# TET PROTEINS SAFEGUARD BIVALENT PROMOTERS FROM DE NOVO METHYLATION IN HESCS

A Dissertation
Presented to the Faculty of the Weill Cornell Graduate School
of Medical Sciences
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

By
Nipun Verma
August 2017

TET proteins safeguard bivalent promoters from *de novo* methylation in hESCs

Nipun Verma, Ph.D.

Cornell University 2017

The TET enzymes oxidize 5-methylcytosine to 5-hydroxymethylcytosine, which can lead to DNA demethylation. However, direct connections between TET-mediated DNA demethylation and transcriptional output are difficult to establish due to challenges of distinguishing global versus locus-specific effects. Here we show that *TET1/2/3* triple knockout (TKO) human embryonic stem cells (hESCs) exhibit prominent bivalent promoter hypermethylation without overall corresponding gene expression decrease in the undifferentiated state. Focusing on the bivalent *PAX6* locus, we find that increased DNMT3B binding is associated with promoter hypermethylation, which precipitates a neural differentiation defect and failure of *PAX6* induction during differentiation. dCas9-mediated locus-specific demethylation and global inactivation of *DNMT3B* in TKO hESCs partially reverses the hypermethylation at the *PAX6* promoter and improves differentiation to neuroectoderm. Taken together with further genome-wide methylation and TET1 and DNMT3B ChIP-Seq analysis, we conclude that TET proteins safeguard bivalent promoters from silencing by *de novo* methylation to ensure robust lineage-specific transcription upon differentiation.

## BIOGRAPHICAL SKETCH

Nipun Verma was born in New Delhi, India. She moved to the United States when she was four and grew up in Miami, Florida. She attended Cornell University and graduated in 2010 after majoring in Biological Sciences and minoring in Chemistry. She moved to New York City to attend medical school at Weill Cornell Medical College (WCMC). After her first year in medical school Nipun volunteered in Dr. Danwei Huangfu's lab as part of the WCMC Clinical and Translational Science internship program. Then after returning to medical school for a year, Nipun decided to work full time in the Dr. Huangfu's lab as part of the Howard Hughes Medical Institute Fellowship Program. During the fellowship year Nipun transferred into the Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD program and worked as a PhD graduate student in Dr. Huangfu's lab for 4 more years. While in Dr. Huangfu's lab Nipun has focused on the role of the TET proteins in hESC pluripotency and differentiation. In this project Nipun has found that the TET proteins are necessary to protect bivalent promoters from abnormal hypermethylation in human embryonic stem cells. This enables the differentiation-associated genes that typically have bivalent promoters to be efficiently activated and expressed during differentiation and development. This establishes a crucial new function of the TET proteins during development and opens up questions on whether they play a similar function in other stem cell populations in which they are abundant. After completing her thesis and finishing medical school Nipun will embark on a career that combines clinical practice and biomedical research.

ACKNOWLEDGEMENTS

Thanks to my mentor, Dr. Danwei Huangfu, for excellent and unique ideas that have been fun to ponder and investigate during these past five years. Thank you for your guidance, support and optimism and for pushing me to become a better scientist. And also thank you for helping me improve my skills in presenting and communicating my work. I am grateful for your consistent and attentive mentorship throughout my PhD.

I want to thank my committee members Dr. Lorenz Studer, Dr. Todd Evans and Dr. Kitai Kim for your advice and support throughout my years as a graduate student. Your insights have been invaluable for my research and having you on my committee has given me additional wonderful scientific role models.

Thank you to the members of Dr. Danwei Huangfu's lab both past and present. I have learned a lot from each and every one of you. Thank you all for your kind support and your useful criticisms that have been vital for me in completing my PhD. Thank you to Federico for sharing your vast knowledge on research and always being very generous with your time. Big thanks to Virginia, Qing and Daniela, I am very grateful to have met and worked with you. You have helped me get through some of the toughest points of the PhD with your generosity, kindness and support.

The most profound thanks to my family. To my siblings, Nupur and Neil, and my parents, Usha and Ashok. Thank you for your love and support throughout my PhD. Thank you for pushing me to work hard but also

reminding me to have fun and enjoy what I am doing. Thank you for always believing in me. I could not have done it without you all.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH ..... iii
ACKNOWLEDGEMENTS ..... iv
LIST OF FIGURES ..... ix
LIST OF TABLES ..... xi
CHAPTER 1: Introduction ..... 1
1.1 EPIGENETICS ..... 1
1.1.1 Noncoding RNA ..... 2
1.1.2 Histone Modifications ..... 3
1.1.3 DNA Methylation ..... 4
1.2 DNA METHYLATION AND DEMETHYLATION PROTEINS ..... 6
1.2.1 DNA Methylation ..... 6
1.2.2 DNA Demethylation ..... 7
1.3 TET-MEDIATED DNA DEMETHYLATION ..... 9
1.3.1 Passive Replication-Dependent Dilution ..... 9
1.3.2 Active DNA Demethylation Through DNA Repair ..... 11
1.3.3 Enzymatic Decarboxylation of 5caC ..... 13
1.3.4 Dehydroxylation by DNMT Enzymes ..... 13
1.4 DNA METHYLATION AND TRANSCRIPTIONAL REGULATION ..... 14
1.4.1 Promoters ..... 14
1.4.2 Gene Bodies ..... 18
1.4.3 Enhancers ..... 18
1.4.4 Insulators ..... 19
1.5 ADDITIONAL FUNCTIONS OF DNA METHYLATION ..... 23
1.5.1 Transposons ..... 23
1.5.2 Pericentromeric Repeats ..... 23
1.5.3 X-Chromosome Inactivation ..... 23
1.5.4 Imprinted Genes ..... 24
1.6 5HMC MAPPING IN THE MAMMALIAN GENOME ..... 25
1.6.1 Promoters ..... 27
1.6.2 Gene Bodies ..... 27
1.6.3 Enhancers ..... 28
1.7 DNA METHYLATION DURING DEVELOPMENT ..... 28
1.7.1 Demethylation in the Early Zygote ..... 29
1.7.2 Methylation Changes in the Early Embryo ..... 31
1.7.3 Methylation Changes in Primordial Germ Cells ..... 34
1.8 PHENTOTYPES OF DNMT- AND TET-DEFICIENT MICE ..... 36

1.9 PHENTOTYPES OF DNMT- AND TET-DEFICIENT MESCS .....39
1.10 ROLE OF TET PROTEINS IN REPROGRAMMING .....42
1.11 THESIS AIMS .....44
CHAPTER 2: TET Protein Requirement and Function of Oxidized Methylcytosines in hESCs .....46
2.1 INTRODUCTION .....46
2.2 RESULTS .....40
2.2.1 Bivalent Promoter Hypermethyaltion in TKO hESC……40
2.2.2 Hypermethylation of the *PAX6* P0 Promoter……48
2.2.3 Impaired Neural Differentiation of TKO hESCs .....74
2.2.4 Promoter Hypermethylation Hinders *PAX6* Expression Upon Differentiation .....84
2.2.5 De novo methylation causes *PAX6* promoter hypermethylation .....97
2.3 DISCUSSION .....110
2.4 MATERIALS & METHODS .....113
2.4.1 hESC Culture .....113
2.4.2 Generation of TET mutant hESCs .....114
2.4.3 hESC Differentiation .....117
2.4.4 Dot Blot .....120
2.4.5 Mass Spectrometry Quantification of 5mC and 5hmC .....121
2.4.6 Pluripotency and Lineage Marker Staining .....121
2.4.7 FACS Analysis .....122
2.4.8 Quantitative RT-qPCR .....123
2.4.9 Epimark .....125
2.4.10 RNA-Seq .....126
2.4.11 5hmC Profiling .....126
2.4.12 Whole Genome Bisulfite Sequencing (WGBS) and Enhanced Reduced Representation Bisulfite Sequencing (ERRBS). .....127
2.4.13 ChIP-Seq .....130
2.4.14 5mC Methylation Analysis by MassArray .....131
2.4.15 PAX6 Overexpression Lentivirus Construct and Generation of PAX6 Overexpression hESCs .....134
2.4.16 Colony Forming Assay .....135
2.4.17 Cloning dCas9 TET1 Catalytic Domain Fusion and Site-Directed Mutagenesis .....136
2.4.18 Targeting dCas9-TET1CD into TKO hESCs .....136
2.4.19 Generation of TET1-3xFlag tagged hESCs .....138
2.4.20 ChIP-qPCR .....138
2.4.21 Statistical Analysis .....139
2.4.22 Data Availability .....139
CHAPTER 3: Conclusions & Perspectives .....140
3.1 SUMMARY .....140

3.2 FUTURE DIRECTIONS ..... 142
3.2.1 Identify Proteins that may be Involved in the Recruitment of DNMT3B and TET1 to Bivalent Promoters ..... 142
3.2.2 Investigate Changes in Chromatin Environment Upon TET Inactivation ..... 144
3.2.3 Investigate the Mechanisms Underlying Heterogeneity in the Gain of Methylation after TET Inactivation ..... 144
3.3 CONCLUSIONS ..... 146
CHAPTER 4: Generation of Pluripotency Reporter hESCs ..... 115
4.1 INTRODUCTION ..... 115
4.2 RESULTS ..... 117
4.2.1 CRISPR/Cas-Mediated Targeting of the OCT4 Locus ..... 117
4.2.2 Selection-free Targeting of the OCT4 Locus using a Mini-vector Donor.. ..... 120
4.2.3 Selection-free Targeting of the NANOG Locus ..... 125
4.3 DISCUSSION ..... 128
4.4 MATERIALS & METHODS ..... 131
4.4.1 Generation of constructs ..... 131
4.4.2 Cell culture ..... 133
4.4.3 In vitro transcription of gRNAs ..... 135
4.4.4 Electroporation ..... 135
4.4.5 Transfection ..... 136
4.4.6 Assessment of indel mutations using SURVEYOR nuclease ..... 137
4.4.7 Southern Blot ..... 138
4.4.8 Immunofluorescence staining ..... 139
4.5 CONCLUSIONS ..... 139
5.0 APPENDIX ..... 150
5.1 REFERENCES ..... 164

LIST OF FIGURES

Figure 2.1……50
Figure 2.2……51
Figure 2.3……54
Figure 2.4……55
Figure 2.5……59
Figure 2.6……61
Figure 2.7……65
Figure 2.8……66
Figure 2.9……68
Figure 2.10……71
Figure 2.11……72
Figure 2.12……77
Figure 2.13……78
Figure 2.14……80
Figure 2.15……82
Figure 2.16……83
Figure 2.17……86
Figure 2.18……87
Figure 2.19……90
Figure 2.20……91
Figure 2.21……94
Figure 2.22……96
Figure 2.23……99

Figure 2.24……100
Figure 2.25……101
Figure 2.26……104
Figure 2.27……106
Figure 2.28……109

LIST OF TABLES

Table 2.1……..115
Table 2.2……..117
Table 2.3……..124
Table 2.4……..131
Table 2.5……..134
Table 2.6……..135
Table 4.1……..158
Table 4.2……..164
Table 4.3……..165
Table 4.4……..165
Table 4.5……..165

# CHAPTER 1: Introduction

## 1.1 EPIGENETICS

The phenotype of a cell and ultimately the entire organism is determined by the underlying DNA sequence of its genes. Epigenetics spans the divide between genotype and phenotype by influencing the final outcome of a genetic locus without changing its DNA sequence. Despite sharing the same genetic information, epigenetic controls enable different cell types within an organism to express unique collections of genes, leading to their specialized morphology and function. Furthermore, stabilization of these gene expression patterns allows them to be recapitulated in progeny of the same lineage.

Eukaryotic chromatin consists of genomic DNA, proteins and RNA packaged into a series of layers. At the base of this organization are nucleosomes, which consist of 147 bp of DNA wrapped around a histone octamer (2 histone proteins each of H2A, H2B, H3 and H4). The DNA and histone proteins can be epigenetically modified by the addition of certain chemical groups, which then affects how compact and inaccessible the chromatin is (Felsenfeld and Groudine, 2003). Histone variants (H2A.Z) can also replace the core histones and nucleosome positioning along the chromatin fiber can be modified to further alter DNA accessibility (Zlatanova and Thakar, 2008). Lastly the chromatin is organized into a 3-dimensional structure that enables the interaction between distal genomic sites and compartmentalizes chromatin regions into certain areas of the nucleus (Bonev and Cavalli, 2016).

Epigenetics is critical for the proper development of specialized cells from the fertilized zygote. Gene expression is directed by the binding of transcription factors but is also regulated by heritable, epigenetic modifications to DNA and histones that impact developmental decisions. Thus development and differentiation processes have become an exceptional setting in which to study how epigenetic processes can influence cell fate conversion (Cantone and Fisher, 2013). Previous work has focused primarily on three modes of epigenetic regulation: non-coding RNAs, covalent and non-covalent modifications of histone proteins and DNA methylation. Although often divided into separate categories it is important to note that there is significant interaction between these different methods of epigenetic regulation (Molina-Serrano et al., 2013). Also, unlike changes to the DNA sequence, epigenetic modifications are reversible and the modifications present are influenced by the cellular context and environment. At the moment, it is clear that proper epigenetic control is necessary for successful development; however, questions on how these modifications are interpreted and inherited, what genomic regions they act upon and how they impact gene expression before, during and after cell-fate transitions still remain.

### *1.1.1 Noncoding RNA*

Non-coding RNAs (ncRNA) are RNAs that do not encode for proteins. Short ncRNAs (typically <30 nucleotides) have been found to regulate imprinting, promote heterochromatin formation and silence genes by post-transcriptional RNA interference pathways (RNAi). Longer ncRNAs (>200 nucleotides), can recruit chromatin modifying proteins to produce dramatic

changes in chromatin structure (Kowalczyk et al., 2012). This is seen most clearly with the *Xist* lncRNA, which recruits epigenetic proteins that compact chromatin, leading to long-term silencing of the X chromosome that can be inherited after cell division (Plath et al., 2002).

### *1.1.2 Histone Modifications*

Histone proteins can be epigenetically modified by the addition or removal of covalent groups including the phosphorylation, acetylation, methylation (mono, di- and tri-) and ubiquitination of specific histone residues. Charge-altering modifications, such as phosphorylation and acetylation, can change the physical properties of the chromatin molecule and thus affect its overall structure (Zhou et al., 2011). Histone modifications can also promote the recruitment of additional chromatin modifiers that can reinforce a particular chromatin state. For example H3K9 methylation is recognized by heterochromatin protein 1 (HP1), which then promotes further H3K9 methylation, DNA methylation and gene silencing (Lomberk et al., 2006). Previous work has produced extensive genome-wide profiles of histone modifications in a variety of cell types. These maps have enabled us to identify regulatory elements in the genome that impact gene expression, and draw correlations between the distribution of these histone marks and the specific gene expression profiles of pluripotency, cellular differentiation and disease states (Consortium, 2012).

Non-covalent histone modifications include chromatin remodeling and the incorporation of histone variants into nucleosomes. ATP- dependent chromatin remodelers alter the position of nucleosomes along the chromatin fiber. Areas of the chromatin that become depleted of nucleosomes typically

become more open and accessible, whereas nucleosome rich areas become more compact (Jiang and Pugh, 2009). Histone variants, such as H3.3 and H2A.Z, can be exchanged for the core nucleosomal histones and often carry their own covalent modification patterns (Kamakaka and Biggins, 2005). Subunits involved in nucleosome remodeling and incorporation of histone variants can occur in the same effector protein, suggesting a mechanistic link between these two processes (Jin et al., 2005). Together covalent and non-covalent modifications of the histone proteins produce a wide array of possible epigenetic states and enable fine-tuned control of transcription.

### *1.1.3 DNA Methylation*

DNA methylation refers to the methylation of the fifth position of cytosine to form 5-methylcytosine (5mC). This methylation primarily occurs at CpG dinucleotides and is present in plant, fungal and animal species (Goll and Bestor, 2005). The distribution of DNA methylation is described in further detail below, but generally DNA methylation follows a bimodal distribution in mammalian cells. Isolated CpGs are generally hypermethylated (60-90% depending on cell type), whereas clusters of CpGs together (200-500 base pair regions of high CpG density) called CpG islands are predominantly hypomethylated. The effect of DNA methylation on gene expression is context dependent, as methylation of CpG islands has clearly been shown to lead to reduced gene expression whereas the transcriptional outcome of methylation of isolated CpGs is less understood (Jones, 2012) (Figure 1.1).

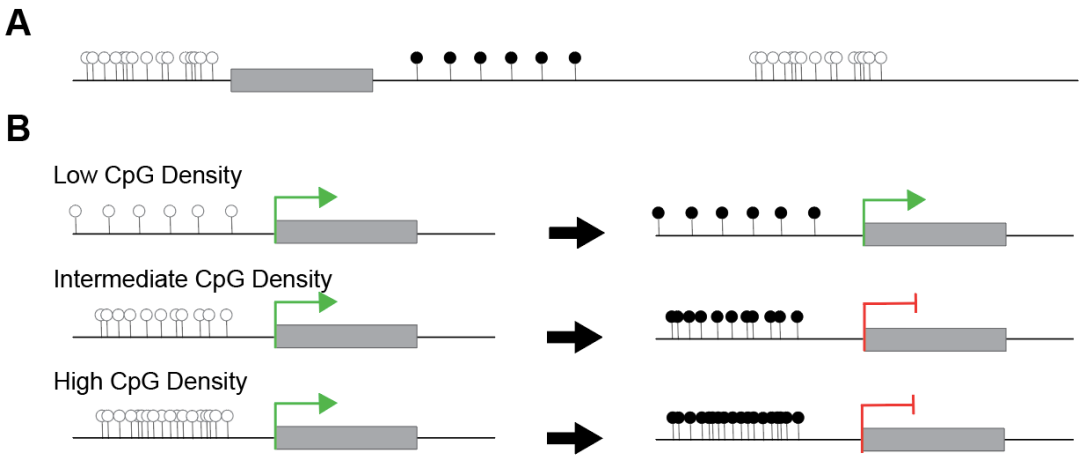



**Figure 1.1 Genomic distribution of CpG dinucleotides and the transcriptional effect of DNA methylation.** (A) Within the genome CpG dinucleotides (depicted as lollipop objects in the diagram) can be present in clusters called CpG islands (CGIs). CGIs can occur within promoters that are upstream of genes (represented by the grey box) or away from genes as “orphan” CGIs. CGIs typically have low levels of DNA methylation (shown as empty circles on the lollipop objects). CpGs dinucleotides that occur outside of clusters are called isolated CpGs. Isolated CpGs typically show high levels of DNA methylation (shown as filled circles on the lollipop objects). (B) Three different promoter types (low CpG density, intermediate CpG density and high CpG density) are shown. Generally DNA methylation of intermediate and high CpG density promoters is associated with transcription repression. In contrast DNA methylation of low CpG density promoters does not typically affect gene transcription.

Similar to other epigenetic modifications, DNA methylation can be inherited following cellular division and dynamically altered by demethylation processes. During development as cells are directed towards certain differentiation lineages, DNA methylation serves as an epigenetic barrier to stabilize the acquired cell fate and prevent regression to undifferentiated cells. DNA methylation is also necessary for X-chromosome inactivation to maintain sex chromosome dosage, the repression of retrotransposons and the coordinated expression of imprinted genes (Smith and Meissner, 2013).

## 1.2 DNA METHYLATION AND DEMETHYLATION PROTEINS

### *1.2.1 DNA Methylation*

There are 3 DNA-methyltransferases, DNMT1, DNMT3A and DNMT3B, which catalyze the transfer of a methyl group from S-adenosyl-l-methionine to the fifth position of cytosine nucleotides in DNA. DNMT1 functions as a maintenance methyltransferase that reproduces methylation marks on nascent DNA strands following DNA replication (Goll and Bestor, 2005). DNA methylation typically occurs at CpG sites, which are palindromes and are symmetrically methylated. Thus, DNA replication produces two new DNA strands that each have hemimethylated CpG sites. The protein UHRF1 recognizes these hemimethylated sites and recruits DNMT1, which then methylates the reciprocal CpG site on the newly-synthesized DNA and thus maintains DNA methylation patterns following cellular replication (Sharif et al., 2007). Due to its function in maintaining methylation patterns DNMT1 is almost ubiquitously expressed in all cell types.

DNMT3A and DNMT3B are identified as the *de novo* methyltransferases (Okano et al., 1999). Although primarily responsible for establishing novel methylation patterns, studies in mouse embryonic stem cells (mESCs) have also found that they are involved in DNA methylation maintenance (Chen et al., 2003; Liang et al., 2002). During early development DNMT3A is maternally provided and is highly expressed in oocytes and early preimplantation embryos (Kaneda et al., 2004). DNMT3B first appears during zygotic gene activation and is then robustly expressed during the blastocyst stage and later in the epiblast. The expression of DNMT3A and DNMT3B varies depending on the cell type during later development and in the different somatic lineages (Watanabe et al., 2002).

DNMT3L lacks the N-terminal catalytic domain found in the other DNA methyl-transferases (Bourc'his et al., 2001). DNMT3L functions as a stimulating cofactor for DNMT3A/DNMT3B mediated methylation and has been found to be required for the DNA methylation of imprinted genes in gametes (Bourc'his and Bestor, 2004).

### *1.2.2 DNA Demethylation*

Recently the TET (Ten-Eleven Translocation) proteins were discovered and speculated to play a role in DNA demethylation. TET1, the founding member of the TET family, was discovered as the fusion partner of Mixed Lineage Leukemia 1 (MLL1) in rare cases of acute myeloid and lymphocytic leukemia that harbor the ten-eleven translocation (t(10;11)(q22;q23)) (Lorsbach et al., 2003). A later computational screen searching for mammalian analogues of JBP1, a trypanosome protein that modifies DNA, identified the three mammalian TET proteins, TET1, TET2

and TET2, as potential DNA modifiers (Iyer et al., 2008). All three TET proteins are Fe2+ and 2-oxoglutarate-dependent dioxygenases that successively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in DNA (Ito et al., 2011; Tahiliani et al., 2009). The possible pathways for DNA demethylation after these oxidation steps are described in further detail below.

The TET proteins are expressed during early mouse development and are believed to play a role during several stages of demethylation. The expression of the individual TET proteins varies during embryonic development and in adult tissues. For example, *TET3* is present in the zygote after fertilization and is involved in the global demethylation of the paternal nucleus (Gu et al., 2011b). Following this, *TET1* and *TET2* increase in expression in the blastocyst and epiblast stages and function in modifying *de novo* methylation patterns (Ruzov et al., 2011). *TET1* and *TET2* are also expressed in developing germ cells where they may play a role in the genome-wide demethylation and loss of imprint methylation marks (Hajkova et al., 2010). Within the adult, expression of the individual *TET* genes can vary depending on the specific tissue, for instance *TET2* is the most highly expressed *TET* gene in hematopoietic cells and is frequently mutated in hematological malignancies (Abdel-Wahab et al., 2009).

The three oxidized forms of 5mC (5hmC, 5fC and 5caC) are found in numerous tissues; however, their abundance can vary between different cell types. 5hmC is most abundant in Purkinje neurons where it's level is approximately 40% of 5mC (Kriaucionis and Heintz, 2009), it is also found in embryonic stem cells (5-10% of 5mC) (Tahiliani et al., 2009) and some

immune cell populations (1% of 5mC) (Ko et al., 2010). 5fC and 5caC are found in similar cell types as 5hmC but typically at much lower levels. For instance, in embryonic stem cells 5fC and 5caC represent .03% and .01% of 5mC in embryonic stem cells (Ito et al., 2011). It is possible that this low abundance could result from the active removal of these oxidation products.

## 1.3 TET-MEDIATED DNA DEMETHYLATION

The TET proteins have generated great interest, in part because they may play a role in DNA demethylation. The TET proteins oxidize 5mC to 3 successive oxidation products: 5hmC, 5fC and 5caC. These oxidation products could then be removed and replaced by unmodified cytosines, producing demethylation. There are four postulated mechanisms through which TET proteins could mediate DNA demethylation.

### *1.3.1 Passive Replication-Dependent Dilution*

DNA demethylation can occur through passive means without the TET proteins. During cell division mitotic cells can lose DNA methylation by downregulating or excluding the maintenance methyltransferase (DNMT1) or its recruitment factor (UHRF1) from the nucleus (Figure 1.2). This passive demethylation process is believed to be involved in the demethylation of the maternal nucleus after fertilization as well as contribute to the genome wide demethylation of developing germ cells (Kagiwada et al., 2013; Ratnam et al., 2002).

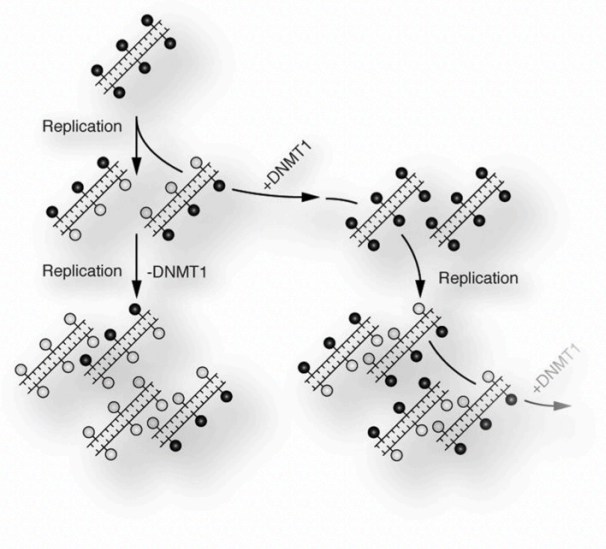



**Figure 1.2 Passive demethylation.** DNMT1 is responsible for copying methylation patterns onto the newly synthesized DNA strand following DNA replication. In the absence of DNMT1 methylation will be passively lost following each round of DNA replication. Adapted from: Messerschmidt et al., 2014.

Although passive demethylation is not dependent on the TET proteins, they could facilitate passive demethylation by producing 5hmC. Binding of UHRF1 to 5hmC is tenfold less efficient than binding to 5mC (Hashimoto et al., 2012). Furthermore the activity of recombinant DNMT1 in-vitro is reduced 12-50 fold at sites of hemi-5hmC versus hemi-5mC (Valinluck and Sowers, 2007). It is important to note that this passive demethylation is restricted to cells that are mitotically active. Thus it cannot account for the rapid loss of DNA methylation in slowly dividing or non-dividing cells.

### *1.3.2 Active DNA Demethylation Through DNA Repair*

Active removal of DNA methylation through DNA repair has been proposed to function through two pathways that are independent of DNA replication and cell division. Both pathways use TET-mediated oxidation and base excision repair (BER) (Figure 1.3). The first proposes that the 5fC and 5caC oxidation products are removed by the glycosylase TDG. The resulting empty site is then filled with an unmodified cytosine residue via BER, resulting in demethylation.

Binding of TDG to 5caC has been confirmed by electrophoretic mobility shift assays and structural studies (Zhang et al., 2012). Furthermore the depletion of TDG in embryonic stem cells results in a 2-10 fold increase in the levels of 5fC and 5caC, suggesting it is acting to remove these products. On the other hand the overall levels of 5fC and 5caC are very low in ESCs even when TDG is depleted, suggesting that this pathway has limited function in these cell types (Shen et al., 2013).

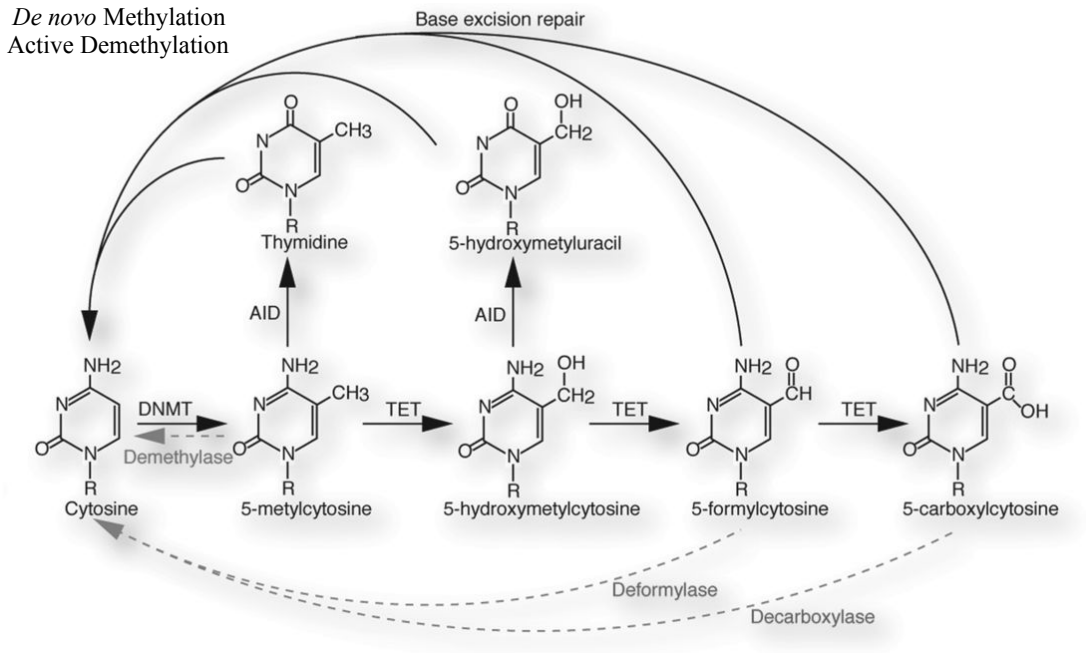



**Figure 1.3 Active demethylation.** Possible DNA demethylation pathways are presented. Pathways for which there is experimental evidence are shown by black arrows. These include conversion of 5mC (5-methylcytosine) to thymidine by AID followed by BER (base excision repair) to replace with cytosine. The TET proteins can oxidize 5mC to 5hmC (5-hyrdoxymethylcytosine), 5fC (5-formylcytosine) and 5caC (5-carboxylcytosine), which can be replaced with unmodified cytosine directly by BER or through AID conversion to 5hmU (5-hyrdoxymethyluracil) followed by BER. Pathways that currently have limited or no evidence are shown in grey. These include direct removal of 5mC, 5fC or 5caC by either a demethylase, a deformylase or a decarboxylase, respectively. Adapted from: Messerschmidt et al., 2014

The second active demethylation process involves conversion of 5hmC to 5-hydroxyuracil (5hmU) by AID and APOBEC family enzymes. 5hmU is then removed by SMUG1 (Single-strand-selective Monofunctional Uracil DNA Glycosylase 1) or the TDG glycosylase and replaced with an unmodified cytosine by BER (Rai et al., 2008) (Figure 1.3). Compared to the first mechanism this pathway is more controversial, as AID enzyme typically acts on single stranded DNA and recombinant AID and APOBEC enzymes have negligible activity on 5hmC (Nabel et al., 2012). On the other hand, TDG can excise 5hmU:G mismatches in vitro, and primordial germ cells derived from AID- deficient mice do show increased methylation at CpG island promoters (Cortazar et al., 2011; Cortellino et al., 2011; Nabel et al., 2012). Thus it is difficult to rule out that these individual pathways function in DNA demethylation but their contribution *in vivo* is still unclear.

### *1.3.3 Enzymatic Decarboxylation of 5caC*

It is possible that 5caC may be decarboxylated to produce an unmodified cytosine. Currently the proteins involved in this pathway are unknown and evidence consists of a single report in which an oligonucleotide containing 5caC was incubated with ES cell lysate and showed a small but detectable direct conversion of 5caC to cytosine without BER (Schiesser et al., 2012).

### *1.3.4 Dehydroxylation by DNMT Enzymes*

In-vitro experiments have indicated that the DNMT enzymes can remove the hydroxymethyl group of 5hmC to produce the unmodified cytosine. The catalytic function of DNMT3A depends on the reaction

conditions, with reducing conditions favoring the typical methyltransferase activity and oxidizing conditions promoting dehydroxymethylation (Chen et al., 2012; Liutkeviciute et al., 2009). It is unknown whether this process also occurs in vivo.

## 1.4 DNA METHYLATION AND TRANSCRIPTIONAL REGULATION

The mammalian genome contains approximately 28 million CpGs, of which 60-80% are methylated. 10% of CpGs occur together in CpG dense regions called CpG islands (Smith and Meissner, 2013). CpG islands are prevalent at transcription start sites and are found in approximately 60% of genes (Jones, 2012). In contrast to isolated CpGs, the CpGs within CpG islands are hypomethylated (Smith and Meissner, 2013). During development bulk genomic methylation patterns are typically static across different tissues with the exception of the genome-wide demethylation events in the fertilized zygote and developing primordial germ cells. Specific localized changes in DNA methylation are observed between different tissues depending on the activation or repression of certain genes. The distribution and function of DNA methylation patterns in the genome is incompletely understood, but I will describe the current views below (Figures 1.3, 1.4).

### *1.4.1 Promoters*

As described above, approximately 60% of human genes contain a CpG island (CGI) in their promoter. These include active, bivalent and silent promoters. Active promoters are associated with expressed genes and their typical features are: hypomethylated CGIs, nucleosome-depletion,

trimethylation of histone H3 at lysine 4 (H3K4me3), and the presence of the histone variant H2A.Z (Kelly et al., 2010) (Figures 1.4, 1.5). The level of gene expression is controlled by the binding of transcription factors and gene expression can be repressed by loss of transcription factor binding, the acquisition of particular histone modifications (H3K27me3) or by DNA methylation (Illingworth and Bird, 2009) (Figure 1.6).

Inactive promoters that have a CGI are held in a stable repressed state by H3K27me3 and H3K9me3 histone marks and often DNA methylation (Figure 1.4). DNA methylation leads to the recruitment of methylated DNA binding proteins and histone deacetylases leading to further compaction (Bartke et al., 2010).

In ESCs approximately 20% of promoters with CGIs are marked by H3K4me3 and H3K27me3 histone marks and low DNA methylation. Although the genes associated with these bivalent promoters are not expressed these promoters are termed poised as they contain both activating (H3K4me3) and repressive (H3K27me3) chromatin marks (Figure 1.4). They are enriched for developmental regulators and upon differentiation to the associated lineage the repressive histone mark is usually removed and the gene is expressed (Voigt et al., 2013).

The CpGs within CGIs are typically hypomethylated regardless of the expression level of the gene (Jones, 2012; Smith and Meissner, 2013). Stably repressed inactive promoters, such as the promoters of germ-cell specific genes, are the exception as they can show DNA methylation at their CGIs (Farthing et al., 2008). Interestingly the methylation of genomic regions surrounding CGIs (termed CGI shores) is found to be far more dynamic than within the CGIs themselves, and is also associated with gene repression

(Irizarry et al., 2009). At the moment it is clear that methylation of CGIs and CGI shores leads to gene repression, but it is unclear how prevalent this mechanism of repression is during normal development.

It is also unclear whether DNA methylation is the initial silencing event. Previous work has shown that *de novo* methylation is inhibited by the H3K4me2 and H3K4me3 marks (Balasubramanian et al., 2012; Ooi et al., 2007), and the H2A.Z variant (Zilberman et al., 2008), all of which are present at the CGIs of active and bivalent promoters. Furthermore DNA methylation requires the presence of nucleosomes, which are typically depleted in active promoters (Jiang and Pugh, 2009). Lastly, both active and bivalent promoters show enrichment of TET1 binding which may oxidize and ultimately remove any spurious 5mC (Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). Thus it seems difficult for DNA methylation to occur at active or bivalent promoters and then precipitate gene silencing. On the other hand, there have been observations that suggest that these molecular protectors of DNA hypomethylation may break down. Bivalent promoters, in particular, can show *de novo* methylation during immortalization (Jones et al., 1990) and oncogenic transformation (Ohm et al., 2007; Schiesser et al., 2012; Widschwendter et al., 2007).

In contrast to promoters with CGIs, approximately 80% of promoters lacking CGIs are DNA methylated in ESCs (Jones, 2012; Smith and Meissner, 2013). These promoters do show substantial fluctuations in the promoter methylation levels among different cell types (Farthing et al., 2008). The consequences of methylation changes on transcription are unclear as there are relatively few CpG dinucleotides present in these promoters and previous genome-wide studies correlating methylation and

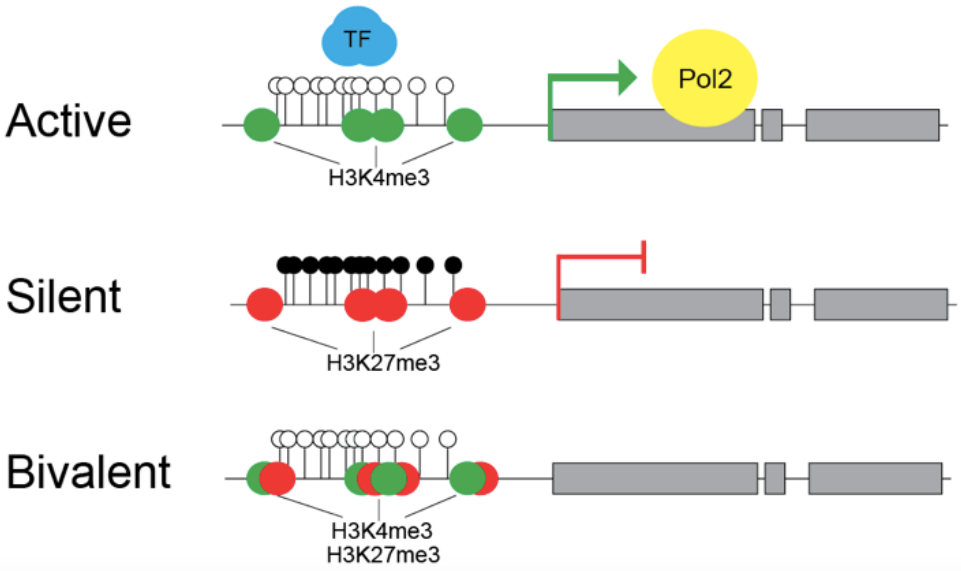



**Figure 1.4 Epigenetic features of different promoter types.** Active, silent and bivalent promoters are three promoter types found in numerous cells. Active promoters are associated with genes that show productive gene transcription. They show low levels of DNA methylation and have the activating histone mark H3K4me3. Silent promoters are associated with genes that have negligible gene transcription. Silent promoters are characterized by high levels of DNA methylation and the repressive histone mark H3K27me3. Bivalent promoters are found in numerous stem cell and progenitor cell populations. These promoters are enriched for development or differentiation associated genes. Bivalent promoters show negligible gene transcription and are characterized by both activating (low DNA methylation and H3K4me3 histone mark) and repressive (H3K27me3 histone mark) epigenetic features.

gene expression have produced contradictory conclusions (Gal-Yam et al., 2008; Weber et al., 2005). The current understanding is that methylation at these promoters will not repress transcription.

### *1.4.2 Gene Bodies*

Gene bodies are generally CpG-poor, extensively methylated and typically this DNA methylation correlates with higher gene expression. Some gene bodies can contain CGIs, which may represent alternative promoters; however, methylation of these CGIs does not impede transcription elongation even if these methylated CGIs are marked by H3K9me3 and bound by methyl-CpG-binding proteins (Hellman and Chess, 2007; Wolf et al., 1984). This emphasizes that the effects of DNA methylation depend on the genomic context. Within mammalian cells transcription initiation not transcription elongation is sensitive to DNA methylation, as methylation of CGIs in promoters is inversely correlated with gene expression whereas methylation within the gene body is generally positively correlated with gene expression (Figure 1.5).

### *1.4.3 Enhancers*

Enhancers are genomic regions that can recruit transcription factors, RNA polymerase II and chromatin modifiers to impact gene expression at distally located promoters. Enhancers are identified by the presence of certain histone marks (H3K4me1, H3K27ac) and co-activators (p300) and the depletion of H3K4me3 (Shlyueva et al., 2014). Enhancers are generally CpG-poor and have variable methylation levels (Stadler et al., 2011). This may be due to dynamic cycles of methylation followed by demethylation or

inefficient maintenance of methylation following DNA replication. Previous studies have found differentially methylated regions (DMRs) within the enhancers of differentiation genes, in which methylation of the enhancer correlated with suppression of gene expression (Schmidl et al., 2009) (Figures 1.5, 1.6). In addition 5hmC and TET proteins can be detected in these regions suggesting they may be sites of dynamic DNA methylation and demethylation (Ficz et al., 2011; Williams et al., 2011). Currently how DNA methylation of CpG poor enhancer regions is regulated and how it affects transcription factor binding and gene expression requires further study.

### *1.4.4 Insulators*

Insulators are genomic regions that block the interaction between distally located enhancers and promoters. These include genomic sequences that are bound by the CTCF protein, such as a site within the *IGF2-H19* locus. Methylation of the insulator within this locus blocks CTCF binding which leads to a disruption in the enhancer-promoter interaction (Bell and Felsenfeld, 2000). Methylation of other CTCF sites, such as within exon 5 of the *CD45* gene, can also affect CTCF binding and affect gene splicing (Shukla et al., 2011). In other cases CTCF binding in CpG poor locations are not altered by DNA methylation levels and rather induce local DNA demethylation upon binding (Stadler et al., 2011). At the moment there are no universal conclusions on the effect of DNA methylation on insulator function, CTCF binding and gene transcription.

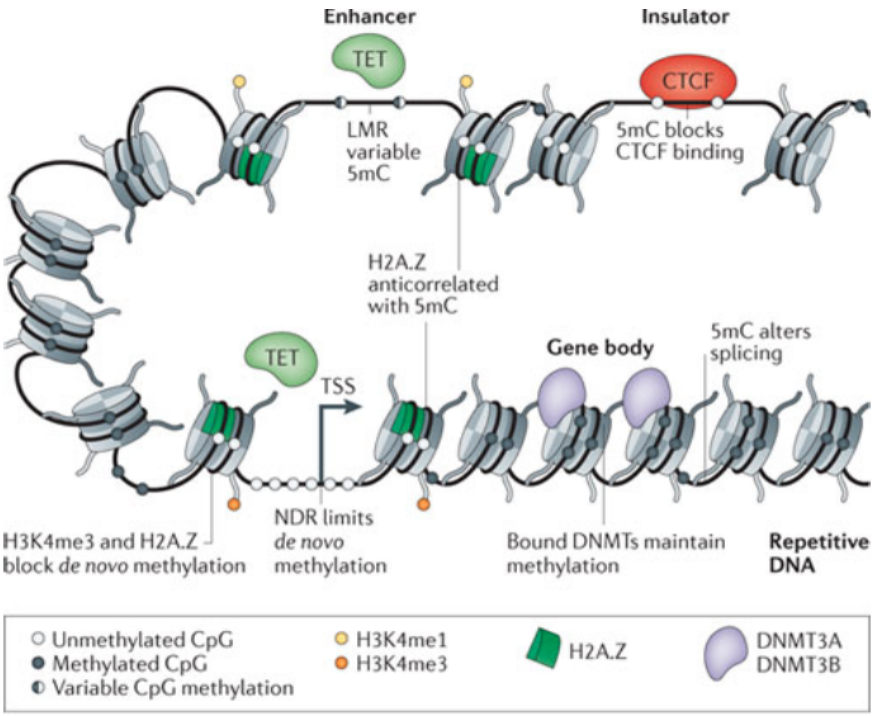



**Figure 1.5 DNA methylation at genomic regulatory regions.** Actively expressed genes with CGIs in their promoters typically have nucleosome deleted regions (NDR), low DNA methylation, H3K4me3 and the H2A.Z histone variant. In the gene body DNA methylation correlates with gene transcription and could function to regulate gene splicing. Enhancers are typically CpG poor with variable and changing DNA methylation possibly due to TET activity. In insulator regions DNA methylation may function to inhibit CTCF binding and thus enable interaction between distal regulatory regions and promoters. Adapted from: Jones et al., 2012.

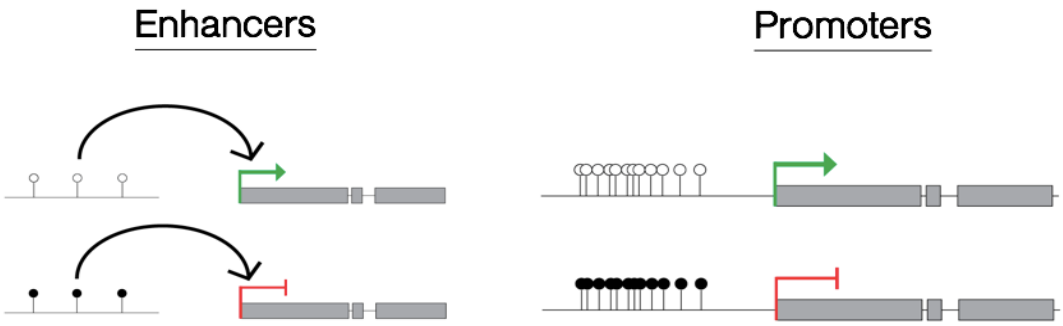



**Figure 1.6 DNA methylation is associated with repression of gene transcription.** Enhancers are distally located from the genes that they regulate. Enhancers typically show low CpG density but the methylation of these CpGs is associated with repression of gene transcription. Promoters are directly upstream of the genes that they regulate. Promoters typically have higher CpG density than enhancers and approximately 60% of promoters have a CGI. DNA methylation of promoters is also associated with repression of gene transcription.

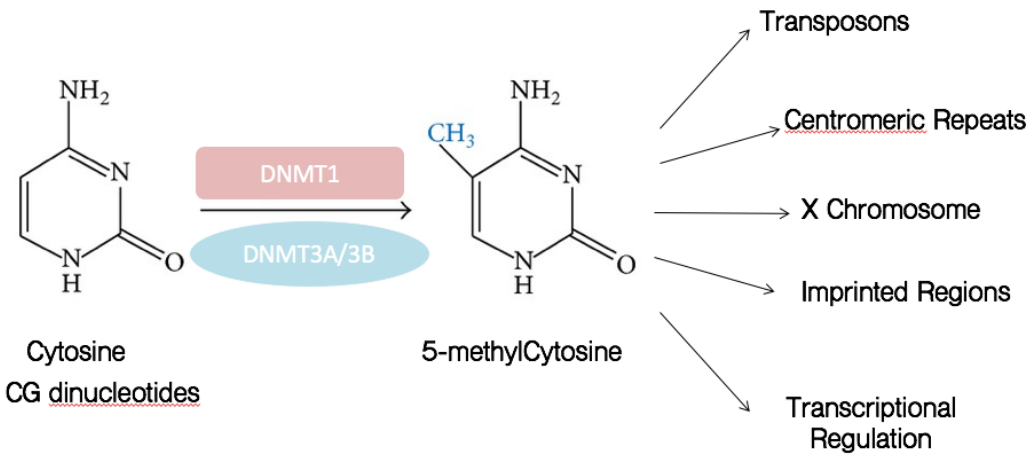



**Figure 1.7 Additional functions of DNA methylation.** In addition to regulating gene transcription DNA methylation is involved in silencing of transposons and centromeric repeats, X chromosome inactivation and the regulation of imprinted regions.

## 1.5 ADDITIONAL FUNCTIONS OF DNA METHYLATION

### *1.5.1 Transposons*

Approximately 40% of the mammalian genome consists of endogenous transposable elements of three major classes: long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and long terminal repeat (LTR)-containing endogenous retroviruses. The promoters for these transposable elements are hypermethylated, which is believed to prevent transcriptional expression and retrotransposition as well as changes in the expression of neighboring genes (Figure 1.8) (Lander et al., 2001; Walsh et al., 1998).

