-mitotic cells (Jin et al., 2011; Lian et al., 2012).
Figure 2.1 Differential enrichment of DNA methylation and hydroxymethylation in the intestinal epithelium. (A) Dot blot shows global 5hmC levels in the mouse adult intestinal epithelium compared to brain and embryonic stem cells (ESCs). (B-C) Immunofluorescence staining of mouse jejunum with 5mC (B), 5hmC (C) and Ki67 antibodies as indicated. Scale bars: 50µm.

Dynamic changes of 5hmC distribution during intestinal epithelial differentiation

Next, we examined dynamic genomic 5hmC abundance during ISC differentiation. First, we isolated Lgr5-expressing ISCs (Lgr5⁺) from Lgr5-EGFP-IRES-CreERT2 mice (Barker et al., 2007) by FACS sorting and differentiated villus epithelial cells by careful scraping of the luminal side of the gut tube (Figure 2.2; see detailed
description in Experimental Procedures). Genomic DNA was then isolated from both cell populations. To investigate genome-wide changes of the 5hmC mark between Lgr5+ and differentiated cells, we performed hydroxymethylated DNA immunoprecipitation followed by next-generation sequencing (hMeDIP-seq; see Experimental Procedures).

Figure 2.2 Isolation of Lgr5+ intestinal stem and villus differentiated cells. (A) FACS plot to isolate Lgr5-EGFP+ cells and RT-qPCR validation for Lgr5 expression in GFP+ cells. (B) H&E staining image of scraped differentiated villus cells and RT-qPCR validation for cell-specific marker genes (Creb3l3: Enterocyte, Vimentin: Mesenchymal cells, Hemoglobin: red blood cells).
Figure 2.3 Differential 5hmC distributions in intestinal stem and differentiated cells.
(A–C) Screenshots of bisulfite sequencing (with measures 5mC and 5hmC) and hMeDIP-sequencing (5hmC) at marker genes in Lgr5+ intestinal stem cells (A), Wnt target genes (B) and differentiated villus cells (C). Data are presented as % methylation for the bisulfite-sequencing data, and as reads per million mapped reads (RPM) for the 5hmC data.
Remarkably, we discovered striking differences in 5hmC levels marker genes for both ISCs and mature enterocytes (Figure 2.3). For example, high levels of 5hmC were present near promoters and throughout the gene bodies of ISC-marker genes such as *Lgr5*, *Olfm4* and *Msi1*, and Wnt target genes such as *Axin2*, *c-Myc* and *Sox9* in Lgr5\(^+\) cells but were decreased dramatically after differentiation into villus epithelial cells (Figure 2.3A-B, Figure 2.4A-B). Conversely, genes expressed at high levels only in differentiated cells, such as *Alpi*, *Sis* and *Fabp1*, were hydroxymethylated only in differentiated cells (Figure 2.3C, Figure 2.4C).

**Figure 2.4. Marker gene expression level in Lgr5\(^+\) stem and differentiated cells.** (A) RNA-seq results for intestinal stem cell marker genes (B) RNA seq results for Wnt target genes (C) RNA-seq results for differentiated villus cell marker genes
Having established that the 5hmC mark is dynamic for the nine genes described above, we performed a genome-wide analysis of hydroxymethylation. We identified a total of 33,183 5hmC-enriched regions (peaks) in Lgr5+ cells and 100,707 peaks in differentiated cells using HOMER (Heinz et al., 2010) (Figure 2.5A), and about eight-fold increase of 5hmC signal in differentiated cells relative to the 5hmC signal in Lgr5+ cells (Figure 2.5B). When comparing 5hmC peaks between the two cell populations, we identified 10,450 differentially hydroxymethylated regions (DhMRs) (≥5 fold, \( p < 0.0001 \)), and divided them into two groups – higher 5hmC in Lgr5+ DhMRs (n=2,537) and higher 5hmC in differentiated cells (n=7,903) (Figure 2.6A). The two groups of DhMRs displayed distinct genomic distribution patterns. While Lgr5+ DhMRs were enriched in intergenic regions, those present in differentiated cells were more likely to be located in gene bodies. We identified the closest gene for each DhMR and performed gene ontology analysis for the differentially hydroxymethylated gene sets (Figure 2.6B). The genes associated with higher 5hmC levels in Lgr5+ cells were involved in ‘developmental process’ and ‘cell differentiation’, primary functions of stem cells. Conversely, the genes associated with higher 5hmC levels in differentiated cells were enriched for genes that control metabolic processes and nutrient transport, major functions of enterocytes. Thus, overall, high levels of hydroxymethylation correlated positively with the gene sets expressed in either Lgr5+ stem cells or post-mitotic, differentiated villus epithelial cells.
Figure 2.5 Overall 5hmC levels in Lgr5+ stem cells and differentiated cells. (A) Total number of 5hmC enriched regions (peaks) number counting and genomic distribution in Lgr5+ stem and differentiated cells (≥ 4-fold enrichment over local tag count, p-value < 0.001). (B) The relative 5hmC signal of hMedIP-seq in Lgr5+ stem and differentiated cells. The ratio of hMeDIP/Input reads numbers in DIFF was set to 100%.

Figure 2.6 Dynamics of 5hmC and gene expression changes during intestinal stem cell differentiation. (A) Genomic distribution of differentially hydroxymethylated regions (DhMRs). Upper pie chart: regions with 5hmC levels that are higher in Lgr5+ stem cells than in differentiated epithelial cells (DIFF) (at least 5 fold, p-value < 0.0001). Lower pie chart: regions with 5hmC levels that are higher in differentiated epithelial cells (DIFF) than in Lgr5+ stem cells (at least 5 fold, p-value < 0.0001). Circle size represents
the total number of DhMRs in each group. (B) Gene ontology search results for genes closest to DhMRs. (C) Differential gene expression of differentially hydroxymethylated genes (≥1.5 fold change) during intestinal epithelial differentiation. RNA-seq data for both cell populations are represented as reads per kilobase per million mapped read (RPKM).

Next, we examined the correlation between differential hydroxymethylation and differential gene expression, determined by RNA-seq (Sheaffer et al., 2014). Genes enriched for the 5hmC mark in Lgr5\(^+\) stem cells (n=1,181), including ISC-marker genes such as *Olfm4*, *Lgr5* and *Msi1*, showed higher gene expression level in Lgr5\(^+\) stem cells than in differentiated villus epithelial cells (Figure 2.6C, left). Conversely, genes enriched for the 5hmC mark in differentiated cells (n=1,559) including villus differentiated cell marker genes such as *Alpi*, *Sis* and *Fabp1*, exhibited increased expression in differentiated cells (Figure 2.6C, right). Overall, gene expression changes correlated strongly with 5hmC level variation during ISC differentiation. However, we also noted some exceptions where gene expression levels were anti-correlated with 5hmC status.

5hmC has been reported to have a role in transcriptional repression in specific cases including poised enhancers and promoters (Choi et al., 2014; Pastor et al., 2011; Wu et al., 2011).

**Tet1 is required for postnatal intestinal development**

Having determined that hydroxymethylation is highly dynamic during differentiation of ISCs into functional villus epithelial cells, we investigated which of the
Tet enzymes is most likely controlling this process. We measured Tet1, 2 and 3 mRNA expression levels in Lgr5⁺ stem and in differentiated villus cells by RT-qPCR (Figure 2.7). Tet1 was expressed at much higher levels in Lgr5⁺ stem cells than in differentiated cells, while Tet2 and 3 showed the opposite pattern. These data suggested that Tet1 might have an essential role in Lgr5⁺ ISC function.

![Graphs of Tet1-3 mRNA expression](image)

**Figure 2.7 Differential expression of Tet1-3 in Lgr5⁺ stem and differentiated cells.** Relative mRNA expression level of Tet1 (A), Tet2 (B) and Tet3 (C) in Lgr5⁺ stem and differentiated cells. Relative to Tbp.

Next, we phenotyped Tet1 null mutant mice (Dawlaty et al., 2011) during postnatal development, when gastro-intestinal function first becomes relevant (Table 2.3). Strikingly, we found that Tet1-deficient mice were growth-retarded during the early postnatal period (Figure 2.8A), exhibited a much smaller body size (Figure 2.8B), and showed significant postnatal lethality (Figure 2.8C). In fact, viability of Tet1⁻/⁻ mice was already significantly impaired by postnatal day 3 (Figure 2.8D).
Table 2.3 Previous and current studies of Tet1 null mice

<table>
<thead>
<tr>
<th>Tet1 ablation studies</th>
<th>Allele</th>
<th>Background</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dawlaty et al., 2011)</td>
<td>exon 4 deletion</td>
<td>129Sv, C57BL/6</td>
<td>viable, fertile, impaired memory extinction</td>
</tr>
<tr>
<td>(Rudenko et al., 2013)</td>
<td></td>
<td>Mixed</td>
<td></td>
</tr>
<tr>
<td>(Yamaguchi et al., 2012)</td>
<td>gene-trap cassette</td>
<td>6X backcrossed to C57BL/6</td>
<td>partial embryonic lethal, reduces female germ-cell and fertility</td>
</tr>
<tr>
<td></td>
<td>into the 1st intron</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Zhang et al., 2013)</td>
<td>exon 11-13 deletion</td>
<td>129Sv, C57BL/6</td>
<td>viable, fertile, impaired learning and short-term memory</td>
</tr>
<tr>
<td>(Xin et al., 2015)</td>
<td></td>
<td>Mixed</td>
<td></td>
</tr>
<tr>
<td>(Kang et al., 2015)</td>
<td>exon 9-10 deletion</td>
<td>C57BL/6</td>
<td>partial embryonic lethal, hydrocephalus, degenerated ovaries</td>
</tr>
<tr>
<td>Current work</td>
<td>exon 4 deletion</td>
<td>5x backcrossed to C57BL/6</td>
<td>partial postnatal lethal, intestinal defects</td>
</tr>
</tbody>
</table>

Figure 2.8 Postnatal growth retardation and lethality of Tet1 null mice. (A) Tet1 null mice are growth-retarded. Weight curve during the postnatal period, ** P-value < 0.01 by
t-test. (B) Representative Tet1\(^{+/+}\) and Tet1\(^{-/-}\) mice on postnatal day 11 (P11). (C) Survival curve during postnatal period. *** P-value < 0.001 by a Log-rank (Mantel-Cox) test. (D) Reduced perinatal viability of Tet1 homozygous mutants. Mice were quantified on postnatal day 3. *** P-value < 0.0005 (chi-square test)

The intestines of Tet1-deficient mice were thinner and shorter than those of wild-type littermates (Figure 2.9A). On histological examination, we found both villus length and proliferating cell number per crypt to be significantly reduced in Tet1-null mice compared to wild-type littermates (Figure 2.9B-D). Thus, Tet1 plays an important role in the maturation of the postnatal intestine.

---

**Figure 2.9 Developmental defects during postnatal intestinal maturation.** (A) Representative image of intestinal tracts of Tet1\(^{+/+}\) and Tet1\(^{-/-}\) mice at one week of age. (B) Histological analysis of Tet1\(^{+/+}\) and Tet1\(^{-/-}\) jejunum of one week-old mice.
Proliferating cells are marked by immunohistochemistry for Ki67 (dark brown signal). Scale bars: 50µm (C) Villus length comparison between Tet1<sup>+/+</sup> and Tet1<sup>−/−</sup> one week-old jejunum (*** P-value < 0.001). (D) Ki67<sup>+</sup> cell number per crypt is reduced in one week-old Tet1<sup>−/−</sup> jejunum, compared littermate controls. (*** P-value < 0.001).

**Wnt target genes and Lgr5 expression in the postnatal intestine is dependent on Tet1 function**

Since Wnt signaling is essential for intestinal maturation and proliferation, we reasoned Wnt target gene expression might be affected in the Tet1-deleted intestine. To examine Wnt target gene expression levels, we performed immunostaining for Wnt molecules in one week-old Tet1<sup>−/−</sup> jejunum compared to wild-type littermate. Expression of Wnt target genes such as Axin2, Sox9 and Cyclin D1 were clearly reduced in mutant intestine (Figure 3.10).
Figure 3.10 Decreased level of Wnt target gene expression in the postnatal Tet1 null mice intestine. Immunofluorescence stains of Wnt target proteins such as (A) Axin2 (B) Sox9 and (C) Cyclin D1 in postnatal day 7 Tet1+/+ and Tet1-/- mice. Scale bars: 100µm

To evaluate the consequences of Tet1-deficiency on ISCs, we crossed Tet1 mutant mice with Lgr5-EGFP-IRES-CreERT2 mice, which allows for the easy detection of Lgr5+ stem cells via expression of EGFP. We found a dramatic reduction in the frequency of Lgr5-EGFP positive cells in the crypt of the postnatal and adult Tet1-null jejunum by immunofluorescence staining and FACS (Figure 3.11). Thus, Tet1 is required for maintaining Wnt target gene expression and Lgr5 expression during intestine maturation.
**Figure 3.11 Reduction of Lgr5 expression in Tet1 null mice intestine.** (A) Decreased Lgr5-EGFP expression levels in two-week old and two-month old Tet1 null jejunum as detected by immunofluorescence stains of GFP (green). Scale bars: 20µm (B) FACS analysis demonstrates loss of Lgr5-EGFP+ cells in two-month old Tet1−/− mice. (C) Quantification of Lgr5-EGFP positive cell number in two-month old Tet1+/+ and Tet1−/− mice. *** P-value < 0.001 (n=4)

**Decreased 5hmC levels at Wnt target loci leads to spheroid formation**

Next, we investigated stem cell function using intestinal crypt culture. We found significantly decreased mature organoid-forming efficiency and reduced level of Wnt target/ISC marker gene expression in cultures derived from Tet1-deficient intestine (Figure 3.12). This finding suggests that the state of postnatal crypt from Tet1−/− mice is more close to fetal crypt than adult crypt (Fordham et al., 2013; Mustata et al., 2013).
Figure 3.12 Mature organoid-forming potential is reduced in the absence of Tet1. Images of intestinal organoid formation from two-week old Tet1+/+ and Tet1−/- jejunal crypts. Scale bars: 10µm. (B) Quantification of intestinal organoid volume after nine days of culture. ** P-value < 0.01 (n=3). (C) Ratio of mature budding organoids over total organoids. * P-value < 0.05 (n=3). (D) Differential gene expression in Tet1−/- premature spheroid compared to Tet1+/+ mature organoid.

To ascertain if the decrease in Wnt target genes and Lgr5 expression in Tet1−/- mice was caused by altered hydroxymethylation levels, we analyzed 5hmC levels at their loci in control and mutant mice. We found significantly decreased 5hmC levels at their promoter region in Tet1 null mice by 5hmC immunoprecipitation followed by qPCR analysis (Figure 3.13), which correlated with decreased Wnt target genes and Lgr5 expression levels (Figure 3.10 and 3.11). This finding suggests that Tet1 is critical for hydroxymethylation followed by demethylation to allow for full activation of Wnt target and stem cell genes.
Figure 3.13 Differential 5hmC level at the promoter of Wnt target genes. hMeDIP-qPCR results at the promoter of Lgr5, Axin2, Sox9 in two-week old Tet1+/+ and Tet1−/− crypts. * P-value < 0.05 (n=3).
2.5 Discussion

In this study, we document dynamic changes of 5hmC at key gene loci during ISC differentiation and report that Tet1-mediated hydroxymethylation is essential for ISC function. 5hmC was enriched at promoters and gene bodies of highly expressed genes in each cell type and may reverse the repressive effect of 5mC for gene activation. Tet1 was expressed in ISCs, not in differentiated villus cells.