### *1.5.2 Pericentromeric Repeats*

Pericentromeric repeats are required for centromeric assembly during cell division and enable proper chromosome alignment and segregation. The pericentromeric minor and major satellites extend from the centromere as tens of thousands tandem copies. These elements have latent transcriptional activity that is suppressed by DNA methylation (Figure 1.8). DNMT3B in particular is recruited to these regions to produce DNA methylation, and is even found at pericentromeric repeats in metaphase chromosomes, suggesting a continued requirement for silencing during mitosis. Missense mutations in DNMT3B lead to the immunodeficiency, centromeric instability and facial anomaly (ICF) syndrome in which transcription from pericentromeric repeats causes chromosome misalignments during mitosis (Xu et al., 1999).

### *1.5.3 X-Chromosome Inactivation*

X chromosome inactivation in females is necessary for gene dosage control. X chromosome inactivation is triggered by the noncoding RNA *Xist*, which coats the X chromosome, displaces transcription factors and induces chromatin changes. Ultimately DNA methylation occurs as a final step to stabilize silencing (Figure 1.8) (Hellman and Chess, 2007).

### *1.5.4 Imprinted Genes*

Imprinted genes show parent-of-origin specific expression, which is dictated by differential DNA methylation at imprinting control regions (ICRs) of the maternal and paternal alleles. This results in allele specific gene expression in which either the maternal or paternal allele is expressed. Differential methylation at ICRs is introduced during gametogenesis. During the genome-wide demethylation events that occur upon fertilization methylation at ICRs are preserved. DNA methylation at ICRs is only removed during primordial germ cell formation and then re-established in an allele-specific manner in developing gametes. DNMT3A, DNMT3B and DNMT3L are involved in the generation of genomic imprinting marks and dysregulation of these imprinting patterns is associated with numerous developmental defects (Bartolomei, 2009; Ferguson-Smith, 2011).

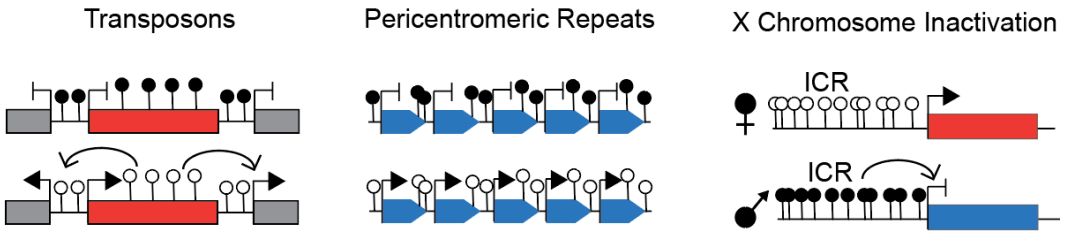



**Figure 1.8 Additional functions of DNA methylation.** DNA methylation functions to repress transcription at transposons (de-repression can also result in the activation of neighboring genes), centromeric repeats, and reinforce silencing of one X chromosome.

## 1.6 5HMC MAPPING IN THE MAMMALIAN GENOME

The generation of 5hmC genome wide maps is in many ways more challenging then whole genome mapping of 5mC. 5hmC is approximately 14-fold less abundant than 5mC and the levels can vary dramatically between different tissue types. The traditional method to investigate 5mC distribution, bisulfite sequencing, cannot distinguish 5mC from 5hmC and thus cannot be used for 5hmC mapping. Several methods for 5hmC profiling have been developed. The first category consists of capture-based technologies such as antibody-based hydroxymethylated DNA immunoprecipitation followed by sequencing (hMeDIP-Seq) and enrichment by hydroxymethyl selective chemical labeling (hMeSeal). Although widely used, the key disadvantage of these techniques is that they produce low-resolution 5hmC maps. In contrast whole genome oxidative bisulfite in combination with conventional bisulfite sequencing (WG Bis/OxBis-seq) and TET-assisted bisulfite sequencing (TAB-seq) can produce base pair resolution maps of 5hmC distribution (Booth et al., 2012; Song et al., 2012; Yu et al., 2012).

In general 5hmC profiling maps from ESCs have found that 5hmC is enriched at promoter CGIs and gene bodies of actively expressed genes. 5hmC is also present at the bivalent promoters of developmental regulators and within CpG poor enhancers (Pastor et al., 2011). TET protein binding does not overlap perfectly with 5hmC peaks but in general is found at the TSS of CpG rich promoters (Deplus et al., 2013; Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). I have emphasized observations taken from ESCs below, but it is important to note that 5hmC distribution can vary in different cell types. Because 5hmC is produced from 5mC this variability

may be due to changes in the distribution of 5mC or due to changes in the abundance of the individual TET proteins.

### *1.6.1 Promoters*

5hmC is enriched at CGIs within promoters in human and mouse ESCs (hESCs and mESCs), mouse neural progenitor cells, mouse neurons and adult mouse cerebellum. The 5hmC at promoters typically shows bimodal peaks, with a 5hmC peak occurring 500-2000 bases before and after the TSS (Szulwach et al., 2011). Notably 5hmC is depleted at the TSS itself. 5fC and 5caC are also found at the same promoters as 5hmC and their levels increase upon TDG depletion indicating that some active demethylation is occurring at these loci (Shen et al., 2013). TET1 binding, which has been most extensively investigated in ESCs, shows enrichment at the TSS of CpG rich promoters. TET1 peaks are located at the TSS of these promoters and are then flanked on either side by the 5hmC peaks (Williams et al., 2011; Wu et al., 2011; Xu et al., 2011). TET2 and TET3 also show enrichment at CpG rich promoters (Chen et al., 2013; Deplus et al., 2013). The association of TET1 and TET3 with CGIs could be due to their CXXC DNA binding domain. Interestingly previous studies have found a repressive role for TET1 in gene transcription in mESCs due to its association with the repressive histone deacetylase complex MBD3-NURD and the repressive complex SIN3A (Wu et al., 2011; Yildirim et al., 2011).

### *1.6.2 Gene Bodies*

In almost all cell types studied, including mESCs and hESCs, 5hmC is enriched within gene bodies (Wu and Zhang, 2011). Presence of 5hmC in

gene bodies shows a positive correlation with gene expression in mESCs but not in hESCs (Mellen et al., 2012). The mechanisms by which 5hmC within the gene body affects RNA elongation are currently unclear.

### *1.6.3 Enhancers*

5hmC is strongly enriched in enhancer regions in mammalian cells (Stroud et al., 2011; Szulwach et al., 2011). Enhancers are typically CpG-poor regions that show high levels of DNA methylation, which could possibly explain the enrichment of 5hmC. Base pair resolution 5hmC profiling has shown that 5hmC is enriched at the borders of transcription factor binding sites in enhancers. Notably 5hmC is depleted at the exact transcription factor binding site, possibly due to exclusion of DNMTs and TET proteins by transcription factor binding (Yu et al., 2012). It is unknown whether 5mC oxidation functions to open up the chromatin at enhancers to make them more accessible to transcription factor binding or if the transcription factors themselves recruit TET proteins to these loci. It is possible that a combination of these two mechanisms are involved. Similar to the results at promoter regions, 5fC and 5caC are enriched within enhancers and their levels increase upon TDG depletion in ESCs. This indicates that active demethylation is occurring at enhancers, but the extent of demethylation seems to be less than seen at promoters (Shen et al., 2013). Upon differentiation of ESCs the 5hmC levels at poised enhancers increases and precedes or coincides with the appearance of the active enhancer mark, H3K27 acetylation (Serandour et al., 2012).

## 1.7 DNA METHYLATION DURING DEVELOPMENT

DNA methylation levels and distribution are dramatically changed during embryonic development. Following fertilization both the maternal and paternal nucleus undergo genome-wide demethylation to erase previous methylation marks, with the exception of methylation at imprinting control regions. As a result, by the time of the early blastocyst stage (32-64 cells) methylation levels in the embryo are at the lowest levels. This is followed by a progressive increase in methylation levels and the establishment of new methylation patterns as the embryo develops into the inner cell mass and then the epiblast. A second wave of genome-wide demethylation occurs in the developing primordial germ cells (PGCs). Demethylation in developing PGCs is much more absolute than in the zygote, as now even imprinting control regions lose methylation. Notably the TET proteins are abundant and may play a role during each of these global DNA demethylation events. The TET proteins are also present in the late blastocyst, inner cell mass and epiblast stages as new DNA methylation configurations are established, suggesting that they may play a role in fine-tuning this process (Messerschmidt et al., 2014).

### *1.7.1 Demethylation in the Early Zygote*

Immediately after fertilization the maternal and paternal DNA undergo global DNA demethylation. The rates of demethylation for the maternal and paternal DNA, and the mechanisms underlying them appear to be different. The paternal DNA undergoes more rapid demethylation, which involves 5mC oxidation by maternally derived TET3 (Figure 1.5). Immunostaining of the paternal nucleus shows the presence of 5hmC and the rapid loss of 5mC (Mayer et al., 2000). Notably this 5hmC is lost upon siRNA knockdown

of *TET3* and in *TET3* deficient oocytes (Gu et al., 2011b; Wossidlo et al., 2011), indicating that TET3 is uniquely necessary among the TET proteins for this oxidation event. It is unclear whether TET3 mediated oxidation leads to active or passive DNA demethylation. Some loss of 5mC is observed prior to any replication event and BER has been observed in the male pronucleus, suggesting that active DNA demethylation in which TDG removes the oxidized product and BER inserts an unmodified cytosine may be occurring (He et al., 2011). It is unlikely, however, that active DNA demethylation with BER accounts for the majority of DNA demethylation as substantial levels of 5hmC, 5fC and 5caC are retained at least through the 8-cell stage (Inoue and Zhang, 2011). Furthermore increased frequency of BER in the early embryo would be considered deleterious to survival as it may increase rates of mutation. More likely the oxidized 5hmC is lost by passive dilution following DNA replication. This is supported by the observation that maintenance methylation is inefficient in the early embryo as DNM1O (a splice variant of DNMT1) is excluded from the nucleus (Ratnam et al., 2002). In this case TET3 mediated oxidation serves to accelerate the removal of 5mC from the paternal DNA. TET3 deficient oocytes fertilized with wild-type sperm produce embryos that have delayed activation of the paternal alleles for genes involved in the establishment of the pluripotent epiblast. Approximately half of these embryos develop normally, but the rest fail to develop past E11.5 (Gu et al., 2011b). Thus it seems that TET3 oxidation of the paternal DNA accelerates demethylation and enhances survival of the embryo, but is not absolutely necessary.

In contrast the maternal DNA does not undergo TET mediated oxidation. The maternal DNA loses methylation through passive means, as

maintenance methylation is inefficient due to the cytoplasmic localization of DNM1O. The maternally derived DPPA protein (also known as STELLA) protects maternal DNA and some imprinted loci in the paternal genome from TET3 mediated oxidation. In the absence of DPPA the maternal genome shows 5mC oxidation and the developing embryos show abnormalities at the 2 or 4-cell stage (Nakamura et al., 2007).

### *1.7.2 Methylation Changes in the Early Embryo*

In the early preimplantation embryo the sequence of methylation changes occur as follows: global demethylation of the paternal pronucleus, global demethylation of the maternal pronucleus and then progressive *de novo* methylation beginning at the blastocyst stage (Figure 1.9). This sequence is conserved in numerous mammalian species including humans. The beginning of *de novo* methylation coincides with the appearance of DNMT1, DNMT3A and DNMT3B within the nucleus. The TET proteins, in particular TET1 and TET2, are also present and are believed to regulate the addition of new methylation marks. The initiation of *de novo* methylation also coincides with the appearance of the first two distinct cell lineages: the inner cell mass (ICM) and the trophectoderm. The outer cell layers of the blastocyst are biased towards forming the trophectoderm, which will ultimately develop into the placenta. The inner cell layers are biased towards forming the inner cell mass, which will later develop into the embryo proper. Although DNA methylation occurs in both cell populations it is interesting to note that greater DNA methylation occurs in the future ICM cells, whereas the cells of the future TE remain relative hypomethylated (Messerschmidt et al., 2014; Pastor et al., 2011). The unequal levels of DNA methylation may

indicate the first epigenetic barrier between two different cell lineages.

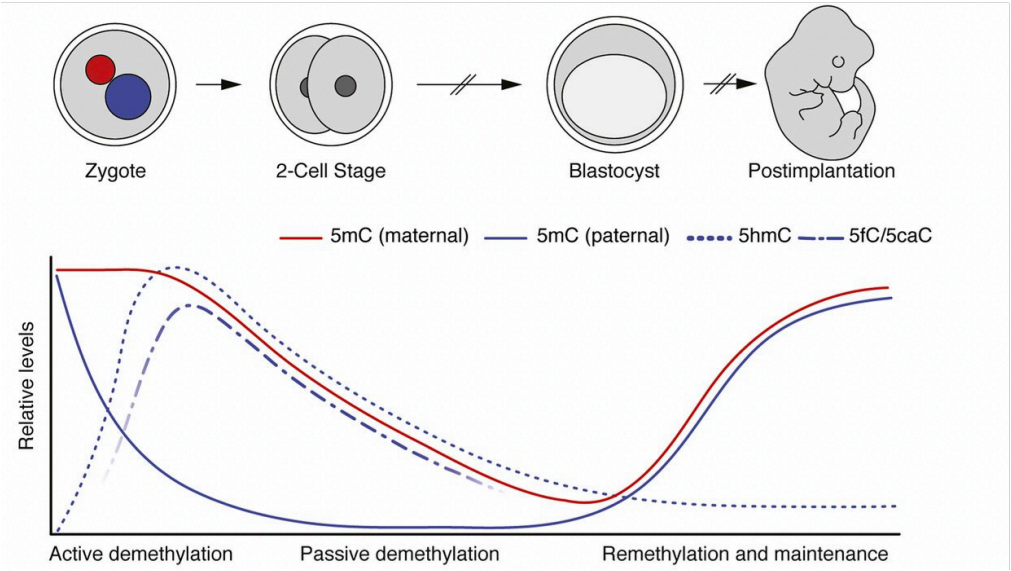



**Figure 1.9 Demethylation in the early embryo.** There are distinct patterns of demethylation of the maternal and paternal genomes following fertilization. The maternal genome undergoes passive demethylation whereas the paternal genome is actively demethylated with the appearance of oxidized derivatives (5hmC, 5fC/5caC). DNA methylation increases during the formation and implantation of the blastocyst. Adapted from: Messerschmidt et al., 2014.

### *1.7.3 Methylation Changes in Primordial Germ Cells*

The second global DNA demethylation event occurs in the developing primordial germ cells (PGCs). In mice PGCs are first identified at E6.5 in the proximal epiblast. Beginning on E7.5, PGCs begin to migrate along the embryonic and extra-embryonic boundary towards the developing gonad (McLaren and Lawson, 2005). Prior to their migration the future PGCs display methylation patterns that are similar to other neighboring somatic cells. They become demethylated between E9.5 (as determined by immunohistochemical analysis) and E10.5-E11.5 (by locus-specific bisulfite sequencing). Genome-wide analysis of PGCs show that the DNA demethylation is almost completely global and includes erasure of imprinted marks. Furthermore germline-specific genes that are hypermethylated in somatic cells become demethylated during PGC specification and are ultimately expressed (Hayashi et al., 2007).

The role of the TET proteins during this process is controversial. *TET1* and *TET2* are expressed in developing PGCs and there is significant conversion of 5mC to 5hmC. There is no detectable 5fC or 5caC indicating that 5hmC is ultimately lost by passive dilution instead of active removal (Hackett et al., 2013). Notably *TET1* knockout mice show full germline competence, which may be due to redundant effects of *TET2* or due to alternative mechanisms for DNA demethylation in PGCs (Dawlaty et al., 2011). During PGC development expression of *UHRF1* is decreased, indicating that maintenance methylation may be inefficient. Furthermore *de novo* methylation is decreased as DNMT3B levels are reduced (Kagiwada et al., 2013). Recent studies have indicated that the TET proteins may not be

necessary for the global DNA demethylation observed but instead play a role in facilitating the activation of germline-associated genes and erasure of imprinting marks during PGC formation, as 5hmC is particularly enriched at the promoters of these genes and at imprinting loci (Cortellino et al., 2011) (Figure 1.10). Deamination by AID may also play a role as AID deficient PGCs show elevated DNA methylation, although they are still dramatically demethylated when compared to their cell of origin (Popp et al., 2010). Currently the importance of inefficient maintenance methylation, decreased *de novo* methylation and hydroxymethylation for demethylation in PGCs is unclear.

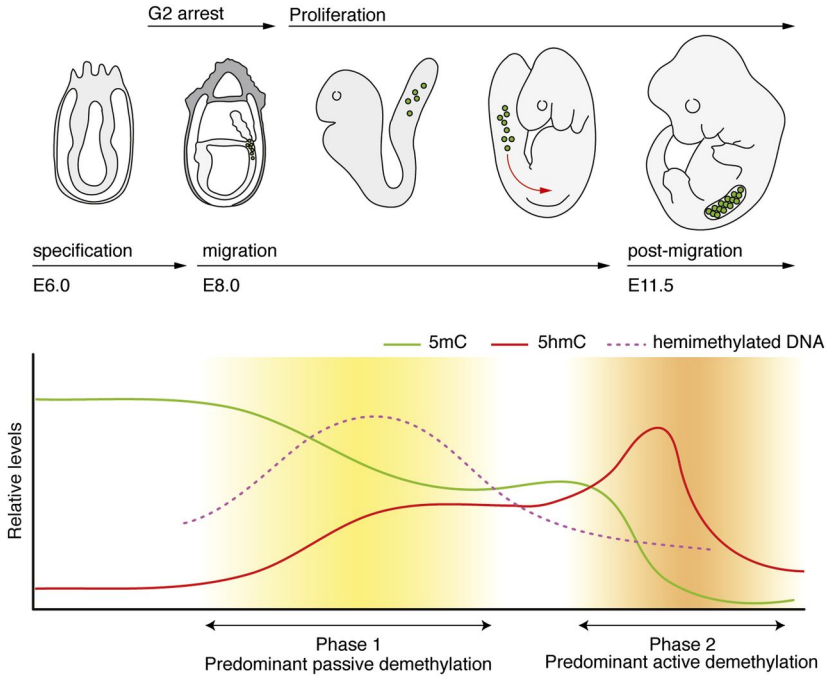



**Figure 1.10 Methylation dynamics in developing primordial germ cells (PGCs).** PGCs emerge at approximately E6.5 with high methylation levels, similar to other somatic cells. In this model there are two stages of demethylation, the first is passive demethylation possibly due to decreased expression of *UHRF1*. This is followed by active demethylation as oxidized derivatives such as 5hmC increase. Adapted from: Messerschmidt et al., 2014.

## 1.8 PHENTOTYPES OF DNMT- AND TET- DEFICIENT MICE

Clearly DNA methylation during embryonic development is quite dynamic and suggests a crucial role for the DNA methylation and demethylation machinery. Previous studies have sought to understand the developmental importance of the DNMT and TET proteins by inactivating these genes in mouse embryos. Deletion of *Dnmt1* in mouse embryos is embryonic lethal with defects arising around gastrulation as a result of global loss of DNA methylation; this is also seen with depletion of its recruiting cofactor *Uhrf1* in mouse embryos (Bostick et al., 2007; Li et al., 1992). *Dnmt3a*-/- mice developed to term and appeared normal at birth, but then died at about 5 weeks after birth. Embryos with deletion of *Dnmt3b*-/- developed normally until E9.5 but then showed multiple developmental defects, including growth retardation, and did not produce any viable mice at birth (Okano et al., 1999). As expected combined deletion of *Dnmt3a* and *Dnmt3b* led to more severe embryonic defects that were similar to those observed in *Dnmt1*-/- embryos. These mice showed smaller size and abnormal morphology by E8.5, shortly after gastrulation, and died before E11.5 (Okano et al., 1999). Inactivation of *Dnmt3l*, the non-catalytic cofactor for DNMT3A and DNMT3B proteins, in mouse embryos produced viable mice; however, due to defects in *de novo* methylation of the germline, male *Dnmt3l*-/- mice were sterile and female *Dnm3tl*-/- mice produced embryos that died early in development (Bourc'his and Bestor, 2004; Bourc'his et al., 2001).

*Tet1*, *Tet2* and *Tet3* have been individually and in combination inactivated in mouse embryos. The phenotype of *Tet1*-/- mice is controversial. The earliest works of *Tet1*-/- and *Tet1*-/-*Tet2*-/- embryos

showed that viable and fertile mice could be obtained (Dawlaty et al., 2013; Dawlaty et al., 2011). However a more recent study found that targeting the N-terminus of TET1 led to promiscuous expression of differentiation genes within the epiblast, morphological embryonic defects as early as E8.0 and ultimately embryonic lethality by E9.5 in 129P2/OlaHsd mice. Intriguingly this study found that the severity of embryonic defects was dependent on the strain of mice used, with reduced penetrance in inbred C57BL/6 (B6) embryos and complete lethality in non-inbred mice, 129P2/OlaHsd (Khoueiry et al., 2017). The reasons underlying the differences in phenotype based on mice strain are currently unknown. *Tet1* has also been shown to be necessary for proper gamete formation. Female *Tet1* gene trap mutants have smaller litters. This is due to arrest of developing female gametes (E16.6-E18.5) because of defects in meiotic synapsis. As a result these mice have only half the normal number of gametes, which could explain their reduced fertility (Yamaguchi et al., 2012). Inactivation of the paternal *Tet1* led to aberrant hypermethylation of paternally expressed imprinted genes during gametogenesis.

*Tet2* inactivation in C57BL/6 and non-inbred strains produce viable and fertile mice. These offspring show clear hematological abnormalities such as increased number of hematopoietic stem cells with increased proliferation rates (Quivoron et al., 2011). One strain of these *Tet2*−/− mice developed a condition similar to human chronic myelomonocytic leukemia (Li et al., 2011; Moran-Crusio et al., 2011), which is interesting considering the frequency of *TET2* mutations in human hematological malignancies. Inactivation of *Tet1* and *Tet2* in mice of the C57BL/6 × 129 strain can produce viable, fertile offspring but approximately half of the litter die

perinatally for unknown reasons (Dawlaty et al., 2013). Approximately half of embryos derived from *Tet3*-/- oocytes show early embryonic lethality. This could be due to hypermethylation of pluripotency genes in the paternal allele, which utilizes TET3 to promote DNA demethylation after fertilization. Interestingly the *Tet3*-/- mice that survive embryonic development die perinatally for unknown reasons (Gu et al., 2011b).

In order to remove any possible compensation among the TET proteins investigators have generated *Tet* TKO embryos in which all three *Tet* genes are inactivated. A recent study found that *Tet* TKO embryos displayed dramatic gastrulation defects. They found that excessive Nodal signaling produced an aberrant increase in mesendoderm formation from the epiblast. The promoter or enhancer of the signaling antagonists for *Nodal*, *Lefty1* and *Lefty2* were hypermethylated, leading to their reduced expression (Dai et al., 2016).

## 1.9 PHENTOTYPES OF DNMT- AND TET- DEFICIENT MESCS

ESCs are used as a platform to study early development in a cell culture system. ESC lines have been generated from cells of the inner cell mass in mouse and human embryos. Despite having originated from similar stages in development, mESCs and hESCs show differences in growth, genetic requirements and epigenetic features. mESCs are characterized as displaying a “naïve” pluripotency that is similar to the inner cell mass of the preimplantation blastocyst. In contrast hESCs reflect a “primed” pluripotency stage that is similar to the post-implantation blastocyst. Regardless of these differences both mESCs and hESCs display the properties of indefinite self-renewal and the ability to differentiate into all somatic cell types (Ginis et al.,

2004; Schnerch et al., 2010). Thus they are an ideal platform to study the role of the TET proteins in the cell fate conversion that occurs during the formation of somatic tissues.

The *DNMT* genes have been inactivated in mESCs and hESCs. Inactivation of *DNMT3A* and *DNMT3B* individually or in combination has no effect on self-renewal or pluripotency marker expression in mESCs and hESCs. *Dnmt3a*-/-*Dnmt3b*-/- mESCs progressively lose DNA methylation. Interestingly early passage *Dnmt3*a-/-*Dnmt3b*-/- mESCs, which still contain significant levels of DNA methylation are able to differentiate into the germ layer and form teratomas. This ability is lost in later passages, upon greater loss of DNA methylation, indicating that although DNA methylation is not required for ESC proliferation, a certain level is required for proper differentiation of mESCs (Okano et al., 1999). Inactivation of *DNMT1* has different phenotypes in mESCs versus hESCs. *Dnmt1*-/- mESCs are significantly hypomethylated but are still viable (Lei et al., 1996). In fact *Dnmt1*-/-*Dnmt3a*-/-*Dnmt3b*-/- mESCs have been generated that display normal stem cell proliferation and euploidy (Tsumura et al., 2006). In contrast knockout of *DNMT1* in hESCs is lethal and produces global DNA demethylation, DNA damage and cell cycle arrest (Liao et al., 2015). It is possible that because hESCs show properties of a later developmental stage they are more sensitive to loss of DNMT1. As all somatic cells require DNMT1 for survival hESCs may represent the first developmental period at which DNA methylation maintenance becomes essential.

The three *Tet* genes have been inactivated individually and in different combinations in mESCs. Thus far none of the *TET* genes has been inactivated in hESCs. Notably inactivation of the individual *Tet* genes, or all

3 *Tet* genes together, has no impact on the self-renewal capacity of mESCs. However inactivation of the *Tet* genes does produce differentiation defects that become more severe as a greater number of *Tet* genes are mutated (Dawlaty et al., 2014; Dawlaty et al., 2013; Dawlaty et al., 2011; Lu et al., 2014). Depletion of *Tet1* singly or together with *Tet2* showed mild differentiation defects. For example teratomas from these cells contain tissues from all three germ layers but are also enriched for extraembryonic trophoblastic cells. This suggests that in wildtype cells TET1 may function to suppress extra embryonic differentiation. *Tet1* depletion leads to increased expression of key extra embryonic markers such as caudal-type homeobox 2 (*CDX2*), eomesodermin (*EOMES*) and E74 Like ETS Transcription Factor 5 (ELF5). At the same time *Tet1*-/- and *Tet1*-/-*Tet2*-/- mESCs contribute efficiently to chimeras when injected into blastocysts showing that the increased extraembryonic potential can be suppressed in the normal developmental environment. Furthermore upon *in-vitro* differentiation, *Tet1*-/- and *Tet1*-/-*Tet2*-/- mESCs can form all three germ layers (Dawlaty et al., 2013; Dawlaty et al., 2011). It is possible that in these DKO mESCs, *Tet3* can compensate for the loss of *Tet1* and *Tet2*. In fact TKO mESCs, in which all three *Tet* genes have been inactivated, show strong impairment in their ability to differentiate by spontaneous embryoid body differentiation and teratoma formation assays. Furthermore TKO mESCs show poor chimeric contribution to embryos and cannot support embryonic development by tetraploid complementation assays (Dawlaty et al., 2014). The overall conclusion from these studies is that the TET proteins may not be necessary to support ESC proliferation and self-renewal but are required for the proper activation of differentiation programs.

The mechanisms underlying the observed differentiation defect of TKO mESCs are currently unclear. Measurement of total 5mC show a modest increase in total methylation upon *Tet* inactivation in TKO mESCs (Dai et al., 2016; Dawlaty et al., 2014; Lu et al., 2014). Although the overall increase in methylation is small it is possible that concentration of this hypermethylation in regulatory regions could explain the dramatic differentiation defects of TKO mESCs. Genome-wide methylation analysis of TKO mESCs has thus been performed to uncover specific loci that gain methylation upon *Tet* inactivation. Interestingly a number of regulatory regions gained hypermethylation in TKO mESCs including enhancers and promoters. Hypermethylation of enhancers leads to a decrease in expression of associated genes (Lu et al., 2014). A more recent study found that TET1 and 5hmC localize to similar developmental enhancers in xenopus, zebrafish and mouse embryos during early embryonic development. This 5hmC signal is associated with increased chromatin accessibility and activation of enhancers as determined by H3K27ac and p300 enrichment. Knockdown of *tet1* in zebrafish was associated with increased methylation at these particular enhancers and reduced chromatin accessibility (Bogdanovic, 2017).

## 1.10 ROLE OF TET PROTEINS IN REPROGRAMMING

Somatic cells can be converted into cells with similar characteristics as ESCs, termed induced pluripotent stem cells (iPSCs), by the transfection of certain combinations of transcription factors. The classic combination includes OCT4, SOX2, KLF4 and MYC (OSKM) (Takahashi et al., 2007). For successful reprogramming the promoters and enhancers of pluripotency

genes, which are typically hypermethylated in somatic cells, need to be demethylated to enable activation of expression (Lee et al., 2014). This is supported by previous investigations in which 5-azadeoxycytidine, a compound that inhibits DNMT activity, promotes reprogramming and reduces the number of transcription factors required (Huangfu et al., 2008; Mikkelsen et al., 2008). Supporting the role of demethylation, previous work has shown that TET proteins could potentially be involved with reprogramming. One such study found that the promoter of the pluripotency gene *Nanog* is hydroxymethylated and demethylated early during reprogramming and that depletion of *Tet2* impedes reprogramming into iPSCs (Doege et al., 2012). Another investigation found that TET1 and TET2 interact with Nanog in ESCs, and co-transfection of *Nanog* with either *Tet1* or *Tet2* increased the efficiency of reprogramming using the original OKSM cocktail, possibly by demethylating the endogenous *Oct4* and *Esrrb* genes (Costa et al., 2013). A final study found that the addition of *Tet1* alone could increase the efficiency of reprogramming using the OSKM combination (Gao et al., 2013). These previous studies indicate that demethylation of pluripotency associated loci increases the efficiency of reprogramming of somatic cells to iPSCs. Notably the most recent study suggested a different mechanism by which the TET proteins regulate reprogramming. The investigators found that *Tet* deficient MEFs could not be reprogrammed into iPSCs due to impairment of the mesenchymal-to-epithelial (MET) transition. This impairment was partly due to reduced demethylation and reduced expression of miRNAs that promote MET (Hu et al., 2014).

## 1.11 THESIS AIMS

Previous studies have inactivated the three *Tet* genes in mESCs. *Tet* mutant mESCs showed defects in differentiation ability, which increased in severity as a greater number of *Tet* genes were inactivated. TKO mESCs also showed hypermethylation at regulatory regions such as promoters and enhancers. It is clear from these studies that the *Tet* genes are required for proper differentiation of mESCs; however, the molecular mechanisms underlying the differentiation defects are unclear.

**Aim 1: Inactivate all three *TET* genes in hESCs.**

None of the *TET* genes have previously been mutated in hESCs. In contrast to mESCs, hESCs represent a later stage in development, the post-implantation epiblast. This is believed to underlie differences in growth requirements, proliferation rate and pluripotency marker expression between mESCs and hESCs. Furthermore mESCs and hESCs show different genetic requirements, with some genes (e.g. *DNMT1*) that are readily inactivated in mESCs proving to be lethal in hESCs. Our first aim was to inactivate all three *TET* genes in hESCs (TKO hESCs) and evaluate the self-renewal and pluripotency characteristics of these cells.

**Aim 2: Investigate methylation changes between WT and TKO hESCs using genome-wide approaches.**

Our second aim was to analyze methylation changes between WT and TKO hESCs on a genome-wide level. We decided to complete this aim using two approaches: whole genome bisulfite sequencing and enhanced reduced representation bisulfite sequencing. Our goal was to determine

whether we could identify robust and reproducible methylation changes between WT and TKO hESCs that could possibly explain the differentiation defect of TKO hESCs.

**Aim 3: Investigate whether a differentiation defect of TKO hESCs into a particular lineage can be attributed to a specific methylation change between TKO and WT hESCs.**

In this aim we were interested in 3 questions: whether a locus-specific aberrant hypermethylation in TKO hESCs affected expression of the associated gene upon differentiation, if restoring expression of this gene could rescue differentiation into the associated lineage and if targeted demethylation of the hypermethylated locus in TKO hESCs would restore proper expression of the associated gene after differentiation.

**Aim4: Determine the mechanisms underlying the aberrant gain of methylation in TKO hESCs.**

This aim involved determining which of the DNMT proteins were responsible for hypermethylation in TKO hESCs and whether variability in the gain of methylation between loci was reflected by variability in the recruitment or activity of the DNMT proteins.

# CHAPTER 2: TET Protein Requirement and Function of Oxidized Methylcytosines in hESCs

## 2.1 INTRODUCTION

DNA methylation during early embryonic development is a very dynamic process and involves the erasure of established DNA methylation marks and the generation of new patterns. Although the most dramatic examples are the global demethylation that occurs after fertilization and during the development of primordial germ cells, DNA demethylation also occurs during the formation and implantation of the blastocyst. DNA demethylation during this period is believed to fine-tune the establishment of new DNA methylation patterns. These patterns will allow the proper differentiation into the embryonic germ layers and then ultimately into all of the somatic cells of the adult organism (Jones, 2012; Pastor et al., 2013; Smith and Meissner, 2013).

During the formation and implantation of the blastocyst both the *de novo* DNMTs and the TET proteins are abundant and there are high levels of 5mC and its oxidized derivatives. Mouse embryonic stem cells (mESCs), which are derived from the pre-implantation blastocysts, are used to study this developmental stage *in vitro*. The *Tet* genes have been inactivated previously in mESCs (TKO mESCs). Generally these studies found that loss of the TET proteins led to differentiation defects that became more severe as more TET proteins were depleted (Dawlaty et al., 2014; Dawlaty et al., 2013; Dawlaty et al., 2011). Moreover inactivation of *Tet* genes in mouse embryos led to defects in gastrulation, which is when the three germ layers

are specified (Dai et al., 2016). Whole genome methylation analysis of TKO mESCs showed that these cells had locus-specific hypermethylation that was enriched at regulatory regions such as promoters and enhancers (Lu et al., 2014). Another recent study found that enhancers that typically undergo DNA demethylation during development also show TET1 binding and 5hmC enrichment. Morpholino depletion of TET1 in zebrafish led to hypermethylation of such enhancers, reduced chromatin accessibility and reduced gene expression (Bogdanovic, 2017).

We were interested in whether changes in the methylation state of promoters could also explain the differentiation defects of hESCs after *TET* inactivation. Approximately 60% of human genes contain a CpG island within their promoter. These promoters are typically DNA hypomethylated and DNA methylation correlates with gene repression. Active promoters are protected from aberrant DNA methylation by the chromatin changes and protein binding that accompanies transcription (Jones, 2012; Smith and Meissner, 2013). However, it is unclear how the DNA hypomethylation of non-expressed promoters, such as bivalent promoters, is preserved. Bivalent promoters are particularly significant because they are enriched for lineage regulators and signaling components that are critical for proper development.

Here we have inactivated the three *TET* genes in hESCs, individually and in various combinations. To our knowledge this is the first study that has genetically knocked out any of the *TET* genes in hESCs. Fortunately we were able to generate and propagate *TET* TKO hESCs in which all three *TET* genes have been inactivated. Although there was complete depletion of 5hmC in these cells we did not observe any differences in total 5mC. Upon

whole genome methylation analysis, however, we did observe locus specific hypermethylation in TKO hESCs. Furthermore bivalent promoters showed the greatest gain in methylation after *TET* inactivation. We found that this bivalent promoter hypermethylation contributes to the differentiation defect of TKO hESCs and is the result of aberrant hypermethylation by the *de novo* methyltransferases. Our study suggests that the DNA methylation state of these promoters is critically important for the proper activation of bivalent genes upon differentiation. Furthermore the TET proteins and *de novo* DNMTs compete to regulate the methylation level at these promoters.

## 2.2 RESULTS

### *2.2.1 Bivalent Promoter Hypermethylation in TKO hESCs*

Since all three *TET* genes are expressed in hESCs (Figure 2.1A), and none has been genetically deleted previously, we used the iCRISPR platform developed in our lab (Gonzalez et al., 2014) to generate a panel of *TET1*/*2*/*3* knockout lines in the HUES8 hESC background. The CRISPR gRNAs we used targeted the beginning of the catalytic domain for each of the 3 *TET* genes and were efficient at producing indel mutations (Figure 2.1B-D). hESCs in which all 3 *TET* genes have been inactivated (TKO hESCs) were karyotypically normal (Figure 2.1E) and had no detectable 5hmC signal by immunofluorescence (Figure 2.1F) and dot blot (Figure 2.1G). TKO hESCs showed no difference in morphology, self-renewal capacity or pluripotency marker expression when compared to wild-type (WT) hESCs (Figure 2.2A-C). However, TKO hESCs showed a complete inability to form teratomas and diminished expression of differentiation

genes upon spontaneous embryoid body differentiation, suggesting that TET proteins may be particularly important for the regulation of cellular differentiation (Figure 2.2D-E).

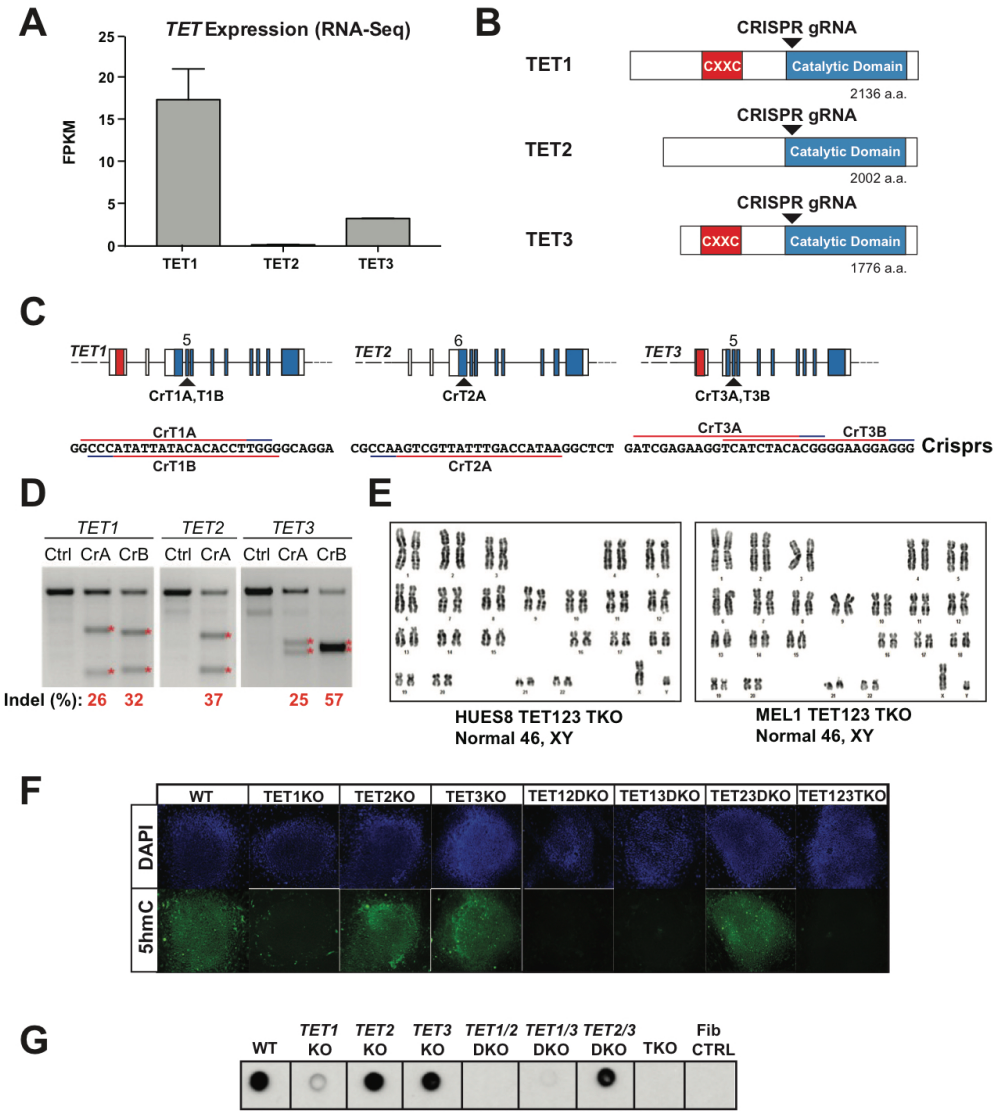



**Figure 2.1 Generation of *TET* TKO hESCs.** (A) *TET* expression in WT hESCs by RNA-Seq. *n*= 2 independent experiments. (B) *TET* knockout mutants were generated using CRISPR gRNAs (arrowheads) that target the beginning of the catalytic domain of TET1, TET2 and TET3. (C) Targeting sequence for *TET* gRNAs. All *TET* gRNAs targeted the beginning of the TET catalytic domain. Red indicates the putative DNA binding domain and blue indicates the catalytic domain. The HUES8 TKO line was generated using CrT1A, CrT2A, CrT3A. The MEL1 TKO line was generated using CrT1B, CrT2A, CrT3B. (D) T7 Endonuclease showing indel formation by the *TET* gRNAs. (E) Karyotyping results for HUES8 and MEL1 TKO lines. (F) Immunofluorescence for 5hmC in HUES8 *TET* mutants. (G) Analysis of 5hmC levels in WT and knockout hESCs by 5hmC dot blot. Human fibroblasts (Fib CTRL) are used as a negative control.

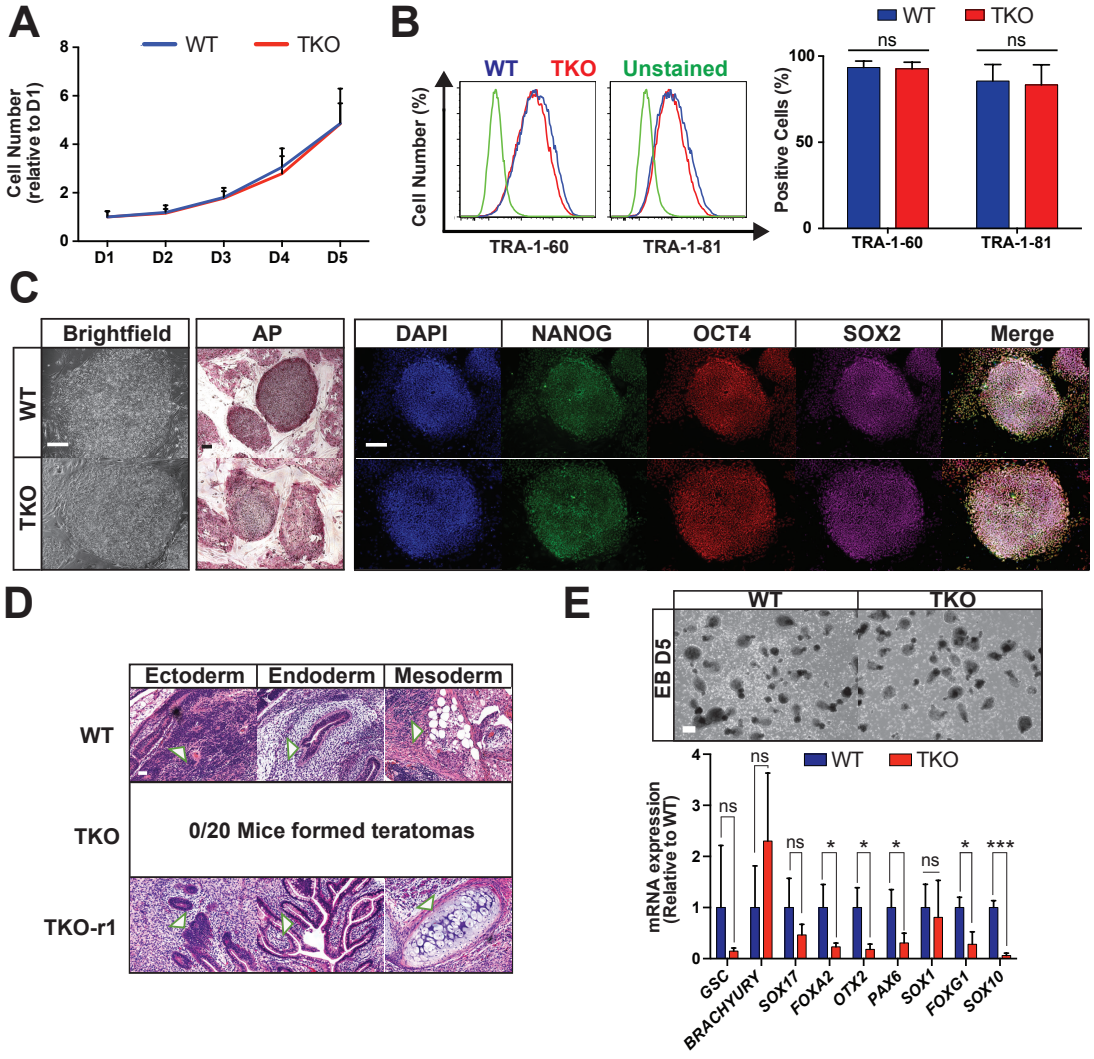



**Figure 2.2 Characteristics of *TET* TKO hESCs.** (A) Growth curves for WT and TKO hESCs. *n*= 3 independent experiments. (B) FACS analysis for pluripotency associated surface markers, TRA-1-60 and TRA-1-81, in WT and TKO hESCs; fluorescence intensity (left panel) and percent positive cells (right panel) are shown. *n*= 3 independent experiments. (C) Brightfield and alkaline phosphatase (AP) staining in WT and TKO hESCs. The scale bar indicates 100 μm. (D) Hematoxylin and eosin staining of teratoma sections from WT teratomas. Arrows point to representative tissues for the respective germ layers. No teratomas were obtained 6 months after injection of TKO hESCs. (E) Phase-contrast images of embryoid body formation by WT and TKO hESCs at D5 (top panel). Expression for markers of mesoderm (*GSC*, *BRACHYURY*), endoderm (*SOX17*, *FOXA2*), neuroectoderm (*OTX2*, *PAX6*, *SOX1*, *FOXG1*) and neural crest (*SOX10*) at D12 of spontaneous embryoid body differentiation. Data are mean ± STD. ns, not significant by Student's t test (B, E).

Loss of the *TET* genes results in locus-specific hypermethylation rather than a global gain of methylation. Mass spectrometry analysis (performed by Louis Dore) showed a negligible amount of 5hmC in TKO hESCs but did not show a difference in 5mC levels between TKO and WT hESCs (Figure 2.3A) similar to previous reports in mESCs (Dawlaty et al., 2014; Lu et al., 2014). Instead, we identified 3,523 hypermethylated differentially methylated regions (regions with at least 5 hypermethylated CpGs, that have at least 10% methylation difference between samples, hyper-DMRs) by whole genome bisulfite sequencing (WGBS) when comparing TKO to WT hESCs (computational analysis performed by Heng Pan). We also observed a similar number (3,832) of hypomethylated differentially methylated regions (hypo-DMRs) (Figure 2.3B). These hypo-DMRs primarily occurred at CpGs outside of CpG islands (CGI) (Figure 2.3D). They could be a direct result of *TET* inactivation or a secondary effect, possibly due to the redirection of the DNMT proteins to novel sites in TKO hESCs. On the other hand, CGIs are enriched in regulatory regions, and they showed increased methylation in TKO hESCs (Figure 2.3D). Additionally hyper-DMRs were enriched for regulatory regions such as promoters and enhancers (Figure 2.2C), with bivalent promoters showing the greatest magnitude of methylation increase (Figure 2.3D) and a greater fraction of bivalent promoters (compared to other regulatory regions) were associated with hyper-DMRs (Figure 2.3E). This hypermethylation was also observed at characteristic bivalent genes such as *HOXA7* and *HOXA9* (Pan et al., 2007)(Figure 2.4A). Indeed significant gain of methylation in TKO hESCs was found to center around genomic sites with both H3K4me3 and

H3K27me3 marks (Figure 2.3F). Finally, we performed genome wide 5hmC profiling through 5hmC-Seal (Song et al., 2011) (performed by Louis Dore and computational analysis by Heng Pan) and found that DNA hypermethylation in TKO hESCs was detected most strongly in regions that had 5hmC peaks in WT hESCs (Figure 2.3F), supporting our hypothesis that loss of TET is directly responsible for the gain of methylation.