*Tet1<sup>−/−</sup>* mice on a mixed genetic background (129Sv, C57BL/6) were previously reported to be viable and fertile with no morphological and growth abnormalities (Dawlaty et al., 2011; Zhang et al., 2013), but adult *Tet1<sup>−/−</sup>* mice exhibited defective self-renewal of neural progenitor cells (Rudenko et al., 2013; Xin et al., 2015; Zhang et al., 2013) and hematopoietic stem cells (Cimmino et al., 2015). However, after backcrossing onto the C57BL/6 background, we observed decreased postnatal viability and striking intestinal defects in these mice. Recently, Kang and colleagues also reported that *Tet1*-deficient C57BL/6 mice exhibit partial embryonic lethality and growth retardation (Kang et al., 2015); however, intestinal health was not examined in the prior study. We found that *Tet1<sup>−/−</sup>* mice showed shortened intestinal length, reduced proliferating cell number and villus height in the postnatal intestine, which is similar to the phenotype seen in mice with conditional deletion of Dnmt1 in the developing intestine using Dnmt1<sup>fl/fl</sup>; VillinCre mice (Elliott et al., 2015; Yu et al., 2015).

Lgr5 is known as one of Wnt target gene and an adult stem cell marker gene in the intestine (Barker et al., 2007), but a limited number of Lgr5<sup>+</sup> cells are also presented
in the intervillus zone of the developing intestine at E 16.5 (Kinzel et al., 2014). During intestinal maturation in the postnatal period, these Lgr5+ cells are rapidly cycling at the bottom of the crypt and clearly exhibit stem-cell properties at this stage (Kim et al., 2012). Here, we show that 5hmC enrichment at Wnt target loci in Lgr5+ stem cells is required for full activation of gene expression, and homozygous deletion of Tet1 causes significantly decreased 5hmC levels at these loci, resulting in a dramatically reduced Wnt target genes and Lgr5 expression during postnatal intestinal maturation. Importantly, loss of Tet1 also impairs mature organoid-forming capacity and thus stem-cell function of small intestinal crypts. In culture, intestinal spheroids were generated from fetal intestine, and exhibited low level of Wnt target gene expression (Mustata et al., 2013), which is similar to what we generated from Tet1-deleted postnatal intestine. Our data demonstrate that Tet1-mediated conversion from 5mC to 5hmC is an important epigenetic mechanism regulating expression of Wnt target genes including Lgr5 during postnatal intestinal maturation.

The definitive role of Tet1-mediated DNA hydroxymethylation in the intestine warrants further investigation using conditional gene ablation models. Nevertheless, our results demonstrate that 5hmC is an essential epigenetic mark required for ISC function. DNA hydroxymethylation could be a switch to activate key genes in stem cells and represents a novel mechanism of gene regulation during stem cell differentiation.
2.6 Reference


Fordham, R.P., Yui, S., Hannan, N., Soendergaard, C., Madgwick, A., Schweiger, P.J.,
Expanded Fetal Intestinal Progenitors Contributes to Colon Regeneration after Injury.
Cell Stem Cell.


Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre,
transcription factors prime cis-regulatory elements required for macrophage and B cell

analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44-57.

(2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-
carboxylyctosine. Science 333, 1300-1303.

Jin, S.-G.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q.,
but its levels do not correlate with IDH1 mutations. Cancer Res 71, 7360-7365.

van Es, J.H., Schuijers, J., de Wit, E., de Laat, W., *et al.* (2013). DNA methylation
dynamics during intestinal stem cell differentiation reveals enhancers driving gene

Kang, J., Lienhard, M., Pastor, W.A., Chawla, A., Novotny, M., Tsagaratou, A., Lasken,
of the methylytosine oxidases Tet1 and Tet3 increases transcriptome variability in early
embryogenesis. Proc Natl Acad Sci USA 112, 45.


CHAPTER 3

5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation

I. Choi*, R. Kim*, H.-W. Lim, K.H. Kaestner†, K.-J. Won†.

*Equal contribution

BMC Genomics 2014, 15:670 *Equal contribution
3.1 Summary

Recent mapping of 5-hydroxymethylcytosine (5hmC) provides a genome-wide view of the distribution of this important chromatin mark. However, the role of 5hmC in specific regulatory regions is not clear, especially at enhancers. We found a group of distal transcription factor binding sites highly enriched for 5-hydroxymethylcytosine (5hmC), but lacking any known activating histone marks and being depleted for nascent transcripts, suggesting a repressive role for 5hmC in mouse embryonic stem cells (mESCs). 5-formylcytosine (5fC), which is known to mark poised enhancers where H3K4me1 is enriched, is also observed at these sites. Furthermore, the 5hmC levels were inversely correlated with RNA polymerase II (PolII) occupancy in mESCs as well as in fully differentiated adipocytes. Interestingly, activating H3K4me1/2 histone marks were enriched at these sites when the associated genes become activated following lineage specification. These putative enhancers were shown to be functional in embryonic stem cells when unmethylated. Together, these data suggest that 5hmC suppresses the activity of this group of enhancers, which we termed “silenced enhancers”. Our findings indicate that 5hmC has a repressive role at specific proximal and distal regulatory regions in mESCs, and suggest that 5hmC is a new epigenetic mark for silenced enhancers.

Keywords

5hmC, GROseq, PolII, eRNA, mESC, Enhancer


3.2 Background

5-hydroxymethylcytosine (5hmC) is an epigenetic mark that arises from oxidation of 5-methylcytosine (5mC) by Ten-eleven translocation (Tet) enzymes (Tahiliani et al., 2009; Williams et al., 2011). The 5hmC mark has been studied in several cell types, such as mouse embryonic stem cells (mESCs) (Tahiliani et al., 2009; Yu et al., 2012), neuronal cells (Kriaucionis and Heintz, 2009; Mellen et al., 2012; Szulwach et al., 2011) and adipocytes (Serandour et al., 2012). 5hmC is enriched at promoters marked bivalently by H3K4me3 and H3K27me3 in mESCs (Matarese, 2011), but depleted at promoters in the brain (Szulwach et al., 2011). 5hmC is also enriched at specific transcription factor binding sites (TFBSs) in human and mouse ESCs (Ficz et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Szulwach et al., 2011; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). Specifically, in mESCs, 5hmC is depleted at Sox2 and Oct4 binding sites, but enriched for Esrrb and Tcfcp2l1 occupancy (Wu et al., 2011). In human embryonic stem cells (hESCs), 5hmC is highly enriched at CTCF, Nanog, and Oct4 binding sites (Stroud et al., 2011). Another study in hESC observed that the 5hmC profile showed a bimodal distribution at Oct4, Sox2, TAF1 and p300 binding sites (Szulwach et al., 2011). While these studies suggest a possible regulatory role for 5hmC at promoters and TFBSs, its function at these regulatory regions remains unclear.

Here, we report on a new repressive role for 5hmC at specific regulatory regions in mESCs. We show that 5hmC negatively correlates with nascent transcripts, especially at TFBSs. Interestingly, we discovered that a group of distal TFBSs displays a new
an epigenetic signature; these sites are exclusively enriched for 5hmC, depleted for activating histone modification marks (H3K4me1 and H3K27ac), and significantly reduced for nascent transcripts or enhancer RNAs (eRNAs). The expression of the genes close to these TFBSs was significantly lower than that of genes close to other classes of TFBSs. In addition, we found that a fraction of these TFBSs becomes enriched for activating histone marks (H3K4me1/2) in neural progenitor cells (NPCs) or endomesoderm cells. RNA polymerase II (PolII) chromatin interaction analysis with paired-end tagging (ChIA-PET) (Zhang et al., 2013) showed that the target genes of these regulatory regions were indeed significantly upregulated in NPCs. Enhancer/luciferase reporter assays demonstrated that these regions function as in gene activation when 5hmC is removed for these sites. Together, our findings suggest that 5hmC is as a novel marker for transcriptional silent enhancers in mESCs for regulatory regions that are activated during development.
3.3 Methods

Experimental procedures

We used genome-wide GROseq maps (Min et al., 2011) and ChIP-seq data for chromatin status (Meissner et al., 2008; Mikkelsen et al., 2007), PolII occupancy (Mikkelsen et al., 2007), 5mC (Ficz et al., 2011), and Tet1 occupancy (Ficz et al., 2011) in mESCs for our integrated analysis. We employed H3K4me1/2 data from NPC (Mikkelsen et al., 2007) and endomesoderm cells (Yu et al., 2013) to analyze the fate of our novel 5hmC regions after differentiation. We also included 5hmC from various independent studies (Pastor et al., 2011; Tan et al., 2013; Williams et al., 2011; Wu et al., 2011; Wu and Zhang, 2011; Xu et al., 2011) for our analysis. Table 3.1 summarizes all genome-wide datasets we used in our study.

All ChIP-seq data were normalized to 10 reads per kilobase per million mapped reads (RPKM) (Mortazavi et al., 2008). For clustering analysis we used Mev V4.8 (Saeed et al., 2006) and applied the K-means clustering algorithm using the Pearson correlation with absolute distance as a metric. To cluster distal TFBs in mESCs, we used the H3K4me1/2/3, H3K27ac, H3K27me and 5hmC levels and generated applied clustering (K = 10). We showed other epigenetic marks and GROseq and PolII next to the identified clusters.

To study the functional roles of 5hmC in various regulatory regions, we employed binding site data of 13 TFs (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1 and CTCF) in mESC (Chen et al., 2008).
To investigate 5hmC and nascent RNA levels across genes, we divided the genes into promoter (from -1Kbp to 500 bp around the annotated start site), 3' end (from -500 bp to 500 bp around the annotated termination site), and gene body regions (500 bp from the annotated start site to -500 bp from the annotated termination site). For transcription levels, we calculated RPKM using GROseq reads from 500 bp of the annotated start site to the annotated termination site in order not to include transcriptional pausing at promoters (Core et al., 2008; Rahl et al., 2010).

**Luciferase reporter assay**

Genomic DNA was prepared from R1 mouse embryonic stem cells (Nagy et al., 1993). About 600 bp genomic fragments for five distal TFBSs in cluster 2 were amplified by PCR with dNTPs and the PCR products ligated into the pGL3-SV40 luciferase vector (Promega). Empty vector (control) or cloned vectors were transfected directly into R1 mESC, together with the pRL-tk vector (Promega) as internal control, using Lipofectamine LTX (Life Technologies). At 24 h after transfection, cells were harvested and lysates subjected to the dual-luciferase reporter assay (Promega). Firefly luciferase activity was measured and normalized to the internal control, Renilla luciferase activity.

**Table 3.1. Datasets.**

<table>
<thead>
<tr>
<th>Cell type</th>
<th>Data type</th>
<th>GEO accession number</th>
<th>Reference</th>
</tr>
</thead>
</table>

75
<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Assay</th>
<th>ChIP-seq Peaks</th>
<th>GEO accession numbers</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>mESCs (E14)</td>
<td>ChIP-seq (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1 and CTCF)</td>
<td>GSE11431</td>
<td>(Chen et al., 2008)</td>
<td></td>
</tr>
</tbody>
</table>
| mESCs (E14) | MeDIP-seq : 5mC
ChIP-seq : H3K27ac | GSE38596
GSE36114 | (Yu et al., 2012)
(Xiao et al., 2012) |
| mESC (E14)
NSC, NPC | ChIA-PET | GSE44067 | (Zhang et al., 2013) |
| mESC(E14)-Endomesoderm | ChIP-seq : H3K4me1, H3K4me2 | GSE36114
GSE38596 | (Xiao et al., 2012) |
| mESCs(E14TG2a.4) | 5hmC (Williams) | GSE24841 | (Williams et al., 2011) |
| mESCs(E14Tg2A) | Bisulfite-seq : 5hmC | GSE36173 | (Yu et al., 2012) |
| mESCs(J1/E14) | TET1,MeDIP-Seq: 5hmC | GSE28500 | (Cortellino et al., 2011) |
| mESCs (J1) | MeDIP-Seq: 5mC, 5hmC | EBI: EPR000570 | (Cortellino et al., 2011) |
| mESCs (V6.5) | GROseq | GSE27037 | (Min et al., 2011) |
| mESCs (V6.5) | ChIP-seq : H3K4me1, H3K4me2, H3K4me3, H3K27me3, PolII | GSE11172
GSE12241 | (Meissner et al., 2008; Mikkelsen et al., 2007) |
| mESCs (V6.5) | 5hmC(Pastor)-GLIB
5hmC(Pastor)-CMS | GSE28682 | (Pastor et al., 2011) |
| mESCs(V6.5) | RNA-seq : Tet1kd, Tet2kd | GSE50198 | (Huang et al., 2014) |
| mESCs (Tdgnfl/fl) | 5fC | GSE41545 | (Song et al., 2013) |
| mESCs(46C) | MeDIP-Seq : 5hmC (Tan) | GSE40810 | (Tan et al., 2013) |
| mNPCs | MeDIP-Seq : 5hmC (Tan) | GSE40810 | (Tan et al., 2013) |
| Mouse adipocyte (3T3L1) | ChIP-seq : PPARγ, PolII | GSE13511 | (Nielsen et al., 2008) |
| hESCs (H1) | Bisulfite-seq : 5hmC | GSE36173 | (Yu et al., 2012) |
| | H3K4me1, H3K27ac, DNasel | ENCODE | (Bernstein et al., 2010) |
| | GROseq | GSE41009 | (Sigova et al., 2013) |
3.4 Results

A group of 5hmC-enriched distal TFBSs is lacking activating histone marks and nascent RNA transcription

A recent survey had found 5hmC enriched at TFBSs in hESC (Stroud et al., 2011), mouse neuronal cells, and adipocytes (Serandour et al., 2012). Therefore, we investigated 5hmC levels (Xu et al., 2011) at the binding sites of 13 key transcription factors (TFs) (Nanog, Oct4, STAT3, Smad1, Sox2, Zfx, c-Myc, n-Myc, Klf4, Esrrb, Tcfcp2l1, E2f1 and CTCF) in mESC (Chen et al., 2008). We confirmed previous results (Stroud et al., 2011; Wu et al., 2011) that 5hmC was generally depleted at the core of the proximal (within 2 kb to transcription start sites (TSSs)) TFBSs, but relatively high in the regions neighboring (±2 kb) the core (Figure 3.1A). We also confirmed that 5hmC is highly enriched at the core of distal binding sites of many TFs, such as Zfx and Esrrb (Figure 3.1B; (Stroud et al., 2011; Wu et al., 2011)).
Figure 3.1 5hmC profile at promoters and enhancers. (A) The average profiles of 5hmC at promoter-proximal TFBSs. 0 represent the center of the TFBSs. (B) The average profile of 5hmC at 13 TFBSs in promoter-distal (>2kbps) regions.

To further investigate the role of 5hmC in gene regulation in conjunction with other epigenetic marks, we performed an integrative analysis using 5hmC, 5mC (Xu et al., 2011), Tet1 (Ficz et al., 2011), H3K4me1/2/3, H3K27me3, RNA polymerase (Pol) II occupancy (Mikkelsen et al., 2007) and nascent RNAs from global run-on sequencing (GROseq) (Min et al., 2011) data. We found that 5hmC levels were inversely correlated with nascent RNA transcription and Pol II occupancy at proximal TFBSs (Figure 3.2). We confirmed the levels of 5hmC positively correlated with the levels of the repressive H3K27me3 histone mark at proximal TFBSs (Matarese, 2011; Wu et al., 2011).