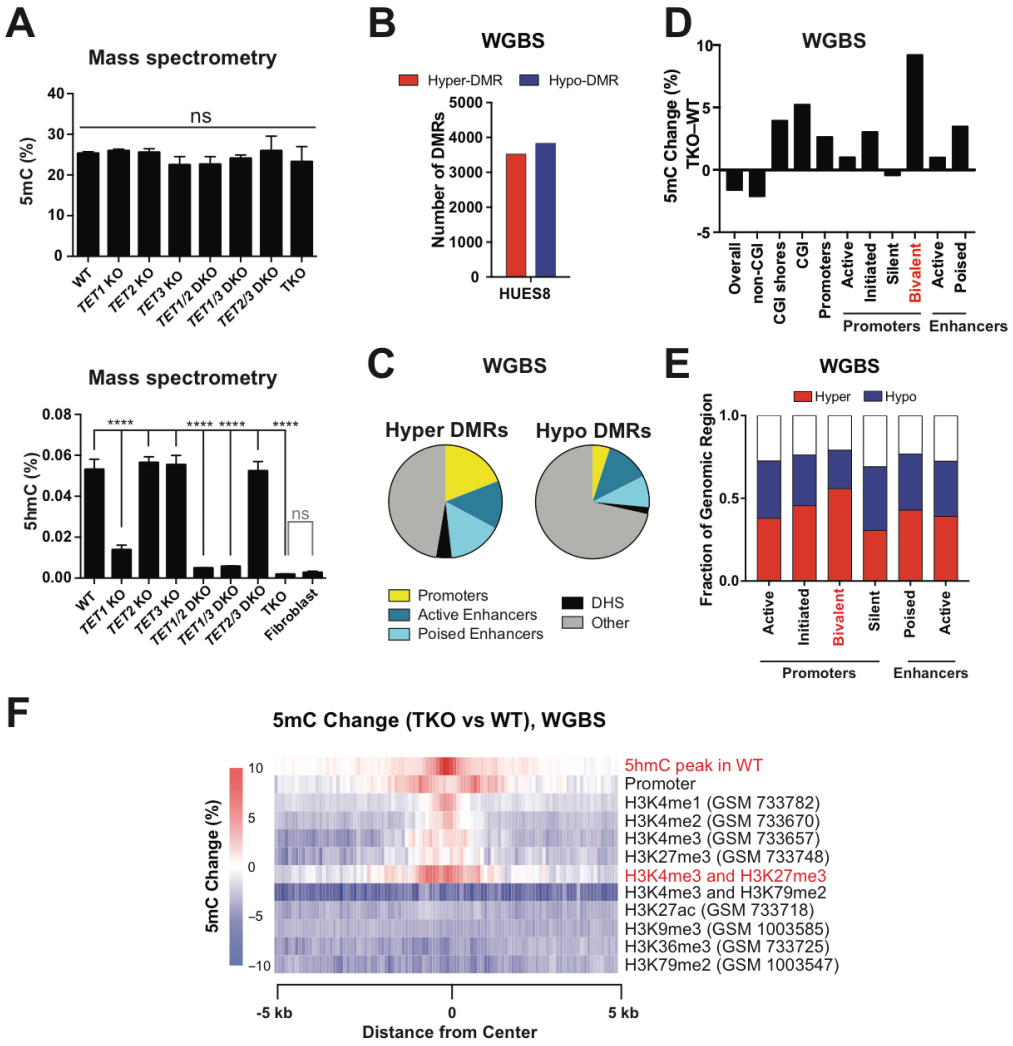



**Figure 2.3 Bivalent promoters gain methylation in TKO hESCs** (A) Analysis of 5mC (top) and 5hmC (bottom) in HUES8 WT and *TET* knockout hESCs by mass spectrometry. For all mass spectrometry analysis, 2 mutant lines were used for all KO genotypes except for TKO. For TKO lines, 2 different passages of the same line were used for mass spectrometry measurements. For significance tests black lines indicate comparisons to WT. Human fibroblasts were used as a negative control for mass spectrometry of 5hmC. (B) Total number of hyper-DMRs and hypo-DMRs (TKO vs. WT) for HUES8 TKO hESCs compared to HUES8 WT hESCs by whole genome bisulfite sequencing. (C) Enrichment of various regulatory regions in hyper-DMRs and hypo-DMRs by WGBS. (DHS) DNase I Hypersensitive sites. The definitions for the different genomic regions can be found in the Methods section. (D) Average percent DNA methylation change between HUES8 TKO and WT hESCs by whole genome bisulfite sequencing (WGBS) at different regulatory regions. The definitions for the different genomic regions can be found in the Methods section. (E) Fraction of genomic regions that show >5% increase in 5mC (Hyper) or a >5% decrease in 5mC (Hypo) in methylation between TKO and WT hESCs by WGBS of HUES8 WT and TKO hESCs. (F) Heat map of the average 5mC level differences between HUES8 TKO and WT hESCs at the center of the annotated histone modifications.

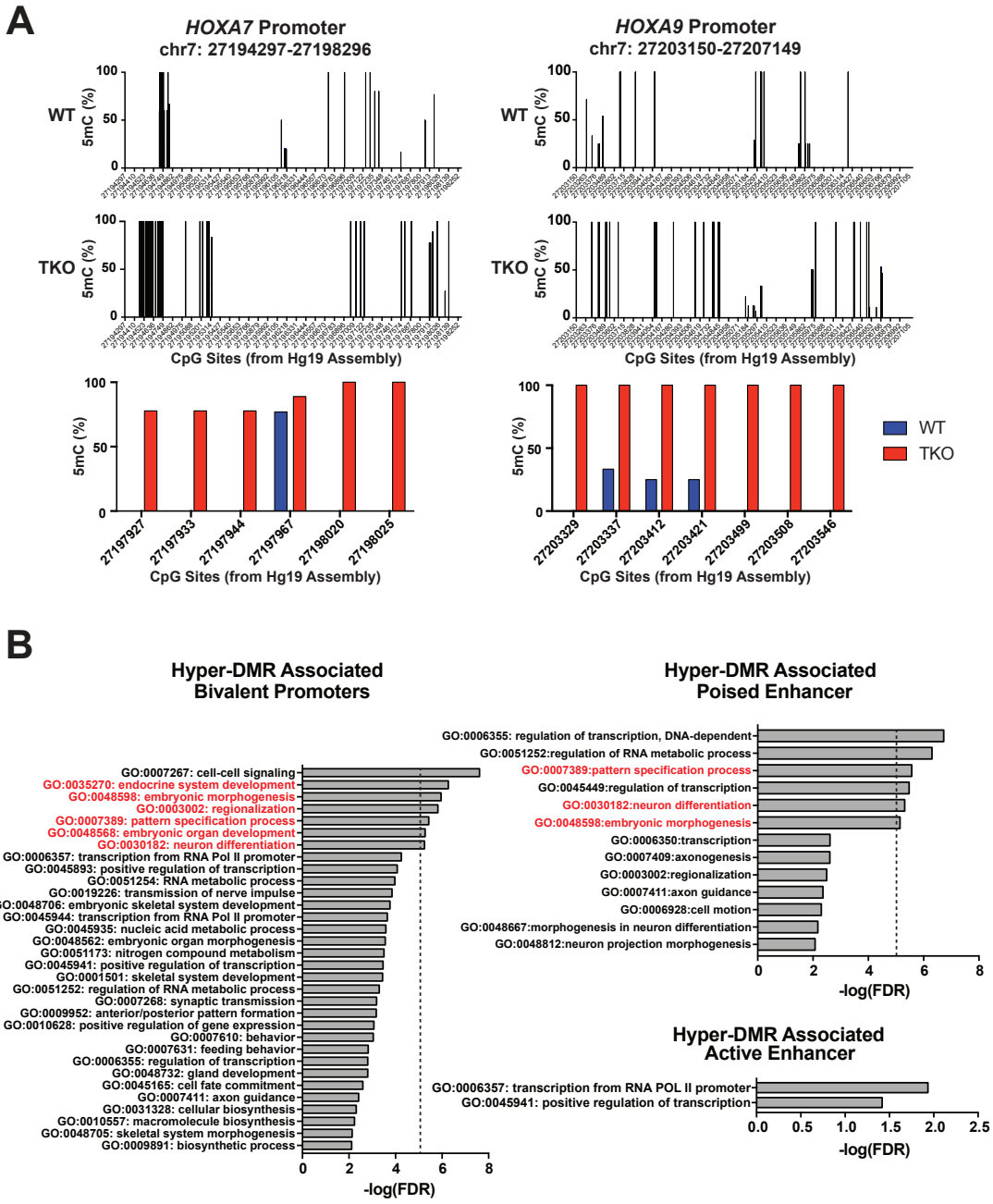


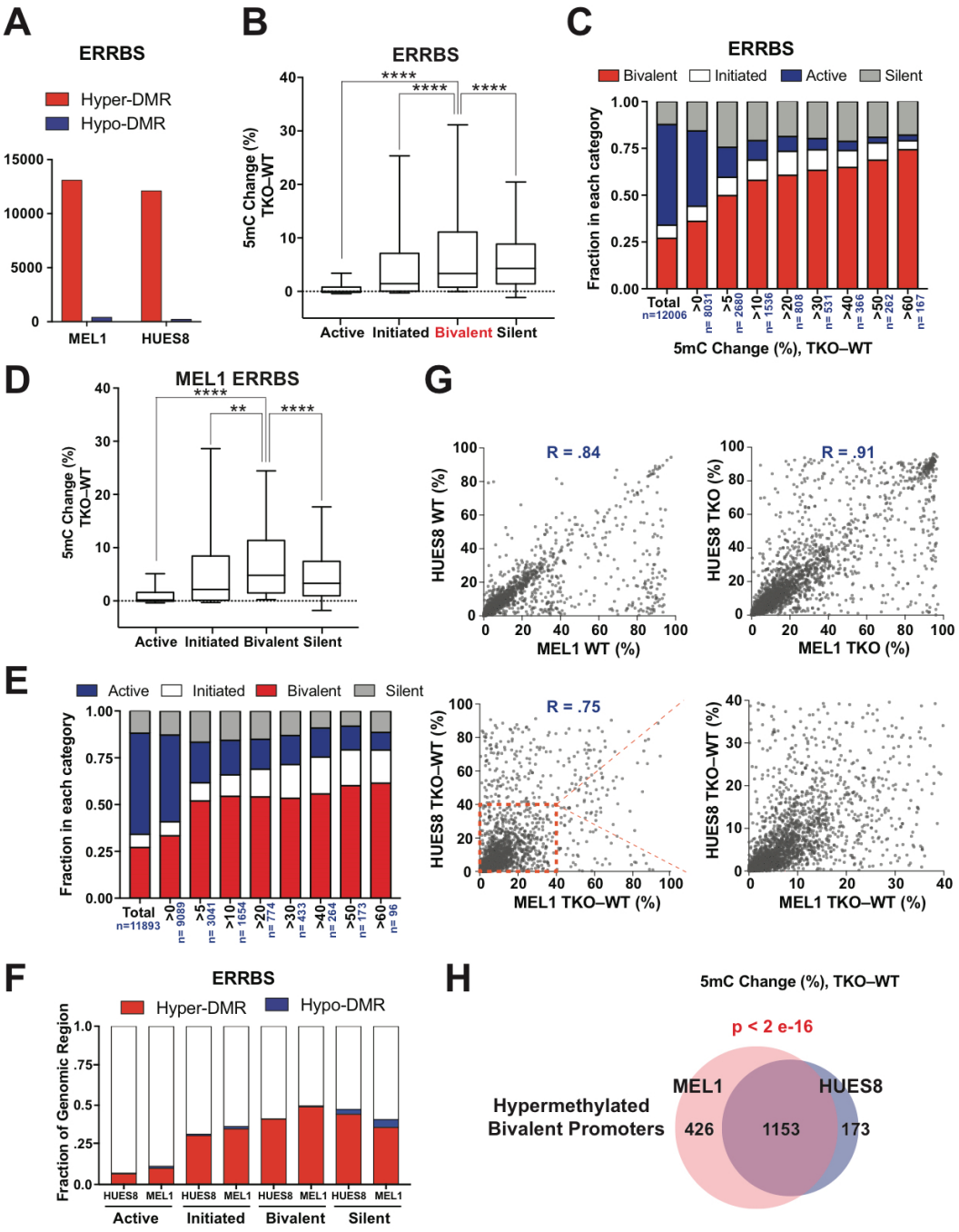

**Figure 2.4. Regulatory regions for developmental genes become hypermethylated in TKO hESCs.** (A) Methylation level of *HOXA7* and *HOXA9* promoters in WT and TKO hESCs by WGBS. Methylation at a few specific CpGs (for which measurements in both WT and TKO hESCs are available) is shown in the bottom bar graph. The x-axis numbers indicate the position of CpG sites in the hg19 genome assembly. (B) Gene ontology analysis for genes associated with hypermethylated bivalent promoters (n=1693), genes associated with hypermethylated poised enhancers (n= 3570) and genes associated with hypermethylated active enhancers (n= 3805). Developmental categories are in red, FDR < .01 was set as a cutoff for bivalent promoters and poised enhancers, FDR < .05 was set as a cutoff for active enhancers.

Gene Ontology (GO) analysis revealed that hypermethylated bivalent promoters in particular, but also hypermethylated poised enhancers (marked by H3K4me1 but not H3K27ac), were associated with developmental categories, suggesting that these methylation aberrations could be responsible for the differentiation defects of TKO hESCs (Figure 2.4B).

To further investigate the methylation changes at bivalent promoters we performed enhanced reduced representation bisulfite sequencing (ERRBS), which focuses on CGIs and thus increases sequencing coverage of promoter regions (Gu et al., 2011a) (computational analysis was performed by Heng Pan). We also generated another TKO clone in a different hESC background, MEL1, using different CRISPR gRNAs for *TET1* and *TET3*, to determine if the methylation changes we observed were reproducible in an independent TKO clone (Tables 2.1, 2.2). ERRBS analysis showed ~12,000 hyper-DMRs and an insignificant number of hypo-DMRs in TKO lines compared to WT hESCs (Figure 2.5A). We further analyzed the methylation change based on promoter types: bivalent promoters (marked by H3K4me3 and H3K27me3), active promoters (marked by H3K4me3 and H3K79me2), initiated promoters (marked by H3K4me3 only) and silent promoters (based on the absence of H3K4me3) (Lu et al., 2014). Compared to other promoter types, bivalent and silent promoters showed greater magnitude of methylation increase in HUES8 and MEL1 TKO lines. Importantly we observed increasing fractions of bivalent promoters among promoters with greater methylation changes (Figure 2.5B-E). Bivalent promoters constituted 60-74% of all promoters with >60% increase in DNA methylation, greatly exceeding the overall ~27% composition of bivalent promoters evaluated by ERRBS (Figures 2.5C,

2.5E). Overall about half of bivalent promoters showed hypermethylation by ERRBS and WGBS analysis (Figure 2.3E and Figure 2.5F). Although WGBS also showed that a fraction of promoters became hypomethylated after *TET* inactivation, this was not observed by ERRBS. The hypomethylation detected by WGBS is likely due to hypomethylation of non-CGI CpGs that are technically within the promoter regions. We observed similar results in both HUES8 and MEL1 TKO hESCs, with 1,326 and 1,579 bivalent promoters gaining methylation in each respective line. Individual bivalent promoters showed similar methylation changes between these two lines (Figure 2.5G) and 87% of the 1,326 hypermethylated bivalent promoters in HUES8 TKO hESCs also gained methylation in MEL1 TKO hESCs (Figure 2.5H). Overall these results indicate that the TET proteins are critical to preserve hypomethylation at bivalent promoters and in their absence a reproducible subset of bivalent promoters become aberrantly hypermethylated.

**Figure 2.5 Hypermethylation at bivalent promoters is dramatic, robust and significant.** (A) Total number of hyper-DMRs (TKO vs. WT) for HUES8 and MEL1 TKO lines compared to HUES8 and MEL1 WT lines by enhanced reduced representation bisulfite sequencing (ERRBS, right). (B) DNA methylation change between HUES8 TKO and WT hESCs by ERRBS at different promoter types. Box and whisker plots were generated using the methylation change at individual promoters. The error bars show 10 and 90 percent confidence intervals and the bar at the center of the box and whisker plot indicates the median. The promoters are divided into four groups based on histone modification patterns. The details of promoter definitions can be found in the Methods section. (C) Representation of bivalent promoters among promoters that show different degrees of methylation change between HUES8 TKO and WT hESCs by ERRBS, n indicates the total number of promoters in each DNA methylation change group. (D) DNA methylation change between MEL1 TKO and WT hESCs by ERRBS at different promoter types. (E) Representation of bivalent promoters among promoters that show different degrees of methylation change between MEL1 TKO and WT hESCs by ERRBS, n indicates the total number of promoters in each DNA methylation change group. (F) Fraction of genomic regions that show >5% increase in 5mC (Hyper) or a >5% decrease in 5mC (Hypo) in methylation between TKO and WT hESCs by ERRBS of HUES8 and MEL1 WT and TKO hESCs. (G) Correlation between methylation level at bivalent promoters in HUES8 WT, MEL1 WT, HUES8 TKO and MEL1 TKO hESCs. (H) Overlap of the bivalent promoters that show greater than 5% methylation change in HUES8 and MEL1 TKO lines compared to HUES8 and MEL1 WT lines. The p-value for the overlap between HUES8 and MEL1 hypermethylation at bivalent promoters is given (Fisher's exact test). Data are mean ± STD. ns, not significant. *$P$<0.05, **$P$<0.01, ***$P$<0.001, ****$P$<0.0001 by one-way ANOVA (B, D).


A
ERRBS
Hyper-DMR
Hypo-DMR
15000
10000
5000
0
MEL1
HUES8
B
ERRBS
5mC Change (%) TKO–WT
40
30
20
10
0
****
****
****
Active Initiated Bivalent Silent
C
ERRBS
Bivalent
Initiated
Active
Silent
Fraction in each category
1.00
0.75
0.50
0.25
0.00
Total n=12006
>0 n= 8031
>5 n= 2680
>10 n= 1536
>20 n= 808
>30 n= 531
>40 n= 366
>50 n= 262
>60 n= 167
5mC Change (%), TKO–WT
D
MEL1 ERRBS
5mC Change (%) TKO–WT
40
30
20
10
0
****
**
****
Active Initiated Bivalent Silent
G
HUES8 WT (%)
MEL1 WT (%)
R = .84
HUES8 TKO (%)
MEL1 TKO (%)
R = .91
E
Active
Initiated
Bivalent
Silent
Fraction in each category
1.00
0.75
0.50
0.25
0.00
Total n=11893
>0 n= 9089
>5 n= 3041
>10 n= 1654
>20 n= 774
>30 n= 433
>40 n= 264
>50 n= 173
>60 n= 96
HUES8 TKO–WT (%)
MEL1 TKO–WT (%)
R = .75
HUES8 TKO–WT (%)
MEL1 TKO–WT (%)
F
ERRBS
Hyper-DMR
Hypo-DMR
Fraction of Genomic Region
1.0
.75
0.5
.25
0.0
HUES8 MEL1 HUES8 MEL1 HUES8 MEL1 HUES8 MEL1
Active
Initiated
Bivalent
Silent
H
5mC Change (%), TKO–WT
p < 2 e-16
MEL1
HUES8
Hypermethylated Bivalent Promoters
426
1153
173

We also compared our data with previous mESC TKO data (Lu et al., 2014). Out of the 732 bivalent promoters that become hypermethylated (>20% increase) in TKO mESCs, 517 of these 732 promoters are also bivalent in hESCs, out of which 289 (~%56) were associated with hyper-DMRs in TKO hESCs. The remaining 215 promoters are not bivalent in hESCs, and only 55 (~25.6%) were associated with hyper-DMRs in TKO hESCs (Figure 2.6A). This indicates that bivalent promoters found in both mESCs and hESCs tend to show similar methylation changes after *TET* inactivation.

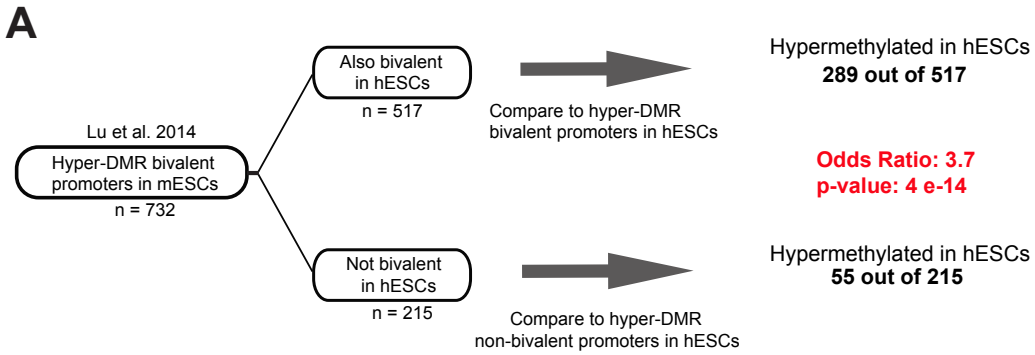


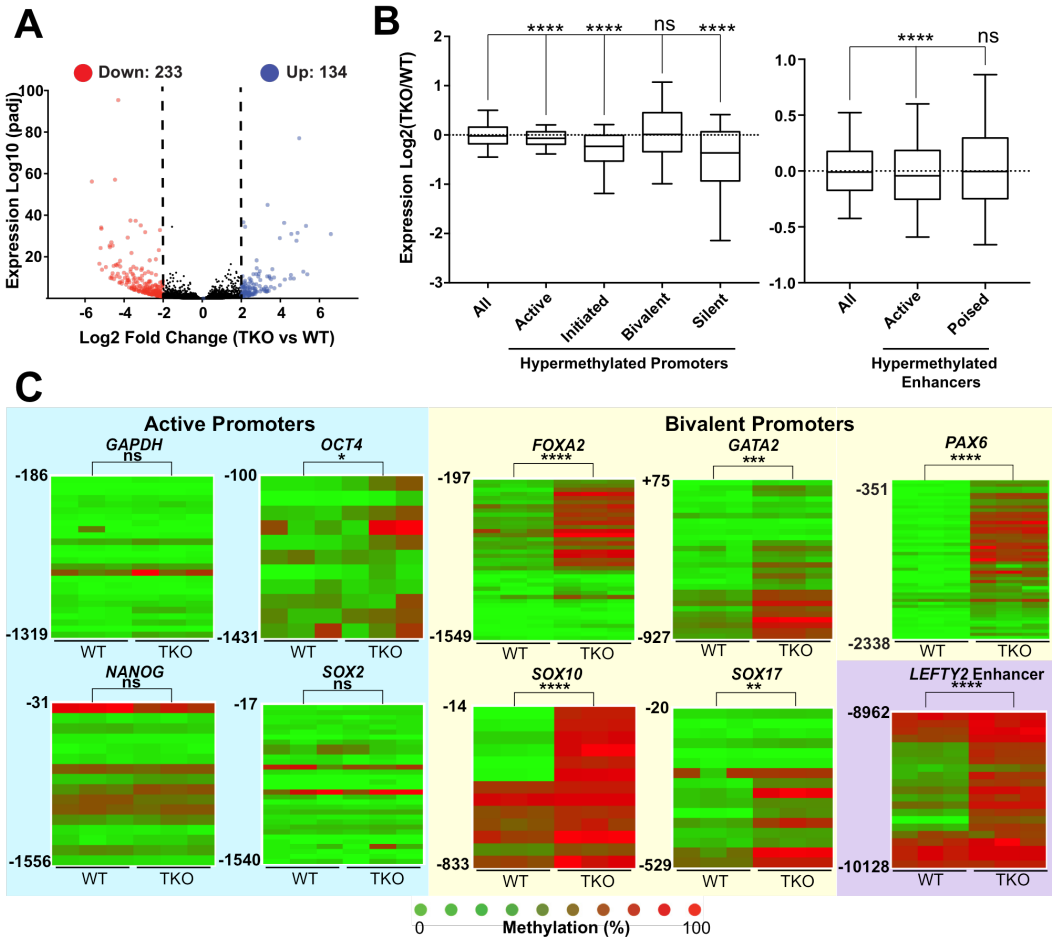

**Figure 2.6 Bivalent promoters in TET TKO hESCs and mESCs show similar methylation changes.** (A) Overlap between all hyper-DMR associated bivalent promoters in mESCs and their human counterparts in hESCs. 517 of the bivalent mESC promoters are also bivalent in hESCs and 215 of the bivalent mESC promoters are not bivalent in hESCs. We compared the overlap of mESC hyper-DMRs to hESC hyper-DMRs within these two groups of promoters, the odds ratio and p-value are given. mESC data was taken from Lu et al., 2014.

### *2.2.2 Hypermethylation of the PAX6 P0 Promoter*

TKO hESCs showed relatively few transcriptional changes compared to WT cells (Figure 2.7A) (RNA-Seq computational analysis was performed by Heng Pan). We found that hypermethylation of active promoters, initiated promoters, silent promoters and active enhancers was associated with a decrease in gene expression. In contrast, genes associated with hypermethylated bivalent promoters and poised enhancers did not show an overall change in expression in TKO hESCs (Figure 2.7B). 5mC MassArray Epityper analysis confirmed that the bivalent promoters of selected developmental genes (*FOXA2*, *GATA2*, *PAX6*, *SOX10*, *SOX17*) showed significant hypermethylation in TKO hESCs, whereas the active promoters for housekeeping and pluripotency genes did not (Figure 2.7C). As expected, the hypermethylated *LEFTY2* enhancer (an active enhancer marked by H3K4me1 and H3K27ac) was associated with a significant decrease in gene expression, as has been previously described in TKO mouse embryos (Dai et al., 2016); whereas no expression change was detected for housekeeping and pluripotency genes (Figure 2.7D). Notably hypermethylation of the bivalent promoters was not associated with a consistent decrease in gene expression. For instance no change in *PAX6* expression was observed. A few bivalent promoter genes showed up- or downregulation, but the expression levels were generally low (Figure 2.7D). We believe that the change for some bivalent genes may be due to increased propensity of TKO cells hESCs to spontaneously differentiate into extra-embryonic tissues, as has been previously reported in single and double *Tet* knockout mESCs (Dawlaty et al., 2013; Dawlaty et al., 2011). We performed immunofluorescence of WT and TKO hESCs after 4 days of

culture. We found that TKO hESCs showed an increased number of cells that were positive for differentiation markers such as *GATA6* and *BRACHYURY* (Figure 2.8A). Spontaneous differentiation typically occurs at the boundaries of hESC colonies and in fact the GATA6 or BRACHYURY positive TKO cells occurred at the boundary of TKO hESC colonies. Notably we did not observe GATA6 or BRACHYURY positive cells within the TKO hESC colonies. Thus we believe that the spontaneous differentiation of a minority of TKO cells is responsible for the increase in expression of some of the differentiation markers that we observed by RNA-Seq.

**Figure 2.7 Gene expression changes in TKO hESCs.** (A) Volcano plot of RNA-Seq data illustrates transcriptional changes in TKO compared to WT hESCs. Upregulated genes are shown as red dots and downregulated genes are shown as blue dots. Genes showing a counts difference larger than twofold were considered differentially expressed. *n*= 2 independent experiments. (B) Left: Expression change for genes of hyper-DMR associated promoters. Significance tests are comparing active, initiated, bivalent and silent promoters to all the promoters together. The details of each of these classifications are provided in the Methods section. Right: Expression change for genes of hyper-DMR associated active enhancers and for genes of hyper-DMR associated poised enhancers. Significance tests are comparing genes associated with hyper-DMR enhancers to all the enhancers together. The details of each of these classifications are provided in the Methods section. (C) Heat map of MassArray analysis of 5mC at different active promoters (*GAPDH*, *OCT4*, *NANOG* and *SOX2*) and bivalent promoters (*FOXA2*, *GATA2*, *SOX10*, *SOX17* and *GATA6*). Each row of the heat map represents either an individual CpG or a few CpGs that are located close together. The location of the CpGs with respect to the TSS is shown to the left of each heat map. For each cell line three biological replicates are shown as three columns. *n*= 3 independent experiments. (D) Expression analysis by RNA-Seq for particular genes. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 by one-way ANOVA (B) and Student's t test (C, D).


A
Down: 233
Up: 134
Expression Log10 (padj)
Log2 Fold Change (TKO vs WT)
B
Expression Log2(TKO/WT)
All
Active
Initiated
Bivalent
Silent
Hypermethylated Promoters
Poised
Hypermethylated Enhancers
C
Active Promoters
Bivalent Promoters
GAPDH
OCT4
FOXA2
GATA2
PAX6
NANOG
SOX2
SOX10
SOX17
LEFTY2 Enhancer
WT
TKO
Methylation (%)
D
FPKM
GAPDH
OCT4
FOXA2
GATA2
PAX6
NANOG
SOX2
SOX10
SOX17
LEFTY2

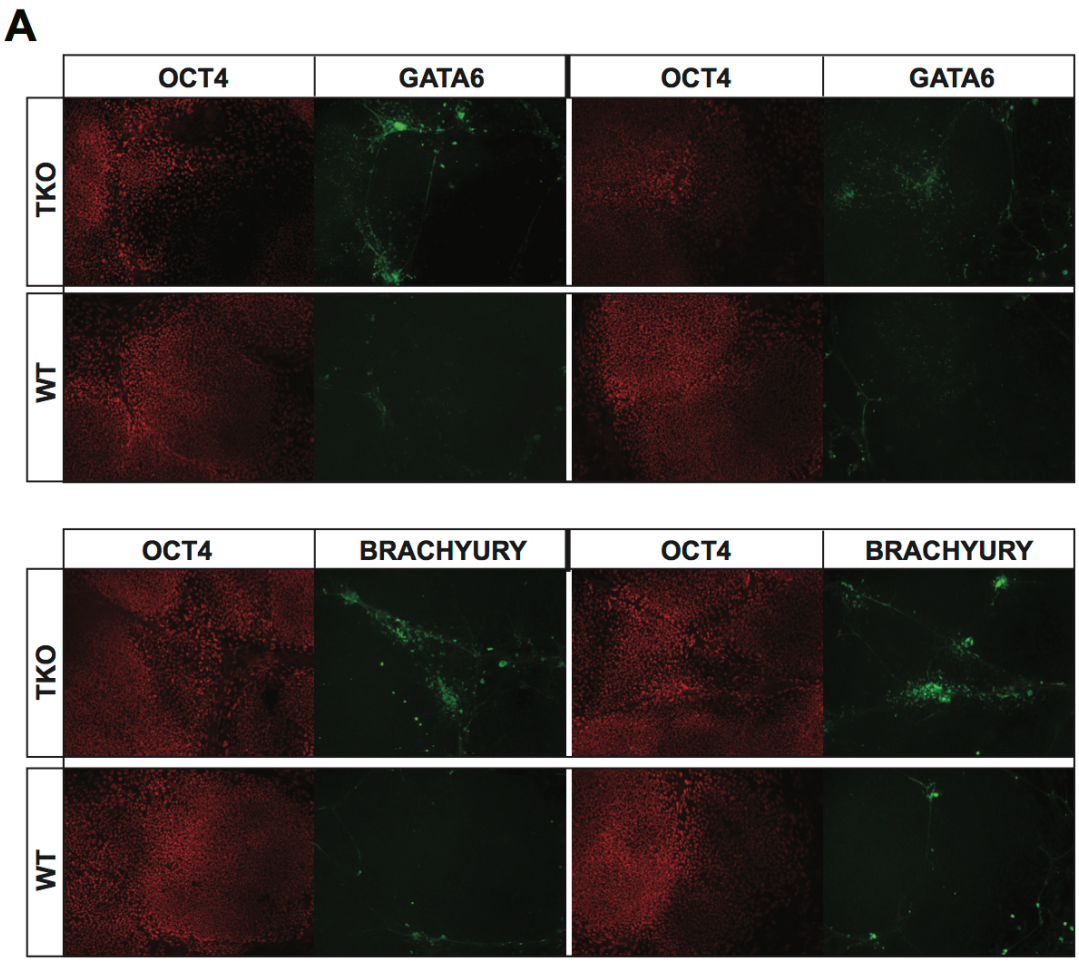
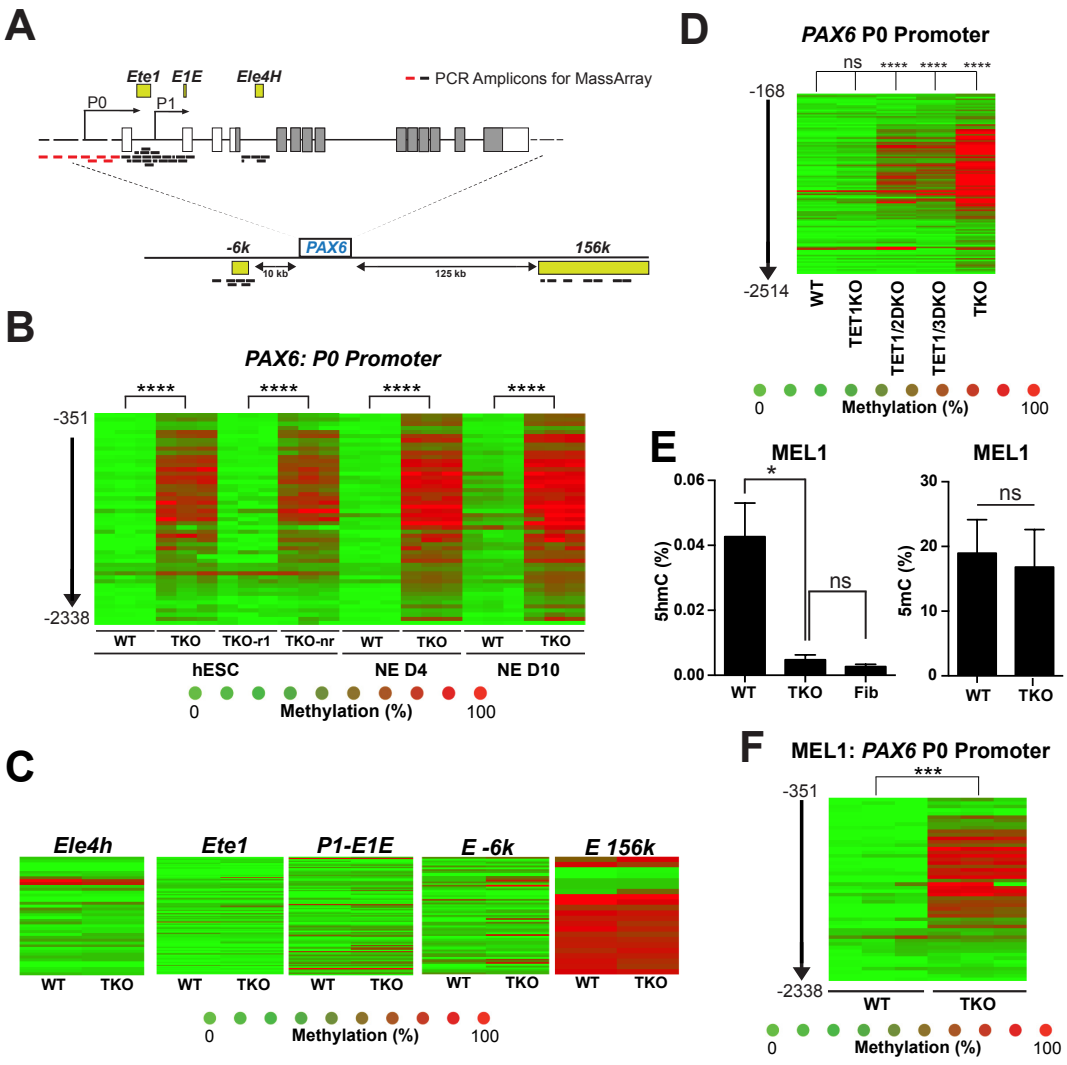



**Figure 2.8 Expression of GATA6 and BRACHYURY in WT and TKO hESCs.** (A) Immunofluorescence for OCT4 and GATA6 or OCT4 and BRACHYURY in WT and TKO hESCs grown in standard hESC culture.

**Figure 2.9 Hypermethylation of the PAX6 P0 promoter.** (A) Diagram of the *PAX6* locus and the associated regulatory regions. Boxes on the horizontal line represent exons. Grey boxes represent the coding sequence. The two promoters that produce a full-length protein (P0 and P1) are shown by arrows. The 5 annotated enhancers are depicted as yellow boxes. (B) Heat map of MassArray analysis of 5mC at the *PAX6* P0 promoter. Each row of the heat map represents either an individual CpG or a few CpGs that are located close together. The location of the CpGs with respect to the P0 TSS is shown to the left of the heat map. For each cell line three biological replicates are shown as three columns. *n*= 3 independent experiments. (NE D4) Neuroectoderm differentiation day 4; (NE D10) Neuroectoderm Differentiation day 10. (C) Heat map of MassArray analysis of 5mC at the regulatory regions of the *PAX6* locus. *n*= 1 independent experiment. (D) Heat map of MassArray analysis of 5mC at the *PAX6* P0 promoter for WT, *TET1* KO, *TET1/2* DKO, *TET1/3* DKO and TKO hESCs. *n*= 1 independent experiment. (E) Left: Analysis of 5hmC levels in MEL1 WT and MEL1 TKO hESCs by mass spectrometry. For MEL1 WT, 2 different lines were used for mass spectrometry measurements. For MEL1 TKO, 2 different passages of the same line were used for mass spectrometry measurements. Human fibroblasts were used as a negative control for mass spectrometry of 5hmC. Right: Analysis of 5mC by mass spectrometry. (F) Heat map of MassArray analysis of 5mC at the *PAX6* P0 promoter in MEL1 WT and TKO hESCs. For each cell line three biological replicates are shown as three columns. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 by one-way ANOVA (D, E) and Student's t test (B, F).


A
Ete1
E1E
Ele4H
PCR Amplicons for MassArray
P0
P1
-6k
PAX6
156k
10 kb
125 kb
B
PAX6: P0 Promoter
-351
-2338
****
WT
TKO
TKO-r1
TKO-nr
hESC
NE D4
NE D10
0
Methylation (%)
100
C
Ele4h
Ete1
P1-E1E
E -6k
E 156k
D
PAX6 P0 Promoter
ns
-168
-2514
TET1KO
TET1/2DKO
TET1/3DKO
E
MEL1
5hmC (%)
5mC (%)
*
ns
Fib
F
MEL1: PAX6 P0 Promoter
***

We were intrigued that despite promoter hypermethylation, bivalent promoter genes such as *PAX6* did not show a change in expression at the hESC stage (Figure 2.7D). Noticing that *PAX6* (as well as *FOXA2* and *SOX10*) failed to be upregulated upon spontaneous embryoid body differentiation (Figure 2.2E), we speculated that the hypermethylation at bivalent promoters could affect activation of gene expression following exposure to differentiation signals. We chose to focus on *PAX6* as it is a critical gene for human neural development and is highly expressed during in vitro differentiation of hESCs into the neural lineage (Li et al., 2005; Pankratz et al., 2007; Zhang et al., 2010), thus allowing us to use neural differentiation to track the effects of *TET* deletion on *PAX6* expression as well as cellular differentiation. *PAX6* has well annotated promoters and enhancers (Figure 2.9A) (Anderson et al., 2002; Kammandel et al., 1999; Plaza et al., 1999; Xu and Saunders, 1998; Zheng et al., 2001). 5mC MassArray analysis of these regions revealed that only the bivalent P0 promoter of *PAX6* showed hypermethylation (Figure 2.9B-C). The severity of this hypermethylation showed a positive correlation with the loss of 5hmC (Figure 2.9D) and was recapitulated in the MEL1 TKO line (Figure 2.9E-F). We further generated two “TKO-repaired” lines through CRISPR/Cas9-mediated homology directed repair for comparison with isogenic WT and TKO hESCs to rule out potential CRISPR off-target effects (Figure 2.10A). Repair of one *TET1* allele to the WT sequence in TKO hESCs was sufficient to restore 5hmC to near WT levels by mass spectrometry (performed by Louis Dore) (Figure 2.10B), and importantly, it also reversed the *PAX6* P0 promoter hypermethylation (Figure 2.9B). Notably the TKO-repaired line was

able to form teratomas, indicating a rescue in their differentiation capacity (Figure 2.2D).

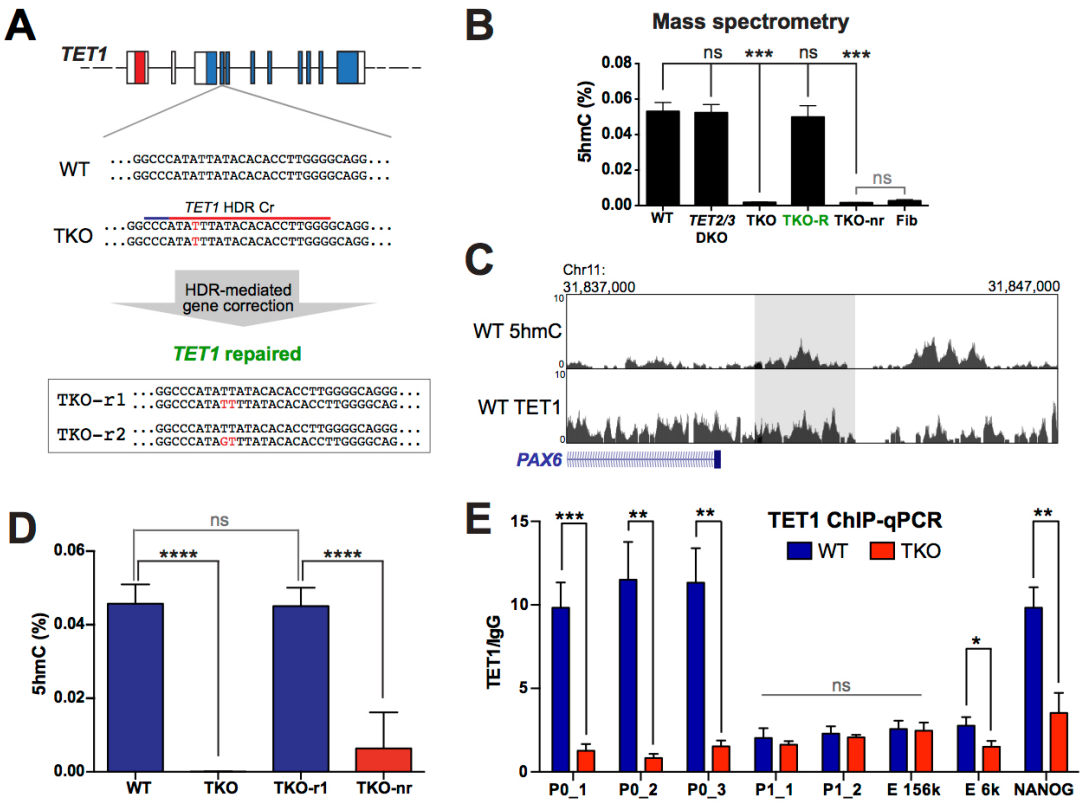



**Figure 2.10 TET1 and 5hmC are present at the *PAX6* P0 promoter.** (A) Diagram of homology-directed repair (HDR) of the *TET1* locus in TKO hESCs. The red line indicates the gRNA targeting sequence and the blue line is the adjacent PAM sequence. Red letters indicate mutations of the WT sequence. The sequences of the two repaired lines (TKO-r1, TKO-r2) are shown below. (B) Mass spectrometry analysis of 5hmC levels in WT, *TET* KO mutant lines, lines in which one allele of *TET1* has been repaired (TKO-r) and lines which underwent HDR targeting but retained the TKO mutations in the *TET1* locus (TKO-nr). For all mass spectrometry analysis, 2 mutant lines were used for all genotypes except for TKO. For TKO lines, 2 different passages of the same line were used for mass spectrometry measurements. Human fibroblasts were used as a negative control for mass spectrometry of 5hmC. For significance tests black lines indicate comparisons to WT. (C) Top panel: Analysis of 5hmC peak at the *PAX6* P0 promoter by hMe-Seal in WT hESCs. The height of the 5hmC peak above the x-axis reflects the normalized tag count. Bottom panel: Analysis of TET1 peak at the *PAX6* P0 promoter by TET1 ChIP-Seq in WT hESCs. The height of the 5hmC peak above the x-axis reflects the normalized tag count. Shaded area represents the region of the *PAX6* P0 promoter assayed for 5hmC by Epimark (D) and TET1 binding by ChIP-qPCR (E). The predicted P0 PAX6 transcript is shown on the bottom. (D) Analysis of percent 5hmC at the *PAX6* P0 promoter by Epimark. *n*= 3 independent experiments. (E) ChIP-qPCR for TET1 in WT and TKO hESCs. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 by one-way ANOVA (B, D) and Student's t test (E).