Figure 3.2 5hmC and other epigenetic modifications in ESCs. (A) Correlation between 5hmC and various marks. The TFBSs were sorted based on the 5hmC levels in
±2 K regions relative to the center of the binding sites. 5hmC levels at promoter-proximal TFBSs were positively correlated with H327me3 levels and inversely correlated with GROseq and PolII levels. Transcription levels of the genes associated with the promoter were calculated using GROseq. In the sorted list, we averaged the transcription levels of the adjacent 100 genes. (B) Clustering results of 5hmC with other epigenomic data at distal (>2kbp from known TSSs) TFBSs. Cluster 1, 8 and 10 are enriched for H3K4me3 and GROseq, showing the properties of promoters. Cluster 5 and 9 display high levels of H3K27ac, indicative of active enhancers. Cluster 2 is enriched for 5hmC and 5fC, has very low GROseq levels, and lacks all investigated histone marks.

To study the epigenetic landscapes surrounding distal TFBSs, we applied the K-means algorithm (K = 10) and found clusters marked by various epigenetic modifications (Figure 3.2B). Clusters 1, 8 and 10 showed the properties of active promoters: H3K4me2/3 enrichment with relatively low levels of H3K4me1 and the presence of nascent RNA transcripts. These clusters thus likely represent the promoters of long intergenic non-coding RNAs (Guttman et al., 2009) or un-annotated promoters of protein-coding genes. Clusters 5 and 9 showed H3K4me1 and H3K27ac enrichment, indicating active enhancers. These clusters, as well as clusters 3, 4, 6, and 7, showed only a small amount of nascent transcripts or enhancer RNAs (eRNAs), which have been known to correlate with the gene transcription levels of adjacent genes (Core et al., 2008; Wang et al., 2011). The presence of eRNAs in these clusters suggest that the TFBS at these clusters have an activating role.

We were especially interested in cluster 2, which was enriched for 5hmC, but was depleted of eRNAs. Strikingly, this cluster had no activating histone marks such as
H3K4me1 or H3K27ac (Creyghton et al., 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2011), even though TFs bind at these sites (Figure 3.2B and Figure 3.3). 5mC was depleted at the core of the TFBS, consistent with the previous observation in hESCs (Lister et al., 2009). Compared with other clusters, cluster 2 was characterized by low levels of eRNAs and low PolII occupancy. To confirm the enrichment for 5hmC, we investigated the profile of sequencing data from other independent studies (Pastor et al., 2011; Tan et al., 2013; Williams et al., 2011; Wu et al., 2011; Wu and Zhang, 2011; Xu et al., 2011). Cluster 2 was enriched for 5hmC consistently for all four independently measured datasets (Figure 3.4). We also examined TAB-seq, which provides base-resolution sequencing of 5hmC in mESC (Yu et al., 2012). The TAB-seq profile also confirmed enrichment for 5hmC at the core of TFBSs for cluster 2 regions for both strands (Figure 3.5). Together, these data suggest that 5hmC combined with absence of H3K4me1 at distal TFBSs marks inactive enhancers.
Figure 3.3 Comparison of the characteristics of each cluster. We plotted the levels of H3K4me1/2/3, H3K27ac, H3K27me3, 5mC, 5hmC, GROseq, PolII and 5fC for all clusters. Cluster2 is significantly enriched with 5hmC and 5fC, but depleted for enhancer marks (H3K4me1/2, H3K27ac), eRNA (GROseq) and PolII. The high level of 5mC is due to the enrichment for 5mC at the regions flanking the TFBSs.

Figure 3.4 Comparison of the 5hmC patterns for each cluster. 5hmC data were obtained from (Pastor et al., 2011; Tan et al., 2013; Williams et al., 2011; Wu et al., 2011; Wu and Zhang, 2011; Xu et al., 2011). Cluster 2 shows consistent enrichment of 5hmC for all four independent studies.
Figure 3.5 The 5hmC profile of cluster 2 using TAB-Seq. The average profiles of 5hmC were generated using TAB-Seq data in mESC. 5hmC is enriched at the center of TFBSs for both strands.

Surprisingly, cluster 2 is also highly enriched for 5-formylcytosine (5fC) compared with other clusters (Figure 3.2B). Both 5fC and 5hmC are involved in the active demethylation pathway (Ito et al., 2011; Maiti and Drohat, 2011). Previous genome-wide study using 5fC revealed that 5fC is enriched at enhancers, especially at poised enhancers marked by H3K4me1 without H3K27ac (Song et al., 2013). However, the properties of the cluster 2 regions are novel, as they lack the H3K4me1 mark. This strongly suggests that 5hmC as well as 5fC mark a novel type of “poised” or silenced enhancer at distal regulatory regions where active histone modification marks are absent.

Next, we interrogated the state of the 5hmC mark in other cell types. In hESCs, we also identified a cluster enriched for 5hmC (Yu et al., 2012) but depleted for both H3K4me1 and H3K27ac at distal DNaseI hypersensitive sites (DHSs) (Figure 3.6; (Bernstein et al., 2010)). As in mESCs, GROseq levels in hESCs (Sigova et al., 2013)
were significantly weaker in this cluster (p-value = 1.7e-14). In mature adipocytes, we observed 5hmC (Serandour et al., 2012) enriched at over 20% of PPARγ binding sites (Figure 3.7; (Nielsen et al., 2008)). Surprisingly, PolII occupancy (Nielsen et al., 2008) was depleted when 5hmC was enriched (Figure 3.7). These data indicate that 5hmC can be a repressive mark at distal regulatory regions regardless of cell type or differentiation state.

Figure 3.6 The 5hmC clusters in hESCs. We performed clustering using 5hmC (Yu et al., 2012), H3K4me1 and H3K27ac at distal DHSs in hESCs (Bernstein et al., 2010) and identified 2 groups. Among 72,395 distal DHSs, we identified 8,034 sites (Group 1) enriched for 5hmC, while H3K4me1 and H3K27ac were depleted. GROseq levels (Sigova et al., 2013) were significantly weak in this cluster.
Figure 3.7 The 5hmC clusters in mature adipocytes. We clustered 5hmC (Serandour et al., 2012) with PolII (Nielsen et al., 2008) using the K-means clustering (K=4) algorithm at distal PPARγ binding sites (Nielsen et al., 2008). 5hmC and PolII were exclusively observed at these enhancers.

Table 3.2 lists the number of binding sites for each TF in cluster 2 in mESCs. The majority of the cluster 2 regions were bound by CTCF, Tcfcp2l1 or Esrrb. Fewer binding sites for Oct4, Sox2, and Nanog, the master regulators for self-renewal and pluripotency in ESCs, were observed in cluster 2 (Boyer et al., 2005). This is consistent with the observation that 5hmC is depleted at highly active enhancers in ESCs. We further investigated if ChIP intensity is lower for the TFBSs in cluster 2. We did not find statistical differences, even though the average profiles of the TFBSs in cluster 2 were slightly lower compared with the TFBSs in other clusters (Figure 3.8).
Table 3.2 The frequency of transcription factor occupancy in cluster 2.

Hyper-geometric p-values were calculated.