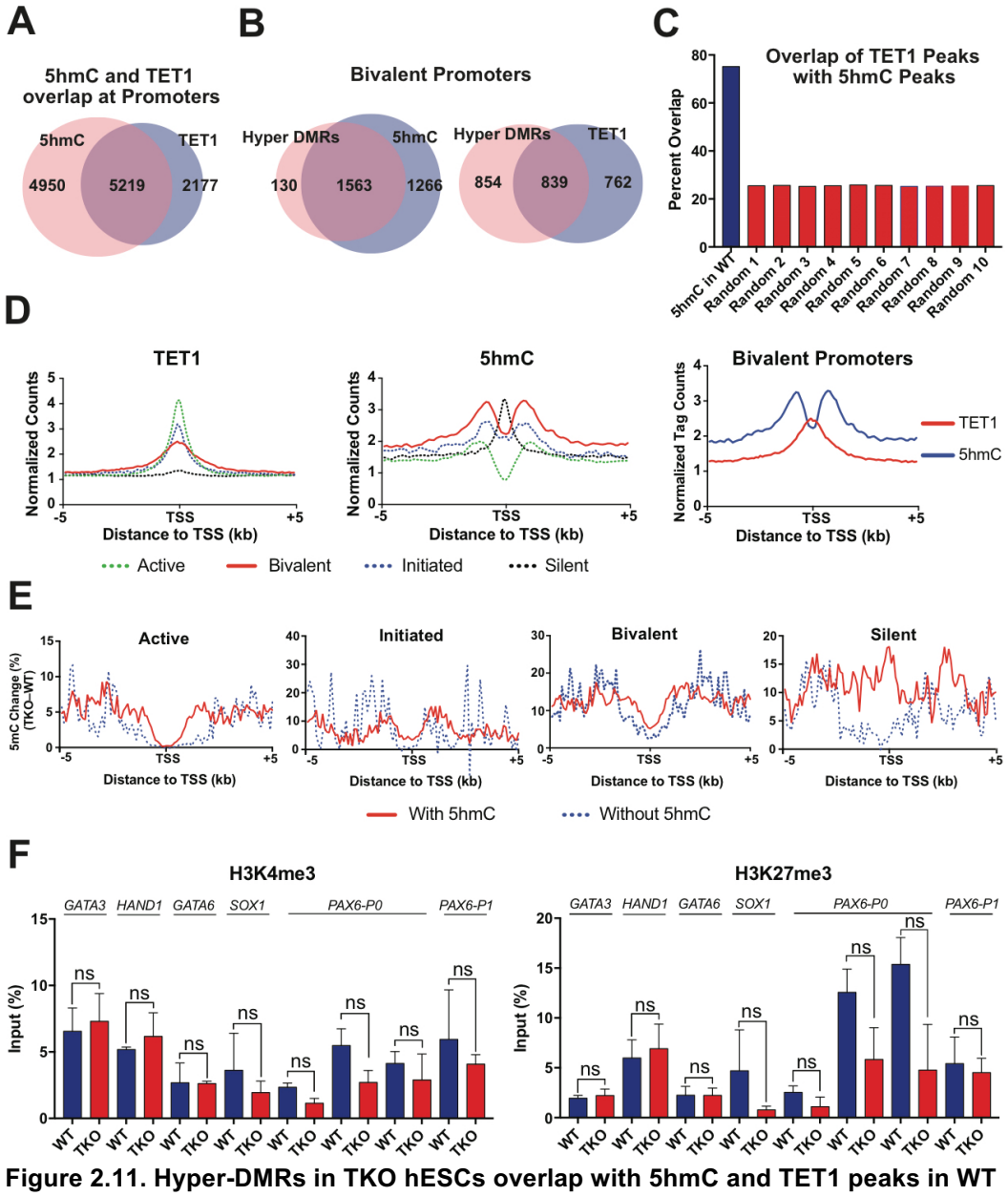



**Figure 2.11. Hyper-DMRs in TKO hESCs overlap with 5hmC and TET1 peaks in WT hESCs.** (A) Overlap of 5hmC and TET1 peaks found at promoters. (B) Overlap between hyper-DMRs (TKO vs. WT) that occur at bivalent promoters with TET1 (left) and 5hmC (right) peaks at bivalent promoters in WT hESCs. (C) Percentage of TET1 peaks overlapping with 5hmC peaks in WT hESCs compared to randomly generated 5hmC peaks of equal number and height. (D) Analysis of TET1 (left) and 5hmC (middle) peaks at active, initiated, bivalent and silent promoters in WT hESCs. On the right are 5hmC and TET1 peaks at bivalent promoters. The height above the x-axis reflects the normalized tag count. (E) Percent DNA methylation change (TKO – WT) in active, initiated, bivalent and silent promoters that have 5hmC peaks compared to promoters that don't have 5hmC peaks. (F) ChIP-qPCR for H3K4me3 and H3K27me3 in WT and TKO hESCs. **P*<0.05, ***P*<0.01, ****P*<0.001 by Student's t test (F).

To confirm that the hypermethylation of the *PAX6* P0 promoter is a direct consequence of losing the TET proteins, we performed locus-specific and genome-wide 5hmC profiling by the Epimark 5hmC Analysis Kit (Stroud H, 2011) and hMe-Seal (Song et al., 2011): both showed 5hmC enrichment at the *PAX6* P0 promoter in WT hESCs (Figure 2.10C-D). We also detected TET1 binding at the *PAX6* P0 promoter in WT hESCs by ChIP-Seq (computational analysis performed by Heng Pan) (Figure 2.10C) and ChIP-qPCR (performed by Abhijit Shukla) (Figure 2.10E). We analyzed our TET1 ChIP-Seq and 5hmC profiling further to determine if the locus specific hypermethylation that we observe is a direct consequence of the loss of the TET proteins. We found that approximately 51 percent of 5hmC peaks overlapped with TET1 binding (Figure 2.11A). The reasons why there may not be complete overlap between 5hmC and TET1 are explained further in the methods section. Moreover we found that approximately 92% and 50% of hyper DMRs found at bivalent promoters in HUES8 TKO hESCs overlapped with 5hmC and TET1 peaks, respectively (Figure 2.11A-B). The greater overlap between hyper-DMRs and 5hmC peaks, than between hyper-DMRs and TET1, is likely due to the production of 5hmC by TET2 and TET3. Globally, TET1 binds to bivalent, active and initiated promoters at the transcription start site. Binding of TET1 overlaps with 5hmC signals, which extends into the promoter and gene body (Figure 2.11D). The functional relevance of TET binding and 5hmC signal is supported by a greater methylation increase in TKO hESCs associated with bivalent promoters that have 5hmC peaks compared to bivalent promoters that don't have 5hmC peaks (Figure 2.11E) Overall these findings support the conclusion that TET1 binding at bivalent promoters (such as the *PAX6* P0 promoter) leads

to 5hmC production, and *TET* inactivation causes 5hmC depletion and aberrant promoter hypermethylation. Finally we performed ChIP-qPCR to determine whether loss of the TET proteins or hypermethylation at the locus affected the bivalent histone mark at the *PAX6* P0 promoter. Although we did observe a decrease in both the H3K4me3 and H3K27me3 marks in TKO hESCs versus WT hESCs the results were not significant (Figure 2.11F).

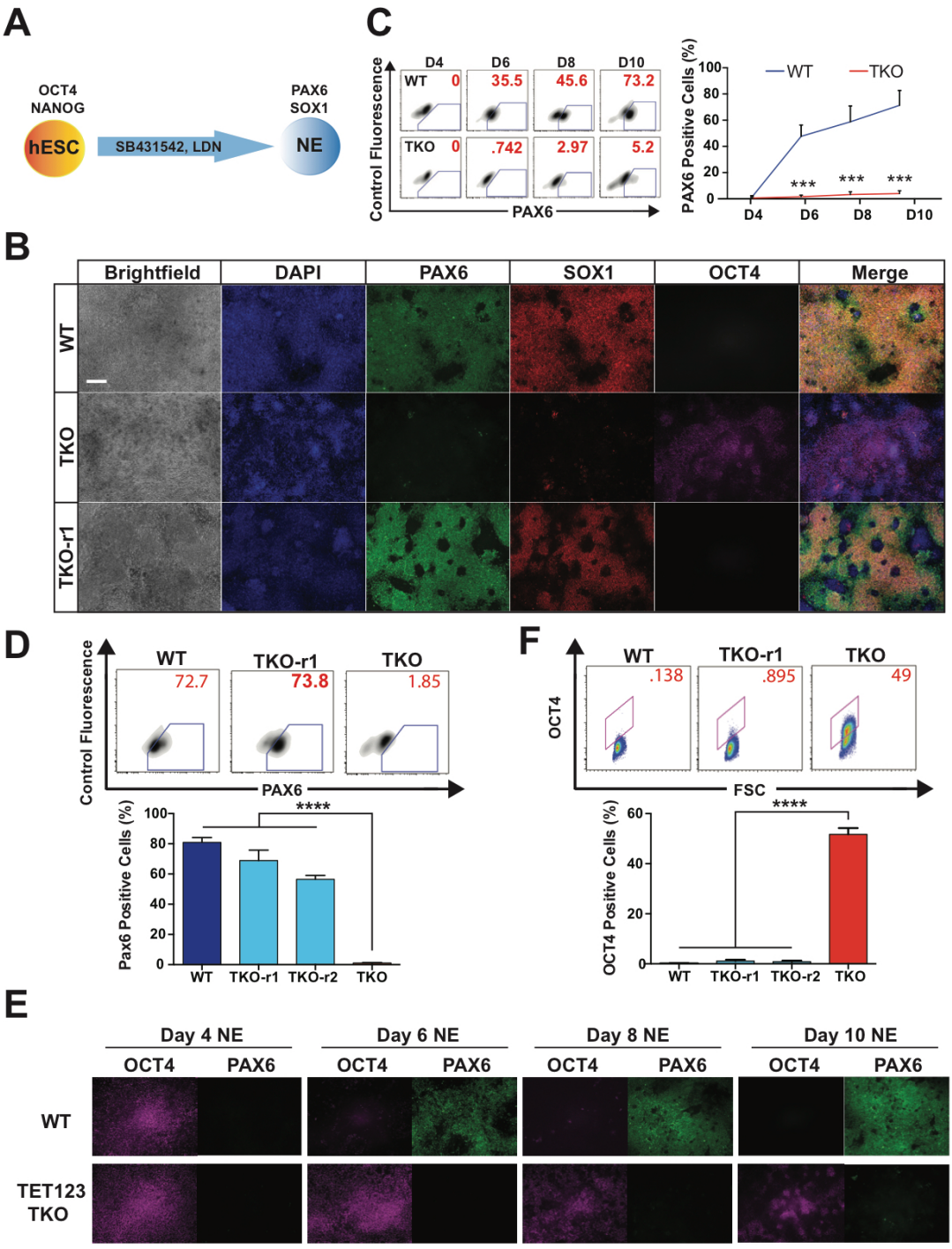
### *2.2.3 Impaired Neural Differentiation of TKO hESCs*

Our observation of bivalent promoter hypermethylation prior to differentiation suggests that hypermethylation of bivalent promoters at the ESC stage could have a direct impact on differentiation. Since *PAX6* is expressed in early neuroectoderm (NE) derived from hESCs (Li et al., 2005; Pankratz et al., 2007), and it is shown to be both necessary and sufficient for NE formation from hESCs (Zhang et al., 2010), we speculate that hypermethylation of the *PAX6* P0 promoter in TKO hESCs may impede hESC differentiation into the neural lineage. We investigated this further using the dual SMAD inhibition protocol for NE differentiation (Chambers et al., 2009) (Figure 2.12A).

We were able to efficiently differentiate WT and TKO-repaired lines into PAX6 and SOX1-positive NE cells with appropriate downregulation of pluripotency markers such as OCT4 (Figure 2.12B). In contrast, TKO hESCs failed to form significant numbers of PAX6-positive cells at any point during the differentiation (Figure 2.12B-D) suggesting impaired acquisition rather than maintenance of the NE fate. Notably whereas WT cells lost OCT4 expression by day 6 (Figure 2.12E), ~40% of TKO cells remained OCT4-positive after 10 days of differentiation as shown by immunostaining and

FACS analysis (Figure 2.12E-F).

**Figure 2.12 TKO hESCs show a defect in neuroectoderm differentiation.** (A) Schematic for neuroectoderm (NE) differentiation. (B) Immunofluorescence of PAX6, SOX1 and OCT4 at the endpoint of differentiation (D10) of WT, TKO and TKO-r1 cells. Unless otherwise noted the scale bar for all immunofluorescence images indicates 100 μm. (C) Representative FACS plots of PAX6 staining at D4, D6, D8 and D10 of WT and TKO cells (left panel). On the right is the quantification of PAX6-positive cells at D4, D6, D8 and D10 of NE differentiation. *n*= 3 independent experiments. (D) Representative FACS plots of PAX6 staining at D10 of WT, TKO-r1 and TKO cells. On the bottom is the quantification of PAX6-positive cells at D10 of NE differentiation. *n*= 3 independent experiments. (E) OCT4 immunofluorescence in WT and TKO cells at different timepoints of NE differentiation (D4, D6, D8 and D10 of NE differentiation). (F) Representative FACS plots of OCT4 staining at D10 of NE differentiation for WT, TKO-r1 and TKO cells. On the bottom is the quantification of OCT4-positive cells. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (D, F) and by Student's t test (C).


A
OCT4
NANOG
hESC
SB431542, LDN
PAX6
SOX1
NE
C
D4
D6
D8
D10
WT
0
35.5
45.6
73.2
TKO
0
.742
2.97
5.2
Control Fluorescence
PAX6
PAX6 Positive Cells (%)
WT
TKO
***
B
Brightfield
DAPI
PAX6
SOX1
OCT4
Merge
WT
TKO
TKO-r1
D
WT
TKO-r1
TKO
72.7
73.8
1.85
Control Fluorescence
PAX6
Pax6 Positive Cells (%)
****
WT
TKO-r1
TKO-r2
TKO
F
WT
TKO-r1
TKO
.138
.895
49
OCT4
FSC
OCT4 Positive Cells (%)
****
E
Day 4 NE
Day 6 NE
Day 8 NE
Day 10 NE
OCT4
PAX6
WT
TET123
TKO

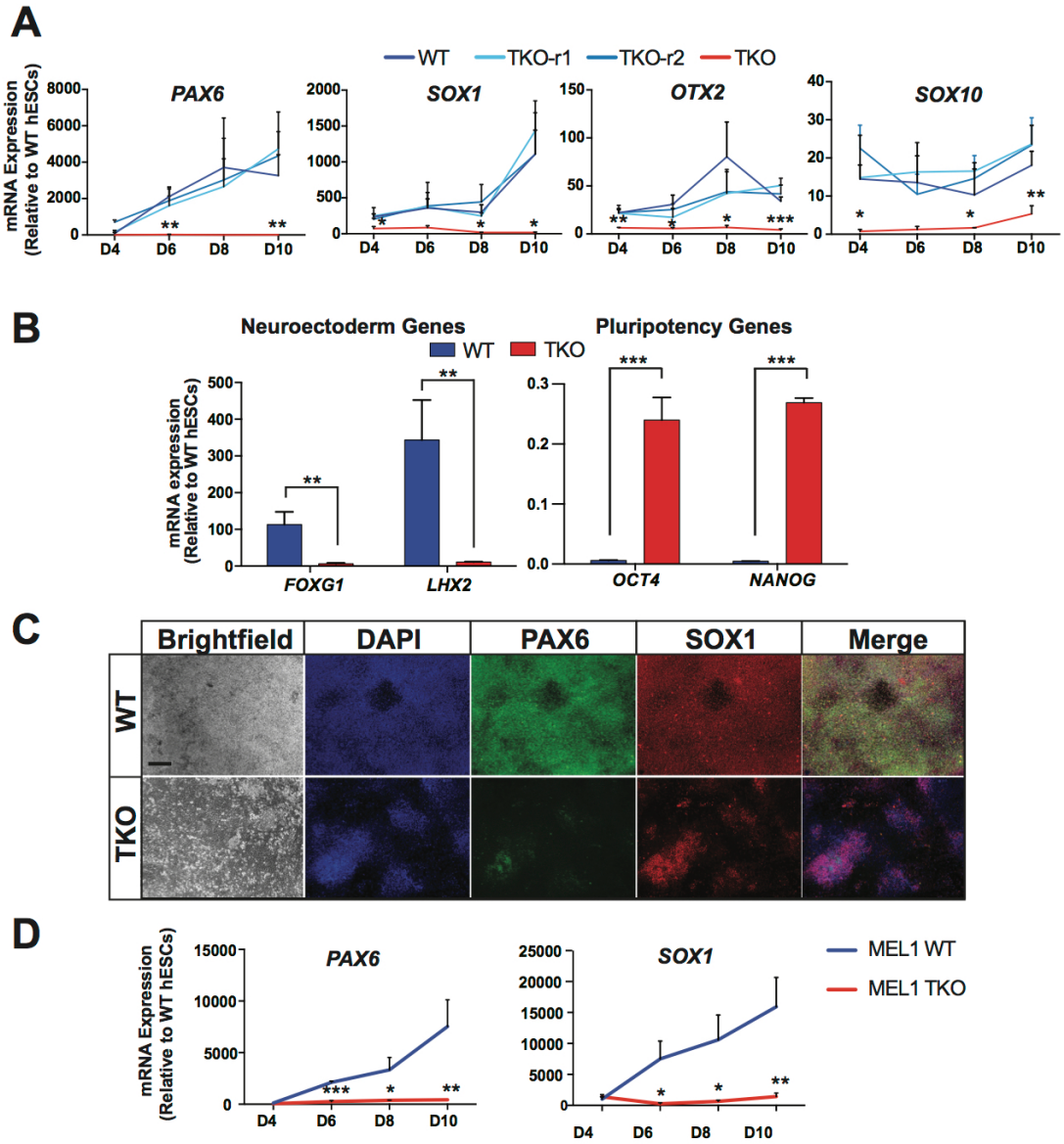



**Figure 2.13 Neural differentiation defect of TKO hESCs.** (A) qPCR analysis for epiblast (*OTX2*), neuroectoderm (*PAX6*, *SOX1* and *OTX2*) and neural crest (*SOX10*) markers in WT, TKO-r1, TKO-r2, and TKO-nr cells at four different time points of differentiation (D4, D6, D8, D10). (B) qPCR analysis for neural (*FOXG1* and *LHX2*) and pluripotency (*OCT4* and *NANOG*) markers in WT and TKO cells at D10 of NE differentiation. (C) Immunofluorescence of PAX6, SOX1 and OCT4 at D10 of NE differentiation in MEL1 WT and MEL1 TKO cells. (D) Expression of neuroectoderm (*PAX6* and *SOX1*) markers during NE differentiation of MEL1 WT and MEL1 TKO lines. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 by Student's t test (B, D).

Analysis of additional markers as well as the MEL1 TKO line confirmed the NE differentiation defect of TKO hESCs (Figure 2.13A-D). We performed immunofluorescence for neural crest (SOX10 and TFAP2 positive) and non-neural ectoderm (SOX10 negative and TFAP2 positive) cells. Interestingly TKO hESCs showed an absence of SOX10 positive cells but formed a similar number of non-neural ectoderm as WT hESCs. RT-qPCR of GATA3 showed a similar expression of *GATA3* at day 10 of differentiation of WT and TKO cells, supporting that non-neural ectoderm differentiation may not be affected in TKO hESCs (Figure 2.14A-B). At day 10 of NE differentiation TKO cells showed increased expression for endoderm (*GATA6*) and mesoderm (*BRACHYURY*) compared to WT cells. This suggests that TKO cells may differentiate into these lineages instead of neuroectoderm; however, the upregulation of these markers in differentiated TKO cells (in comparison to WT hESCs) was low indicating limited differentiation into these lineages (Figure 2.14B). During differentiation TKO cells showed greater apoptosis than WT cells (Figure 2.14C), possibly because TKO hESCs were unable to differentiate into NE and cannot survive in NE differentiation media. Finally the NE phenotype of TKO hESCs could not be rescued by altering the starting density, indicating that this defect is not due to proliferation or colony forming differences between WT and TKO hESCs (Figure 2.14D). Notably, comparison of TKO and *TET* single and double knockout lines showed that the severity of the NE differentiation defect depended on the *TET1/2/3* gene dosage. Loss of TET1 had the largest effect on bulk 5hmC levels (Figure 2.3A) as well as NE differentiation as determined by FACS, immunostaining, and RT-qPCR analysis for *PAX6* and *SOX1* expression (Figure 2.15A-C).

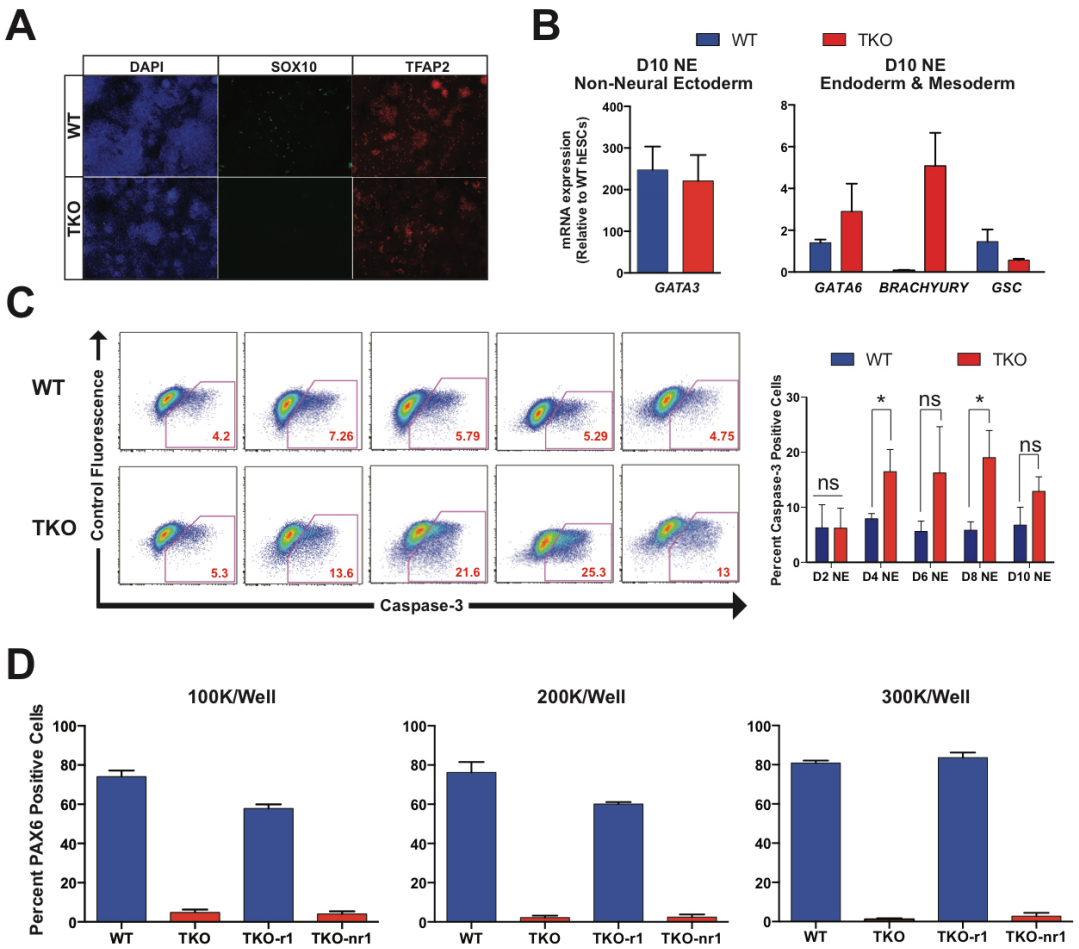
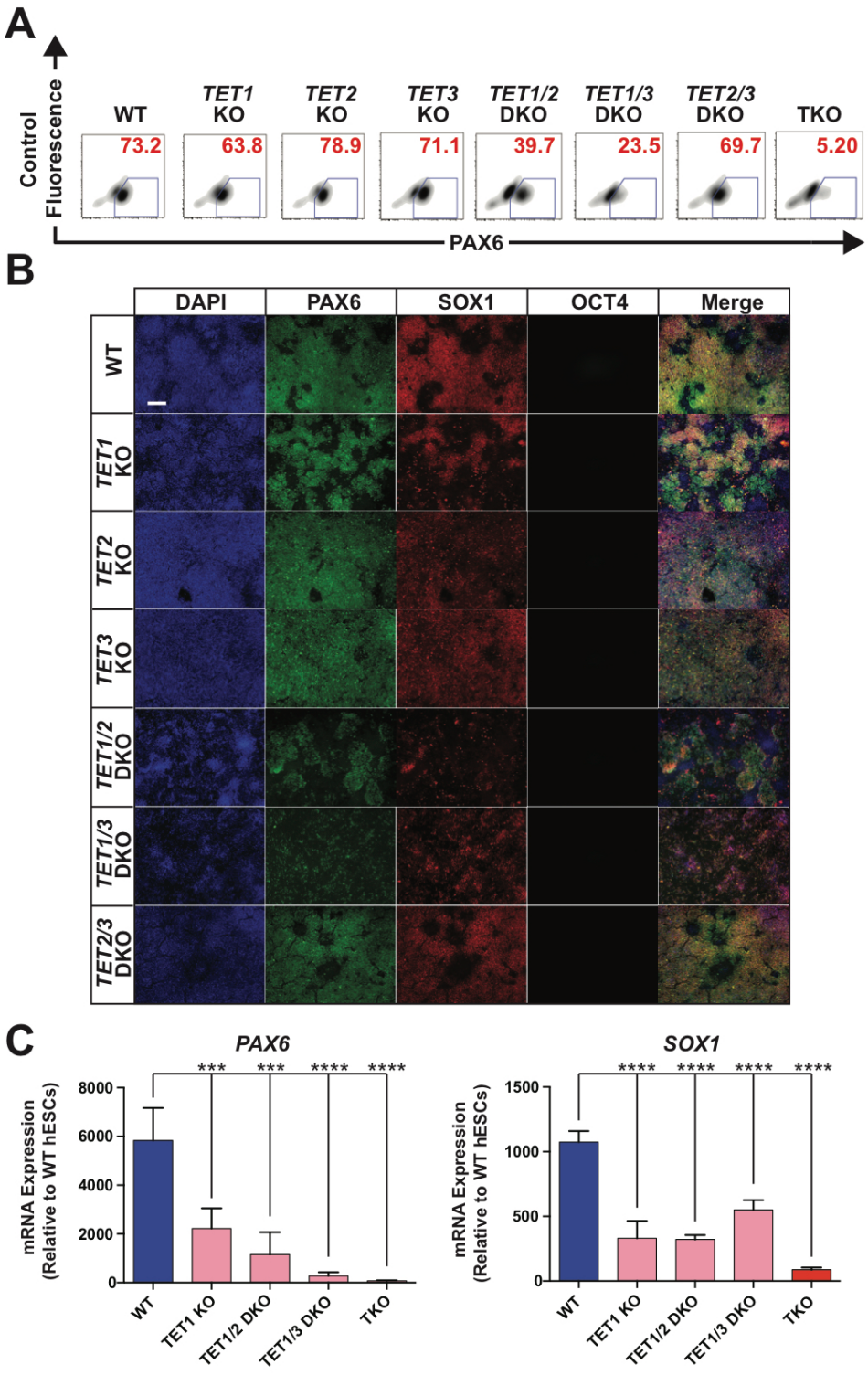



**Figure 2.14 Neural differentiation defect of TKO hESCs.** (A) Immunofluorescence of a neural crest (SOX10 and TFAP2 positive) and a non-neural ectoderm (SOX10 negative and TFAP2 positive) at day 10 of NE differentiation in WT and TKO cells. (B) qPCR analysis for non-neural ectoderm (*GATA2*), endoderm (*GATA6*), mesoderm (*BRACHYURY*) and mesendoderm (*GSC*) markers in WT and TKO cells at D10 of NE differentiation. (C) Apoptosis analysis by FACS for Caspase-3. Representative FACS plots of Caspase-3 staining at D4, D6, D8 and D10 of NE differentiation in WT and TKO cells. On the right is the quantification of Caspase-3 positive cells. *n*= 3 independent experiments. (D) PAX6 positive cells at D10 of NE differentiation of WT, TKO, TKO-r1 and TKO-nr cells. 3 different starting densities are shown. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by Student's t test (C).

**Figure 2.15 Neuroectoderm differentiation is sensitive to the *TET* gene dosage.** (A) Representative FACS plots of PAX6-positive cells at D10 of NE differentiation for *TET* KO mutants. (B) Immunofluorescence of PAX6, SOX1 and OCT4 at the endpoint of differentiation (D10) of *TET* KO mutants. (C) Expression of neuroectoderm (*PAX6* and *SOX1*) markers at D10 of NE differentiation. For significance tests all comparisons are to WT. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (C).


A
Control Fluorescence
WT
TET1 KO
TET2 KO
TET3 KO
TET1/2 DKO
TET1/3 DKO
TET2/3 DKO
TKO
73.2
63.8
78.9
71.1
39.7
23.5
69.7
5.20
PAX6
B
DAPI
PAX6
SOX1
OCT4
Merge
WT
TET1 KO
TET2 KO
TET3 KO
TET1/2 DKO
TET1/3 DKO
TET2/3 DKO
C
PAX6
mRNA Expression (Relative to WT hESCs)
8000
6000
4000
2000
0
***
***
****
****
WT
TET1 KO
TET1/2 DKO
TET1/3 DKO
TKO
SOX1
mRNA Expression (Relative to WT hESCs)
1500
1000
500
0
****
****
****
****
WT
TET1 KO
TET1/2 DKO
TET1/3 DKO
TKO

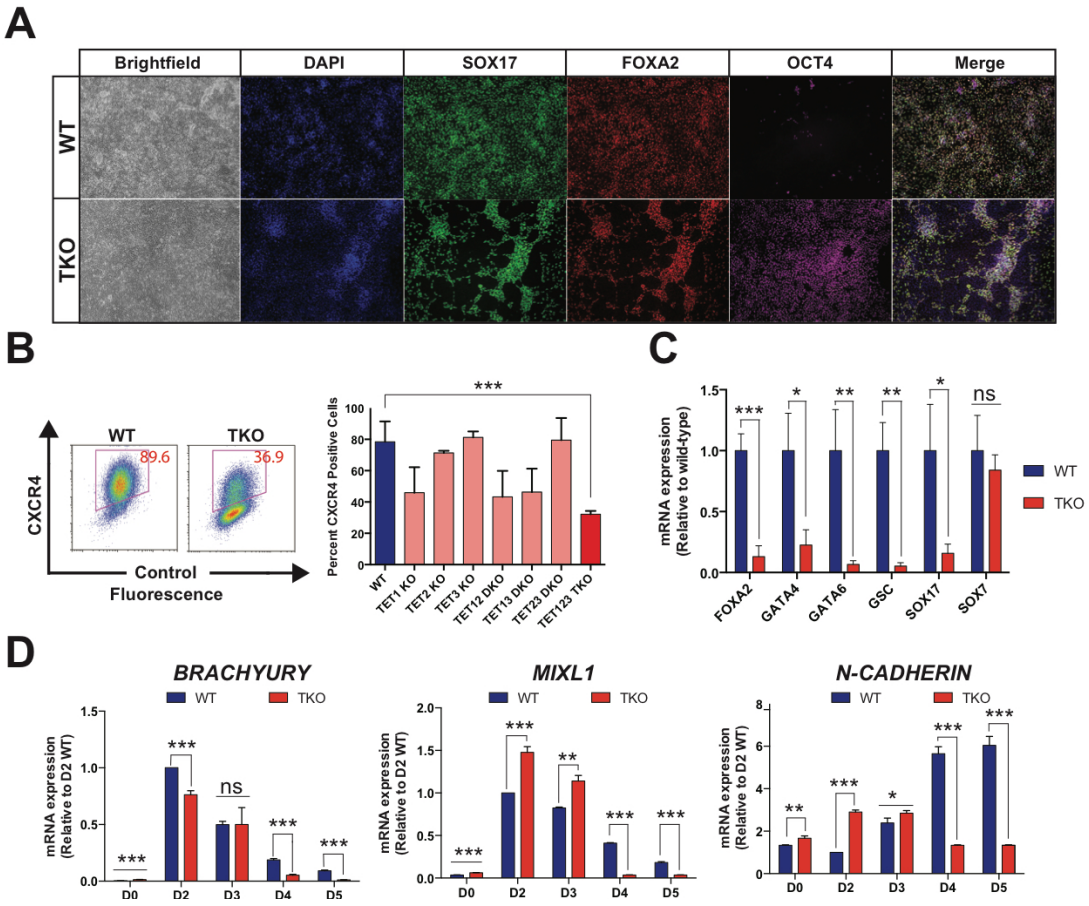



**Figure 2.16 Differentiation of TKO hESCs into definitive endoderm and mesoderm lineages.** (A) Immunofluorescence of SOX17, FOXA2 and OCT4 at day 5 of definitive endoderm differentiation. (B) FACS analysis for surface marker CXCR4 associated with definitive endoderm at day 5 of definitive endoderm differentiation. *n*= 3 independent experiments. (C) Expression of definitive endoderm (*FOXA2*, *GATA4*, *GATA6*, *GSC*, *SOX17*) and primitive endoderm (*SOX7*) markers at day 5 of definitive endoderm differentiation for WT and TKO cells. *n*= 3 independent experiments. (D) Expression of mesoderm (*BRACHYURY, MIXL1* and *N-CADHERIN*) markers at different timepoints of mesoderm differentiation. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (B) or Student's t test (C, D).

Finally we wanted to investigate whether TKO hESCs showed defects upon directed differentiation into additional lineages. TKO hESCs showed a differentiation defect into definitive endoderm (DE) by immunofluorescence (Figure 2.16A), flow-cytometry (Figure 2.16B) and RT-qPCR (differentiations were performed with assistance of Qing Li) (Figure 2.16C). Similar to the results with NE differentiation, upon differentiation to DE TKO cells showed reduced expression of differentiation markers (*SOX17*, *FOXA2*, *CXCR4*, *GATA4*, *GATA6* and *GSC*) and greater number of OCT4 positive cells. Directed differentiation into mesoderm (performed by Chan-Jung Chan) showed a different result than with NE and DE. Interestingly TKO cells showed similar or even increased expression of mesoderm markers (*BRACHYURY*, *MIXL1*, *N-CADHERIN*) at the early time points of differentiation (day 2 and day 3) and reduced expression at later time points (day 4 and day 5) (Figure 2.16D). We did not observe an increase in methylation at *BRACHYURY* and *MIXL1* bivalent promoters in our TKO hESCs. As a result is possible that TKO hESCs are able to increase expression of these genes upon the initiation of differentiation and allow initial differentiation into the mesoderm lineage. Other hypermethylated genes that are expressed at the later time points of differentiation (day 4 and day 5) may hinder further differentiation and lead to a decrease in expression of mesoderm markers.

### *2.2.4 Promoter Hypermethylation Hinders PAX6 Expression Upon Differentiation*

*PAX6* is expressed at a very low level in hESCs, and the P0 promoter hypermethylation in TKO hESCs had no effect on *PAX6* gene expression in

hESCs (Figure 2.7D). We hypothesized that the hypermethylation of the *PAX6* P0 promoter prevents activation of *PAX6* expression upon differentiation and leads to the NE differentiation defect in TKO hESCs. Supporting this hypothesis, 5mC MassArray analysis showed aberrant hypermethylation at the *PAX6* P0 promoter in TKO hESCs both before and during NE differentiation (Figure 2.9B). To establish direct causality we needed to determine whether loss of *PAX6* expression accounts for the NE differentiation defect of TKO hESCs and then investigate whether the P0 promoter hypermethylation is responsible for the loss of *PAX6* expression. We first performed a rescue experiment in which we expressed the *PAX6* transgene under the control of a doxycycline inducible promoter in TKO cells during NE differentiation (Figure 2.17A). We targeted the doxycycline inducible *PAX6* vector into TKO hESCs (TKO-PAX6) using a lentiviral construct (Figure 2.17B) and confirmed that doxycycline treatment induced *PAX6* expression (Figure 2.17C-D). We treated TKO-PAX6 hESCs with doxycycline for different intervals during NE differentiation and found that continuous overexpression of PAX6 during the entire 10 days of NE differentiation was the most effective interval for rescuing the NE defect (Figure 2.17E). TKO cells exposed to doxycycline were able to upregulate NE markers *SOX1* and *FOXG1* and downregulate the pluripotency markers *OCT4* and *NANOG*. However, *SOX10* and endogenous *PAX6* expression were not restored (Figure 2.18A-B). By ERRBS we found that compared to the promoters of *SOX1* and *FOXG1*, the *SOX10* and *PAX6* promoters showed a much greater methylation increase in TKO hESCs, which may prevent their expression even when the *PAX6* transgene is overexpressed.

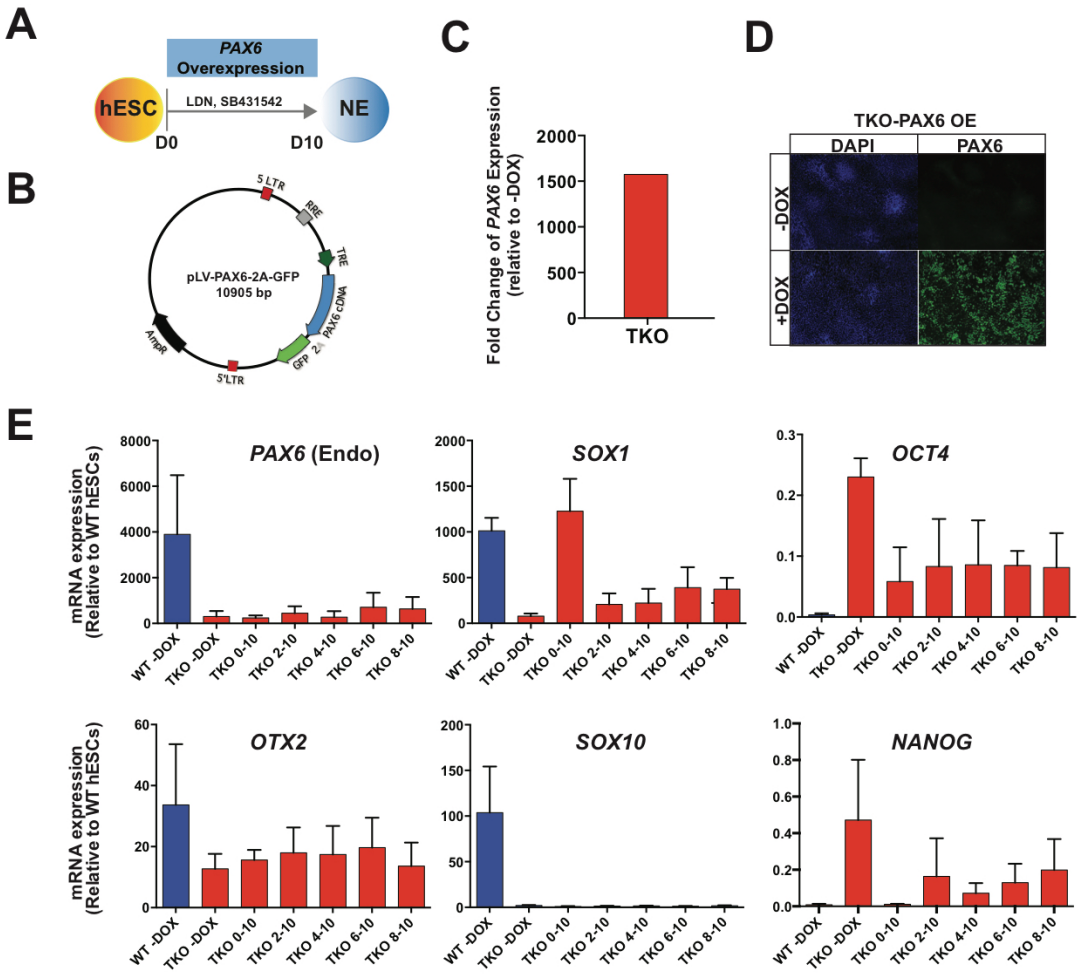



**Figure 2.17. Validation of *PAX6* overexpression TKO line.** (A) Schema for rescue of the NE differentiation defect in TKO hESCs using *PAX6* overexpression. (B) Lentiviral construct to overexpress *PAX6* in TKO cells. (C) PAX6 induction in TKO hESCs that had been infected with the *PAX6* overexpression lentivirus after doxycycline (DOX) treatment. (D) Immunofluorescence for PAX6 in TKO hESCs that had been infected with the PAX6 overexpression lentivirus after doxycycline (DOX) treatment. (E) TKO hESCs were infected with a lentiviral construct for PAX6 overexpression. Infected cells were differentiated to NE using the standard NE differentiation protocol. Cells were treated with doxycycline (DOX) to induce *PAX6* for variable intervals during differentiation. 0-10: DOX treatment from D0 to D10, 2-10: DOX treatment from D2 to D10, 4-10: DOX treatment from D4 to D10, 6-10: DOX treatment from D6 to D10, 8-10: DOX treatment from D8 to D10.

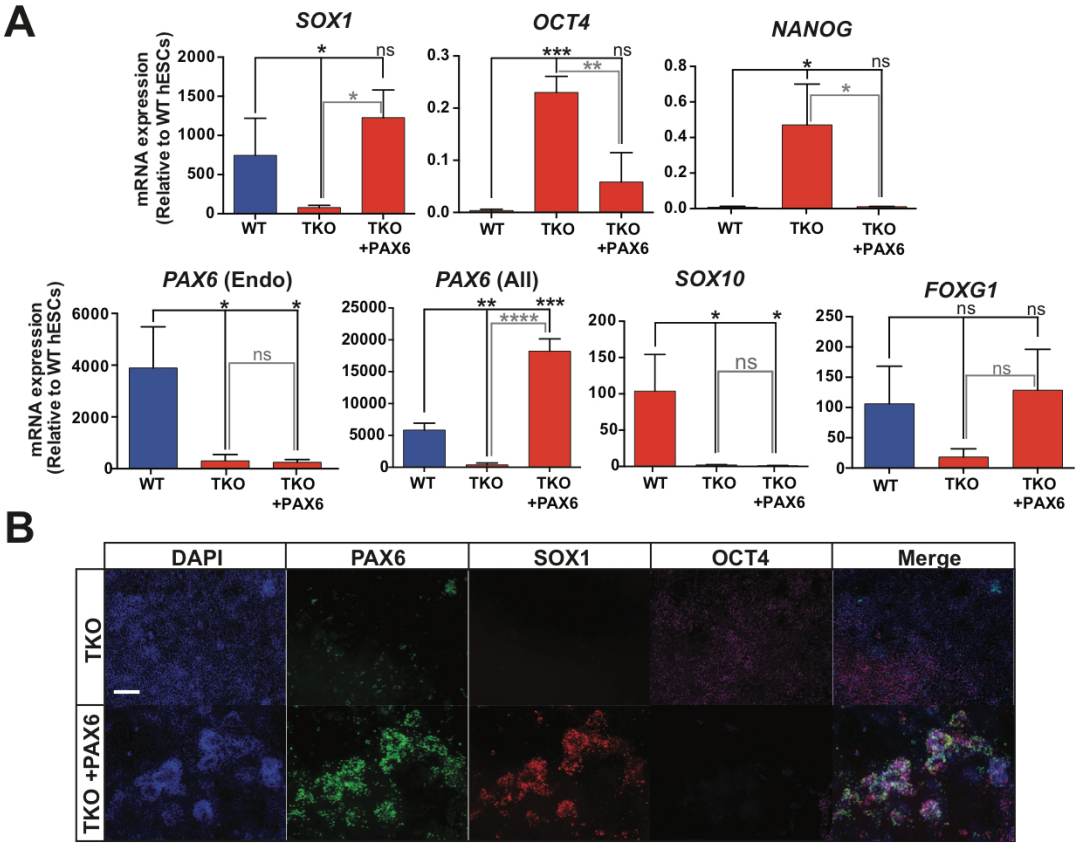



**Figure 2.18 Overexpression of *PAX6* can partly rescue neuroectoderm differentiation in TKO cells.** (A) qPCR analysis of neuroectoderm (*SOX1*, *PAX6*, *FOXG1*), neural crest (*SOX10*) and pluripotency (*OCT4*, *NANOG)* markers in WT and TKO cells without doxycycline (TKO) and with doxycycline treatment for 10 days (TKO +PAX6) at D10 of NE differentiation. For significance tests black lines indicate comparisons to WT. *n*= 3 independent experiments. (B) Immunofluorescence of PAX6, SOX1 and OCT4 at the endpoint of differentiation (D10) of TKO cells without doxycycline treatment (TKO) and with doxycycline treatment for 10 days (TKO + PAX6). Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (A).

The above findings suggest that failure of *PAX6* induction is largely responsible for the impaired NE differentiation observed in TKO hESCs. The failure of *PAX6* induction could be due to hypermethylation of the *PAX6* P0 promoter. Alternatively, it could result from other causes, such as the failure to induce upstream transcription factors that activate *PAX6* expression or perhaps a failure of exiting the pluripotency network. Thus we investigated whether reversing hypermethylation specifically at the *PAX6* P0 promoter could rescue *PAX6* induction during NE differentiation. We developed a targeted demethylation strategy by first fusing the TET1 catalytic domain (TET1CD) to a nuclease "dead" Cas9 (dCas9) (Figure 2.20A) and expressing this fusion protein in TKO hESCs. The procedure to generate TKO hESCs expressing the dCas9-TET1CD fusion are described in detail in the Methods section (Figure 2.19A-E). With the appropriate gRNAs this dCas9-TET1CD fusion protein can be recruited to target genomic sites where the TET1 catalytic domain can oxidize 5mC to 5hmC and induce DNA demethylation as shown in recent studies using similar strategies (Choudhury et al., 2016; Liu et al., 2016; Morita et al., 2016; Xu et al., 2016). We first tested 12 gRNAs that target the *TET1* and 5hmC peak present at the *PAX6* P0 promoter (Figure 2.20B) for their ability to induce indel mutations (Figure 2.20C). This was used as a way to screen effective gRNAs that we would then test using the dcas9-TET1CD fusion for their ability to induce DNA demethylation. We identified 7 gRNAs that were the most efficient at producing indel mutations and these gRNAs were transiently transfected into the TKO hESCs expressing the dCas9-TET1CD fusion to test their efficiency for targeted demethylation (Figure 2.20D). Bisulfite sequencing (Figure 2.20E) and 5mC MassArray (Figure 2.21A)

identified 3 gRNAs, Cr6, Cr7 and Cr9, that produced a significant decrease in methylation at the *PAX6* P0 promoter in TKO hESCs compared to the non-transfected controls.

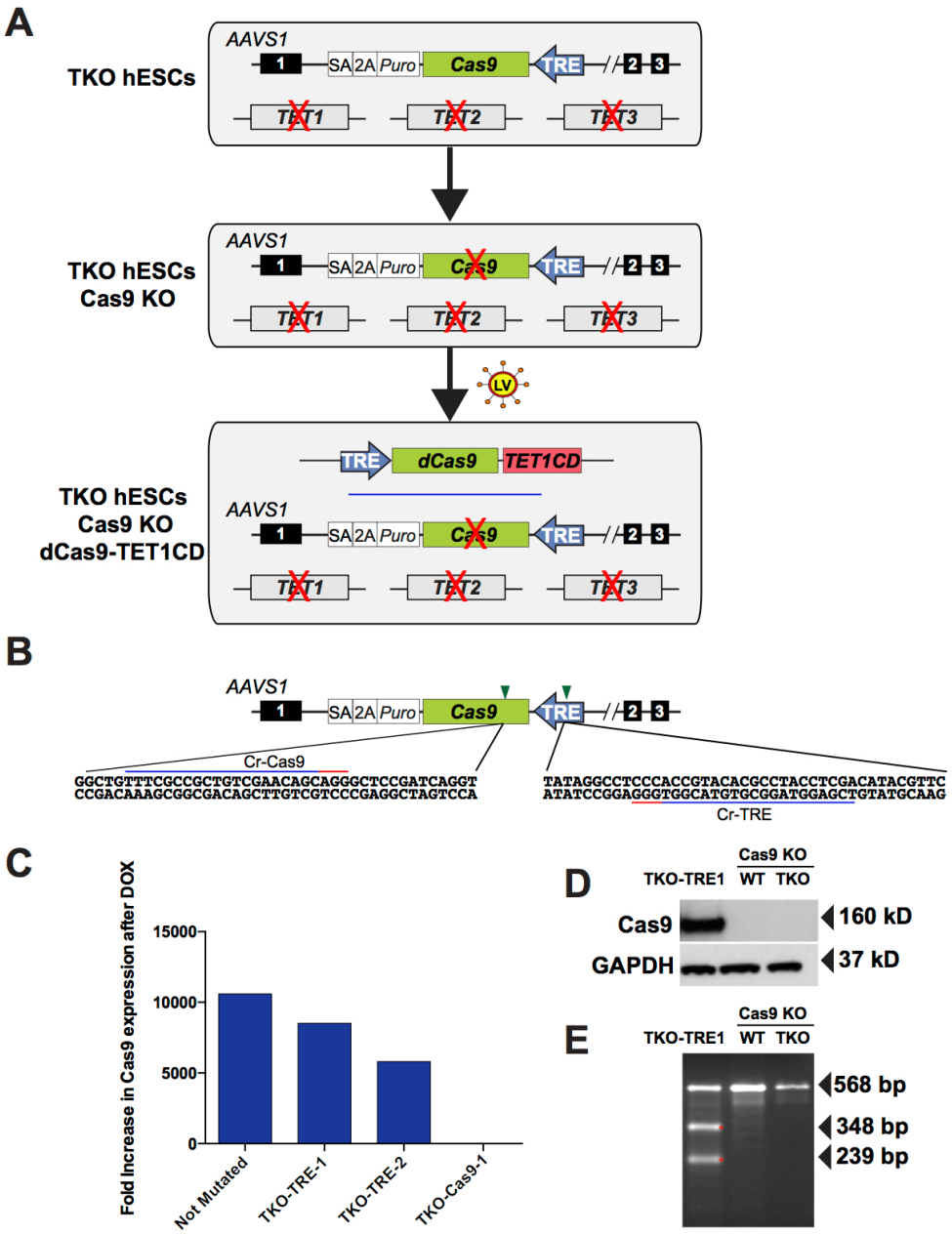



**Figure 2.19. Targeting dCas9-TET1CD into TKO hESCs.** (A) Schema to generate TKO hESCs that have doxycycline inducible expression of dCas-TET1CD. Because TKO hESCs were generated from the iCRISPR system it was necessary to first knock out the targeted Cas9 protein. After this Cas9 was inactivated the TKO hESCs were targeted with a lentivirus that contains a construct for doxycycline inducible expression of dCas9-TET1CD. (B) Schema to inactivate Cas9 in TKO hESCs. One CRISPR gRNA targeted the TRE promoter and the other targeted the beginning of the Cas9 protein. (C) Expression of Cas9 in TKO clones that had mutations in either the TRE promoter (TKO-TRE-1 and TKO-TRE-2) or the beginning of the Cas9 protein (TKO-Cas9-1). (D) Western blot of Cas9 in TKO clones that had mutations in either the TRE promoter (TKO-TRE-1) or the beginning of the Cas9 protein (Cas9 KO- TKO). (E) T7 endonuclease assay after transfection of TET2 gRNA in TKO clones that had mutations in either the TRE promoter (TKO-TRE-1) or the beginning of the Cas9 protein (Cas9 KO-TKO).

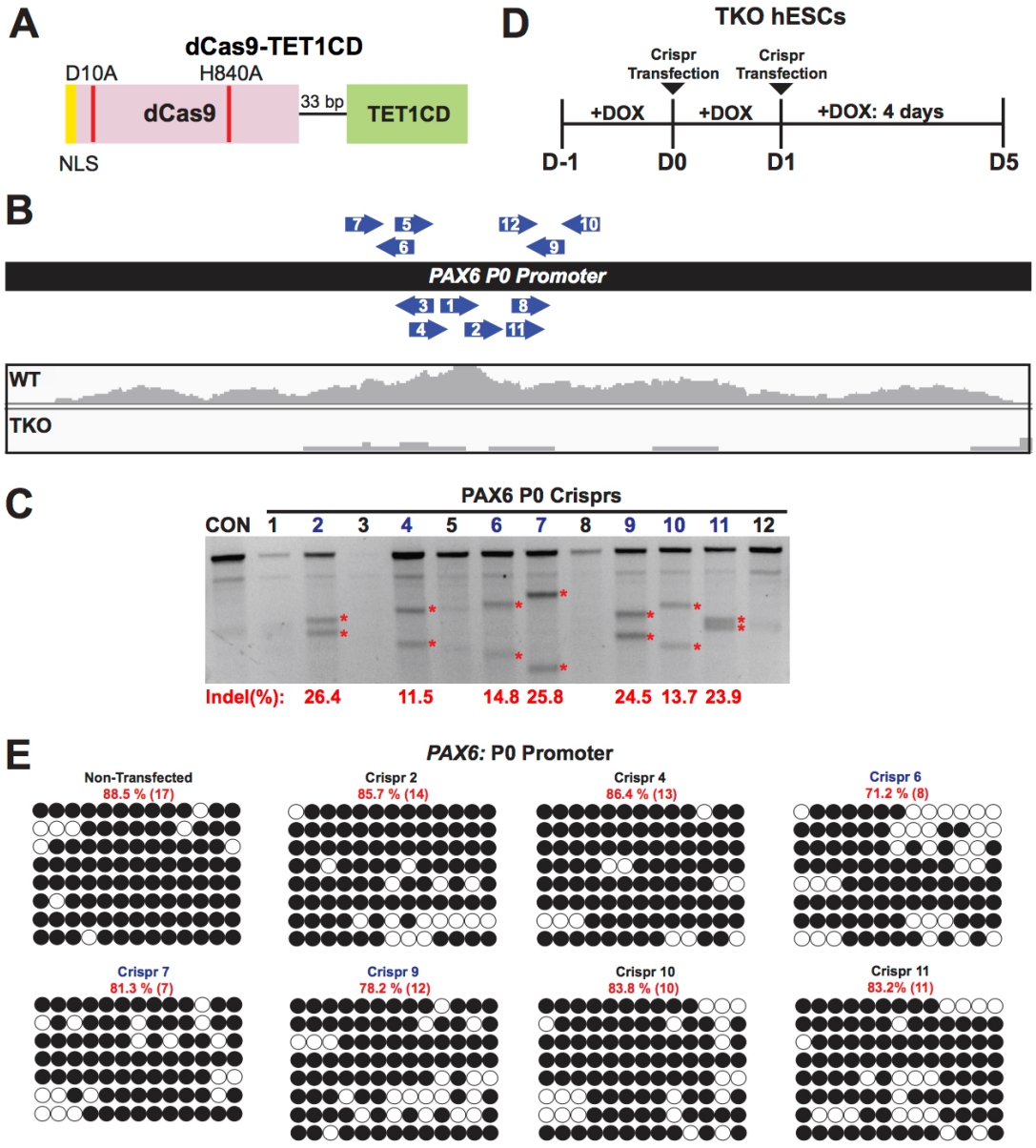
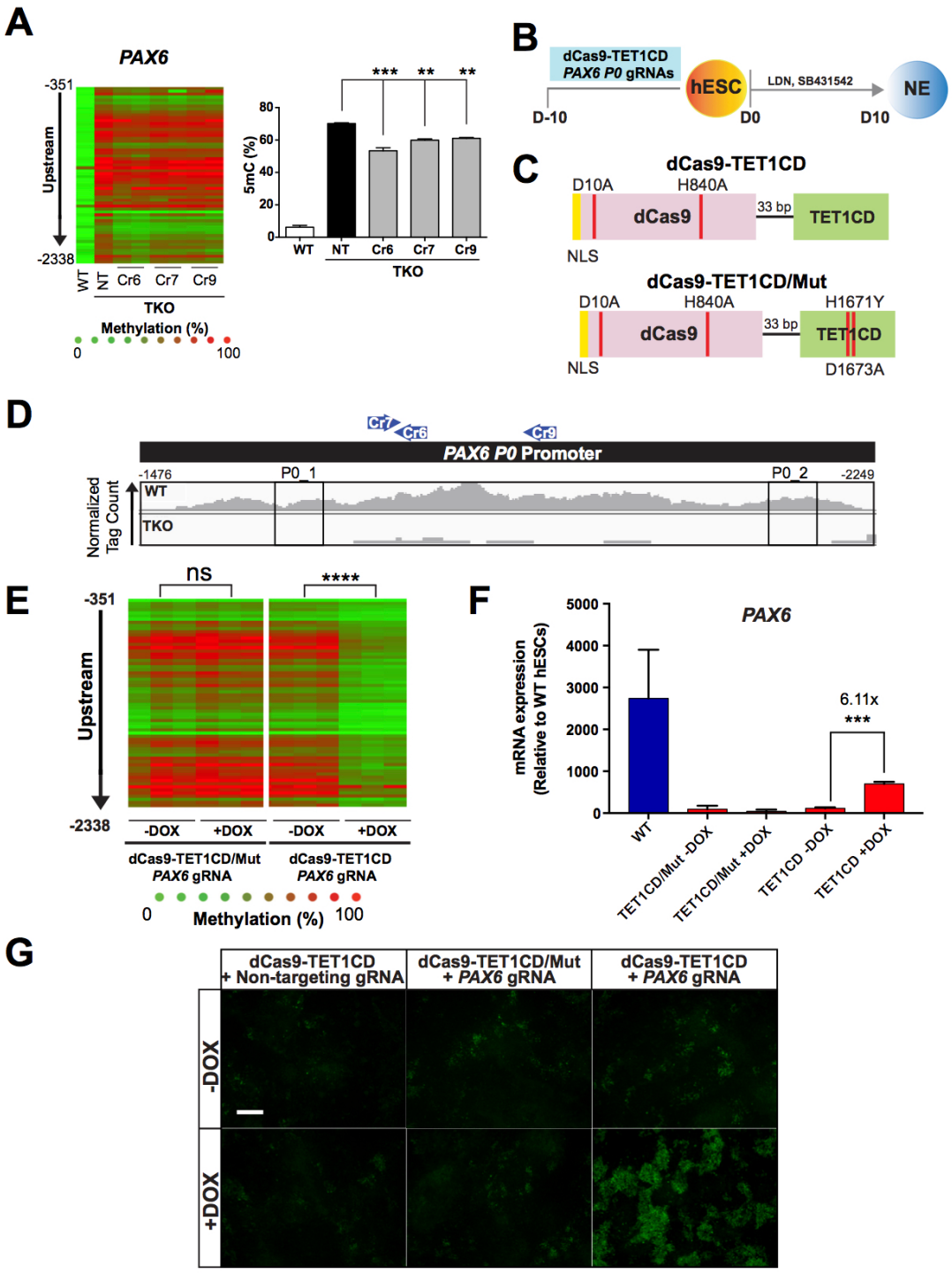



**Figure 2.20. Targeted demethylation of the *PAX6* P0 promoter using dCas9-TET1CD.** (A) Diagram of the dCas9 TET1 catalytic domain fusion protein (dCas9-TET1CD). (B) Targeting regions for CRISPR gRNAs tested for *PAX6* P0 promoter demethylation. The position of each gRNA in relation to the 5hmC peak is shown. The arrowhead indicates the first base pair of the gRNA and the base of the arrow indicates the 20$^{th}$ base pair that is adjacent to the PAM sequence. (C) T7 Endonuclease to test the efficiency of each CRISPR gRNA to generate indel mutations. Only gRNAs that could efficiently produce indel mutations were screened for targeted demethylation using the dCas9-TET1CD line. (D) Schema for testing the efficiency of targeted demethylation using the dCas9-TET1CD line. Genomic DNA was isolated on day 5 (D5) for methylation analysis by bisulfite sequencing. (E) Methylation at the PAX6 P0 promoter using bisulfite sequencing. Numbers in red indicate the average methylation in the PCR fragment, the number in parenthesis indicates the total number of sequences analyzed. CRISPR gRNAs written in blue were used further for targeted demethylation in TKO dCas9-TET1CD hESCs.

To enable persistent demethylation, we used a lentiviral vector to constitutively express the Cr6, Cr7 and Cr9 gRNAs in TKO hESCs with the doxycycline inducible dCas9-TET1CD fusion protein. Our strategy was to treat TKO hESCs with doxycycline prior to NE differentiation in order to express the TET1CD fusion protein and produce targeted demethylation at the PAX6 P0 promoter (Figure 2.21B). A number of controls were performed, including a no doxycycline control, a non-targeting gRNA control, and a control using dCas9-TET1CD mutated fusion protein (dCas9-TET1CD/Mut), in which the TET1 catalytic domain has been inactivated by targeted mutagenesis (Figure 2.21C). After 10 days of doxycycline treatment we performed 5mC MassArray analysis at the *PAX6* P0 promoter and observed demethylation at the *PAX6* P0 promoter in TKO hESCs that expressed both the dCas9-TET1CD and the 3 *PAX6* P0 gRNAs but not in the mutagenized dCas9-TET1CD/Mut control (Figure 2.21D-E). NE differentiation was then performed without further doxycycline treatment. dCas9-TET1CD TKO hESCs expressing *PAX6* P0 gRNAs and treated with doxycycline prior to NE differentiation showed upregulation of *PAX6* expression after NE differentiation, while the controls did not (Figure 2.21F-G). Thus the *PAX6* promoter hypermethylation observed in undifferentiated TKO hESCs is directly responsible for the failure of *PAX6* induction following NE differentiation.

**Figure 2.21 Increased methylation at *PAX6* P0 bivalent promoter leads to a failure of *PAX6* induction in TKO hESCs undergoing neuroectoderm differentiation.** (A) Heat map of MassArray analysis of 5mC in non-transfected TKO dCas9-TET1CD hESCs (NT) and TKO dCas9-TET1CD hESCs transfected with gRNAs targeting the *PAX6* P0 promoter (Cr6, Cr7, Cr9). Each row of the heat map represents either an individual CpG or a few CpGs that are located close together. The location of the CpGs with respect to the TSS is shown to the left of the heat map. The binding site of gRNA Cr6 is shown by the arrowhead on the left side of the heatmap. The graph on the right shows the quantification of methylation in the region depicted in the heat map. For significance tests all comparisons are to the non-transfected control (NT). *n*= 2 independent experiments. (B) Schema for rescue of the NE differentiation defect in TKO hESCs using targeted demethylation of the *PAX6* P0 promoter. (C) Diagram of the dCas9 TET1 catalytic domain fusion protein (dCas9-TET1CD) and the dCas9 TET1 catalytic domain fusion protein in which the TET1 catalytic domain has been mutated (dCas9-TET1CD/Mut). (D) Diagram of the gRNAs (shown as blue arrows) designed to target the *PAX6* P0 promoter in the region surrounding the 5hmC peak found in WT hESCs. Arrowhead indicates the 5' end of the targeting gRNA. Regions previously analyzed for TET1 binding by ChIP-qPCR are enclosed by black rectangles. (E) Heat map of MassArray analysis of 5mC at the *PAX6* P0 promoter for TKO hESCs that express *PAX6* targeting gRNAs with either a dCas9-TET1CD/Mut (left) or a dCas9-TET1CD (right) fusion protein. Methylation analysis at the *PAX6* P0 promoter was analyzed for these cell lines with and without doxycycline treatment. *n*= 3 independent experiments. (F) qPCR of *PAX6* expression on D10 of NE differentiation for TKO hESCs that express *PAX6* targeting gRNAs with either a dCas9-TET1CD/Mut or a dCas9-TET1CD fusion protein. *PAX6* expression was analyzed for these cell lines with and without doxycycline treatment prior to differentiation. *n*= 3 independent experiments. (G) Immunofluorescence of PAX6 on D10 of NE differentiation for TKO hESCs that express *PAX6* targeting gRNAs with either a dCas9-TET1CD/Mut or a dCas9-TET1CD fusion protein. Also TKO hESCs expressing the dCas9-TET1CD fusion and a non-targeting gRNA were used as a control. PAX6 immunofluorescence was analyzed for these cell lines with and without doxycycline treatment prior to differentiation. *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 *****P*<0.0001 by one-way ANOVA (A) or student's t test (E, F).


A
PAX6
-351
Upstream
-2338
WT
NT
Cr6
Cr7
Cr9
TKO
Methylation (%)
0
100
5mC (%)
80
60
40
20
0
***
**
**
WT
NT
Cr6
Cr7
Cr9
TKO
B
dCas9-TET1CD
PAX6 P0 gRNAs
hESC
LDN, SB431542
NE
D-10
D0
D10
C
dCas9-TET1CD
D10A
H840A
dCas9
33 bp
TET1CD
NLS
dCas9-TET1CD/Mut
D10A
H840A
H1671Y
dCas9
33 bp
TET1CD
NLS
D1673A
D
Cr7
Cr6
Cr9
PAX6 P0 Promoter
-1476
P0_1
P0_2
-2249
Normalized Tag Count
WT
TKO
E
ns
****
-351
Upstream
-2338
-DOX
+DOX
-DOX
+DOX
dCas9-TET1CD/Mut
PAX6 gRNA
dCas9-TET1CD
PAX6 gRNA
0
Methylation (%)
100
F
PAX6
mRNA expression
(Relative to WT hESCs)
5000
4000
3000
2000
1000
0
6.11x
***
WT
TET1CD/Mut -DOX
TET1CD/Mut +DOX
TET1CD -DOX
TET1CD +DOX
G
dCas9-TET1CD
+ Non-targeting gRNA
dCas9-TET1CD/Mut
+ PAX6 gRNA
dCas9-TET1CD
+ PAX6 gRNA
-DOX
+DOX

We further tested dCas9-TET1CD mediated targeted demethylation on the hypermethylated bivalent promoter of *SOX10*, and the hypermethylated enhancer of *LEFTY2* in TKO hESCs. We found that targeted demethylation of the hypermethylated *LEFTY2* enhancer was able to increase expression of *LEFTY2* in TKO hESCs. In contrast targeted demethylation of the hypermethylated *SOX10* bivalent promoter had no effect on *SOX1*0 expression at the hESCs stage but increased *SOX10* expression after differentiation, as seen with *PAX6* (Figure 2.22A). These results support the conclusion that hypermethylation of a bivalent promoter impairs gene expression upon differentiation. Although the targeted demethylation induces significant increases in gene expression to similar degrees as reported in previous studies (Choudhury et al., 2016; Liu et al., 2016; Morita et al., 2016; Xu et al., 2016), the expression did not reach WT levels. This may be due to incomplete demethylation. Additional chromatin changes at the target locus or elsewhere in TKO hESCs could also affect the target gene expression directly or indirectly.

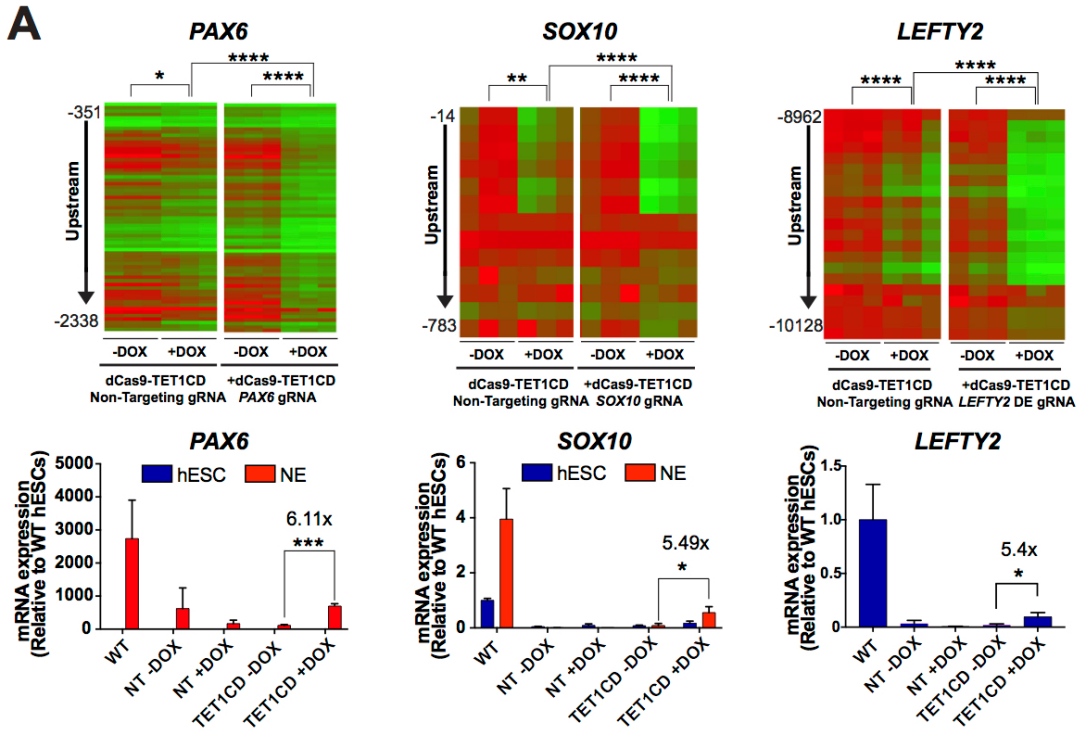



**Figure 2.22. Targeted demethylation of *PAX6* and *SOX10* promoters and LEFTY2 enhancer.** (A) Targeted demethylation at the bivalent promoters of *PAX6* (left) and *SOX10* (middle) and the enhancer of *LEFTY2* (right). For each locus the dCas9-TET1CD expressing TKO hESCs also expressed a non-targeting gRNA (NT) as a control. Top panel: Heat map of MassArray analysis of 5mC at bivalent promoters of *PAX6* and *SOX10* bivalent promoters and the enhancer of *LEFTY2* after dCas9-TET1CD mediated demethylation either with non-targeting or targeting gRNAs. –DOX indicates that the cell lines were not treated with doxycycline and thus do not express dCas9-TET1CD. Each row of the heat map represents either an individual CpG or a few CpGs that are located close together. The location of the CpGs with respect to the TSS is shown to the left of each heat map. For each cell line and condition three biological replicates are shown as three columns, *n*= 3 independent experiments. Bottom panel: qPCR analysis of targeted gene expression either in hESCs (blue) or after NE differentiation (red), *n*= 3 independent experiments. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001 *****P*<0.0001 by student's t test.

### *2.2.5 De novo methylation causes PAX6 promoter hypermethylation*

The DNA methyltransferases, DNMT1, DNMT3A and DNMT3B, are responsible for cytosine methylation. Thus we speculated that the hypermethylation of the *PAX6* P0 promoter and the resulting NE differentiation defect could be due to increased expression or activity of the DNMTs in TKO hESCs. We first examined the effects of DNMT inhibitor 5-Azadeoxycytidine (5-Aza) treatment on TKO hESCs (Figure 2.23A). Although we were limited to one day of 5-Aza treatment due to its cytotoxicity, TKO cells exposed to 10 nM 5-Aza prior to NE differentiation expressed NE marker *SOX1* and downregulated pluripotency markers *OCT4* and *NANOG* upon NE differentiation. However, no induction of *PAX6*, *FOXG1* or *SOX10* was observed (Figure 2.23B-C). We further confirmed that there was no significant increase in PAX6 positive cells by FACS analysis (Figure 2.24A). We believe this is because the limited 5-Aza dosage and exposure time is not sufficient to decrease methylation enough at loci such as *PAX6*, to allow transcriptional activation. This is supported by bisulfite sequencing of the *PAX6* P0 promoter that did not show a significant decrease in methylation (Figure 2.24B). Interestingly WT cells showed a decrease in efficiency of NE differentiation at the highest 5-Aza dosage, possible due to toxicity effects (Figure 2.23B-C).

Encouraged by these findings, we sought to determine which DNMT protein is responsible for the hypermethylation at the *PAX6* P0 promoter in TKO hESCs. We speculated that the hypermethylation of the *PAX6* P0 promoter and the resulting NE differentiation defect could be due to increased expression or activity of the DNMTs in TKO hESCs. There were no differences in the expression of the *DNMT* genes between WT and TKO

hESCs (Figure 2.25A); however, ChIP-qPCR analysis showed increased binding of DNMT3B (Figure 2.25B), but not DNMT1 or DNMT3A (Figure 2.25C), at the *PAX6* P0 promoter in TKO hESCs compared to WT hESCs.

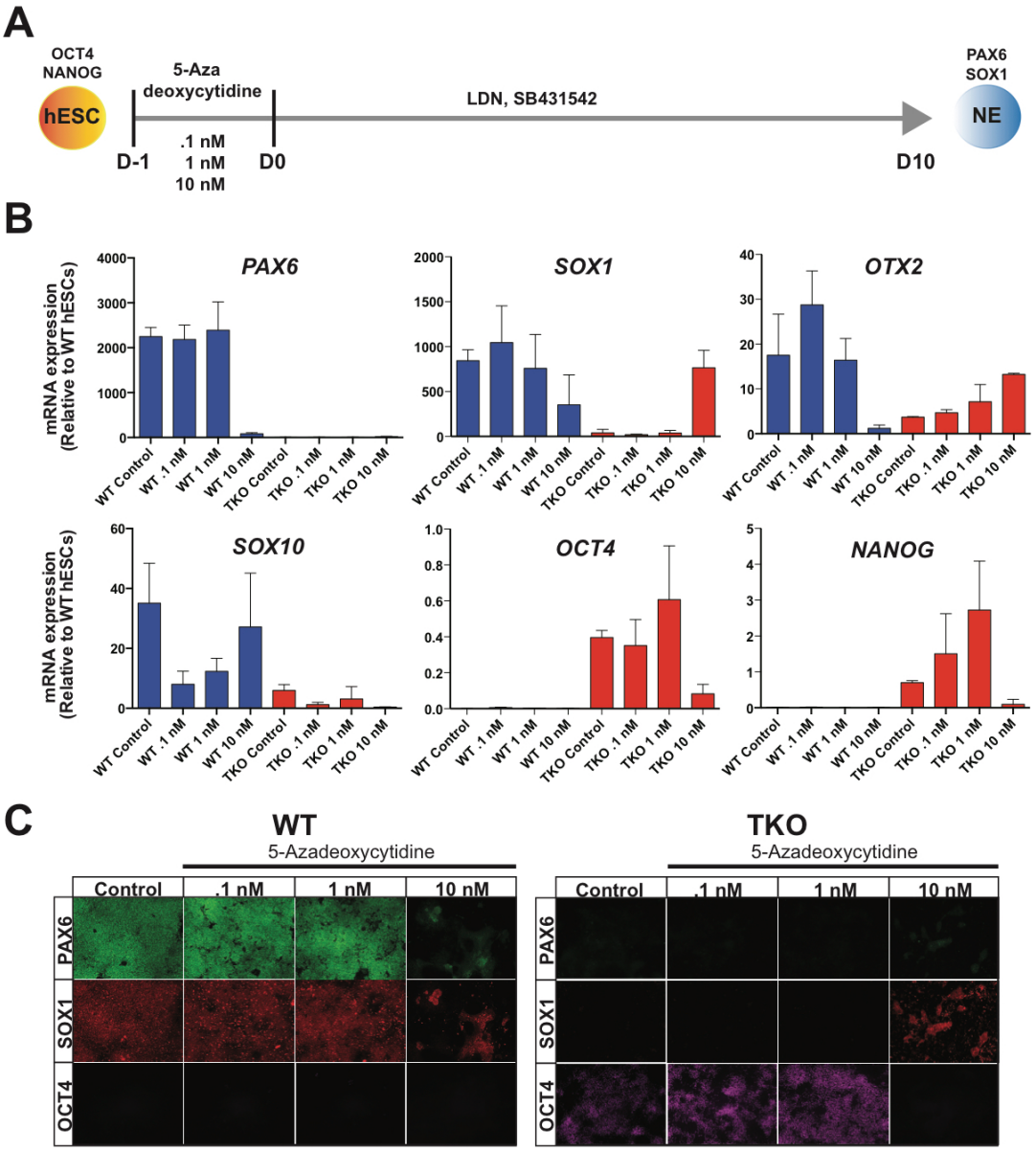



**Figure 2.23 Inhibition of DNMT by 5-Azadeoxycytidine partly rescues NE differentiation of TKO hESCs.** (A) Schema for rescue of the NE differentiation defect in TKO hESCs using 5-Azadeoxycytidine (5-Aza) treatment. (B) Expression analysis at D10 of NE differentiation for neuroectoderm (*PAX6*, *SOX1*, *OTX2*), neural crest (*SOX10*) and pluripotency (*OCT4*, *NANOG*) markers. WT and TKO hESCs were treated with different concentrations of 5-Aza prior to NE differentiation. (C) Immunofluorescence of PAX6, SOX1 and OCT4 at D10 of NE differentiation. WT and TKO hESCs were treated with different concentrations of 5-Aza prior to NE differentiation. *n*= 3 independent experiments. Data are mean ± STD.

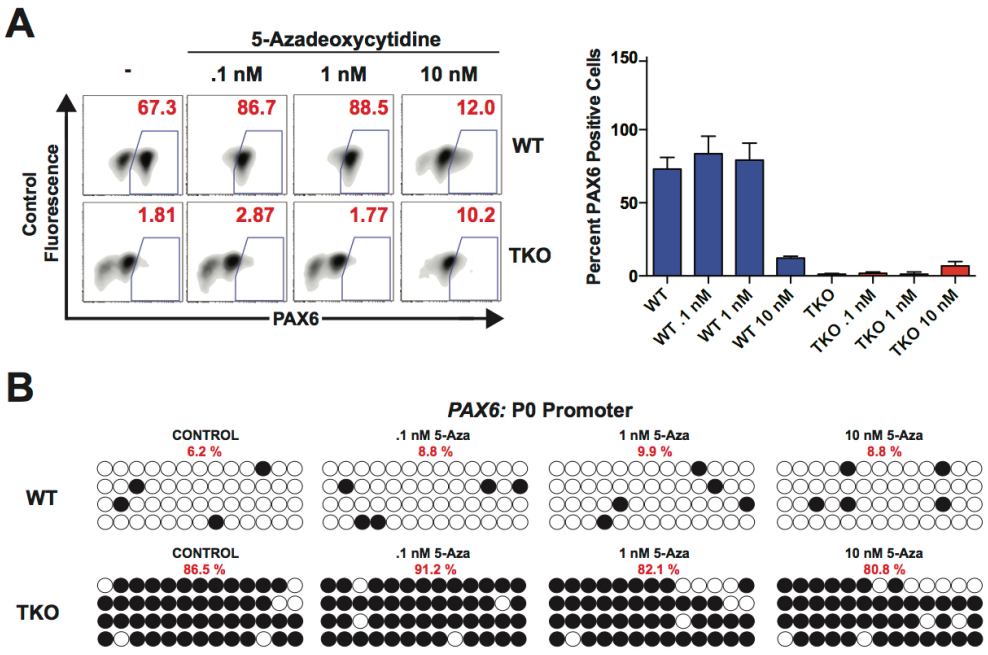



**Figure 2.24. *PAX6* expression and methylation at the P0 promoter after 5-Azadeoxycytidine treatment.** (A) Left: Representative FACS plots of PAX6 positive cells after treatment of WT and TKO hESCs with variable amounts of 5-Azadeoxycytidine (5-Aza) prior to NE differentiation. Right: Quantification of PAX6 positive cells after treatment of WT and TKO hESCs with variable amounts of 5-Aza prior to NE differentiation. *n*= 3 independent experiments. (B) Methylation analysis of PAX6 P0 promoter by bisulfite sequencing after treatment with variable amounts of 5-Aza. The 4 sequences that were analyzed are shown.

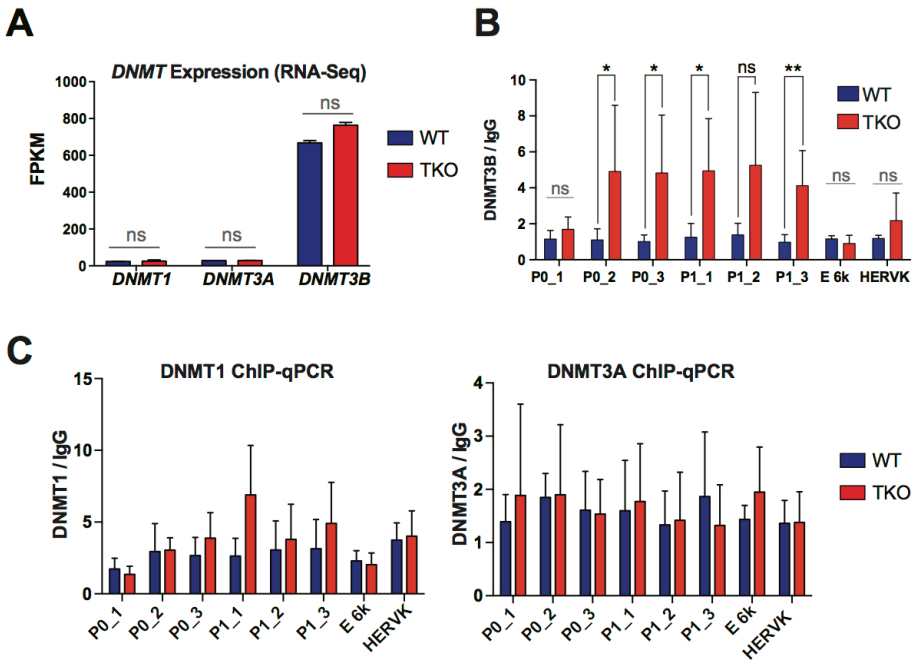
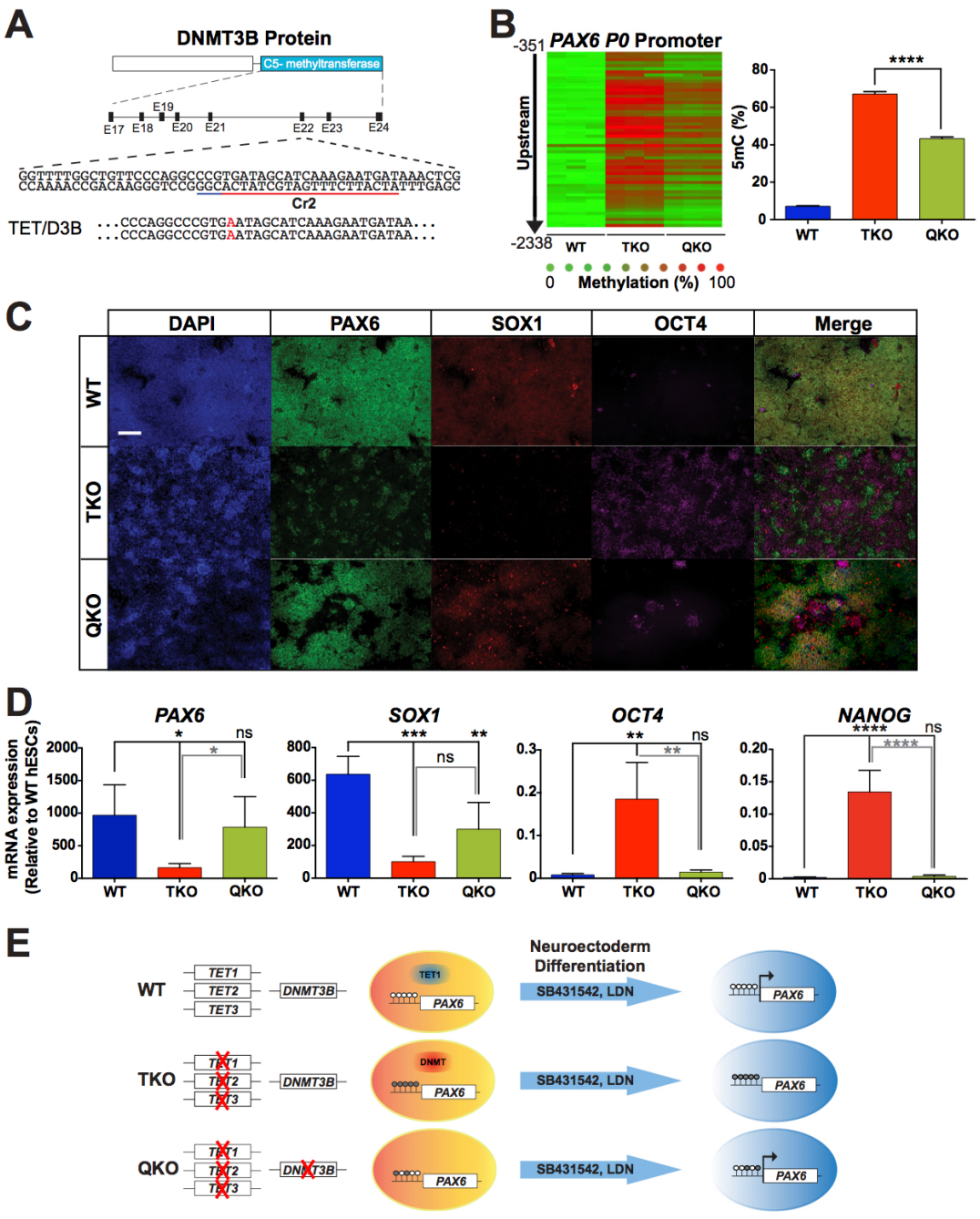



**Figure 2.25. Targeting DNMT proteins rescues NE differentiation.** (A) Expression of *DNMT1*, *DNMT3A* and *DNMT3B* in WT and TKO hESCs by RNA-Seq analysis. *n*= 2 independent experiments. (B) ChIP-qPCR for DNMT1 (left) and DNMT3A (right) at the *PAX6* locus in WT and TKO hESCs. Primers are the same ones as used for Fig. 2f. *n*= 3 independent experiments. (C) Schema for *DNMT3B* targeting in TKO hESCs. The CRISPR was designed to target the C5-methyltransferase enzymatic domain. The red line indicates the gRNA targeting sequence and the blue line is the adjacent PAM sequence. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by Student's t test (A, B).

Thus we decided to further investigate whether DNMT3B is responsible for the hypermethylation at the *PAX6* P0 promoter in TKO hESCs through genetic deletion. By applying a CRISPR gRNA that targets the sequence corresponding to the cytosine C5-methyltransferase domain of DNMT3B (Figure 2.26A) in TKO hESCs, we generated a quadruple *TET/DNMT3B* knockout (QKO) line. QKO hESCs had a ~35% reduction in the methylation at the *PAX6* P0 promoter compared to TKO hESCs (Figure 2.26B). Furthermore, there was a significant rescue of the NE differentiation phenotype compared to passage matched TKO hESCs: QKO cells formed PAX6 and SOX1 double positive cells (Figure 2.26C) and few cells remained OCT4-positive after 10 days of NE differentiation. RT-qPCR analysis also showed rescue of NE markers *PAX6* and *SOX1* and proper downregulation of pluripotency markers *OCT4* and *NANOG* (Figure 2.26D).

Our results suggest that DNMT3B activity at the *PAX6* P0 promoter is responsible for the hypermethylation and NE differentiation phenotypes observed in TKO hESCs. In WT hESCs the TET proteins or the resulting 5hmC marks actively block *de novo* methylation by DNMT3B to maintain a hypomethylated *PAX6* P0 promoter and enable NE differentiation (Figure 2.26E).

**Figure 2.26 Genetic inactivation of *DNMT3B* can rescue the neuroectoderm differentiation defect of TKO hESCs.** (A) ChIP-qPCR for DNMT3B at the *PAX6* locus in WT and TKO hESCs. *n*= 3 independent experiments. (B) Heat map of MassArray analysis of 5mC at the *PAX6* P0 promoter in WT, TKO and QKO hESCs. WT and TKO hESCs are passage-matched with the QKO hESCs. Each row of the heat map represents either an individual CpG or a few CpGs that are located close together. The location of the CpGs with respect to the TSS is shown to the left of the heat map. For each cell line three biological replicates are shown as three columns. Quantification of the percent methylation is shown on the right. *n*= 3 independent experiments. (C) Immunofluorescence of PAX6, SOX1 and OCT4 at D10 of NE differentiation in WT, TKO and QKO cells. (D) qPCR analysis of neuroectoderm (*PAX6* and *SOX1*) and pluripotency (*OCT4* and *NANOG*) markers in WT, TKO and QKO cells at D10 of differentiation. For significance tests black lines indicate comparison with WT. *n*= 3 independent experiments. (E) Model for TET and DNMT3B competition to regulate methylation at bivalent promoters. In WT hESCs TET1 or 5hmC functions to antagonize DNMT3B at the *PAX6* bivalent promoter and prevents DNA hypermethylation. Upon NE differentiation *PAX6* expression is activated and leads to the production of PAX6 and SOX1 double positive cells that are negative for the pluripotency marker OCT4. In TKO hESCs increased DNMT3B binding at the *PAX6* bivalent promoter leads to increased DNA methylation. As a result *PAX6* expression is not activated and ultimately few PAX6 or SOX1-positive cells are produced and a large number of cells still express the pluripotency marker OCT4. Data are mean ± STD. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (D) and by Student's t test (A, B).


A
DNMT3B Protein
C5- methyltransferase
E19
E17 E18 E20 E21 E22 E23 E24
GGTTTTTGGCTGTTCCCAGGCCCGTGATAGCATCAAAGAATGATAAACTCG
CCAAAACCGACAAGGGTCCGGGCACTATCGTAGTTTCTTACTATTTGAGC
Cr2
TET/D3B ...CCCAGGCCCGTGAATAGCATCAAAGAATGATAA...
...CCCAGGCCCGTGAATAGCATCAAAGAATGATAA...
B
PAX6 P0 Promoter
-351
Upstream
-2338
WT TKO QKO
0 Methylation (%) 100
5mC (%)
80 60 40 20 0
****
WT TKO QKO
C
DAPI PAX6 SOX1 OCT4 Merge
WT
TKO
QKO
D
mRNA expression (Relative to WT hESCs)
PAX6
SOX1
OCT4
NANOG
WT TKO QKO
ns
E
WT
TKO
QKO
TET1
TET2
TET3
DNMT3B
PAX6
DNMT
Neuroectoderm Differentiation
SB431542, LDN

We performed ERRBS for passage-matched WT, TKO and QKO lines to investigate whether our findings at the *PAX6* P0 promoter applies to other bivalent promoters (computational analysis was performed by Heng Pan). We found that QKO hESCs showed a global decrease in methylation, both within and outside of CpG islands, and in all promoter types. Bivalent promoters showed a larger decrease in methylation between QKO and TKO hESCs than active and initiated promoter types, but less than silent promoters (Figure 2.27A-C). Prior studies have suggested that DNMT3A and DNMT3B have redundant sites of action. We investigated *DNMT3A* expression in QKO hESCs to determine whether its expression may increase in order to compensate for loss of *DNMT3B*; however, we did not find a significant change in its expression (Figure 2.27D). Bivalent previously gained methylation in the TKO hESCs (compared to WT hESCs). Conversely, ~57% of the bivalent promoters that gained methylation after TET inactivation lost methylation after *DNMT3B* was mutated (Figure 2.28C). Thus continuous DNMT3B activity is needed for the hypermethylation phenotype in TKO hESCs.

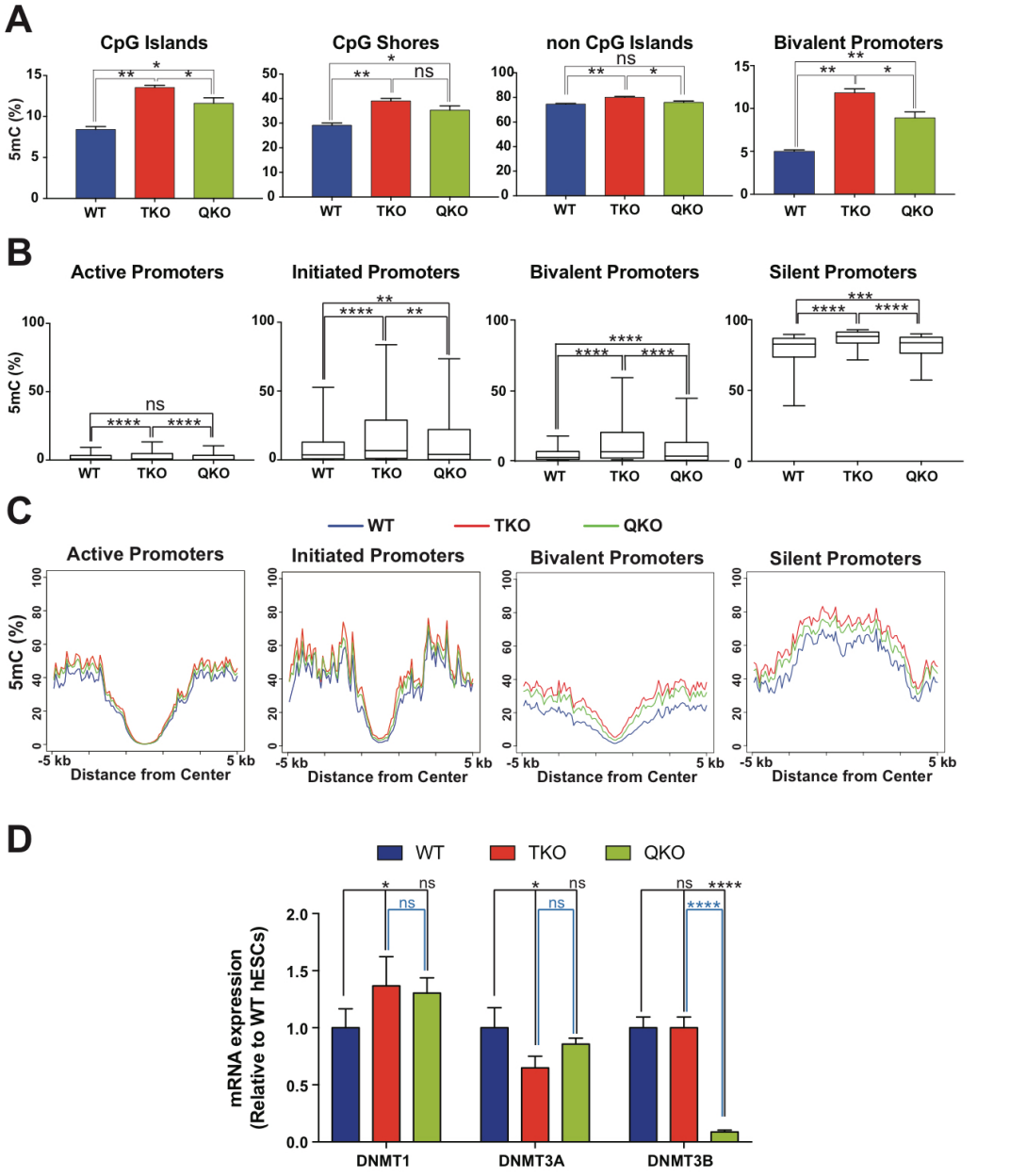
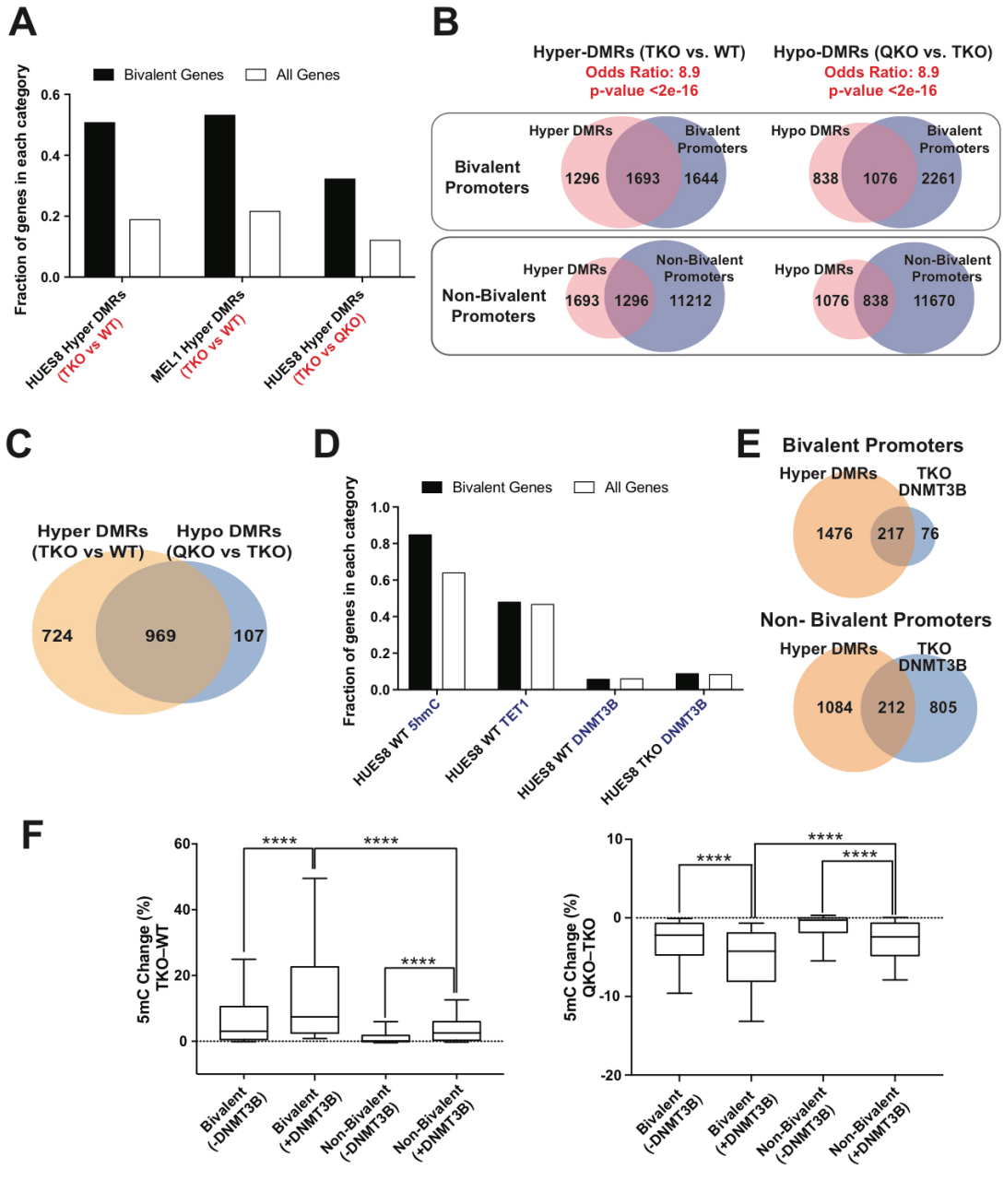



**Figure 2.27 DNMT3B inactivation reduces methylation at bivalent promoters in TKO hESCs.** (A) Average methylation at different genomic regions and bivalent promoters for WT, TKO and QKO hESCs by enhanced reduced representation bisulfite sequencing (ERRBS). Data are mean ± STD. *n* = 2 independent experiments. (B) Percent methylation in WT, TKO and QKO hESCs for active, initiated, bivalent and silent promoters. Error bars show 10 and 90 percent confidence intervals and the bar at the center of the box and whisker plot indicates the median. *n* = 2 independent experiments. (C) Percent methylation in WT, TKO and QKO hESCs in 10kb region surrounding active, initiated, bivalent and silent promoters. (D) *DNMT1*, *DNMT3A*, *DNMT3B* expression in WT, TKO and QKO hESCs by RT-qPCR. *n* = 3 independent experiments. ns, not significant. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA (A, B, D).