<table>
<thead>
<tr>
<th></th>
<th>Number of TFBSs in cluster 2</th>
<th>Total number of distal TFBSs in mESCs</th>
<th>Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCF</td>
<td>2,728</td>
<td>27,051</td>
<td>10.08%</td>
<td>&lt;1e-130</td>
</tr>
<tr>
<td>E2f1</td>
<td>73</td>
<td>9,730</td>
<td>0.75%</td>
<td>1</td>
</tr>
<tr>
<td>Esrrb</td>
<td>1,004</td>
<td>14,102</td>
<td>7.12%</td>
<td>4e-24</td>
</tr>
<tr>
<td>Klf4</td>
<td>108</td>
<td>5,752</td>
<td>1.88%</td>
<td>1</td>
</tr>
<tr>
<td>cMyc</td>
<td>5</td>
<td>861</td>
<td>0.58%</td>
<td>1</td>
</tr>
<tr>
<td>nMyc</td>
<td>24</td>
<td>2,158</td>
<td>1.11%</td>
<td>1</td>
</tr>
<tr>
<td>Nanog</td>
<td>128</td>
<td>8,912</td>
<td>1.44%</td>
<td>1</td>
</tr>
<tr>
<td>Oct4</td>
<td>33</td>
<td>2,674</td>
<td>1.23%</td>
<td>1</td>
</tr>
<tr>
<td>Sox2</td>
<td>29</td>
<td>3,919</td>
<td>0.74%</td>
<td>1</td>
</tr>
<tr>
<td>Smad1</td>
<td>2</td>
<td>1,032</td>
<td>0.19%</td>
<td>1</td>
</tr>
<tr>
<td>STAT3</td>
<td>33</td>
<td>1,867</td>
<td>1.77%</td>
<td>1</td>
</tr>
<tr>
<td>Tcfcp2l1</td>
<td>1,026</td>
<td>17,881</td>
<td>5.73%</td>
<td>0.0017</td>
</tr>
<tr>
<td>Zfx</td>
<td>85</td>
<td>3,881</td>
<td>2.19%</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 3.8 The average profiles of TFs at cluster 2. The average profiles of 13 TFs in cluster 2 and other clusters. The p-value from t-test was calculated using the signals around ±100bps. The significant p-values for Smad1 is because of the low number (2) of binding sites and for E2f1 the flanking signals around the binding sites.
5hmC-enriched distal TFBSs are associated with developmental genes

Next, we analyzed the correlation between 5hmC levels and transcriptional activity of the genes closest to the TFBSs for each cluster. To calculate gene transcription levels, we calculated the reads per kilobase per million mapped reads (RPKM) from GROseq (see Methods). The genes mapping to the TFBSs in cluster 2 had strikingly reduced transcription levels compared to the genes in all other clusters (p-value <1.3e-20), even compared to clusters 8 and 10, where the repressive H3K27me3 mark was relatively enriched (Figure 3.9A). GO analysis of the genes closest to the TFBSs in cluster 2 using GREAT (McLean et al., 2010) revealed that the genes in this cluster were enriched for developmental functions, such as “muscle cell development” (p-value = 3.4e-14)” and “foregut morphogenesis” (p-value = 5.8e-9) (Figure 3.9B). This is consistent with the fact that these genes are silent in ESCs and are only activated once differentiation commences.
Figure 3.9 Functional analysis for TFBSs with 5hmC. (A) Comparison of the transcription levels of the nearest genes between cluster 2 and the other 9 clusters of distal TFBSs identified in Figure 3.2B. The transcription levels (RPKM) were calculated using GROseq data. (B) GO analysis for the genes close to TFBSs in cluster 2. Organ development terms are enriched.

A snapshot in Figure 3.10 shows the enrichment for 5hmC at the Klf4 and the Esrrb binding sites located in the first intron of Sorcs2. Sorcs2 is highly expressed in the developing and mature murine central nervous system (Rezgaoui et al., 2001). We observed that Sorcs2 is silent in mESC, and its promoter is bivalently marked by H3K4me3 and H3K27me3 (Mikkelsen et al., 2007). In mouse neural progenitor cells (NPCs), however, Sorcs2 is highly expressed (Mikkelsen et al., 2007). The Klf4 and the Esrrb binding sites are marked by H3K4me1 in NPCs, suggesting an active role of this region as an enhancer during neural development.
Figure 3.10 Chromatin organization at the Sorcs2 gene in mESCs. The Klf4 and the Esrrb binding sites in the intron of Sorcs2 gene are enriched for 5hmC. These binding sites are depleted with H3K4me1 in mESCs, but are enriched in NPCs.

5hmC-enriched distal TFBSs become activated during development in a lineage-specific way

Because of the coordination of high 5hmC levels with low expression of genes in cluster 2, we hypothesized that 5hmC may attenuate enhancer activity in mESCs, which becomes activated later during development. Therefore, we analyzed H3K4me1/2 data in NPCs (Mikkelsen et al., 2007) and endomesoderm cells (Yu et al., 2013). H3K4me1 and H3K4me2 are known to mark enhancers (Heintzman et al., 2007). While H3K4me1/2 enhancer marks were depleted in mESCs, around 9% of distal TFBSs (out of 5,278 TFBSs) showed enriched H3K4me1/2 occupancy in NPCs, and an additional 20% of TFBSs were enriched for H3K4me1/2 in endomesoderm cells (Figure 3.11). Overall, 5hmC levels were significantly decreased in cluster 2 regions after differentiation into NPCs (Tan et al., 2013). This finding suggests that a group of 5hmC-enriched enhancers are repressed in mESCs, but selectively become activated during development towards the neuronal or endomesoderm lineage. This implies that other regions in cluster 2 might be activated when ESCs are differentiated into other lineages such as primordial germ cells.
Figure 3.11 Lineage specific activation of distal TFBSs in cluster 2. (A) The enriched H3K4me1/2 in NPCs or endomesoderm cells suggests the potential lineage specific enhancer activation of the TFBSs in cluster 2 after differentiation. (B) The target genes of cluster 2 in NPCs identified using ChIA-PET become significantly upregulated (p-value: 0.04) when they gained interactions. The expression change in MEF is compared as a control (p-value: 0.12).

To further determine if the TFBSs in cluster 2 have activating roles in a lineage-specific way, we used chromatin connectivity maps from chromatin interaction analysis with paired-end tagging (ChIA-PET) associated with PolII in ESCs and NPCs (Zhang et al., 2013). By using the chromatin interaction information, we mapped the target genes of the TFBSs in cluster 2 that were only selectively activated in NPCs. The target genes in ESCs were slightly downregulated in NPC (as well as in mouse embryo fibroblasts (MEFs)) because only a small portion of them become activated in a lineage-specific manner as shown in Figure 3.11A. In contrast, the target genes in NPCs become
significantly upregulated during the transition from ESCs to NPCs (p-value <0.05). Moreover, the changes were significant compared with the target genes for clusters (all p-values were <0.02) (Figure 3.12). This further supports the notion that TFBSs in cluster 2 become activated in a lineage specific way following embryonic stem cell differentiation.

Figure 3.12 The gene expression change for the target genes for each cluster. Using ChIA-PET, we obtained the target genes in ESCs and in NPCs. The genes with interacting chromatin in ESCs overall down-regulated. The target genes of cluster 2 significantly upregulated compared with other clusters (all p-values<0.02). (* if p-value<0.001 and ** if p-value <0.0001)
Cluster 2 regions show enhancer activity in mESCs when devoid of 5hmC

Next, we directly determined if lack of 5hmC activates enhancer activity of the distal TFBSs in mESCs. We selected five highly 5-hydroxymethylated distal TFBSs regions from cluster 2, including the site in Sorcs2 (TFBS1). This site is also enriched for ChIA-PET (Zhang et al., 2013) reads. We amplified these sequences (±600 bp) by PCR and subcloned them into luciferase reporter plasmids containing a minimal promoter. We found that these DNA sequences indeed possess enhancer activity in mESCs when lacking 5hmC, showing on average 3-fold increased luciferase activity compared to control (Figure 3.13). This in vitro study suggests that 5hmC-enriched distal TFBSs are bona fide enhancers, which are silent in mESCs when marked with 5hmC.

Figure 3.13 Enhancer activity of distal TFBSs lacking 5hmC in mESCs. (A) Schematic diagram of the experimental setup. (B) Luciferase reporter assay for 5hmC-enriched putative enhancer regions (about 600 bp) including distal TFBSs in
mESCs. Control: empty vector, TFBS1: 5hmC-enriched Klf4/Esrrb binding site (Figure 3.10), TFBS2: 5hmC-enriched Esrrb/Tcfcp2I1 binding site, TFBS3: 5hmC-enriched Tcfcp2I1 binding site, TFBS4: 5hmC-enriched E2f1 binding site, TFBS5: 5hmC-enriched Nanog/Sox2. The normalized luciferase activity of control is set as 1. *p-value < 0.05.