Our results thus far indicate that DNMT3B is responsible for the majority of bivalent promoter hypermethylation that occurs after TET inactivation. We performed DNMT3B ChIP-Seq to identify bivalent promoters directly targeted by DNMT3B (computational analysis was performed by Heng Pan). DNMT3B shows relatively insignificant overall binding to promoter regions (Figure 2.28D) in WT and TKO hESCs, which is similar to previous results of DNMT3B ChIP-Seq in WT mESCs (Baubec et al., 2015). This may be due to weak DNMT3B binding, or technical difficulties of DNMT3B ChIP-Seq. Nevertheless, among promoters with DNMT3B peaks in TKO hESCs, 74% of the 293 bivalent promoters and 21% of the 1,017 non-bivalent promoters gained methylation after TET inactivation (Figure 2.28E). Furthermore bivalent promoters with DNMT3B peaks in TKO hESCs showed a greater methylation increase (TKO versus WT) and a greater methylation decrease (QKO versus TKO) than bivalent promoters that lacked DNMT3B peaks in TKO hESCs or non-bivalent promoters with DNMT3B peaks in TKO hESCs (Figure 2.28F). Thus DNMT3B binding correlates with more dynamic changes in DNA methylation at bivalent promoters and suggests that at these promoters the TET and DNMT3B proteins function in a competitive manner.

**Figure 2.28 DNMT3B regulates methylation level at bivalent promoters.** (A) Fraction of bivalent genes and total genes that overlap with hyper-DMRs (HUES8 and MEL1 TKO versus WT hESCs) and hypo-DMRs (HUES8 QKO versus TKO hESCs). Hyper-DMRs and hypo-DMRs are from ERRBS datasets. (B) Top: Overlap of bivalent promoters with hyper-DMRs at promoter regions (TKO vs. WT) and hypo-DMRs at promoter regions (QKO vs. TKO). Bottom: Overlap of non-bivalent promoters with hyper-DMRs at promoter regions (TKO vs. WT) and hypo-DMRs at promoter regions (QKO vs. TKO). The odd's ratio and p-value for a comparison between bivalent and non-bivalent promoters is provided (Fisher's exact test). (C) The overlap between hyper-DMR (TKO vs. WT) associated bivalent promoters and hypo-DMR (QKO vs. TKO) associated bivalent promoters. (D) Fraction of bivalent genes and total genes that overlap with hyper-DMRs (HUES8 and MEL1 TKO versus WT hESCs) and hypo-DMRs (HUES8 QKO versus TKO hESCs). Hyper-DMRs and hypo-DMRs are from ERRBS datasets. (E) Overlap between DNMT3B peaks in TKO hESCs and hyper-DMRs at bivalent promoters (top) and hyper-DMRs at non-bivalent promoters (bottom). (F) Left: Methylation change (TKO – WT) for bivalent and non-bivalent promoters that either have DNMT3B peaks in TKO hESCs (+DNMT3B) or do not have DNMT3B peaks in TKO hESCs (-DNMT3B). Right: Methylation change (QKO – TKO) for bivalent and non-bivalent promoters that either have DNMT3B peaks in TKO hESCs (+DNMT3B) or do not have DNMT3B peaks in TKO hESCs (-DNMT3B). Error bars show 10 and 90 percent confidence intervals and the bar at the center of the box and whisker plot indicates the median. ns, not significant. *$P$<0.05, **$P$<0.01, ***$P$<0.001, ****$P$<0.0001 by one-way ANOVA (F).


A
Bivalent Genes
All Genes
Fraction of genes in each category
0.6
0.4
0.2
0.0
HUES8 Hyper DMRs (TKO vs WT)
MEL1 Hyper DMRs (TKO vs WT)
HUES8 Hyper DMRs (TKO vs QKO)
B
Hyper-DMRs (TKO vs. WT)
Odds Ratio: 8.9
p-value <2e-16
Hypo-DMRs (QKO vs. TKO)
Odds Ratio: 8.9
p-value <2e-16
Bivalent Promoters
Hyper DMRs
Bivalent Promoters
1296
1693
1644
Hypo DMRs
Bivalent Promoters
838
1076
2261
Non-Bivalent Promoters
Hyper DMRs
Non-Bivalent Promoters
1693
1296
11212
Hypo DMRs
Non-Bivalent Promoters
1076
838
11670
C
Hyper DMRs (TKO vs WT)
Hypo DMRs (QKO vs TKO)
724
969
107
D
Bivalent Genes
All Genes
Fraction of genes in each category
1.0
0.8
0.6
0.4
0.2
0.0
HUES8 WT 5hmC
HUES8 WT TET1
HUES8 WT DNMT3B
HUES8 TKO DNMT3B
E
Bivalent Promoters
Hyper DMRs
TKO DNMT3B
1476
217
76
Non- Bivalent Promoters
Hyper DMRs
TKO DNMT3B
1084
212
805
F
5mC Change (%) TKO–WT
60
40
20
0
****
****
****
Bivalent (-DNMT3B)
Bivalent (+DNMT3B)
Non-Bivalent (-DNMT3B)
Non-Bivalent (+DNMT3B)
5mC Change (%) QKO–TKO
10
0
-10
-20
****
****
****
Bivalent (-DNMT3B)
Bivalent (+DNMT3B)
Non-Bivalent (-DNMT3B)
Non-Bivalent (+DNMT3B)

## 2.3 DISCUSSION

Compared to enhancers, the majority of bivalent promoters do not show dramatic changes in DNA methylation during development (Jones, 2012; Smith and Meissner, 2013). As a result it was unclear whether the methylation state of bivalent promoters is actively regulated and if it has functional relevance for cell differentiation. Previous studies have shown that the TET proteins regulate methylation at enhancers in developing mouse embryos and mESCs (Bogdanovic, 2017; Lu et al., 2014). Here we show that the TET proteins are also critical for maintaining a hypomethylated state at bivalent promoters in hESCs. In the absence of the TET proteins, aberrant hypermethylation of bivalent promoters leads to impaired gene activation upon exposure to differentiation signals. Importantly, the alteration of DNA methylation at bivalent promoters does not cause immediate changes in transcription but significantly impacts hESC differentiation. This suggests the need to revise the predominant approach that focuses on DNA methylation and other epigenetic changes with immediate impact on gene expression, as epigenetic changes may predict future cell behavior during embryonic development or adult stem cell differentiation. Broadly, the competitive balance between the TET proteins and de novo methyltransferases at bivalent promoters could facilitate rapid changes of their methylation state to either activate or silence transcription in a cell-lineage and gene dependent manner. Although we focused on lineage regulators, such as *PAX6*, bivalent promoters may also regulate signaling pathways during development. Furthermore, we observed hypermethylation at enhancer regions after *TET* inactivation. TET activity at a variety of regulatory regions may modulate the expression of signaling pathways and

lineage regulators, all of which contribute to proper embryonic development and cellular differentiation.

Previous studies have indicated that genomic regions marked by the H3K4me3 modification are refractory to *de novo* DNA methylation (Guo et al., 2014; Ooi et al., 2007). *TET* deletion causes bivalent promoter hypermethylation without causing apparent changes in H3K4me3 occupancy. Thus TET or additional TET-dependent factors (including 5hmC) that frequently co-occur with H3K4me3 may mediate some of the effects reported to be associated with the H3K4me3 mark. Also intriguingly, this aberrant hypermethylation, once established in TKO hESCs, is not adequately maintained by DNMT1. Upon inactivation of *DNMT3B* in the TKO background we observed a global reduction of DNA methylation at bivalent promoters. Based on the >99.7% overall fidelity of DNMT1 in preserving methylation (Liao et al., 2015), DNMT1 would be expected to largely preserve the hypermethylation seen in TKO hESCs during the approximately 6 passages it took to generate and expand the QKO cells for analysis. The significant reversal of the hypermethylation phenotype at *PAX6* and other bivalent promoters in QKO cells indicates the hypermethylation phenotype requires continuous DNMT3B activity. It also suggests that additional mechanisms, such as transient transcription (Di Ruscio et al., 2013), may inhibit DNMT1 activity at these loci. Alternatively, though not mutually exclusively, DNMT3B may function as both a *de novo* and maintenance methyltransferase at particular loci in ESCs, as previously suggested (Chen et al., 2003; Liang et al., 2002). Previous studies have shown that DNMT3A and DNMT3B share largely overlapping targets (Liao et al., 2015), yet deletion of *DNMT3B* alone in TKO hESCs was sufficient to

partially reverse bivalent promoter hypermethylation along with the associated NE differentiation defect. DNMT3B may have stronger preference or activity compared to DNMT3A at bivalent promoters such as *PAX6*, and such loci have been identified previously in hESCs (Liao et al., 2015). Perhaps more likely, the rescue by *DNMT3B* deletion alone could be due to the relatively low *DNMT3A* expression compared to *DNMT3B* (Supplementary Fig. 9a) and the lack of compensatory increase of *DNMT3A* expression upon *DNMT3B* deletion (Supplementary Fig. 11d). Similarly a recent paper of TKO mouse embryos found that inactivating either *DNMT3A* or *DNMT3B* was able to rescue the gastrulation phenotype (Dai et al., 2016). It is possible that in other cell types DNMT3A, along with or instead of DNMT3B, may counteract the TET proteins to regulate methylation (Rinaldi et al., 2016; Zhang et al., 2016).

Our work highlights the utility of locus-specific epigenome editing tools to directly probe the functional consequences of epigenetic changes and to distinguish direct, locus-specific effects from indirect effects. The observation of specific instead of global hypermethylation in TKO hESCs suggests that the antagonistic relationship between TET and DNMTs may be regulated through recruitment of TETs and/or DNMTs in a locus-specific manner. Such competitive interplay and potential cooperative interactions between TETs and DNMTs may also occur at later stages of differentiation as recently shown in the hematopoietic system (Zhang et al., 2016). The expression of the individual *TET* genes can vary at different stages of development (Rasmussen and Helin, 2016), and each TET protein may have unique targets so that distinct sets of promoters are maintained in a hypomethylated/5hmC rich state depending on the cellular context. Similarly

the recruitment or enzymatic activity of the DNMTs may be locus-specific or influenced by the genomic distribution of transcription factors and histone modifications, which can change dramatically between cell types. It would be of great interest to investigate what factors determine whether the methylation state of a particular regulatory region is dynamically regulated or is relatively static, how different factors cooperate or counteract in different genomic contexts to establish the unique methylation state of regulatory regions, and ultimately how methylation states not only regulate but also predict (in the context of cell differentiation) cell-type specific transcriptional programs.

## 2.4 MATERIALS & METHODS

### *2.4.1 hESC Culture*

HUES8 (NIHhESC-09-0021) and MEL1 (NIHhESC-11-0139) hESCs were cultured on irradiated mouse embryonic fibroblasts (iMEFs) feeder layers in DMEM/F12 medium (Life Technologies, 12500-062) supplemented with 20% KnockOut Serum Replacement (Life Technologies, 10828028), 1X MEM Non-Essential Amino Acids (Life Technologies, 11140050), 1X GlutaMAX (Life Technologies, 35050079), 100U/ml Penicillin and 100 µg/ml Streptomycin (Gemini, 15070063), 0.055 mM 2-mercaptoethanol (Life Technologies, 21985023) and 10 ng/ml recombinant human bFGF (EMD Millipore, GF003AF). Cells were incubated at 37 °C with 5% (vol/vol) $CO_2$, and media was changed daily. Cultures were passaged at a 1:6 - 1:12 split ratio every 4 - 6 days using TrypLE Express (Life Technologies, 12563-029).

5 μM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals, S1049) was added into the culture medium when passaging or thawing cells. Cells are regularly confirmed to be mycoplasma-free by the MSKCC Antibody & Bioresource Core Facility.

### *2.4.2 Generation of TET mutant hESCs*

***In vitro* transcription of gRNAs.** CRISPR gRNAs were designed to target the genomic sequence corresponding to the beginning of the catalytic domain for *TET1*, *TET2*, and *TET3*. For *TET1* and *TET3* the two gRNAs most efficient at producing indel mutations were used for the targeting experiments. For *TET2* only one gRNA was found to be efficient at producing frameshift indel mutations. CRISPR gRNA design and testing was performed by Federico Gonzalez. A CRISPR gRNA used in a previous study was used to target the C5-methyltransferase domain of DNMT3B (Liao et al., 2015). The procedure to generate mutants has been previously described (Zhu et al., 2014). For each CRISPR a 20 bp T7 promoter was added to the 20 bp gRNA target sequence (Table 2.1) followed by a 80 bp constant gRNA backbone to form a 120 bp oligonucleotide. The T7-gRNA oligo nucleotide was amplified by PCR and the T7-gRNA PCR products were used as templates for *in vitro* transcription using the MEGAshortscript T7 kit (Life Technologies, AM1354M). The resulting gRNAs were purified using the MEGAclear kit (Life Technologies, AM1908M), eluted in RNase-free water and stored at -80°C until use.

**Table 2.1. CRISPR gRNA Sequences**

| Gene Targeted | CRISPR gRNA | CRISPR gRNA Target Sequence |
|---|---|---|
| *TET1* | CrT1A | GCCCCAAGGTGTGTATAATA |
| | CrT1B | GGCCCATATTATACACACCT |
| *TET2* | CrT2A | CTTATGGTCAAATAACGACT |
| *TET3* | CrT3A | GTCATCTACACGGGGAAGGA |
| | CrT3B | GATCGAGAAGGTCATCTACA |
| *TET1* (HDR correction of the *TET1* mutation) | TET1 HDR | CCCAAGGTGTGTATAAATAT |
| *DNMT3B* | DNMT3B Cr | TCATTCTTTGATGCTATCAC |
| *Cas9* | Cas9 Cr | TTTCGCCGCTGTCGAACAGC |
| *PAX6 P0* Promoter (targeted demethylation) | Cr6 | CCTAGCTCCCCTCCAGGACC |
| | Cr7 | TTACCAAGAAACTAGGTCAC |
| | Cr9 | CTAGCTGGCCAGTTTCTGTC |
| *SOX10* Promoter (targeted demethylation) | Cr8 | CACAGTAAGAGAGACTTCTC |
| *LEFTY2* Distal Enhancer (targeted demethylation | Cr1 | GAGAAGCAGAGCAGGAAGTC |
| Non-targeting Control (targeted demethylation) | *HBB* | TTTGCATATTCTGGAGACGC |

**Transfection of gRNAs or gRNA + ssDNA.** HUES8 and MEL1 iCas9 hESCs (Gonzalez et al., 2014) or HUES8 TKO hESCs were treated with ROCK inhibitor and doxycycline one day before transfection. For transfection, confluent cells were dissociated using TrypLE (Life Technologies, 12563-029), replated at a 1:6 ratio in iMEF-coated 24-well plates and transfected in suspension with gRNAs or a mixture of gRNA and ssDNA. A second transfection was performed 24 hours later. Transfection was performed using Lipofectamine RNAiMAX (Life Technologies, 13778-150) following the manufacturer's guidelines. For each targeting, *TET1*, *TET2*, *TET3, DNMT3B* gRNAs, each at a 10 nM final concentration, were used. For repair of the TET1 allele in TKO hESCs, gRNAs and ssDNA (of

*TET1* WT sequence) were transfected at a 10nM and 20nM concentration, respectively. Lipofectamine RNAiMAX and gRNA or gRNA + ssDNA were diluted separately in Opti-MEM (Life Technologies, 31985070), then mixed together, incubated for 5 min, and added drop-wise to cultured hESCs.

**Isolation, expansion and identification of mutant clones.** Two days after the last gRNA transfection, hESCs were dissociated into single cells and replated at ~ 2,000 cells per 10-cm dish. Cells were allowed to grow until colonies from single cells became visible (~10 days). At this stage, single colonies were manually picked, mechanically disaggregated and replated individually into 96-well plates. Colonies were amplified and analyzed by Sanger sequencing at the 3 *TET* and the *DNMT3B* loci for the presence of mutations. Clonal cell lines carrying the desired mutations or TKO hESCs in which one *TET1* allele was repaired were amplified and frozen down. For all targeting experiments (TET, DNMT3B and HDR repair of TET1) we also isolated and froze down lines that had undergone the targeting procedure, but whose genomic sequence was not changed. These lines were used as passage-matched controls for methylation analysis and differentiation (Table 2.2).

**Table 2.2. *TET* Knockout Mutant lines**

| hESC Line | *TET1* genotype | *TET2* genotype | *TET3* genotype | CRISPR gRNAs |
|---|---|---|---|---|
| HUES8 WT | WT | WT | WT | CrT1B, CrT2A, CrT3B |
| HUES8 *TET1* KO | -8bp/-8bp | WT | WT | CrT1B |
| HUES8 *TET2* KO | WT | -16bp/-16bp | WT | CrT2A |
| HUES8 *TET3* KO | WT | WT | -4bp/-4bp | CrT3B |
| HUES8 *TET1/2* DKO | -13bp/-1bp | -4bp/-11bp | WT | CrT1B, CrT2A |
| HUES8 *TET1/3* DKO | +1bp/+1bp | WT | -4bp/+1bp | CrT1B, CrT3B |
| HUES8 *TET2/3* DKO | WT | -8bp/+1bp | -4bp/-4bp | CrT2A, Cr3B |
| HUES8 TKO | +1bp/+1bp | +1bp/+1bp | -4bp/+1bp | CrT1B, CrT2A, CrT3B |
| MEL1 WT | WT | WT | WT | CrT1A, CrT2A, CrT3A |
| MEL1 TKO | +1bp/+1bp | -5bp/-1bp | -1bp/+1bp | CrT1A, CrT2A, CrT3A |

### *2.4.3 hESC Differentiation*

**Teratoma.** Confluent hESCs grown on a MEF feeder layer were collected by 1 mg/mL Collagenase type IV (Life Technologies, 17104-019) treatment, and suspended in PBS. One quarter of the cells from a confluent 100 mm dish (approximately 2.5-3 million hESCs) was injected subcutaneously to the dorsal flank of a SCID mouse (NOD.Cg-Prkdc{scid} Il2rg{tm1Wjl}/SzJ (Stock #: 005557), Jackson Laboratory). Palpable tumors were typically observed 1-2 months after injection. Tumor samples were usually collected in 2-3 months, fixed in 4% paraformaldehyde and processed for paraffin embedding and hematoxylin and eosin staining following standard procedures. Mice injected with TKO hESCs were sacrificed 6 months after injection and dissected to confirm that teratomas

had not formed. All mouse procedures were performed following NIH guidelines, and were approved by the local Institutional Animal Care and Use Committee (IACUC), the Institutional Biosafety Committee (IBC) as well as the Embryonic Stem Cell Research Committee (ESCRO).

**Embryoid Body.** hESCs were expanded onto a 10 $cm^2$ dish on an iMEF feeder layer. Once the cells were confluent they were dissociated using 1 mg/mL Collagenase type IV (Life Technologies, 17104-019) and removed using a cell scraper. Cells were spun down at 200 g for 5 minutes and washed with phosphate buffered saline solution (PBS). This was repeated 2 times for a total of 3 washes. The cell clumps were then plated into 6-well low attachment plate in hESC media that lacks bFGF. One 10 cm dish gave enough cells for one 6 well plate. The cells were kept in the 6 well low attachment plate for a total of 5 days after which time they formed embryoid bodies. After 5 days the EBs were transferred to a 24 well dish that has been coated with Geltrex (Life Technologies, A15696) for attachment and differentiation. The cells were kept in the 24 well dish for 7 days total with media change every 2 days. The media remained hESC media without bFGF.

**Neuroectoderm Differentiation.** 90% confluent hESC cultures were disaggregated using TrypLE (Life Technologies, 12563-029) for 5 minutes and washed using hESC media. The cells were plated on Matrigel (BD, 354234) coated dishes in hESC media with ROCK-inhibitor at a density of 180,000–200,000 cells/$cm^2$. After 12 hours, differentiation into neuroectoderm was initiated by switching to knockout serum replacement (KSR) media with 10 μM TGF-β inhibitor SB431542 (Tocris 161410) and 100 nM BMP inhibitor LDN193189 (Axonmedchem, 1509). On day 1 and

day 2 of differentiation, the media was removed and fresh KSR with 10 μM SB431542 and 100 nM LDN193189 was added. Starting on day 4 of differentiation an increasing amount of N2 media was added to the KSR media every two days, while maintaining 10 μM SB431542 and 100 nM LDN193189. On day 4 a 3:1 mixture of KSR/N2 media was added. On day 6 a 1:1 mixture of KSR/N2 media was added and on day 8, a 1:3 mixture of KSR/N2 media was added. The cells were isolated for analysis on day 10 of differentiation. KSR media contains Knockout DMEM (Life Technologies, 10829018), Knockout Serum Replacement (Life Technologies, 10828028), 1X MEM Non-Essential Amino Acids (Life Technologies, 11140050), 1X GlutaMAX (Life Technologies, 35050079), and 2-mercaptoethanol (Life Technologies, 21985023). N2 media contains DMEM/F12 medium (Life Technologies, 12500-062), glucose (Sigma, G8270), sodium bicarbonate (Sigma, S5761), putrescine (Sigma, P5780), progesterone (Sigma, P8783), sodium selenite (Sigma, S5261), apo-transferrin (Sigma, T1147), and insulin (Sigma, I2643).

**Definitive Endoderm Differentiation.** Definitive endoderm differentiation was performed with the assistance of Qing Li. HUES8 hESCs were converted to E8 feeder free culture prior to definitive endoderm differentiation. When cells reached 80-90% confluence they were treated with TRYPLE to dissociate into single cells for passaging. Approximately 0.15 million hESCs were plated in one well of the six-well plates with 10 μM Rho-associated protein kinase (ROCK) inhibitor in 2 ml E8 media. 24 hours later, cells were washed with PBS once and the culture media was changed to Advanced RPMI (Thermo Fisher Scientific 12633012) with penicillin/streptomycin (Thermo Fisher Scientific 15070063), Glutamax

(Thermo Fisher Scientific 35050079), 0.003% BSA (Thermo Fisher Scientific 15260037), 5 μM CHIR-99021(Tocris 4423) and 100ng/ml Activin A (Peprotech 12014E). In the next 2 days, culture medium was changed to Advanced RPMI with penicillin/streptomycin, Glutamax, 0.2% FBS and 100 ng/ml Activin A. Cells were collected for immunofluorescence, RT-qPCR and FACS analysis after 5 days of differentiation.

**Mesoderm Differentiation.** Mesoderm differentiation was performed by Chan-Jung Chang. hESCs were differentiated as spin embryoid bodies (EBs) in APEL medium. Briefly, cells were dissociated into single cells with accutase and plated at 3000 cells per well in round-bottom low-attachment 96-well plates in APEL medium containing 30 ng/ml bone morphogenetic protein 4 (BMP4) and 10nM Y-27632. The plates were centrifuged at 800g for 5 minutes to induce EB aggregation. After 24 hours, the medium was replaced by APEL medium containing 30ng/mL BMP4 and 50ng/mL FGF2. After 2 days, the cytokine cocktail was changed to: 20 ng/ml vascular endothelial growth factor (VEGF), 10 ng/ml FGF2, 100 ng/ml stem cell factor (SCF), 20 ng/ml Flt3 ligand (Flt3L), 20 ng/ml thrombopoietin (TPO), 40 ng/ml IL-3.

### *2.4.4 Dot Blot*

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, 69504) following manufacturer's guidelines and diluted to 200 ng/μl and 100 ng/μl in 20 μl total volume. 5 μl of 0.5 M NaOH was added to each sample and the samples were incubated at 99°C for 5 minutes. Samples were cooled on ice, spun down and neutralized with 2.5 μl of 6.6 M Ammonium Acetate. 2.75 μl of each mixture was spotted on a nitrocellulose

membrane and allowed to air dry. The membrane was baked for 2 hours at 80°C and then incubated in blocking buffer for 2 hours at RT. 5hmC antibody (Active Motif, 39769) at a 1: 10,000 dilution in blocking buffer was added to the membrane and incubated overnight at 4C. The membrane was washed in phosphate buffered saline (PBS) with 0.1% Tween 20 (PBSTw), 3 times for 15 minutes each and then incubated with HRP Goat anti-Rabbit IgG (Active Motif, 15015) in blocking buffer for 2 hours at RT. The membrane was washed again 3 times in PBSTw, and then once in ddH2O for 5 minutes. Luminata Crescendo Western HRP Substrate (Millipore, WBLUR0100) was used for detection.

### *2.4.5 Mass Spectrometry Quantification of 5mC and 5hmC*

Mass spectrometry was performed by Louis Dore. 4 µg of genomic DNA was denatured for 5 minutes at 95°C, then sequentially digested with nuclease P1 (Wako Chemicals, 0218055501), *Crotalus adamanteus* venom phosphodiesterase (Sigma, P3243), and *E.coli* alkaline phosphatase (P5931, Sigma). Digested nucleosides were desalted, filtered and analyzed by UHPLC/MS/MS on an Agilent 1290 with UHPLC coupled to a 6460 Triple Quadrupole LC/MS system using multiple reaction monitoring and the following ion transitions: dC, 228.1→112.1; mdC, 242.1→126.1; hmdC, 248.1→142.1. Molar quantities of dC, mdC, and hmdC were determined by fitting signal intensities to a standard curve of known concentrations. mdC and hmdC quantities were expressed as a molar ratio compared to total dC, mdC, and hmdC present in each sample.

### *2.4.6 Pluripotency and Lineage Marker Staining*

**Alkaline Phosphatase Staining.** The VECTOR Red Alkaline Phosphatase (AP) Substrate Kit (Vector Labs, SK-5100) was used. Cultured hESCs were washed once with PBS and then incubated with substrate working solution for 30 minutes. Cells were washed 3 times with PBS and images were taken.

**Immunofluorescence staining.** Cells were fixed with 4% paraformaldehyde for 10 minutes, washed once with PBS and permeabilized in PBST (PBS + 0.15% Triton-X) for 15 minutes. Blocking was done for 5 minutes at RT with blocking solution (5% donkey serum in PBST). Primary and secondary antibodies were diluted in blocking solution. Primary antibodies were incubated at RT for 1 hr. The following primary antibodies were used at a 1:100 dilution: OCT4 (Santa Cruz, sc-5279), NANOG (CosmobioJapan, RCAB0004P-F), and SOX2 (Santa Cruz, sc-17320). The following antibodies were used at a 1:500 dilution: PAX6 (Covance, PRB-278P) and SOX1 (R&D Systems, AF3369). After primary antibody staining the cells were washed three times with PBST and then incubated with the appropriate Molecular Probes Alexa Fluor dye conjugated secondary antibodies (Life Technologies, AF3369) and DAPI for 1 hour at RT.

### *2.4.7 FACS Analysis*

**Intracellular marker.** NE differentiated cells at the desired time point were disaggregated with TrypLE for 5 minutes and washed with cold FACS buffer: 5% Fetal Bovine Serum (Life Technologies, 261400799) in PBS. Cells were pelleted by centrifugation and washed again with FACS buffer. Each sample was resuspended in FACS buffer and incubated with a

Live/Dead Fixable stain (Life Technologies L34964, 1:1,000) for 30 minutes at RT. Afterwards cells were washed with FACS buffer and resuspended in fixation solution (eBioscience, 00-5523-00) for 1 hour at RT. Cells were washed with permeabilization buffer (eBioscience, 00-5523-00) and then resuspended in permeabilization buffer and the appropriate antibody: OCT4-APC (eBioscience 50-5841-82, 1:25) or PAX6 (Covance PRB-278P, 1:400) for 1 hour at RT. For OCT4 flow analysis the cells were washed with FACS buffer and analyzed by FACS. For PAX6 flow analysis, cells were washed with FACS buffer and then incubated with the appropriate Molecular Probes Alexa Fluor dye conjugated secondary antibody (Life Technologies A21206, 1:500) for 1 hour at RT. Cells were then washed with FACS buffer and analyzed by FACS.

**Surface marker.** Confluent hESCs were disaggregated with TrypLE for 5 minutes and washed with cold FACS buffer. Cells were pelleted by centrifugation and washed again with FACS buffer. Each sample was resuspended in FACS buffer with the appropriate conjugated antibody: Tra1-60-FITC (BD Biosciences, 560380) or Tra1-81-FITC (BD Biosciences, 560194). Cells were incubated in FACS buffer with the antibody for 30 minutes on ice. After staining cells were washed two times with FACS buffer and resuspended in FACS buffer with DAPI and analyzed by FACS.

### *2.4.8 Quantitative RT-qPCR*

Total RNA was isolated with the RNeasy Mini Kit (Qiagen, 74136). DNA was removed from RNA samples using genomic DNA eliminator spin columns. cDNA was produced from RNA using SuperScript III Reverse Transcriptase kit (Life Technologies, 18080051) or High Capacity cDNA

Reverse Transcriptase kit (Life Technologies, 4368813). Quantitative real-time PCR was performed in triplicate using ABsolute QPCR SYBR Green Low ROX Mix (Thermo Scientific, AB4322B). Primers used for qRT-PCR are in Table 2.3. Please note that the primer pair we used for *PAX6* binds to Exon 6, as a result it downstream of the *PAX6* promoters (P0 and P1) and will not differentiate between P0 and P1 transcripts.

**Table 2.3. qPCR Primers**

| Gene or Genomic Region | Forward Primer | Reverse Primer |
|---|---|---|
| *BRACHYURY* | ACCCAGTTCATAGCGGTGAC | CCATTGGGAGTACCCAGGTT |
| *DNMT1* | AGGCGGCTCAAAGATTTGGAA | GCAGAAATTCGTGCAAGAGATTC |
| *DNMT3A* | AGTACGACGACGACGGCTA | CACACTCCACGCAAAAGCAC |
| *DNMT3B* | GACTGCTTGGAATACAATAGGA | GCACCACAAAACATCTTCTTT |
| *FOXA2* | GGGAGCGGTGAAGATGGA | TCATGTTGCTCACGGAGGAGTA |
| *FOXG1* | CCGCACCCGTCAATGACTT | CCGTCGTAAAACTTGGCAAAG |
| *GAPDH* | GGAGCCAAACGGGTCATCATCTC | GAGGGGCCATCCACAGTCTTCT |
| *GATA3* | TCTGACAGTTCGCACAGGAC | AAAATGAACGGACAGAACCG |
| *GATA4* | AAAGAGGGGATCCAAACCAG | TTGCTGGAGTTGCTGGAAG |
| *GATA6* | GTGCCCAGACCACTTGCTAT | TGGAATTATTGCTATTACCAGAGC |
| *GSC* | AACGCGGAGAAGTGGAACAAG | CTGTCCGAGTCCAAATCGC |
| *LEFTY2* | CTCCAACGCCAGGAACC | GACATGGAGGAGCTGGTCA |
| *LHX2* | CCAAGGACTTGAAGCAGCTC | GTAAGAGGTTGCGCCTGAAC |
| *NANOG* | GCCTGGTTGCCTCATGTTATTATGC | CCATGGAGGAAGGAAGAGGAGAGA |
| *OCT4* | TGGTCCGAGTGTGGTTCTGTAA | TGTGCATAGTCGCTGCTTGAT |
| *OTX2* | CATGCAGAGGTCCTATCCCAT | AAGCTGGGGACTGATTGAGAT |
| *PAX6* | TGGGCAGGTATTACGAGACTG | ACTCCCGCTTATACTGGGCTA |
| *PAX6 3'UTR (Endo)* | ATTGTGTTAATTCAGTCAGTG | CCCAGTGGTACAATACAGGAC |
| *SOX1* | AACACTTGAAGCCCAGATGGA | GCAGGCTGAATTCGGTTCTC |
| *SOX7* | CATGCAGGACTACCCCAACT | ACTCACCCCTGTCCTCCTTC |
| *SOX10* | CTTTCTTGTGCTGCATACGG | AGCTCAGCAAGACGCTGG |
| *SOX17* | GGCGCAGCAGAATCCAGA | CCACGACTTGCCCAGCAT |
| Epimark: *PAX6 P0* | GGCAGGAATCATTTTTAGGAGGA | CCTGGAGAGACCTTTGGCCTA |
| ChIP qPCR: *PAX6 P0 Promoter 1* | CCTCCCACTGGCCACTCTAGT | TAGGGGCTTACCAAGAAACTA |

**Table 2.3.**
**(Continued)**

| | | |
|---|---|---|
| ChIP qPCR: *PAX6 P0 Promoter 2* | CCCCGCCTGCTTATCTGCCTG | CTTCTCCCTCGGCGACCCCAG |
| ChIP qPCR: *PAX6 P0 Promoter 3* | CGGGAAGAAGGGCACCGCGGG | AGTCTGTGCTACCCCGGGCTG |
| ChIP qPCR: *PAX6 P1 Promoter 1* | CCTAAGCTGGACTCGGGACTC | CCGCCGCTGTGCTCTGTGTCT |
| ChIP qPCR: *PAX6 P1 Promoter 2* | GTCTGCTCAGTCCACGGAGGC | GGAGTGTACTGAGGTGTGTCC |
| ChIP qPCR: *PAX6 P1 Promoter 3* | CTCGCCTCCACCGCTCCTCAC | GAGAGCGAGCGGTGCATTTGC |
| ChIP qPCR: E *-6k Enhancer* | CATCCTTTCCTTCTTTCTGCT | CGTCCCCGCCGTGCAAAGAGA |
| ChIP qPCR: E *156k Enhancer* | ACTTAAATGCCTCAAACTTTT | TCTTGTAAATGAGTAGCCTAT |
| ChIP qPCR: *NANOG Promoter* | GATGGGGGAATTCAGCTCAGG | GTCTCTCTTAATCAGCACAGT |
| ChIP qPCR: *HERVK* | AGAGGAAGGAATGCCTCTTGCAGT | TTACAAAGCAGTATTGCTGCCCG C |

### *2.4.9 Epimark*

In addition to base level genome-wide 5hmC profiling, 5hmC at the *PAX6* P0 promoter was validated using the Epimark 5hmC and 5-mC Analysis kit (NEB, E3317S) according to the manufacturer's protocol. Briefly genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, 69504) following manufacturer's guidelines. 10 ug of DNA for each sample was divided into 6 tubes. The first 3 tubes were treated with beta-glucosyltransferase (BGT) overnight. 2 of the 6 total tubes were digested with MspI (one tube had been treated with BGT whereas the other had not),

2 of the 6 tubes were digested with HpaI (one tube had been treated with BGT whereas the other had not) and two of the tubes were undigested. After overnight digestion we performed qRT-PCR analysis. The primers for analysis are found in Table 2.3.

### *2.4.10 RNA-Seq*

For RNA-Seq total RNA was isolated with the RNeasy Mini Kit (Qiagen, 74136) from HUES8 WT and TKO hESCs (n=2 each). RNA samples were submitted to the MSKCC Integrated Genomics Core for library prep and sequencing.

RNA-seq data analysis was performed by Heng Pan. RNA-seq data was aligned to the hg19 reference genome using Bowtie. Read counts were derived from HTSeq.scripts.count module in HTSeq-0.6.0 with default parameters in a non-directional model (Anders et al., 2015). Differentially expressed genes were generated by DESeq2-1.4.5 in R (Love et al., 2014). Up regulated genes were decided by logFC > 2 and p-adj < 0.1 (n=134). Downregulated genes were decided by logFC < -2 and p-adj < 0.1 (n=233).

### *2.4.11 5hmC Profiling*

5hmC-Seal was performed by Louis Dore as previously described (Song et al., 2011). 40 µg of genomic DNA was sonicated to ~200-400bp fragments using a Diagenode Bioruptor Sonicator. Sonicated DNA was then labeled with azide glucose in a 1 hour reaction at 37°C catalyzed by recombinant βGT utilizing UDP-6-$N_3$-glucose as the sugar donor. The reactions were cleaned up using a Zymo DNA Clean & Concentrator kit

(Zymo, D4003), then a biotin moiety was added to the azide-labeled DNA via a copper-free click chemistry reaction with DBCO-S-S-PEG3-Biotin in water at 37°C for 1 hr. Reactions were once again cleaned with the Zymo kit and then bound to Dynabeads MyOne Streptavidin C1 beads (Life Technologies, 65001) for 15 minutes at RT and washed 5 times with binding buffer (5mM Tris-HCl, pH7.5, 0.5mM EDTA, 1M NaCl, 0.01% Tween 20). Bound DNA was eluted by reducing the disulfide in the biotin linker with 100mM DTT for 2h at RT with gentle rotation. Eluted DNA was cleaned on a Micro Bio-spin Column (BioRad, 7326204) to remove DNA and then purified by the Zymo kit. Libraries were constructed from eluted DNA by end repair, A-tailing, and adapter ligation, followed by 4 cycles of PCR, size selection via agarose gel electrophoresis, 12 additional PCR cycles, and a final size selection via agarose gel electrophoresis. Library quantity and quality were analyzed by Bioanalyzer prior to sequencing on an Illumina HiSeq 2500. Sequence alignment and peak identification were performed as previously described (Song et al., 2011). Peak calling was performed by macs14 1.4.2 with default parameters.

### *2.4.12 Whole Genome Bisulfite Sequencing (WGBS) and Enhanced Reduced Representation Bisulfite Sequencing (ERRBS).*

For WGBS, genomic DNA was isolated from HUES8 WT and HUES8 TKO hESCs. Genomic DNA was sheared using E220 focused-ultrasonicator (Covaris) to 250-350bp fragments. After end-repair and TruSeq adapter (Illumina) ligation the DNA-libraries were denatured followed by treatment with bisulfite (Zymo Research, D5020) for 30 min at 65C. Single stranded DNA was purified using silica gel columns and amplified using HiFi Uracil+

polymerase (Kapa biosciences, KK2802). The amplified libraries were quantified and mixed equimolarly for sequencing on Illumina HiSeq2500 by the MSKCC Integrated Genomics Core. One replicate for HUES8 WT and HUES8 TKO was submitted for sequencing.

For ERRBS, genomic DNA was isolated from HUES8 WT, HUES8 TKO and HUES8 QKO hESCs. Genomic DNA was submitted to the Weill Cornell Medical College Epigenomics Core for ERRBS. Two independent replicates each for HUES8 WT, TKO and QKO cells were submitted for sequencing.

WGBS and ERRBS data were analyzed by Heng Pan. WGBS data were aligned to the bisulfite-converted hg19 reference genome using Bismark v0.13.0 (Krueger et al., 2011). We extracted the methylation status with the bismark_methylation_extractor script in Bismark (Krueger et al., 2011). Only CpGs with at least 3 covered reads were used for downstream analysis. The WCMC Computational Genomics Core Facility supported alignment and methylation extraction for ERRBS data (Akalin et al., 2012). DMRs were defined as regions containing at least five differentially methylated CpGs (DMCs, false discover rate=20%, Chi square test) and whose total methylation difference was more than 10% (Pan et al., 2015). For Figures 1h and Extended Data Fig. 3b-c hypermethylation was defined as a 5% increase in methylation in order to set a uniform cutoff for comparison between different promoters or cell lines. For whole genome bisulfite sequencing only one replicate was used but the use of five or more DMCs partially overcomes the statistical limitation of individual Chi square tests based on a *n*=1 sample, as the latter should be cautiously interpreted in the absence of multiple measurements from independent samples. DMRs

calling was performed with RRBSeeqer with default parameters. DMRs were annotated using ChIPseeqerAnnotate from ChIPseeqer package (Giannopoulou et al., 2011). Methylation of a specific region was calculated by averaging the methylation levels of all covered CpGs in that region.

Genomic regions for CpGs are defined as the following definitions. CpG islands (CGI) are defined as annotations from refSeq. CGI shores are defined as the region encompassing 1 kb upstream and downstream of known CGIs. Non-CGIs are defined as at least 10 kb away from known CGIs. Promoters are defined as the region encompassing 2 kilobases (kb) upstream and downstream from the transcription start site (TSS) from refSeq. The following characteristics were used to classify promoters as active, initiated, bivalent and silent. Active promoters are associated with H3K4me3 (in 1.5 kb region flanking the TSS) and H3K79me2 (in 5 kb downstream from the TSS). Initiated promoters are associated with H3K4me3 (in 1.5 kb region flanking the TSS). Bivalent promoters are associated with H3K4me3 and H3K27me3 (in 1.5 kb region flanking the TSS). Silent promoters were not associated with H3K4me3 (in 1.5 kb region flanking the TSS). With WGBS (HUES8 TKO vs. HUES8 WT) 6,695 active promoters, 989 initiated promoters, 3,327 bivalent promoters and 4,707 silent promoters were analyzed. With ERRBS (HUES8 TKO vs. HUES8 WT) 6,450 active promoters, 846 initiated promoters, 3,234 bivalent promoters and 1,476 silent promoters were analyzed. With ERRBS (MEL1 TKO vs. MEL1 WT) 6,420 active promoters, 3,222 bivalent promoters, 835 initiated promoters and 1,413 silent promoters were analyzed.

Enhancers are required outside of promoters and exons (RefSeq). Poised enhancers were identified as regions overlapping with H3K4me1

peaks only. Active enhancers were identified as regions overlapping with H3K4me1 and H3K27ac peaks.

Overlap of 5mC change with histone marks was determined using previous ChIP-Seq datasets for histone marks (H3K4me1: GSM 733782; H3K4me2: GSM 733670; H3K4me3: GSM 733657; H3K27me3: GSM 733748; H3K27ac: GSM 733718; H3K9me3: GSM 1003585; H3K36me3: GSM 1003585; H3K79me2: GSM 1003547). Peak calling was performed by ChIPseeqer-2.1 with default parameters (Giannopoulou et al., 2011).

### *2.4.13 ChIP-Seq*

ChIP-Seq was performed for TET1 (WT hESCs) and DNMT3B (WT and TKO hESCs). WT and TKO hESCs were cultured in standard hESC media, as described above. ~5 × $10^7$ cells were fixed, washed and snap-frozen according to the Cell Fixation protocol from Active Motif (http://www.activemotif.com/documents/1848.pdf). Immunoprecipitation and DNA sequencing was done by Active Motif.

ChIP-Seq data analysis was performed by Heng Pan. ChIP-Seq data was aligned to the hg19 genomes using bowtie-0.12.9 with default parameters except --n 2 and --best. Peak calling and read density in peak regions were performed by macs14 1.4.2 with default parameters. The technical success of the ChIP-Seq was confirmed by using standard quality control measures and determining the overlap between TET1 binding and the presence of 5hmC. We observed that 52.6 % of 5hmC peaks associated with gene promoters overlapped with TET1 peaks (Figure 2.11A), which is similar to previous results obtained from TET1 ChIP-Seq in mouse embryonic stem cells (Williams et al., 2011, Xu et al., 2011). Approximately

75% of TET1 peaks overlapped with 5hmC peaks present in HUES8 WT, in contrast only 25% of TET1 peaks overlapped with randomly generated 5hmC peaks (Figure 2.11C). Random peaks were generated with ChIPseeqerCreateRandomRegions in ChIPseeqer-2.1 with default parameters. The lack of a complete overlap between 5hmC and TET1 could be due to a number of factors including: 5hmC production by TET2 and TET3, rapid turnover of 5hmC and reduced binding of TET1 to 5hmC. Peaks from ChIP-Seq were annotated using ChIPseeqerAnnotate from ChIPseeqer package (Giannopoulou et al., 2011).

### *2.4.14 5mC Methylation Analysis by MassArray*

hESCs were disaggregated using TrypLE and genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, 69504) following manufacturer's guidelines. MassArray Epityper analysis was performed using the WCMC Epigenomics core. MassArray Epityper analysis consists of bisulfite treatment of genomic DNA, followed by PCR for the specific region of analysis, base specific cleavage that differentiates between previously methylated and unmethylated DNA and finally MALDI-TOF mass spectrometry (Ehrich et al., 2005). Primers used for mass array 5mC methylation analysis are found in Table 2.4. Genomic coordinates of regions assayed by MassArray are provided in Table 2.5.

**Table 2.4. 5mC MassArray Bisulfite Sequencing Primers**

| Gene or Genomic Region | Forward Primer | Reverse Primer |
|---|---|---|
| *PAX6 P0 Promoter* 1 | AGGAAGAGAGAGTTTTTTTTAAGGTTGGTGGTAGG | CAGTAATACGACTCACTATAGGGAGAAGGCTCCTTCCTAACTAACCAATTTCTATCC |
| *PAX6 P0 Promoter* 2 | AGGAAGAGAGAGTATTTTGTTGGGTTGTAGGGATT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCTCCCACTAACCACTCTAATCTA |
| *PAX6 P0 Promoter* 3 | AGGAAGAGAGATTTTAGGATATTTGAGGTTGGAGG | CAGTAATACGACTCACTATAGGGAGAAGGCTACCAAAAAACAACTCCAAACAAC |
| *PAX6 P0 Promoter* 4 | AGGAAGAGAGTTTTAATTTGAAAGTGATAGTGGTGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACAAAACAAAAAATCAAATCCAA |

**Table 2.4**
**(Continued)**

| | | |
|---|---|---|
| *PAX6 P0 Promoter* 5 | AGGAAGAGAGGGTGTGAGGGAAAAATAGGTATAGAT | CAGTAATACGACTCACTATAGGGAGAAGGCTCTCTAAATAATCACCCTATCCCCTC |
| *E -6k Enhancer* 1 | AGGAAGAGAGTTGTTGTTTGATTTTGAATATGAATAG | CAGTAATACGACTCACTATAGGGAGAAGGCTCCTAAAAACCCCTAAAACAAAAAAA |
| *E -6k Enhancer* 2 | AGGAAGAGAGTTGTTTTTTTGTTTTAGGGGTTTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAACCTTAACCTAACTTACACCCC |
| *E -6k Enhancer* 3 | AGGAAGAGAGGTAGGATTTGTTGGGAAAAAGGTAT | CAGTAATACGACTCACTATAGGGAGAAGGCTCTAAACAAAAAAACCCCTTCTAACC |
| *E -6k Enhancer* 4 | AGGAAGAGAGGAGGTTTTGGTTTAGGTTTTTTTTG | CAGTAATACGACTCACTATAGGGAGAAGGCTACTACCTAACTTTAAATATAAACAATCACA |
| *E -6k Enhancer* 5 | AGGAAGAGAGGTTTAGGTTTTTTTTGATTGGGATT | CAGTAATACGACTCACTATAGGGAGAAGGCTTAAAAAATAAATTTCTTCCCCTCCC |
| *E -6k Enhancer* 6 | AGGAAGAGAGTGTAATTTGTAATTATTGTTTTTTATGTGG | CAGTAATACGACTCACTATAGGGAGAAGGCTTAATCTTCTCCCAACCTCCCTACTA |
| *E -6k Enhancer* 7 | AGGAAGAGAGTTTTTTGTGATAAAGGTTTGTAGTTGTT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCAAACCAAAACCTCCTCAATAC |
| *E -6k Enhancer* 8 | AGGAAGAGAGGGAGTGGAGATTTTTTGGTTTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACCTTTAAACTCAACCCTTACTTTT |
| *E 156k Enhancer* 1 | AGGAAGAGAGTGGGTTAAATTTAGTTTTATGTTGTTG | CAGTAATACGACTCACTATAGGGAGAAGGCTACTTTAAAACACTTTAAACTCCCTTCC |
| *E 156k Enhancer* 2 | AGGAAGAGAGTGTTTTTTAGGTTGGAGTATAGTGA | CAGTAATACGACTCACTATAGGGAGAAGGCTCCACCTTAACCTCCCAAAATACTAA |
| *E 156k Enhancer* 3 | AGGAAGAGAGATTATTTTGGTTGGTTAGGTTGGTT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCTTCCAAAAATAAAATACCCAAAA |
| *E 156k Enhancer* 4 | AGGAAGAGAGGTTTTTAATAGTTTTTGGAAATAAAGATAG | CAGTAATACGACTCACTATAGGGAGAAGGCTAACTCACACCTATAATCTCAACACTTT |
| *E 156k Enhancer* 5 | AGGAAGAGAGGGTTGGATGAGAAAATATGTGTATATAATG | CAGTAATACGACTCACTATAGGGAGAAGGCTCATAAACCAAATCCAACTCCATATT |
| *E 156k Enhancer* 6 | AGGAAGAGAGTTGGTTTTTTATGTTGGTTGTAAAAA | CAGTAATACGACTCACTATAGGGAGAAGGCTAATAAAACCCTATCTCTACTAAAAATTCAA |
| *E 156k Enhancer* 7 | AGGAAGAGAGAAGAAATAGTTTTATTTTAGGTTAGGAAA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACTCAAACAATCCTCCTACCTC |
| *E Ele4h Enhancer* 1 | AGGAAGAGAGTTAATTTTAGAGAATAGGGAGAGGGA | CAGTAATACGACTCACTATAGGGAGAAGGCTTCACAAAATAACAAAACTACCACAAA |
| *E Ele4h Enhancer* 2 | AGGAAGAGAGTTTGTGGTAGTTTTGTTATTTTGTGA | CAGTAATACGACTCACTATAGGGAGAAGGCTCTTTCTAAAACAACCTTTCTTTTTCC |
| *E Ele4h Enhancer* 3 | AGGAAGAGAGGTTTTGGTGAGGGTTTTTTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTTAACTCCCTATTCAACTCTCCTCTC |
| *E Ele4h Enhancer* 4 | AGGAAGAGAGTTTGGTATTAGTAGATTGGGAATTG | CAGTAATACGACTCACTATAGGGAGAAGGCTCTTCTCCAACAAAACAAAACCTAAC |
| *E Ele4h Enhancer* 5 | AGGAAGAGAGTTTTGGTATTAGTAGATTGGGAATTGT | CAGTAATACGACTCACTATAGGGAGAAGGCTCCCACCTCTAAACACAAAAAACTT |
| *E Ete1 Enhancer* 1 | AGGAAGAGAGGTTGTAGATTTGGGAATAGGTAGG | CAGTAATACGACTCACTATAGGGAGAAGGCTTAACCAAATTTAAAAATCACCCTCT |
| *E Ete1 Enhancer* 2 | AGGAAGAGAGTTTATGATTGGAGGAGAAGGTTTAT | CAGTAATACGACTCACTATAGGGAGAAGGCTTATCCCCTTAATAAATCAAAAAAAA |
| *E Ete1 Enhancer* 3 | AGGAAGAGAGTGGAAGTTTATTTTTAGGAGTGGAA | CAGTAATACGACTCACTATAGGGAGAAGGCTACCTAAACACCCTCTTTTCTTATCATTA |
| *E Ete1 Enhancer* 4 | AGGAAGAGAGGGTAGGGGAAGTGGTAGATTTGAT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCCAAAAAAATAACTAACCAACCT |
| *E Ete1 Enhancer* 5 | AGGAAGAGAGATTTTTGTTTTGGTTTTTTTTGGTT | CAGTAATACGACTCACTATAGGGAGAAGGCTCCCTTTAAAACTTCCCTCTTAAAAA |
| *E Ete1 Enhancer* 6 | AGGAAGAGAGTTTTTAAGAGGGAAGTTTTAAAGGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAATCTCCATTCCAAAATAACCAAAC |
| *E Ete1 Enhancer* 7 | AGGAAGAGAGGGTTTGTAGGATATTGATTTGTTGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAACTAACCCTAAAACCCTCCTC |
| *E Ete1 Enhancer* 8 | AGGAAGAGAGGGTTGTTAGTTTGGGTGGTGTTAT | CAGTAATACGACTCACTATAGGGAGAAGGCTAATTAAAACAAACTATTCCCAACCC |

**Table 2.4**
**(Continued)**

| | | |
|---|---|---|
| *E Ete1 Enhancer* 9 | AGGAAGAGAGTTTTTTTATAGAATTTGGATGATTGG | CAGTAATACGACTCACTATAGGGAGAAGGCTTCTCACAAAAAATAACTTCCAAACC |
| *E Ete1 Enhancer* 10 | AGGAAGAGAGTTTAGTTATTTGTGATAGGTGTTGGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAATCTAAATAATAATTCCAATCATCCA |
| *E Ete1 Enhancer* 11 | AGGAAGAGAGTTTTTTAGGTTTTTTGGGAGTATTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAAAAATCCACTTCCCACTAC |
| *E Ete1 Enhancer* 12 | AGGAAGAGAGTGTTTTTTTAGTGAGTTAGGGAAGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAATCCCAACACCTATCACAAAT |
| *E Ete1 Enhancer* 13 | AGGAAGAGAGGGTTGTAAAGTGTAGATGGTTGGAT | CAGTAATACGACTCACTATAGGGAGAAGGCTCCCTTCCCTAACTCACTAAAAAAAC |
| *E Ete1 Enhancer* 14 | AGGAAGAGAGGTTTTGAGTTGGGAGTAGGGG | CAGTAATACGACTCACTATAGGGAGAAGGCTATCCAACCATCTACACTTTACAACC |
| *E Ete1 Enhancer* 15 | AGGAAGAGAGGGATTATGGTAAGGTTTAGGTTTAGA | CAGTAATACGACTCACTATAGGGAGAAGGCTACCAAAAAACAACTCCAAACAAC |
| *E P1 Promoter E1 Enhancer* 1 | AGGAAGAGAGGGGTGTAAGTTTTTGTGTTGTTTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCTCTCCAACCAAAACTATACCTC |
| *E P1 Promoter E1 Enhancer* 2 | AGGAAGAGAGGAGGTATAGTTTTGGTTGGAGAGGT | CAGTAATACGACTCACTATAGGGAGAAGGCTAATAAAACACACACACACACACACA |
| *E P1 Promoter E1 Enhancer* 3 | AGGAAGAGAGGGTAGGGAGGTTGAAATGAAGTAGT | CAGTAATACGACTCACTATAGGGAGAAGGCTACCCCACCTCTAACTAAAACC |
| *E P1 Promoter E1 Enhancer* 4 | AGGAAGAGAGTTGGAGGATGATGATAGAGGTTAGGT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAACATCCTTTCTAATTATCACAACTTC |
| *E P1 Promoter E1 Enhancer* 5 | AGGAAGAGAGTTGTTTTGTATAAAGTAATATTTTGTGTGA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAATAACTACAACCAACACACCT |
| *E P1 Promoter E1 Enhancer* 6 | AGGAAGAGAGGGTGAGGATTTTTTAGGGTTTTTT | CAGTAATACGACTCACTATAGGGAGAAGGCTTATCCTTTCTTACCAACTCCAAAAC |
| *E P1 Promoter E1 Enhancer* 7 | AGGAAGAGAGGGTGAGGATTTTTTAGGGTTTTTTA | CAGTAATACGACTCACTATAGGGAGAAGGCTATCCTAACAACTCCATTCCAACTAC |
| *E P1 Promoter E1 Enhancer* 8 | AGGAAGAGAGGGATTGAGTAGATTTAGGAGAGGGA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAACACAACTCTAACTAAAAAAACCC |
| *E P1 Promoter E1 Enhancer* 9 | AGGAAGAGAGTTTTGGTAATAGTTGAAGGGGAGTT | CAGTAATACGACTCACTATAGGGAGAAGGCTTAAAAAACCCTAAAAAATCCTCACC |
| *E P1 Promoter E1 Enhancer* 10 | AGGAAGAGAGGTTTTGTTTTGGGGTGTGTATAGTA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACAACTCCCCTTCAACTATTACC |
| *E P1 Promoter E1 Enhancer* 11 | AGGAAGAGAGGGTTGTTGGGTTTTAGGTAGGAA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACACACCTCAATACACTCCAAAA |
| *GAPDH* Promoter 1 | AGGAAGAGAGTTTAGAAAGGTAGGGTTAGGGATTG | CAGTAATACGACTCACTATAGGGAGAAGGCTCAATATACCTTTCATTCCATCCAAC |
| *GAPDH* Promoter 2 | AGGAAGAGAGGTTGGATGGAATGAAAGGTATATTG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACAAAAACCTAAAAAAAACCCAT |
| *GAPDH* Promoter 3 | AGGAAGAGAGGAGATGTTAGGAGTTAGGAGATGGG | CAGTAATACGACTCACTATAGGGAGAAGGCTTCCCAAAACTAAACTATAAACAACAAAA |
| *OCT4* Promoter 1 | AGGAAGAGAGGATTAGATTTTGGATTGATTGGG | CAGTAATACGACTCACTATAGGGAGAAGGCTAATTAAAAAACAAAAAACAATCCCC |
| *OCT4* Promoter 2 | AGGAAGAGAGTGTTTTTTAGGAATTTAGGTGTTTGA | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAAAATAAAAACCAACCCCTT |
| *OCT4* Promoter 3 | AGGAAGAGAGGTTTAATGGTGGTGGTAATGGTGTT | CAGTAATACGACTCACTATAGGGAGAAGGCTACAAAACCTAAAAAAATACCAAAAA |
| *NANOG* Promoter 1 | AGGAAGAGAGTGTGGGAGTAAAGTTAGTTGTTTTG | CAGTAATACGACTCACTATAGGGAGAAGGCTTATCTATCCCTCCTCCCAAATAATC |
| *NANOG* Promoter 2 | AGGAAGAGAGTTTTTGTATTTTGTTTTTGGGTTTG | CAGTAATACGACTCACTATAGGGAGAAGGCTCTTAATAACCTTAACAACCCCCACT |
| *NANOG* Promoter 3 | AGGAAGAGAGAATTTTTTGAATTTGGGAGGTAGAG | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAAAACATAACAAATCACCAAACC |
| *NANOG* Promoter 4 | AGGAAGAGAGGGGTATATTTTTTATTGATTTATTTTTGTG | CAGTAATACGACTCACTATAGGGAGAAGGCTTAACTAAAACTACAAACACCCACCA |
| *SOX2* Promoter 1 | AGGAAGAGAGAAAGATTTTAATAAGAGAGTGGAAGGAA | CAGTAATACGACTCACTATAGGGAGAAGGCTAACAAAACCAACCCTAACATTTTC |

**Table 2.4**
**(Continued)**

| | | |
|---|---|---|
| *SOX2* Promoter 2 | AGGAAGAGAGGTGGGATGTTAGGAAGTTGAAATTAT | CAGTAATACGACTCACTATAGGGAGAAGGCTTTAAAAAACCCAAACCTCTATCCTC |
| *SOX2* Promoter 3 | AGGAAGAGAGGGATAGAGGTTTGGGTTTTTTAATTT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAACCAACCTACCAACCACTAAAA |
| *FOXA2* Promoter 1 | AGGAAGAGAGTGGTTTGGATATTTTATAAAGAGGGT | CAGTAATACGACTCACTATAGGGAGAAGGCTAAAATAAACACCCACATAAACTCACA |
| *FOXA2* Promoter 2 | AGGAAGAGAGTTTTTATTTTGTTTTTGGGTGGAA | CAGTAATACGACTCACTATAGGGAGAAGGCTCCTAAAAATTAAAACTCCAAAAAAACC |
| *FOXA2* Promoter 3 | AGGAAGAGAGTGTAGATTTGAGAGTTTTGGGGTTA | CAGTAATACGACTCACTATAGGGAGAAGGCTTCTCTATCCTCTCTATCTTCCAAAAAA |
| *FOXA2* Promoter 4 | AGGAAGAGAGTTTAATAAAATGGAAAGGGAAGGGT | CAGTAATACGACTCACTATAGGGAGAAGGCTTAAAACCCCAACCCCTAAATTCTAC |
| *FOXG1* Promoter 1 | AGGAAGAGAGTAATTGGATTAGGTTTAAAAATGGAA | CAGTAATACGACTCACTATAGGGAGAAGGCTTTAACAAAAACCCCTAAAACTCTCC |
| *FOXG1* Promoter 2 | AGGAAGAGAGGGAGTTGTAGGTTGTTAGTGGTTGT | CAGTAATACGACTCACTATAGGGAGAAGGCTTATCAAAACCTTCTCCTCATCCTTA |
| *FOXG1* Promoter 3 | AGGAAGAGAGGGGAGTGAGGAGATTTTATATAGAGG | CAGTAATACGACTCACTATAGGGAGAAGGCTACAACCACTAACAACCTACAACTCC |

**Table 2.5. Genomic Coordinates for Promoters and Enhancers**

| Region | Genomic Coordinates (GRCh37/hg19) |
|---|---|
| *PAX6* P0 Promoter | Chr11: 31839283-31840731 |
| *PAX6* P1 Promoter | Chr11: 31832705-31834156 |
| *PAX6* 156k Enhancer | Chr11: 31679370-31690371 |
| *PAX6* Ele4h Enhancer | Chr11: 31825691-31825906 |
| *PAX6* E1E Enhancer | Chr11: 31832708-31832765 |
| *PAX6* Ete1 Enhancer | Chr11: 31837446-31839211 |
| *PAX6* -6k Enhancer | Chr11: 31847524-31848122 |
| *SOX10* Promoter | Chr22: 38380465-38381233 |
| *LEFTY2* Promoter | Chr1: 226129154-226131027 |
| *LEFTY2* Enhancer proximal | Chr1: 226132079-226134622 |
| *LEFTY2* Enhancer distal | Chr1: 226138153-226139319 |
| *GAPDH* Promoter | Chr12: 6642271-6643589 |
| *OCT4* Promoter | Chr6: 31138570-31139891 |
| *NANOG* Promoter | Chr12: 7940446-7941959 |
| *SOX2* Promoter | Chr3: 181428186-181429721 |
| *FOXA2* Promoter | Chr20: 22566291-22567649 |
| *FOXG1* Promoter | Chr14: 29233677-29234801 |
| *SOX17* Promoter | Chr8: 55369986-55370476 |
| *GATA2* Promoter | Chr3: 128211937-128212937 |
| *GATA6* Promoter | Chr18: 19747670-19749477 |

### *2.4.15 PAX6 Overexpression Lentivirus Construct and Generation of PAX6 Overexpression hESCs*

To generate the *PAX6* overexpression construct *PAX6* cDNA from D10 of WT NE differentiation was tagged with a 2A-GFP sequence and inserted into a lentiviral backbone. The 2A-GFP sequence was amplified from the OCT4-eGFP-PGK-Puro plasmid (Addgene, plasmid # 31937) using primers that contain BamHI (Forward) and AscI sites (Reverse). The *PAX6* cDNA PCR product was cloned into the pENTR-dTOPO vector (Thermo Fisher Scientific, K240020). Digestion of the PAX6-pENTR-dTOPO vector and the 2A-GFP PCR product with BamHI and Asc1 was followed with ligation to form the PAX6-2A-GFP-pENTR-dTOPO vector. The PAX6-2A-GFP insert was then transferred into a doxycycline inducible lentiviral backbone through a LR reaction. Primers used for cloning are listed in Table 2.6.

**Table 2.6. Primers for cloning PAX6 overexpression vector, dCas9 TET1 catalytic domain fusion (dCas9-TET1CD) and mutagenesis of dCas9-TET1CD**

| Cloning Construct | Forward Primer | Reverse Primer |
|---|---|---|
| 2A-GFP construct | GGATCCGCCACTAACTTCTCCCTGT | GGCGCGCCTTACTTGTACAGCTCGTCCAT |
| PAX6 cDNA construct | CACCGAATTCATGCAGAACAGTCACAGCGGA | GGATCCCTGTAATCTTGGCCAGTATTG |
| TET1 catalytic domain (for dCas9-TET1CD) | ACTGAGGCCGGCCAGCTGCCCACCTGCAGCTGTCTT | TTCTTGGCCGGCCTCAGACCCAATGGTTATAGG |
| TET1 catalytic domain mutagenesis | GGACTTCTGTGCTCATCCCTACAGGGCCATTCACAACAT | ATGTTGTGAATGGCCCTGTAGGGATGAGCACAGAAGTCC |

### *2.4.16 Colony Forming Assay*

hESC cells were washed with PBS, treated with TrypLE for 5min at 37°C, then collected with fresh hESC medium, passed through a 40um cell strainer to make sure cells were single cells, and then counted. Five thousand cells were plated in a 10cm dish previously coated with iMEF with

Rock inhibitor. Cell were grown for two weeks, then stained with alkaline phosphatase (AP) and counted.

### *2.4.17 Cloning dCas9 TET1 Catalytic Domain Fusion and Site-Directed Mutagenesis*

The TET1 catalytic domain was amplified using forward and reverse primers that contain the Fse1 restriction site from the pJFA274 plasmid (Addgene, plasmid # 49239). This PCR product was then digested with Fse1. The dCas9-krab construct was cloned from the pHR-SFFV-dCas9-BFP-KRAB plasmid (Addgene, Plasmid # 46911) into the pENTR-dTOPO vector (Thermo Fisher Scientific, K240020). The KRAB protein was removed by Fse1 digestion and the Fse1-digested TET1 catalytic domain PCR product was ligated to produce the dCas9-TET1 catalytic domain fusion vector (dCas9-TET1CD). This cloning was performed by Abhijit Shukla. Primers used for cloning are listed in Table 2.6.

The TET1 catalytic domain, from the dCas9 TET1 catalytic domain fusion vector (described above), was mutated at two sites (H1671A and D1673A) to produce a catalytic null control (dCas9-TET1CDmut). Mutagenesis was performed using the QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies, 200521) according to the manufacturer's protocol. The primers used for mutagenesis are in Table 2.6.

### *2.4.18 Targeting dCas9-TET1CD into TKO hESCs*

The dCas9-TET1CD was cloned into the pINDUCER lentiviral backbone (Addgene, Plasmid # 46948) under the control of doxycycline inducible promoter. This lentiviral backbone also constitutively expresses

GFP, which allows us to isolate infected cells. Prior to infection of TKO hESCs with pINDUCER-dCas9-TET1CD, the already AAVS1-integrated Cas9 was inactivated through CRISPR targeting of the beginning of the Cas9 protein (Figure 2.19A). The gRNA used to inactivate Cas9 is listed in Supplementary Table 1. Knockout of Cas9 was confirmed using RT-qPCR (Figure 2.19C, performed by Bobbie Pelham-Weber), western blot (Figure 2.19D, performed by Virginia Teijeiro) and T7 Endonuclease assay (Figure 2.19E, performed by Bobbie Pelham-Weber). Cas9 knockout clones were isolated and infected with virus containing the pINDUCER-dCas9-TET1CD construct. Infected cells were isolated by FACS sorting of GFP-positive cells. These cells were expanded and frozen down.

After identifying a CRISPR gRNAs that produce demethylation at the *PAX6* P0 promoter, *SOX10* promoter and *LEFTY2* enhancer we cloned these CRISPR gRNAs into the piCRg Entry vector (Addgene, Plasmid # 58904). For the *PAX6* P0 promoter we identified 3 gRNAs (Cr6, Cr7, Cr9) that were able to produce demethylation. We also cloned a gRNA that targets the *HBB* promoter into the piCRg Entry vector as a non-targeting control. We performed golden gate cloning to attach the gRNAs together in tandem. An LR reaction (Thermo Fisher Scientific, 11791100) was then performed to transfer the U6 promoter and either the tandem array of Cr6, Cr7 and Cr9 gRNAs for the *PAX6* P0 promoter or the individual gRNAs for *SOX10* and *LEFTY2* into a lentiviral backbone containing a hygromycin selection cassette. TKO hESCs containing the dCas9-TET1CD were then infected by the lentivirus. Infected cells were isolated by 4 days of hygromycin selection (40 μg/mL) and then amplified and frozen down. As described in the text targeted clones were treated with doxycycline for 10

days prior to methylation analysis and RNA expression analysis at the hESC stage. For *PAX6* and *SOX10* targeted demethylation the cells were also differentiated using the standard NE differentiation protocol.

### *2.4.19 Generation of TET1-3xFlag tagged hESCs*

We targeted the TET1 N-terminus to generate a TET1-3xFlag line. 4 gRNAs were tested and gRNA 2, which showed the greatest efficiency was used for targeting (Appendix 1A). A 200 base pair single strand DNA donor in which a 69 base pair 3xFlag sequence is inserted immediately after the ATG start and separated from the rest of the coding sequence using a 12 base pair linker sequence. Targeted clones were identified using PCR primers that flank the target site (Appendix 1B). Clones were sequenced to verify there were no mutations in the coding sequence on both TET1 alleles (Appendix 1C) and an immunoprecipitation was performed to ensure that immunoprecipitation of the Flag tag leads to an increase in the overlapping band of TET1 (Appendix 1D).

### *2.4.20 ChIP-qPCR*

Confluent hESCs from a 100 mm dish were used for 2 immunoprecipitations. ChIP was performed using the SimpleChIP Plus Enzymatic Chromatin IP kit (Cell Signaling Technology, 9003S) according to the manufacturer protocols. The antibodies used for ChIP were: H3 (Cell Signaling Technology, 9003S), IgG (Cell Signaling Technology, 9003S), TET1 (Genetex, GTX627420), DNMT1 (Active Motif, 39204). DNMT3A (Abcam, ab2850), and DNMT3B (Novus Biologicals, NB100-56514). Primers for ChIP-qPCR are provided in Table 2.3.

### *2.4.21 Statistical Analysis*

Data are presented as mean ± STD (unless otherwise noted) and were derived from at least three independent experiments. Data on replicates (*n*) is given in figure legends. Statistical analysis was performed using the Student's *t*-test (comparing two groups) or one-way comparison ANOVA (comparing multiple groups against one group). Variance was similar between the groups that are being compared. Distribution of the raw data approximated normal distribution (Kolmogorov–Smirnov normality test) for data with sufficient number of replicates to test for normality. No method of randomization was performed and investigators were not blinded to the genotypes of cell lines. Statistical analysis for whole genome bisulfite sequencing, ERRBS and RNA-Seq data are described in those separate methods section.

### *2.4.22 Data Availability*

All sequencing datasets are available under the Gene Expression Omnibus accession GSE89728.

# CHAPTER 3: Conclusions & Perspectives

## 3.1 SUMMARY

The TET proteins (TET1, TET2 and TET3) catalyze the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), which can serve as an intermediate in DNA demethylation. Previous studies have focused on TET action at enhancers and found clear examples in which loss of the TET proteins causes hypermethylation and dysregulated gene expression. However, the importance of TET activity at promoters remains ambiguous. Bivalent promoters have the H3K4me3 and H3K27me3 marks on the same or adjacent nucleosomes. They are present in embryonic stem cells and adult stem cells, and they typically regulate developmental genes. Similar to active promoters, bivalent promoters are hypomethylated; however, unlike active promoters, which support productive gene transcription, the hypomethylated state at bivalent promoters is associated with negligible gene transcription. Thus at bivalent promoters the importance of DNA methylation for gene expression regulation is not readily apparent. Moreover it is unclear if the DNA methylation state of these promoters is actively regulated and whether this regulation is necessary to maintain developmental regulators in a “poised state”.

The key findings from this project were:

1. ***TET* inactivation leads to hypermethylation of bivalent promoters in human embryonic stem cells.** We inactivated the

*TET* genes in hESCs (TKO hESCs) and observed by whole genome bisulfite sequencing (WGBS) that bivalent promoters showed the greatest methylation increase compared to other genomic regions. Furthermore enhanced reduced representation bisulfite sequencing (ERRBS) in two independent TKO lines showed that a similar subset of approximately half of bivalent promoters become hypermethylated after loss of the TET proteins.

2. **Hypermethylation of bivalent promoters leads to gene repression upon differentiation.** The consequences of bivalent promoter hypermethylation on gene transcription are not known. At the hESC stage, in which expression of bivalent genes is low, we observed that some hypermethylated bivalent genes showed variable changes (both up and down) in gene expression. Upon differentiation hypermethylated bivalent genes (*PAX6*, *SOX10*) critical for the neural lineage failed to be expressed leading to defects in differentiation.
3. **Loss of TET1 and 5hmC is directly responsible for hypermethylation at bivalent promoters.** TET1 ChIP-Seq and global 5hmC profiling showed that TET1 and 5hmC localize to bivalent promoters in WT cells. Furthermore TET1 and 5hmC positive sites overlapped significantly with hypermethylation after *TET* inactivation.
4. **The *de novo* methyltransferase DNMT3B is responsible for the gain of methylation in bivalent promoters.** 57% of the regions that gained methylation in TKO hESCs lost methylation

after inactivation of *DNMT3B* in TKO hESCs. Furthermore ChIP-Seq of DNMT3B in TKO hESCs showed that DNMT3B binding correlated with a greater methylation increase in TKO hESCs and greater methylation loss in QKO (TET1/2/3 TKO, DNMT3B KO) hESCs.

Whereas previous work has focused on the bivalent histone marks for regulating gene expression from bivalent promoters this project showed that the methylation of bivalent promoters is also actively regulated and significantly impacts gene expression upon differentiation. This also expands the functional role of the TET proteins during development. Besides their documented function in the active demethylation of enhancers we show here that they are also critical to maintain hypomethylation at bivalent promoters.

## 3.2 FUTURE DIRECTIONS

### *3.2.1 Identify Proteins that may be Involved in the Recruitment of DNMT3B and TET1 to Bivalent Promoters*

It is unclear why certain genomic loci are more prone to dynamic changes in DNA methylation and what are the molecular mechanisms that underlie competition between the TET and DNMT proteins to determine methylation levels at these loci. Focusing specifically on bivalent promoters it would be interesting to determine whether recruitment of the TET and DNMT proteins by a third protein may explain the competition between TET and DNMT (Appendix 2A). The first candidate we chose to study is the

polycomb repressive complex, PRC2. PRC2 itself is present at bivalent loci as it is responsible for deposition of the H3K27me3 mark. Furthermore previous studies have found interaction between PRC2 and TET1 and PRC2 and DNMT3A/DNMT3B in mESCs (Neri et al., 2013a; Neri et al., 2013b). We generated a 3x flag tagged TET1 hESC line to first perform co-immunoprecipitation experiments (Appendix 2B). We detected interaction between PRC2 and DNMT3B but limited interaction between PRC2 and TET1 (Appendix 2C). It is possible that TET1 is recruited by another protein instead of PRC2. It would thus be useful to perform mass spectrometry analysis of the 3xFlag-TET1 line to identify potential interacting partners that may function to recruit TET1 to specific genomic regions.

We decided to investigate further the interaction between PRC2 and DNMT3B. We had already observed that inactivating DNMT3B in the TKO background reduced methylation at bivalent loci, among other regions. We wanted to determine whether inactivation of PRC2 in the TKO background might provide more precise loss of DNA hypermethylation at bivalent promoters as it may abrogate DNMT3B recruitment specifically to these regions. We decided to target two different components of PRC2, the catalytic component EZH2 and the DNA binding component SUZ12. We generated CRISPRs that could produce indel mutations at the start codons of both *EZH2* and *SUZ12*; however, upon targeting we obtained few clones that had frameshift mutations in both alleles (Appendix 3A). It is possible that inactivating *EZH2* has a lethal phenotype as suggested by previous studies (Collinson et al., 2016). We continued to culture the WT and TKO lines that showed frameshift mutations in both alleles of *SUZ12*. Unfortunately upon further sequencing we found that the mutations in all of

the lines lie in a *SUZ12* pseudogene that shares the first 12 exons of the functioning *SUZ12* gene (Appendix 3B). We are currently generating additional targeting CRISPR gRNAs, to reattempt the inactivation of SUZ12 in the TKO background. We aim to identify interacting partners of TET1 and DNMT3B and determine whether they function to recruit these proteins to bivalent promoters. This may explain the specificity of hypermethylation that we observe in TKO hESCs.

### *3.2.2 Investigate Changes in Chromatin Environment Upon TET Inactivation*

We focused on changes in DNA methylation upon inactivation of the *TET* genes. It is likely however that there are additional changes in the chromatin as there is significant crosstalk between DNA methylation and other epigenetic modifications. We aim to investigate this further using genome-wide ChIP-Seq analysis in WT and TKO hESCs to determine whether the distribution of certain histone marks (H3K4me1, H3K4me3, H3K27me3, H3K27ac) and chromatin modifiers (EZH1, SUZ12, DNMT1, DNMT3A) are altered upon TET inactivation. We will also perform ATAC-Seq to determine if there are changes in the chromatin accessibility for biologically relevant loci.

### *3.2.3 Investigate the Mechanisms Underlying Heterogeneity in the Gain of Methylation after TET Inactivation*

Among bivalent promoters we observed large variability in the gain of methylation. Approximately half of the bivalent promoters in two hESC backgrounds showed greater than 5% increase in methylation between TKO

and WT lines, and among these approximately 10% showed a greater than 60% increase in methylation. It would be interesting to explore why these particular promoters were vulnerable to such large increases in methylation. It is possible that transcription factor binding, local chromatin environment or chromatin 3D structure or configuration may influence whether a particular locus is protected from or vulnerable to aberrant hypermethylation. The TKO hESCs offer a useful platform as we could manipulate these variables in TKO hESCs to explore whether we can cause hypermethylation at previously resistant loci.

### 3.2.4 *Investigate Hypermethylation Patterns and Consequences in Other Progenitor Populations Upon TET Inactivation*

It is unclear whether the observations we made at the hESC stage regarding the hypermethylation of bivalent promoters is also observed in other progenitor cell types. Furthermore it is unknown whether hypermethylation at bivalent promoters in other progenitors would also produce differentiation defects due to impaired activation of the differentiation genes associated with these promoters. hESCs are an ideal platform to study this question because they have the potential to differentiate into a variety of tissues. We are currently working on an inducible system that would enable us to immediately inactivate the *TET* genes at different stages of differentiation and thus observe the effect of the loss of the TET proteins in different cellular contexts. In our first attempt to generate such a platform we infected our iCRISPR hESCs (in which Cas9 is doxycycline inducible) with a lentiviral construct containing the three TET gRNAs in tandem, each under the control of its own individual U6 promoter

(Appendix 4A). After establishing one clone we found that doxycycline treatment led to progressive depletion of 5hmC, consistent with the loss of the TET proteins (Appendix 4B). Upon T7 endonuclease analysis we found that indel mutations were generated in the *TET1* and *TET3* loci, but not at the *TET2* locus (Appendix 4C). We are currently investigating why there is no functional TET2 gRNA and exploring alternative strategies to develop a true inducible TKO line.

We believe the question of whether our observations are reproducible in other progenitor cells is particularly important due to the prevalence of mutations in the DNA methylation and demethylation pathways in a variety of cancers. Ideally we would want to investigate cell populations that are particularly prone to transformation upon disruption of TET or TET-interacting pathways, such as hematopoietic stem cells and neural stem cells. A better understanding of the molecular mechanisms that drive cancer progression could potentially enable the design of more efficacious therapeutics.

## 3.3 CONCLUSIONS

The TET enzymes oxidize 5-methylcytosine to 5-hydroxymethylcytosine, which can lead to DNA demethylation. Previous studies in mice and mESCs have suggested that the TET proteins are critical for proper embryonic development and differentiation. However, the mechanisms underlying these defects have proven elusive to isolate. It is also unclear whether observations found in mice and mESCs will be similar in studies using

hESCs. Finally the mechanisms that produce this locus specific hypermethylation and the factors that lead to the variable gain of methylation between loci are unknown.

Here we show that we can produce viable hESCs after *TET* inactivation. Although the TKO hESCs do not show a global methylation change they do exhibit prominent bivalent promoter hypermethylation without overall corresponding gene expression decrease in the undifferentiated state. Focusing on the bivalent *PAX6* locus, we find that TET1 binding and 5hmC signal present at the P0 promoter in WT hESCs is lost in TKO hESCs. Instead TKO hESCs show increased DNMT3B binding that is associated with *PAX6* P0 promoter hypermethylation, which then precipitates a neural differentiation defect and failure of *PAX6* induction during differentiation. Using recently developed targeted demethylation strategies we can achieve locus-specific demethylation of the PAX6 P0 promoter, which improves expression of PAX6 after differentiation. Furthermore global inactivation of *DNMT3B* in TKO hESCs partially reverses the hypermethylation at the *PAX6* promoter and improves differentiation to neuroectoderm. Notably we found that approximately 90% of the bivalent promoters that gained methylation after *TET* inactivation lose methylation upon depletion of DNMT3B. This suggests that DNMT3B is responsible for the gain of methylation at these bivalent promoters and is even acting to preserve this aberrant hypermethylation during hESC culture.

Finally TET1 and 5hmC ChIP-Seq showed that TET1 or 5hmC is found

in approximately 90% of the bivalent promoters that gain methylation in TKO hESCs. This indicates that the hypermethylation at bivalent promoters is the direct consequence of losing TET1 and 5hmC. Although DNMT3B ChIP-Seq showed limited binding of DNMT3B to promoters, we did find that DNMT3B binding correlates with gain of methylation after TET inactivation. We are currently exploring ways in which DNMT3B may be recruited to bivalent promoters and investigating whether technical limitations with ChIP-Seq prevented us from detecting these events.

Taken together we observed that TKO hESCs show significant bivalent promoter hypermethylation. Using the bivalent *PAX6* P0 promoter we found that this is associated with reduced gene expression of the associated gene upon differentiation. This hypermethylation is lost, both at the *PAX6* promoter and bivalent promoters in general, upon inactivation of *DNMT3B*. Finally this hypermethylation in TKO hESCs correlates with TET1 binding in WT cells and DNMT3B binding in TKO hESCs. From these results we conclude that TET proteins safeguard bivalent promoters from silencing by *de novo* methylation to ensure robust lineage-specific transcription upon differentiation.

## CHAPTER 4: Generation of Pluripotency Reporter hESCs

## 4.1 INTRODUCTION

Human embryonic stem cells (hESCs) are capable of unlimited self-renewal in culture while maintaining the potential to differentiate into any cell type present in the human body and thus provide researchers great opportunities for human developmental studies, disease modeling, and cell-replacement therapies (Zhu and Huangfu, 2013). All these applications benefit from lineage-specific knockin reporters that allow real-time observation of gene-expression dynamics, cell-lineage tracing, and isolation of a specific cell population of interest from a heterogeneous differentiation culture for downstream analysis. However, creating knockin alleles in hESCs is usually a lengthy and technically challenging process. Because of the low efficiency of homologous recombination, the donor vector needs to contain a drug-resistance gene for enrichment of cells with the correct integration. Due to the concern that the insertion of a drug-resistance cassette may interfere with the expression of the reporter gene or neighboring genes, it is usually necessary to remove the drug-resistance cassette through a second electroporation step followed by isolation of clonal lines and further characterization (Davis et al., 2008). Thus, substantial time and effort is needed to generate a knockin reporter hESC line through this two-step procedure.

The development of engineered “genomic scissors” that introduce site-specific DNA double-strand breaks (DSBs), including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs),

and more recently the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system, has greatly facilitated gene targeting in hESCs (Kim and Kim, 2014). Repair of a DSB by non-homologous end joining (NHEJ) often results in insertion and/or deletions (indels) that can be used to knock out a target gene in hESCs (Ding et al., 2013a; Ding et al., 2013b; Gonzalez et al., 2014). Alternatively, homology-directed repair (HDR) can be employed to efficiently incorporate exogenous sequences such as a fluorescent reporter into a specific genomic locus in hESCs (Hockemeyer et al., 2009; Hockemeyer et al., 2011; Merkert et al., 2014). Despite the significant improvement, a drug-resistance cassette is still required for generating knockin reporters of genes that are not expressed in undifferentiated hESCs.

To overcome these limitations, we made use of the CRISPR/Cas system, in which the CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) duplex or a single chimeric guide RNA (gRNA) recognizes a 20-nucleotide (nt) DNA sequence upstream of the 5′-NGG-3′ protospacer adjacent motif (PAM) and directs the DNA endonuclease Cas9 for site-specific cleavage (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013). Based on this, we have developed an efficient genome-editing platform in hESCs, which we named iCRISPR (Gonzalez et al., 2014). Through TALEN-mediated gene targeting in the *AAVS1* locus, we have created hESC lines (referred to as iCas9 hESCs) that allow robust, doxycycline-inducible expression of Cas9. By transfecting iCas9 hESCs with gRNAs, the

iCRISPR system enables efficient NHEJ-mediated gene disruption as well as HDR-mediated precise nucleotide modifications in the presence of short single-stranded DNA (ssDNA) donors (~100 nt).

We reasoned that the iCRISPR system would also facilitate the generation of knockin reporter alleles using longer double-stranded (dsDNA) donors and may further enable the identification of correctly targeted hESC lines without drug selection. Thus, this work explores the utility of iCRISPR for targeting fluorescent reporters into two endogenous loci, *OCT4* (*POU5F1*) and *NANOG*, and demonstrates the generation of knockin hESC lines without drug selection. Further characterization of the *OCT4* (*POU5F1*) reporter lines confirmed the creation of multiple hESC reporter lines with no undesired mutations in the targeted loci or any potential off-target sites analyzed, supporting the broad application of this approach for efficient generation of knockin alleles in hESCs.

## 4.2 RESULTS

### 4.2.1 CRISPR/Cas-Mediated Targeting of the OCT4 Locus

We first evaluated the efficiency of the CRISPR/Cas system for making knockin reporter alleles by targeting the *OCT4* locus using drug selection. HUES8 hESCs were co-electroporated with two plasmids: one expressing Cas9 and a crRNA/tracrRNA duplex targeting *OCT4* and the other containing the fluorescent reporter and a drug-resistance cassette as the HDR template (Appendix 5A). We used a donor plasmid, 2A-eGFP-

PGK-Puro (Hockemeyer et al., 2011), in which the last *OCT4* coding codon is fused in frame with a 2A sequence followed by eGFP (2A-eGFP) and a loxP-flanked (floxed) puromycin-resistance gene expressed from the constitutive PGK promoter (PGK-Puro) (Appendix 5A). This strategy minimizes any potential impact on the endogenous protein and is applicable to targeting both silent and expressed genes. CRISPR gRNAs that target the stop codon of *OCT4* were cloned into the px260 vector (Appendix 5B) and then tested for efficiency in forming indel mutations (Appendix 5C). PCR and Southern blot analysis identified eight correctly targeted clones without random transgene integration from a total of 288 puromycin-resistant clones screened (Appendix 5D and Appendix 6A). The targeting efficiency (2.8%) was comparable to the efficiencies observed with TALENs and ZFNs using similar targeting strategies (Hockemeyer et al., 2009; Hockemeyer et al., 2011). Despite correct targeting, we failed to detect eGFP expression in any of the targeted lines. This is likely caused by the presence of the drug-resistance cassette as observed for other genes (Davis et al., 2008). Indeed, after Cre-mediated excision of the PGK-Puro cassette, all resulting *OCT4-eGFP* lines showed proper co-expression of eGFP with pluripotency markers OCT4, SOX2, and NANOG (Appendix 6B-D). These results highlight the necessity of removing the drug-resistance cassette for proper reporter gene expression.

### 4.2.2 Selection-free Targeting of the *OCT4* Locus using a Mini-vector Donor

To further explore the possibility of making knockin reporter alleles without drug selection, we designed a “mini-vector” donor plasmid, 2A-mOrange, which is similar to 2A-eGFP-PGK-Puro except that there is no PGK-Puro cassette and eGFP was replaced by mOrange (Appendix 7A). We also replaced the crRNA/tracrRNA duplex cr1-dp with the single gRNA cr1 targeting the same sequence (Appendix 7B), as the chimeric version works more efficiently than the original duplex design (Jinek et al., 2012). Similar to the experiment with the 2A-eGFP-PGK-Puro donor, we co-electroporated HUES8 hESCs with a plasmid expressing Cas9/gRNA and the new 2A-mOrange mini-vector (Appendix 7A). In contrast to the absence of fluorescence after integration of the 2A-eGFP-PGK-Puro cassette, integration of the 2A-mOrange cassette resulted in mOrange expression in ~0.001% of cells as detected by fluorescence-activated cell sorting (FACS) (Appendix 7C). One may enrich mOrange-expressing cells for establishing *OCT4* reporter lines. However, this low efficiency is impractical for genes not expressed in undifferentiated hESCs, as one has to rely on randomly picking individual colonies to establish clonal lines.

Our recent study shows that the iCRISPR platform enables efficient gene editing using short ssDNA HDR templates (Gonzalez et al., 2014), prompting us to further optimize the iCRISPR platform for HDR using longer circular dsDNA donor vectors. After optimizing transfection conditions, we

co-transfected doxycycline-treated HUES8 iCas9 cells twice in 2 days with the *OCT4* cr1 gRNA and the 2A-mOrange mini-vector using Lipofectamine 3000 (Appendix 7D, Appendix 8A-B performed by Zhengrong Zhu). FACS analysis identified ~0.4% mOrange-expressing cells from the HUES8 iCas9 targeting and ~.24% mOrange-expressing cells from the MEL1 iCas9 targeting (Appendix 7E), supporting the general utility of this new approach in diverse human pluripotent stem cell (hPSC) backgrounds. In contrast electroporation of the HUES8 iCas9 without doxycycline treatment did not show a detectable number of mOrange positive cells (Appendix 7F). The much-improved efficiencies can be partially attributed to the integration of Cas9 in the genome as a ~5- to 6-fold increase of mOrange+ cells was observed compared to the control condition where iCas9 hESCs (not treated with doxycycline) were transfected with Cas9/gRNA and the donor vector using Lipofectamine 3000 (Appendix 7F). The use of Lipofectamine transfection also substantially increased the targeting efficiency compared to the electroporation method (Appendix 7E-F). There may be other ways to improve the transfection efficiency (e.g., through nucleofection) to achieve similar results with or without the use of iCas9 hESCs (Byrne et al., 2015). We picked ten colonies from individual FACS-isolated mOrange+ cells and identified six correctly targeted clones by PCR and Southern blot analysis (Appendix 9A-B). All six lines co-expressed mOrange with pluripotency markers such as OCT4, SOX2, and NANOG and displayed normal hESC morphology (Appendix 9C). We further examined the *OCT4-mOrange* hESC

reporter lines along with the *OCT4-eGFP* lines for reporter gene expression after differentiation. After 3 days of treatment with BMP4 and SB431542, a TGFβ inhibitor (Hou et al., 2013), hESCs exhibited a differentiated morphology, and eGFP and mOrange expression were downregulated in the respective *OCT4-eGFP* and *OCT4-mOrange* hESC reporter lines with concomitant loss of endogenous OCT4 expression as determined by immunostaining and FACS analysis (Appendix 10A-B). Thus, the *OCT4-eGFP* and *OCT4-mOrange* reporters faithfully reflect endogenous gene expression during the maintenance and differentiation of hESCs.

We next investigated whether the relatively high targeting efficiency was achieved at the expense of undesirable mutations at the *OCT4* locus or any off-target sites. All eight *OCT4-eGFP* and six *OCT4-mOrange* lines examined showed the expected sequence at the junction between the endogenous *OCT4* sequence and the inserted sequence. This is reassuring, as we made sure that the donor template did not contain the CRISPR target sequence to prevent undesired mutagenesis after reporter gene integration. However, Indel mutations were detected in the non-targeted allele in two of the six *OCT4-mOrange* reporter lines examined (Appendix 10C). These findings underscore the necessity of thorough sequence analysis for eliminating clones with undesired mutations in the non-targeted allele, a point not widely recognized with the CRISPR/Cas-mediated targeting strategy. We also sequenced seven predicted off-target sites based on the 12-bp seed sequence important for target recognition (Jiang and Pugh,

2009; Jinek et al., 2012). Examination of six *OCT4-mOrange* and eight *OCT4-eGFP* lines revealed no mutations except that three *OCT4-eGFP* lines carried mutations at the *POU5F1P4* locus, which shares the same 20-nt target sequence with the intended target (Table 4.1).

**Table 4.1 Sequencing analysis of potential off-target sites in *OCT4-eGFP* and *OCT4-mOrange* reporter lines**

| Gene | CRIPSR Target Sequence-PAM | Sequencing results in *OCT4-eGFP* reporter lines | | | | | | | | | Sequencing results in *OCT4-mOrange* reporter lines | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 4 | 5 | 6 | 7 | 9 | 10 | | 3 | 5 | 6 | 7 | 8 | 10 |
| ***OCT4*** | TCTCCCATGCATTCAA ACTG-AGG | | | | | | | | | | | | | | | |
| ***DLG2*** | AAGCTCAGGCATTCAA ACTG-TGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |
| ***GPHN*** | GCCCTCAGGCATTCA AACTG-TGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |
| ***IMMP2L*** | TAGACTTAGCATTCAA ACTG-AGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |
| ***PEMT*** | GCACCCTAGCATTCAA ACTG-TGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |
| ***POLR2J4*** | AAGGAGAAGCATTCAA ACTG-TGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |
| ***POU5F1P4*** | TCTCCCATGCATTCAA ACTG-AGG | - | HET* | - | HOM** | - | HOM** | - | - | | - | - | - | - | - | - |
| ***SLC33A1*** | CAGAAATGGCATTCAA ACTG-CGG | - | - | - | - | - | - | - | - | | - | - | - | - | - | - |

-: Both alleles are wild-type; HET: One allele has a mutation; HOM: Both alleles have mutations;
* 6 bp insertion; ** 4 bp deletion; *** 10 bp deletion.