We also investigated if 5hmC at distal regulatory regions has a repressive role using the Tet1 shRNA suppression experiments in mESCs (Huang et al., 2014). We found that the target genes of cluster 2 were significantly upregulated (p-value < 0.01) after Tet1 gene suppression, suggestive of repressive roles of 5hmC (Figure 3.14). We did not find a similar pattern in Tet2 shRNA-treated mESCs, possibly due to the fact that Tet2 is dominantly associated with the 5hmC present in gene bodies (Huang et al., 2014).

![Figure 3.14 The gene expression changes of the target genes after Tet1 knockdown for each cluster.](image)

Fold changes were calculated using RNAseq data after Tet1 knockdown in mESCs (Huang et al., 2014). After knockdown of Tet1, the changes were significant in cluster 2 (p-value: 0.006). For other clusters, p-value>0.01.
Figure 3.15 The 5hmC in mESC and NPC at the TFBSs in mESCs. TFBS from all clusters (except for cluster 2) were sorted based on the 5hmC levels in NPC. We used 5hmC in mESCs and NPC from (Tan et al., 2013).

Figure 3.16 5hmC at CTCF binding sites in cluster 2. 5hmC levels were shown at the CTCF binding sites in cluster 2. After differentiation into NPCs, 5hmC become depleted.
3.5 Discussion

The field of DNA methylation has expanded recently, with the identification of multiple cytosine variants; 5hmC, 5fC, and 5-carboxylicytosine (5-caC) (He et al., 2011; Ito et al., 2011). Among the cytosine variants, 5hmC has been most extensively studied (Ficz et al., 2011; Pastor et al., 2011; Stroud et al., 2011; Szulwach et al., 2011; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011; Yu et al., 2012). Although there are genome-wide 5hmC maps in several cell types, our understanding about the functional role of 5hmC remains limited.

The contribution of the 5hmC modification to gene regulation is actively debated. Recent studies found that 5hmC gain is accompanied by H3K27me3 loss at promoters and in the gene body during neurogenesis, suggesting an activating role of 5hmC (Hahn et al., 2013). On the other hand, the presence of 5hmC at the promoter of bivalently marked genes (Matarese, 2011; Pastor et al., 2011; Wu et al., 2011) and in vitro transcription studies revealed a repressive role of 5hmC at promoter regions (Robertson et al., 2011). However, the role of 5hmC at enhancers has not been investigated thus far.

We observed that enrichment of 5hmC corresponds with the depletion of eRNAs at distal TFBSs. Considering that eRNAs correlate with gene transcription (Core et al., 2008; Wang et al., 2011), we suggest that low levels of 5hmC at enhancers are required for gene expression. Importantly, we found that a subset of distal TFBSs that carry the 5hmC mark in embryonic stem cells become enriched for the activating histone mark
(H3K4me1/2) following differentiation into neural progenitors or endomesoderm, suggesting that distal TFBSs with 5hmC are repressed in mESC but become active enhancers in a lineage-specific manner. Indeed, using ChIA-PET interaction information (Zhang et al., 2013), we found that those regions that gained connections to their target genes were significantly upregulated during differentiation compared with the target genes in other clusters. This finding suggests that their target genes were repressed in ESCs and become selectively activated in a lineage-specific way.

To ascertain if the proposed “silent enhancers” identified above can indeed function as enhancers we employed luciferase reporter assays. We demonstrated that the novel distal elements, characterized by TF binding, high levels of 5hmC, and absence of the H3K4me1 “enhancer” mark, can indeed function as enhancers in mESCs if they are devoid of the 5hmC modification. This experiment is consistent with the notion that 5hmC could inhibit enhancer activity at a subset of distal TFBSs in mESCs.

Our findings are different from the work of Sérandour and colleagues (Serandour et al., 2012), who had suggested an activating role for 5hmC at distal regulatory regions. They identified 5hmC peaks after differentiation, which were surrounded by the activating H3K4me2 mark. However, more than 50% the 5hmC peaks they identified were located at genic regions, where they are known to be associated with gene activation (Ficz et al., 2011; Hahn et al., 2013; Huang et al., 2014; Kim et al., 2014; Wu et al., 2011). It is also possible that the 5hmC peaks at distal regions are associated with non-coding RNAs such as long non-coding RNAs (lincRNAs) (Guttman et al., 2010). Sérandour and colleagues also identified 5hmC at distal PPARγ binding sites (Nielsen et
al., 2008). Even though Sérandour and colleagues proposed an activating role of 5hmC at these master regulator in adipocytes, only a portion of PPARγ binding sites were enriched for 5hmC (Serandour et al., 2012). We revisited their data and found that 5hmC was only present at sited lacking PolII occupancy (Figure 3.7), indicating that 5hmC at PPARγ binding sites bears repressive roles in mature adipocytes.

In hESCs, we also identified a group of distal DHSs with strong 5hmC but weak H3K4me1 and H3K27ac (Figure 3.6). The GROseq levels were significantly weak for the group with 5hmC (Figure 3.6). These lines of evidences suggest a general repressive role of 5hmC at distal regulatory regions.

In ESCs, poised enhancers have been suggested to exist at sites where both activating marks (H3K4me1) and repressive marks (H3K27me3) are enriched, but H3K27ac is depleted (Creighton et al., 2010; Rada-Iglesias et al., 2011). 5fC is enriched in this type of poised enhancers (H3K4me1[+] and H3K27ac[−]) (Song et al., 2013). In contrast to these poised enhancers, we identify a novel group of enhancers with no activating histone marks (H3K4me1[−] and H3K27ac[−]) but enrichment only for 5hmC. Furthermore, this group is strongly enriched for 5fC, even though cluster 2 lacks the H3K4me1 mark (Figure 3.2). Our results strongly suggest that 5hmC and 5fC can be epigenetic mark for poised or silent enhancers. As shown in our results, many of these enhancers display activating histone marks only after differentiation has occurred (Figure 3.11). The existence of 5hmC and 5fC also show the active oxidation dynamics at these sites.
We found that 5hmC was enriched at distal PPARγ binding sites in fully differentiated adipocytes. These findings suggest 5hmC as a new marker for poised enhancers even in absence of H3K4me1 and H3K27me3. Additionally, we also found enriched 5hmC in NPC at the subset of the active TFBSs (except for cluster 2) in mESCs (Figure 3.15). This may suggest that active enhancers in mESCs are repressed by 5hmC in NPC to remove the enhancer activities in mESCs.

The majority of cluster 2 regions are CTCF binding sites (Table 3.2). In general, 5hmC levels negatively correlated with CTCF occupancy in cluster 2 (Figure 3.16). After differentiation into NPCs, 5hmC became depleted at these sites even though the binding CTCF remained. At these sites, we did not observe activating H3K4me1 and H3K4me2 marks. However, it is difficult to discuss the role of 5hmCs at these sites, because CTCF takes part in various regulatory roles including transcriptional activation, repression, as well as the formation of higher order chromatin structure (Nikolaev et al., 2009). The function of 5hmC in mESCs at CTCF binding sites warrants further study.

We report a new repressive role for 5hmC in gene regulatory regions in mESCs. The TFBSs enriched for 5hmCs were depleted for nascent transcripts and activating histone modification marks in human and mouse ESCs. Furthermore, the 5hmC levels were inversely correlated with PolII occupancy in mESCs as well as in fully differentiated adipocytes. Our findings indicate that 5hmC has a repressive role at specific distal regulatory regions and suggest that 5hmC is a new epigenetic mark for silenced enhancers.
3.6 Reference


Conclusions and Future direction
4.1 TET1-mediated DNA hydroxymethylation is required for the regulation of gene expression in intestinal stem cells.

5-hydroxymethylcytosine (5hmC) is converted from 5-methylcytosine (5mC) by Ten-eleven translocation (TET) enzymes, which can further oxidize 5hmC to 5-formyleytosine and 5-carboxycytosine (Ito et al., 2011; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). These oxidation products are thought to be involved in DNA demethylation in mammals (Hashimoto et al., 2012; He et al., 2011; Maiti and Drohat, 2011; Valinluck and Sowers, 2007). This suggests that 5hmC could be the first step towards reversing the repressive effects of 5mC on gene expression (Ficz et al., 2011; Nestor et al., 2012). Recently, it has been suggested that TET1 and 5hmC are important for gene expression during differentiation of a human colon carcinoma epithelial cell line (Chapman et al., 2015).

In Chapter 2, I investigated the role of 5hmC and TET1 during intestinal stem cell (ISC) maturation and differentiation. To determine the genome-wide map of 5hmC, I isolated intestinal stem cells and differentiated villus cells from adult mice and performed hydroxymethylated DNA immunoprecipitation followed by next-generation sequencing. I found that 5hmC is enriched at highly expressed genes such as the Wnt target genes in ISCs, and at genes encoding metabolic and transport function in differentiated cells. To examine if altered DNA hydroxymethylation leads to intestinal defects, I used Tet1<sup>−/−</sup> mice and found growth retardation and postnatal lethality in the mutant mice. In the Tet1-deleted postnatal intestine, I observed reduced proliferation, and decreased Wnt target
gene expression levels, which correlated with significantly decreased 5hmC levels at the relevant promoter regions in Tet1 null mice. This finding suggests that Tet1-mediated DNA hydroxymethylation is essential for epigenetic regulation of Wnt target and stem cell genes in ISCs.

Although I showed that reduced expression of several Wnt target genes correlated with decreased levels of 5hmC at their promoters in Tet1 null intestinal crypts, in the future it would be interesting to determine differential 5hmC levels and gene expression changes in the Tet1 null crypt compared to wild-type crypts on genome-wide in an unbiased manner using next-generation sequencing. To investigate the details of 5hmC dynamics during intestinal epithelial differentiation further, single-base resolution mapping of 5hmC in ISCs and differentiated cells would be helpful. We have already published single-base resolution map of 5mC in ISCs and differentiated cells (Sheaffer et al., 2014), and have started preparing DNA libraries for 5hmC mapping at single-base resolution. With both 5mC and 5hmC maps at single-base resolution, we can identify DNA methylation and hydroxymethylation levels at each cytosine residue precisely and characterize their relationship to gene expression with our existing RNA sequencing data (Sheaffer et al., 2014).

Whereas genome-wide maps for a few transcription factors and histone modifications in ISCs and differentiated cells are already available (Kim et al., 2012; San Roman et al., 2015; Sheaffer et al., 2014; Verzi et al., 2013), epigenomic information of intestinal differentiation and specific cell types is still limited. We have performed chromatin-immunoprecipitation sequencing for several histone modifications, such as
H3K4me1, H3K4me3 and H3K27me3, and defined open chromatin regions using transposase-accessible chromatin sequencing (ATAC-seq) in ISCs and differentiated cells. With all these datasets, integrative analyses will be required for a comprehensive view of epigenomic regulation during intestinal stem cell differentiation. We have provided the basis for this integrative analysis, and current and future research is directed at this problem.

We found postnatal lethality and intestinal defects in Tet1−/− mutant mice, which is similar to the phenotype seen in mice with conditional deletion of Dnmt1 in the developing intestine using Dnmt1<sup>f/f</sup>; VillinCre mice (Elliott et al., 2015; Yu et al., 2015). This suggests DNA methylation and demethylation are both important for intestinal maturation in the postnatal period. But, still, the definitive role of Tet1-mediated DNA hydroxymethylation in the intestine warrants further investigation using intestinal-specific gene ablation models. We have begun to derive Tet1-3<sup>f/f</sup> mice for conditional gene deletion in the intestinal epithelium.

To ascertain the critical role of Tet1 in intestinal stem cells, we could examine if regenerative capacity of Tet1-deleted intestine after radiation injury is altered in future studies. In the wild-type intestine, intestinal stem cells can repopulate the intestinal epithelium and show robust regenerative capacity after damage (Booth and Potten, 2000). We have preliminary results that intestinal regenerative capacity of Tet1-deleted mice is compromised, which warrants further investigation.
4.2 5hmC can function as a repressive mark at distal transcription factor binding sites

5hmC is not just an intermediate product of the DNA demethylation process. It is abundant in post-mitotic neuronal cells, and relatively stable showing distinct patterns in each tissue of the body (Bachman et al., 2014; Kriaucionis and Heintz, 2009; Li and Liu, 2011), and have specific binding proteins (“readers”) that regulate gene expression (Frauer et al., 2011; Mellén et al., 2012; Spruijt et al., 2013; Yildirim et al., 2011). Therefore, 5hmC could function as a new stable or semi-stable epigenetic mark, and not just as a kinetic intermediate towards demethylation. 5hmC is also present at bivalent promoters of poised developmental genes (Pastor et al., 2011; Wu et al., 2011) and has a role in pausing or silencing gene expression (Robertson et al., 2011; Szulwach et al., 2011).

In Chapter 3, we identified putative enhancer sites, which are transcription factor binding sites (TFBSs) with high levels of 5hmC and no active histone modifications, in mouse embryonic stem cells (ESCs). We performed integrative analyses with genome-wide sequencing datasets of various epigenetic factors, transcription factors, nascent RNAs and RNA Polymerase II in ESCs, and found about distal 5,000 transcription factor binding sites as putative enhancer sites fitting these criteria. A subset of them lost 5hmC and gained active histone marks during ESC differentiation towards a specific lineage. We also determined the enhancer activity of a subset of these elements when they were devoid of 5hmC using reporter assays. These findings suggest that 5hmC has a repressive role at a subset of distal TFBSs, and marks silenced developmental enhancers in ESCs.
*Tet1* and *Tet2* are highly expressed in ESCs and are required for ESC maintenance and proper differentiation. Depletion of *Tet1* and *Tet2* in ESCs causes spontaneous and/or skewed differentiation. (Ito et al., 2010; Koh et al., 2011). It is possible that the silenced developmental enhancers described above could be activated in these conditions due to decreased levels of 5hmC in *Tet1* and *Tet2*-deficient ESCs, and that this aberrant enhancer activation contributes to the mutant phenotype.

To further validate the function of silenced developmental enhancers during ESCs differentiation, we could delete them using clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR associated protein system (Cong et al., 2013) in ESCs and examine the differentiation capacity of mutant ESCs. With this approach, one could determine whether silenced developmental enhancers are actually necessary to be activated by loss of 5hmC and the acquisition of active histone modifications.
4.3 The dual role of 5hmC in gene activation and repression

Targeting the TET1 catalytic domain to specific CpGs in the promoter proximal regions for DNA demethylation leads to target gene activation (Maeder et al., 2013). In chapter 2, high levels of 5hmC in promoters and gene bodies were correlated with active gene expression in intestinal stem and differentiated cells. This suggests that 5hmC is involved in DNA demethylation and maintaining an open chromatin structure for active transcription. However, it has been also shown that 5hmC is enriched at bivalent promoters of poised developmental genes in embryonic stem cells (Pastor et al., 2011; Wu et al., 2011). In chapter 3, enrichment of 5hmC suppresses enhancer activity of a group of distal transcription factor binding sites, which is associated with poised developmental genes in embryonic stem cells. This suggests that 5hmC is not only involved in DNA demethylation for gene activation, but can also contribute to gene repression as a stable epigenetic modification in bivalent promoters and silent enhancers of poised genes in stem cells.
4.4 Reference