### 4.2.3 Selection-free Targeting of the *NANOG* Locus using a Mini-vector Donor

Finally we targeted the *NANOG* locus to generate an individual *NANOG* reporter line and a dual reporter, which has fluorescent proteins, targeted to both the *OCT4* and *NANOG* loci. For the *OCT4* locus we continued to use the 2A-mOrange mini-vector that we previously used (Appendix 7A). For the *NANOG* locus we designed a mini-vector donor plasmid, 2A-ppGFP, which does not have a drug selection cassette and in

which the 2A-ppGFP is flanked by sequences homologous to the *NANOG* loci (Appendix 11A). We tested a number of CRISPR gRNAs that target the stop codon of *NANOG* and identified 3 that efficiently produced indel mutations (Appendix 11B). CRISPR gRNA 330-1 was used for targeting experiments. We tested whether we could target 2A-mOrange and 2A-ppGFP into the *OCT4* and *NANOG* loci respectively using a single transfection that includes *OCT4* and *NANOG* targeting gRNAs and donor vectors. Unfortunately upon FACS Analysis we did not observe any mOrange and ppGFP double positive cells indicating that the procedure is not efficient enough for two integration events in a single cell (Appendix 11C). We decided to isolate GFP positive cells to expand. These cells would be used as single *NANOG*-ppGFP reporters and also used for *OCT4* 2A-mOrange targeting to produce a dual reporter cell line. Genomic DNA for one such GFP positive clone was isolated and upon sequencing it showed correct integration of the 2A-ppGFP sequence into the *NANOG* locus (Appendix 11D). This clone was targeted with the *OCT4* gRNA and the 2A-mOrange donor vector. Isolation, expansion and sequencing of a mOrange and GFP double positive clone showed correct integration of the 2A-mOrange sequence into the *OCT4* locus (Appendix 11E).

## 4.3 DISCUSSION

Here, we demonstrate the generation of hESC reporter lines without

the use of drug selection for both active and silent genes through the use of the iCRISPR system. Selection-free gene targeting eliminates the need for removal of drug-resistance cassette after identification of correctly targeted clones, and mini-vector donors with short homology arms (~500–1,000 bp each) are convenient to make. Thus, this method significantly reduces the time and effort required for establishing hESC reporter lines. Additionally, conventional gene-targeting strategies typically use the Cre-loxP strategy to remove drug-resistance cassettes, which leaves behind a 34-bp loxP “scar” in the endogenous locus. Although not an issue in most cases, this residual sequence could interfere with the expression of the targeted gene in some situations (Meier et al., 2010). In comparison, our strategy eliminates the need for the selection cassette and thereby minimizes the alteration of the endogenous locus. Although we focused on creating promoter-fusion reporters, the same knockin approach can be readily applied to making protein-fusion reporters for visualizing protein subcellular localization, precisely deleting or replacing specific genomic sequences, and introducing or correcting disease-associated mutations.

It is known that gene-targeting efficiencies can vary significantly depending on the target locus, though the exact reason is unclear. Traditional gene targeting relies on drug selection; thus, the relative targeting efficiencies after drug selection depends, at least in part, on the expression of the drug-resistance gene from the targeted locus. Because the expression of drug-resistance gene may differ significantly between

expressed and silent loci, the relative targeting efficiency after drug selection for a lineage-specific gene may appear much lower compared with a pluripotency gene. For certain loci, the drug-resistance gene may be expressed at such low levels that hinder the identification of a correctly targeted clone using the drug-selection method (Rostovskaya et al., 2012). Because our targeting approach obviates drug selection, it may overcome such bias and facilitate the generation of reporter alleles for genes that were previously difficult to target. One may further use this ability to measure absolute targeting frequencies to compare HDR efficiencies across different genomic contexts.

There have been concerns about potential off-target mutagenesis with the CRISPR/Cas system (Cho et al., 2014; Fu et al., 2013; Hsu et al., 2013; Pattanayak et al., 2013). Our analysis so far did not reveal any off-target mutations at sites without perfect complementarity with the CRISPR target sequence. However, we cannot exclude the possibility of off-target mutations elsewhere in the genome, and a thorough analysis may be necessary before the reporter lines are used in future studies. The CRISPR/Cas system is continuously improved with the development of better algorithms for CRISPR design and off-target prediction. It is reassuring that a recent high-coverage whole-genome sequencing study failed to detect significant incidence of off-target mutations in CRISPR-targeted hPSC lines (Veres et al., 2014). On the other hand, we noticed that some correctly targeted clones carried mutations in the non-targeted allele,

though the targeting efficiency is sufficiently high that one could simply discard the minority of clones carrying mutations. One may also target intron regions with low-sequence conservations to further mitigate any concerns associated with Indel mutations in the non-targeted allele.

Our selection-free targeting approach enables rapid generation of knockin reporter lines, though it also requires either using established iCas9 cells or creating new iCas9 lines in a desired hPSC background. The upfront effort for generating iCas9 cells is relatively small due to the efficient TALEN-mediated *AAVS1*-targeting approach, and it is possible to establish an iCas9 line in about 1 month (Gonzalez et al., 2014). Once an iCas9 line is established, it can be used for making different types of reporters. Our previous study has shown that Cas9 activity is tightly regulated by doxycycline treatment, and established iCas9 lines exhibit no apparent chromosomal aberrations or defects in the maintenance of the pluripotent state (Gonzalez et al., 2014). A recent study also observed no adverse effects in constitutive Cas9-expressing mice (Platt et al., 2014). An additional benefit of using iCas9 hPSCs for making reporter lines is that the cells can be conveniently used for a variety of downstream genetic studies using gene-editing approaches we already established (Gonzalez et al., 2014). Thus, we expect this selection-free knockin strategy to further facilitate the use of hESCs for developmental studies, disease modeling, and cell-replacement therapy.

## 4.4 MATERIALS & METHODS

### 4.4.1 Generation of constructs

To generate CRISPR plasmids expressing Cas9 and the crRNA/tracrRNA duplex targeting specific genomic loci, 30-bp protospacer sequences were cloned into the px260 (Addgene: 42229) as previously described (Cong et al., 2013). Briefly, vectors were digested with BbsI, treated with Antarctic Phosphatase, and gel purified. A pair of oligonucleotides containing the 30-bp protospacer sequence was annealed generating BbsI overhangs, and cloned into BbsI-digested, dephosphorylated vectors. The same procedure was also used to generate CRISPR plasmids expressing Cas9 and the chimeric gRNA with the difference that a pair of oligonucleotides containing the 20-bp protospacer sequence was cloned into the piCRg Entry plasmid (Addgene: 58904). The sequences for all oligonucleotides used for generating the CRISPR constructs are listed in Table 4.2.

For generation of *OCT4-eGFP* hESC reporter lines, the OCT4-2A-eGFP-PGK-Puro plasmid (Addgene: 31938) was used. The left homology arm is 697 bp and the right homology arm is 699 bp. The *OCT4-mOrange* hESC reporter lines were made using the OCT4-2A-mOrange targeting vector. To generate the OCT4-2A-mOrange targeting vector, an NheI-2AmOrange-AscI cassette was PCR amplified using the mOrange-pBAD plasmid template (Addgene: 54751) and primers Nh2AOr-F and AscOr-

R. Next, the NheI-2A-mOrange-AscI PCR fragment and the OCT4-2A-eGFP-PGK-Puro plasmid were digested with NheI and AscI and ligated. The NANOG-ppGFP hESC reporter lines were made using the NANOG-2A-ppGFP targeting vector. To generate the NANOG-2A-ppGFP targeting vector, an Nhe1-2AppGFP-BglII cassette was PCR amplified using a pCR2_NhPTVppGFP_BglII, with primers NhPTV_F and Bg_ppGFP_R. *NANOG* homology arms were PCR amplified from HUES8 genomic DNA and cloned into the pBlueScript SKII (+) plasmid to generate the pBS-NANOG plasmid. The left homology arm is 811 bp and the right homology arm is 1014 bp. The reverse primer for the 5' homology arm contained a NheI site and the forward primer for the 3' homology arm contained a BglII site. The Nhe1-2AppGFP-BglII cassette and the pBS-NANOG plasmid were digested by NheI and BglII and ligated to generate the NANOG-ppGFP targeting vector. The primers used for construction of donor plasmids are given in Table 4.3.

**Table 4.2. PCR Primers for generating templates for gRNA *in vitro* transcription**

| Gene | CRISPR target sequence (5' of PAM) | | CRISPR specific forward primer (5' to 3') |
|---|---|---|---|
| ***OCT4*** | **cr1** | TCTCCCATGCATTCAAACTG | **F:** TAATACGACTCACTATAGGGTCTCCCATGCATTCAAACTG |
| ***NANOG*** | **cr1** | ACTCATCTTCACACGTCTTC | **F:** TAATACGACTCACTATAGGGACTCATCTTCACACGTCTTC |
| **Universal reverse primer** | | | **gRNA-R:** AAAAGCACCGACTCGGTGCC |

**Table 4.3. PCR Primers for donor plasmid construction**

| Donor plasmid | Primer sequence (5' to 3') | |
|---|---|---|
| *OCT4* | Nh2AOr-F | TTCTAGCTAGCACCGGTGCCACGAACTTCTCTCT |
| | AscOr-R | CTTATGGCGCGCCTTACTTGTACAGCTCGTCCAT |
| *NANOG* | 5'HA-F | GGTACCTAGCCTGTAGCGAACTCCTG |
| | 5'HA-R | GCTAGCCACGTCTTCAGGTTGCATGTT |
| | 3'HA-F | AGATCTAGATGAGTGAAACTGATATTACT |
| | 3'HA-R | GCGGCCGCCTTCTCCACCCCAACCAAAAA |
| | 2A-ppGFP-F | CAGTGCTAGCGCCACTAACTTCTCCCTGTT |
| | 2A-ppGFP-R | TCAGTCGGCGCGCCTTACTTGTACAGCTCGTCCA |

**Table 4.4. PCR primers for genotyping**

| Gene | Primer sequence (5' to 3') |
|---|---|
| *OCT4* | **F:** AGTCCAAAGCTTGCCCTTGTCACC |
| | **GFP-R:** AGCTCCTCGCCCTTGCTCACC |
| | **mOr-R:** GAGGTGATGTCCAACTTGATGCCGA |
| *NANOG* | **F:** TTTGTAGACACAGTGTTTCCTC |
| | **R:** GCAGTGTCCTCAGCTAATTTC |

**Table 4.5. PCR primers for generating Southern blot probes**

| Gene | Probe | Primer sequence (5' to 3') |
|---|---|---|
| *OCT4* | External | **F:** CCAGTGGGAGTCAGTGGGGCT |
| | | **R:** GTCCGACTCCCAAGAGGTCACAG |
| | Internal Puro | **F:** TGACCGAGTACAAGCCCACGG |
| | Internal mOrange | **R:** TCGTAGAAGGGGAGGTTGC<br>**F:** TGGCCATCATCAAGGAGTTCA<br>**R:** CTTCTTCTGCATTACGGGGCCG |
| *NANOG* | External | **F:** ACCTCCTGGGTTCAAGGGATT |
| | | **R:** GAGGAAACACTGTGTCTACAAA |
| | Internal GFP | **F:** GTTCATCTGCACCACCGG |
| | | **R:** CGCGCTTCTCGTTGGGGT |

### 4.4.2 Cell culture

HUES8 (NIHhESC-09-0021) and MEL-1 (NIHhESC-11-0139) hESCs were cultured on irradiated mouse embryonic fibroblasts (iMEFs) feeder layers in DMEM/F12 medium (Life Technologies) supplemented with 20% KnockOut Serum Replacement (Life Technologies), 1X MEM Non-Essential Amino Acids (Life Technologies), 1X GlutaMAX (Life Technologies), 100U/ml Penicillin and 100 μg/ml Streptomycin (Gemini), 0.055 mM 2-mercaptoethanol (Life Technologies) and 10 ng/ml recombinant human

basic FGF (EMD Millipore). Cells were incubated at 37 °C with 5% (vol/vol) $CO_2$, and media was changed daily. Cultures were passaged at a 1:6 - 1:12 split ratio every 4 - 6 days using 0.05% trypsin/EDTA. 5 μM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals) was added into the culture medium when passaging or thawing cells.

293T cells were maintained in DMEM supplemented with 15% fetal bovine serum, 1X GlutaMAX, 1X MEM NEAA and 1mM Sodium Pyruvate (Life Technologies).

For differentiation of *OCT4* reporter lines, undifferentiated hESCs cultured on iMEF feeder layer were first adapted to the feeder-free E8 culture. Briefly, hESCs were passaged using 0.05% trypsin/EDTA and plated at a 1:3 split ratio on Matrigel (BD Biosciences) coated plates in E8 medium (Stem Cell Technologies). Cells were incubated at 37 °C with 5% (vol/vol) CO2, and media was changed daily. After 2 passages in E8 medium, the cells were differentiated: 2 days after passaging, cells were treated with 10 μM SB431542 (Sigma Aldrich) and 10 ng/mL BMP4 (R&D Systems) to initiate differentiation. After 3 days of this treatment, eGFP and mOrange fluorescence was analyzed by flow cytometry and fluorescence microscopy.

### 4.4.3 In vitro transcription of gRNAs

A T7 promoter was added to gRNA templates by PCR amplification on piCRg Entry vectors using CRISPR-specific forward primers and a

universal reverse primer gRNA-R (Table S2). T7-gRNA PCR products were used as templates for *in vitro* transcription using the MEGAshortscript T7 kit (Life Technologies). The resulting gRNAs were purified using the MEGAclear kit (Life Technologies), eluted in RNase-free water and stored at -80°C until use.

### 4.4.4 Electroporation

HUES8 hESCs were pre-treated with Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals) 24 hours before electroporation. On the day of electroporation, hESCs were disassociated into single cells with 0.25% Trypsin/EDTA and filtered through a 40 μM cell strainer to remove cell clumps. 10 million cells were re-suspended in 800 μL cold PBS and mixed with targeting and donor plasmids (10 μg CRISPR targeting plasmid and 40 μg donor plasmid). Cells were electroporated using the Gene Pulser XCell (Bio-Rad) at 250 V, 500 μF in a 0.4 cm Gene Pulser cuvette (Bio-Rad). Cells were recovered and re-plated on irradiated mouse embryonic fibroblast (iMEF) coated plates with ROCK inhibitor.

For the generation of *OCT4-eGFP* lines, two days after electroporation cells were treated with puromycin (0.5 μg/mL) for three days. Two weeks after electroporation individual colonies were picked and expanded for PCR genotyping, Southern blot analysis and sequencing for the establishment of reporter lines. Three of the correctly targeted clones (#1, #4 and #7) were electroporated with Cre recombinase following the

same procedure described above. After electroporation with 50 μg of Cre recombinase plasmid, the cells were recovered and replated on iMEF with ROCK inhibitor. 3 days after electroporation, GFP-expressing ($GFP^+$) cells were observed. $GFP^+$ cells were isolated by FACS and plated at a low density (2,000 cell/10cm dish) for subsequent picking and expansion of individual clones.

### 4.4.5 Transfection

HUES8 and MEL-1 iCas9 hESCs (Gonzalez et al., 2014) were treated with ROCK inhibitor and doxycycline one day before transfection. Different transfection reagents were tested for efficiency for delivering gRNAs and plasmid DNA into hESCs (Appendix 8). Lipofectamine 3000 was determined to be the most effective reagent for transfection. For transfection, cells were dissociated using TrypLE (Life Technologies), replated at 200,000 hESCs per well in iMEF-coated 12-well plates and transfected in suspension with gRNAs and donor plasmid. A second transfection was performed 24 hours later. Transfection of the gRNAs and donor plasmid into hESCs was performed using Lipofectamine 3000 (Life Technologies) following manufacturer's guidelines. For each targeting, gRNAs at a 10 nM final concentration and 5 μg of donor plasmid were used. Lipofectamine 3000 and gRNA + donor plasmid were diluted separately in Opti-MEM (Life Technologies), mixed together, incubated for 5 min, and added drop-wise to cultured hESCs.

### 4.4.6 Assessment of Indel mutations using the Surveyor Nuclease Assay

80-90% confluent 293T cells in 6-well dishes were transfected using JetPrime transfection reagent (Polyplus) following manufacturer's guidelines. For each well, a total of 2 μg CRISPR plasmid was transfected. After transfection the cells were incubated at 37 °C for 2 days prior to genomic DNA extraction. The Surveyor Kit (Transgenomic) was used to test the efficiency of each crRNA for producing indels. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's protocol. The genomic region flanking the CRISPR target site for each gene was PCR amplified. For Surveyor analysis 16 μL of PCR product (at 25 ng/μL) was denatured and re-annealed in JumpStart buffer to a total volume of 20 μL using the following protocol: 95 °C, 5 min; 95–85 °C at −2 °C/s; 85–25 °C at −0.1 °C/s; hold at 4 °C. 10 μL of hybridized PCR products was treated with 1 μL of Surveyor Enhancer S and 1 μL of Surveyor Nuclease S at 42°C for 60 minutes. The reaction was stopped by adding 1.2 μL of Stop solution to each tube. Products were then analyzed on 2.5% agarose gel and imaged with a Gel Doc gel imaging system (Bio-Rad). Quantification was based on relative band intensities using ImageJ. Indel mutation percentage was determined by the formula: $100 \times (1 - (1 - (b + c) / (a + b + c))^{1/2})$, where $a$ is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

### 4.4.7 Southern Blot Analysis

The external and internal probes were generated by PCR using the PCR DIG Probe Synthesis Kit (Roche). For the external probes we used HUES8 genomic DNA as the template. For internal probe generation we used the OCT4 2A-eGFP-PGK-Puro donor, OCT4-2A-mOrange donor and PDX1-2A-eGFP donor templates. For membrane hybridization, 5 µL of denatured DIG-labeled PCR product was added to 20 mL of hybridization buffer.

To identify correctly targeted hESC lines, genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen). For PCR genotyping the PCR primers listed in Table 4.4 were used. For southern blot analysis 5-10 µg of genomic DNAs was digested overnight with the appropriate restriction enzymes (OCT4 external and mOrange internal: BamHI, puromycin internal: EcoRI, PDX1 external: ApaLI, eGFP internal: ApaLI) and then migrated in 1% agarose gels. The gel was denatured, neutralized, and transferred overnight by capillarity on Hybond-N membranes (GE Healthcare) using 10x SSC transfer buffer. Hybridization with the external or the internal probe was carried out overnight at 65 °C. Probes were detected using an AP-conjugated DIG-Antibody (Roche) using CDP-Star (Roche) as a substrate for chemiluminescence as per manufacturer's instructions. The PCR primers used for generation of southern blot probes are given in Table 4.5.

### 4.4.8 Immunofluorescence Staining

For immunostaining, cells were fixed with 4% paraformaldehyde for 10 minutes, washed once with phosphate buffered saline (PBS) and permeabilized in PBST (PBS + 0.1% Triton) for 15 minutes. Blocking was done for 5 minutes at RT with blocking solution (5% donkey serum in PBST). Primary and secondary antibodies were diluted in blocking solution. Primary antibodies were incubated at RT for 1 hr. The following primary antibodies were used at a 1:100 dilution: OCT4 (mouse monoclonal, Santa Cruz sc-5279); NANOG (rabbit polyclonal, CosmobioJapan REC-RCAB0004P-F); SOX2 (goat polyclonal, Santa Cruz sc-17320), RFP (rabbit polyclonal, Life Technologies R10367), GFP (rabbit polyclonal, Life Technologies A-6455). The PDX1 antibody was used at a 1:500 dilution (goat polyclonal, R&D AF2419). After primary antibody staining the cells were washed three times with PBST and then incubated with the appropriate Molecular Probes Alexa Fluor dye conjugated secondary antibodies (Life Technologies) for 1 hr.

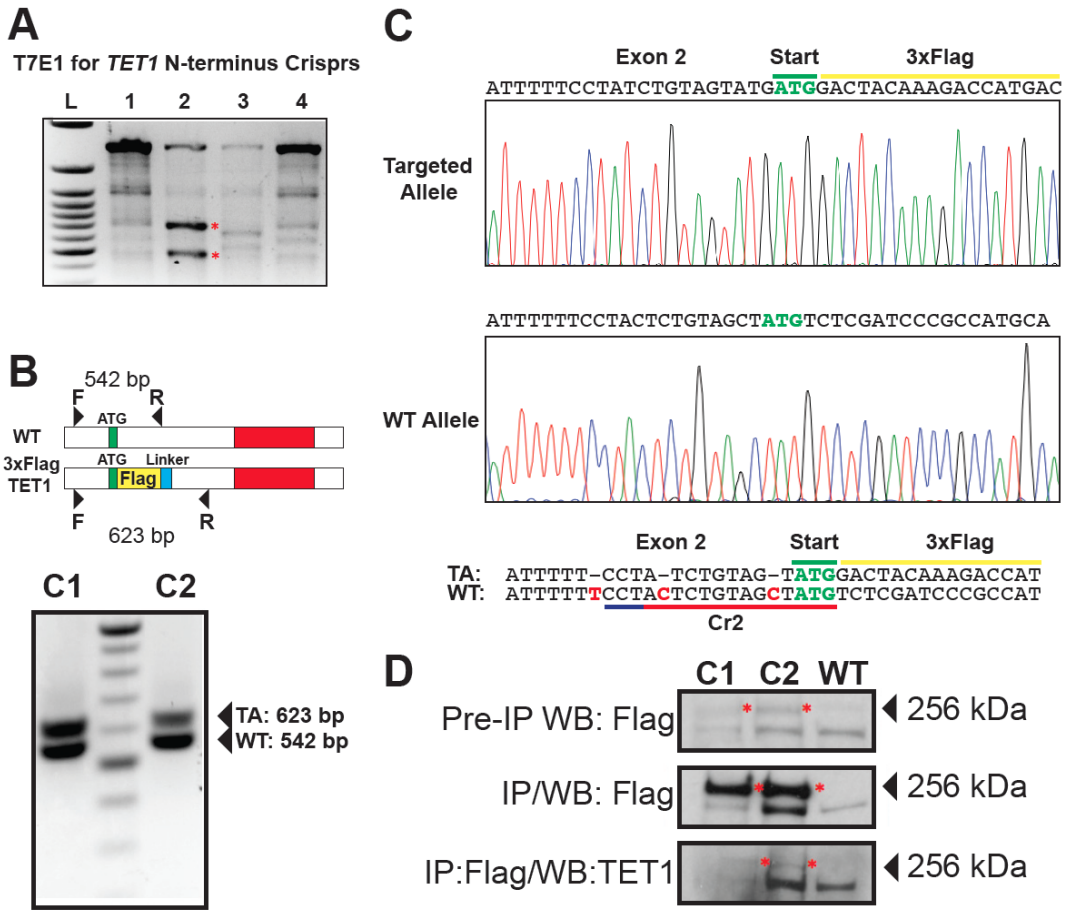
## 4.5 CONCLUSIONS

The development of new gene-editing tools, in particular the CRISPR/Cas system, has greatly facilitated site-specific mutagenesis in human embryonic stem cells (hESCs), including the introduction or correction of patient-specific mutations for disease modeling. However,

integration of a reporter gene into an endogenous locus in hESCs still requires a lengthy and laborious two-step strategy that involves first drug selection to identify correctly targeted clones and then excision of the drug-resistance cassette. Through the use of iCRISPR, an efficient gene-editing platform we recently developed, this study demonstrates a knockin strategy without drug selection for both active and silent genes in hESCs. Lineage-specific hESC reporter lines are useful for real-time monitoring of cell-fate decisions and lineage tracing, as well as enrichment of specific cell populations during hESC differentiation. Thus, this selection-free knockin strategy is expected to greatly facilitate the use of hESCs for developmental studies, disease modeling, and cell-replacement therapy.

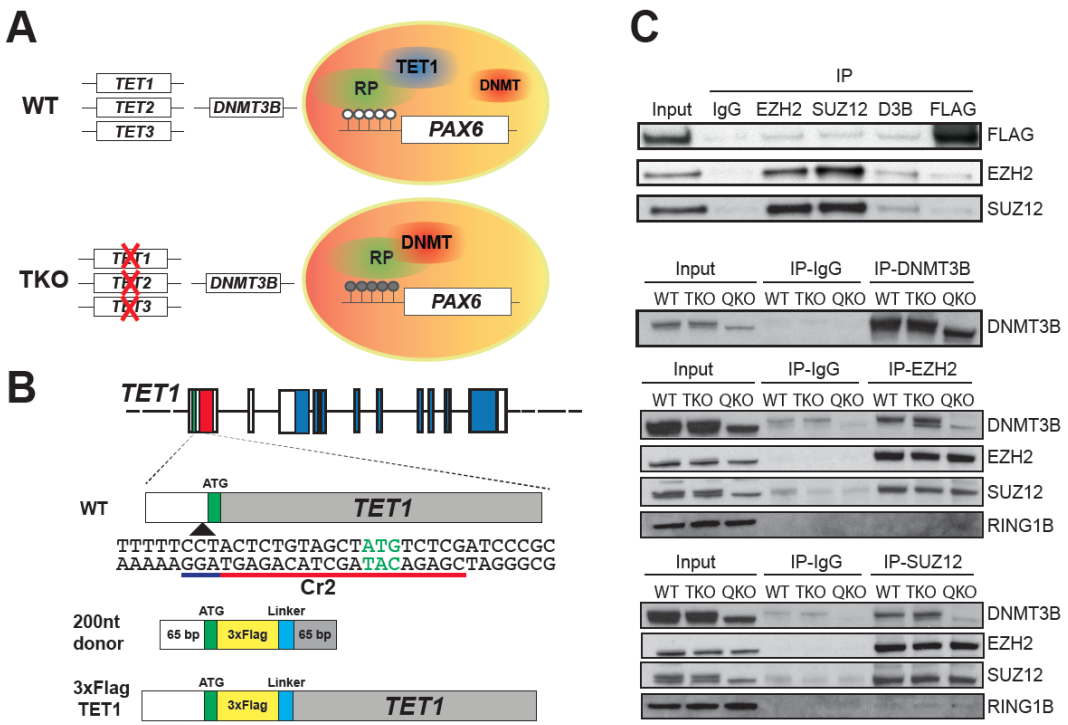
## *5.0 APPENDIX*

### *Appendix 1.*



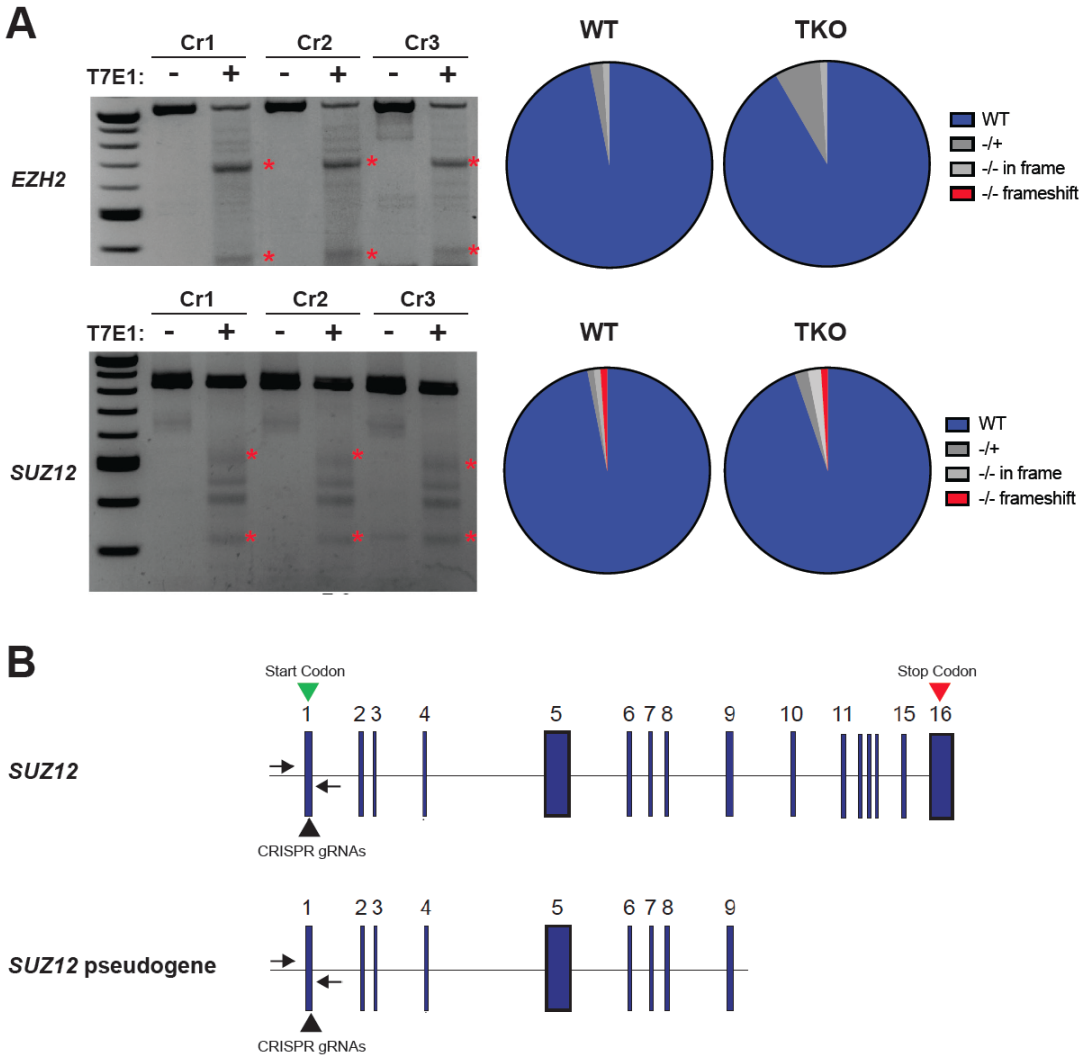

**Appendix 1. Generation and validation of TET1-3XFlag Tag lines.** (A) T7 Endonuclease assay for different CRISPR gRNAs that target the stop codon of *TET1*. CRISPR gRNA 2 was used for targeting. (B) Top: schematic of insertion of the Flag tag into the TET1 locus. The flag tag is inserted immediately after the start codon and separated from the coding sequence by a 12 bp linker sequence. Red box indicates the putative DNA binding domain of TET1. Arrowheads indicate the PCR primer pairs used for genotyping and identification of targeted clones. The WT allele shows a PCR product that is 542 bp while the allele with the Flag tag insertion produces a PCR fragment that is 623 bp long. Bottom: PCR genotyping of two positive clones. (C) Sequencing of PCR fragments from (B) of clone 1 (C1) indicates that there are 3 base pairs that are deleted prior to the start codon in the targeted allele. Similar mutations were observed in the second clone (C2). No mutations were observed in the non-targeted (WT) allele. The targeting sequence of the CRISPR gRNA used (Cr2) is shown. (D) Immunoprecipitation and western blot for the two targeted clones (C1 and C2). Asterisks indicate the expected band.

***Appendix 2.***



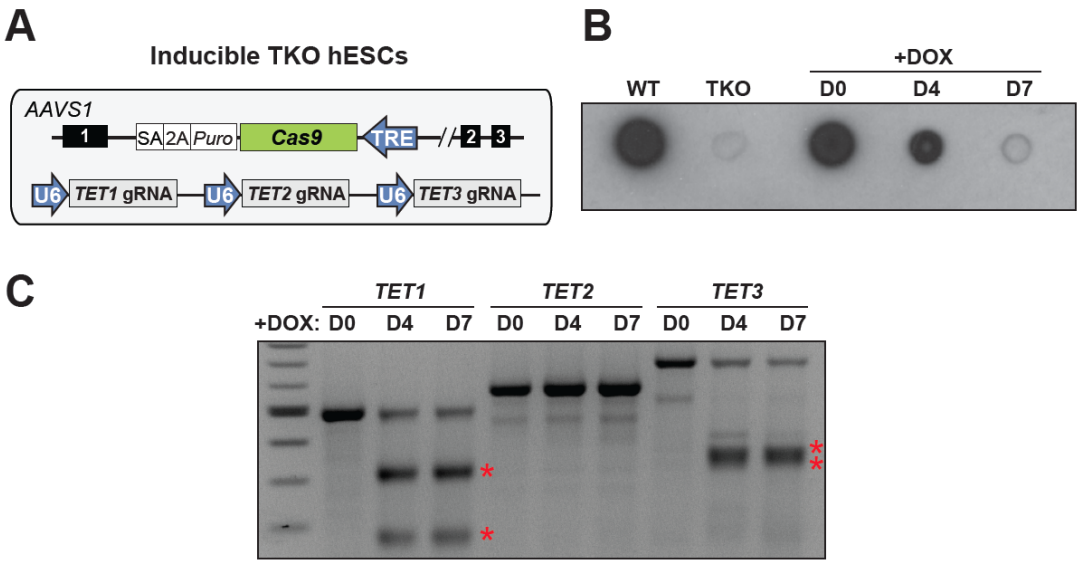

**Appendix 2. Co-immunoprecipitation of PRC2 and DNMT3B** (A) Model for recruitment of TET1 and DNMT3B to bivalent promoters such as the *PAX6* P0 promoter. In WT hESCs TET1 and DNMT3B compete for recruitment to bivalent promoters leading to hypomethylated promoters. After *TET* inactivation DNMT proteins is increasingly recruited to bivalent promoters leading to hypermethylation. (B) Schema for generation of the TET1-3xFlag line. CRISPR gRNA Cr2, which targets the start codon is shown. A 200 nucleotide ssDNA donor was used that contains a 69 bp 3xFlag sequence and 12 bp linker sequence. The bottom shows insertion of the 3xFlag into the endogenous *TET1* locus immediately after the ATG start codon and separated from the remainder of the TET1 coding sequence by a 12 bp linker sequence. Top: schematic of insertion of the Flag tag into the *TET1* locus. The flag tag is inserted immediately after the start codon and separated from the coding sequence by a 12 bp linker sequence. (C) Co-immunoprecipitation and western blot of PRC2 and TET1-Flag and DNMT3B in WT, TKO and QKO hESCs.

*Appendix 3.*



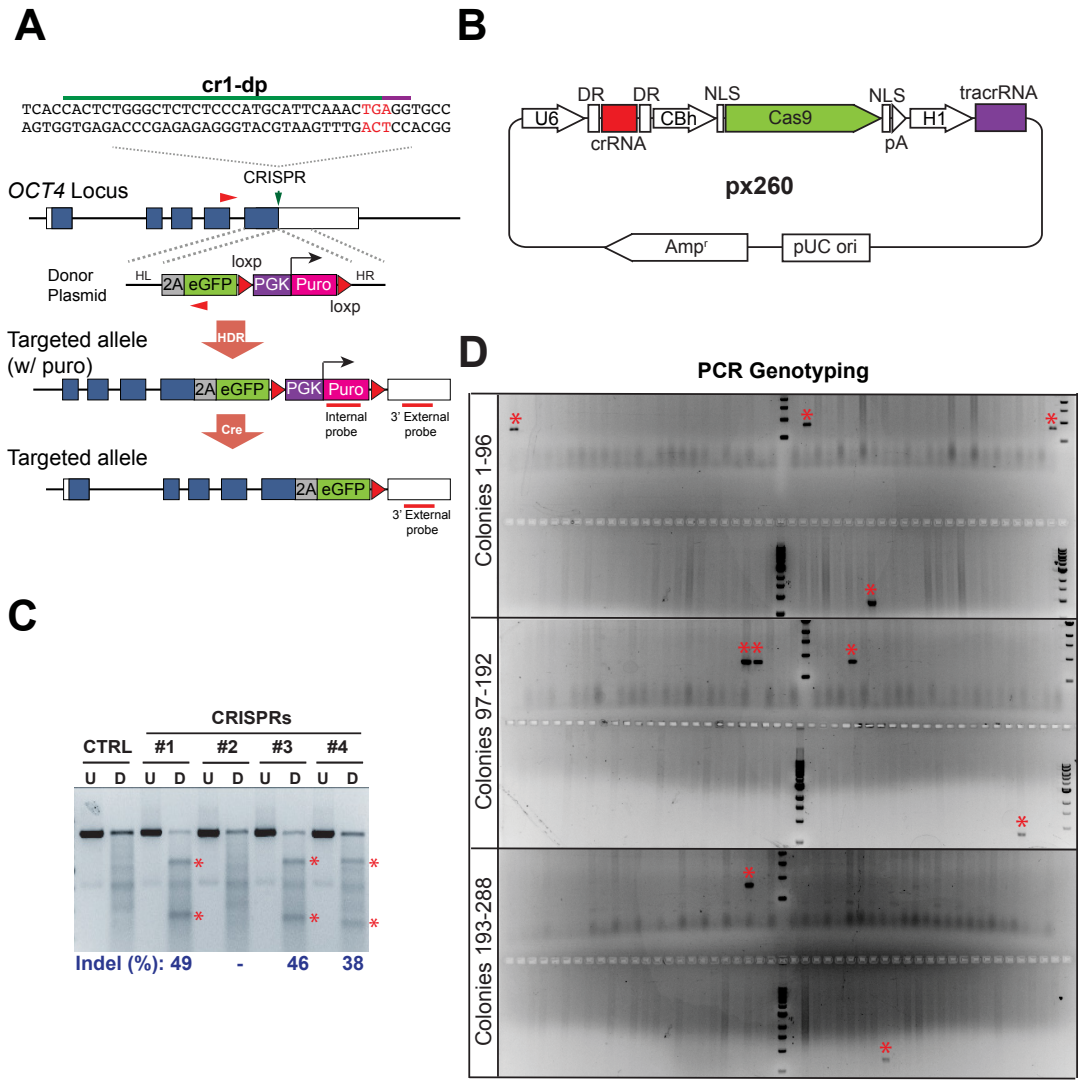

**Appendix 3. CRISPR targeting of *EZH2* and *SUZ12* start codons.** (A) Left: T7 Endonuclease I assay for CRISPR gRNAs that target the genomic sequences of *EZH2* (Top) and *SUZ12* (Bottom) start codons. Right: Generation of mutants after CRISPR gRNA targeting in WT and TKO hESCs. (B) Top: *SUZ12* gene with exons shown as blue box. The start codon and stop codon are indicated along with the CRISPR gRNAs that target exon 1. The PCR primers used for genotyping are indicated as arrows. Bottom: *SUZ12* pseudogene that shares exons 1-9 with the *SUZ12* gene. The binding of the CRISPR gRNAs and PCR genotyping to the *SUZ12* pseudogene are also indicated.

***Appendix 4.***



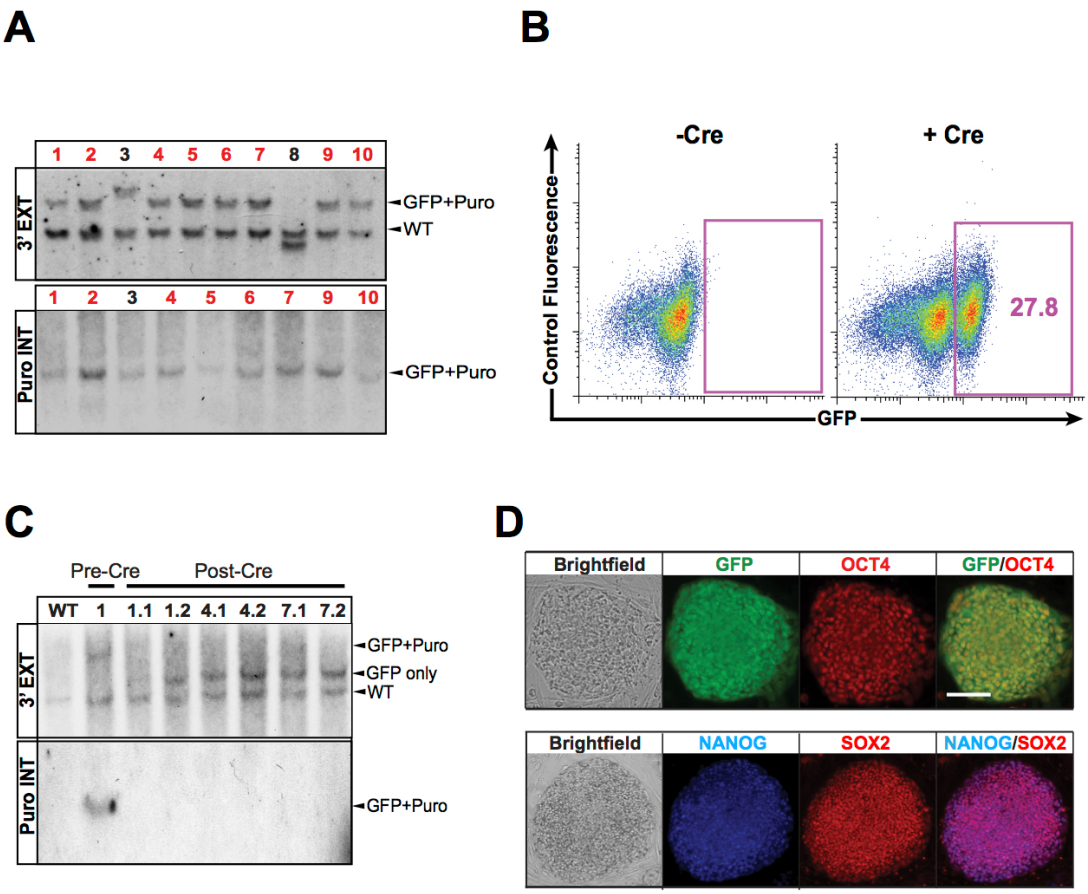

**Appendix 4. Generation of inducible TKO hESCs.** (A) Model of inducible TKO hESCs in which Cas9 is induced by doxycycline treatment and *TET1*, *TET2* and *TET3* gRNAs are expressed constitutively by U6 promoters. (B) 5hmC staining in WT, TKO hESCs and the inducible TKO line before doxycycline treatment (D0), after 4 days (D4) and 7 days (D7) of doxycycline treatment. (C) T7 Endonuclease I assay for indel formation at *TET1*, *TET2* and *TET3* loci in inducible TKO line before doxycycline treatment (D0), after 4 days (D4) and 7 days (D7) of doxycycline treatment.

*Appendix 5.*




**Appendix 5. CRISPR/Cas-Mediated Targeting of the *OCT4* Locus through Drug Selection.** (A) Schematics of the targeting strategy. In the presence of the donor plasmid, HDR results in the replacement of the *OCT4* stop codon with 2A-eGFP-PGK-Puro. The PCR primers (F + GFP-R) used for genotyping are indicated with red arrowheads. *OCT4* cr1-dp (the duplex version) targets a 30-nt sequence (indicated with a green line) upstream of the PAM sequence (indicated with a purple line). In all targeting schematics here and after, boxes are exons, filled blue boxes indicate the coding sequence (CDS), connecting lines are introns, the stop codon (TGA) is labeled in red, HL and HR indicate left and right homology arms, and the Southern blot external and internal probes are indicated with red bars. (B) The vector map of the px260 plasmid for expressing Cas9 and crRNA/tracrRNA. (C) Four crRNAs were designed to target the stop codon of the *OCT4* locus. Each crRNA was cloned into the px260 vector and transfected into 293T cells. Two days after transfection genomic DNA was collected and Surveyor analysis was used to estimate the efficiency of each CRISPR in generating indels. Asterisks indicated the cleavage products and the estimated Indel frequencies were labeled in blue. CTRL: px260 vector control; U: undigested control; D: digestion reaction with Surveyor nuclease. (D) PCR genotyping results showing 10 positive clones (indicated by red asterisks) identified based on the presence of a correct PCR product (811 bp).

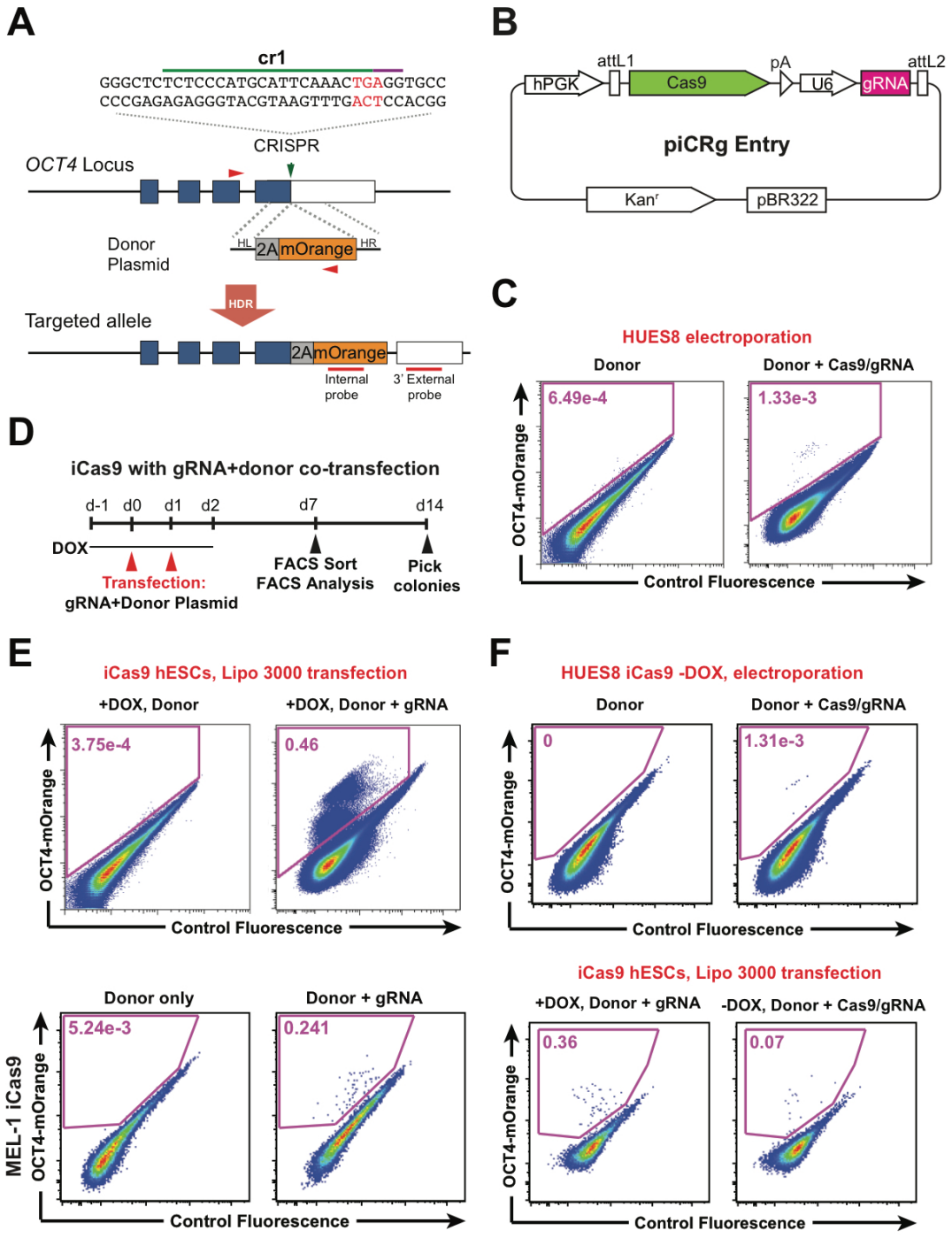
*Appendix 6.*



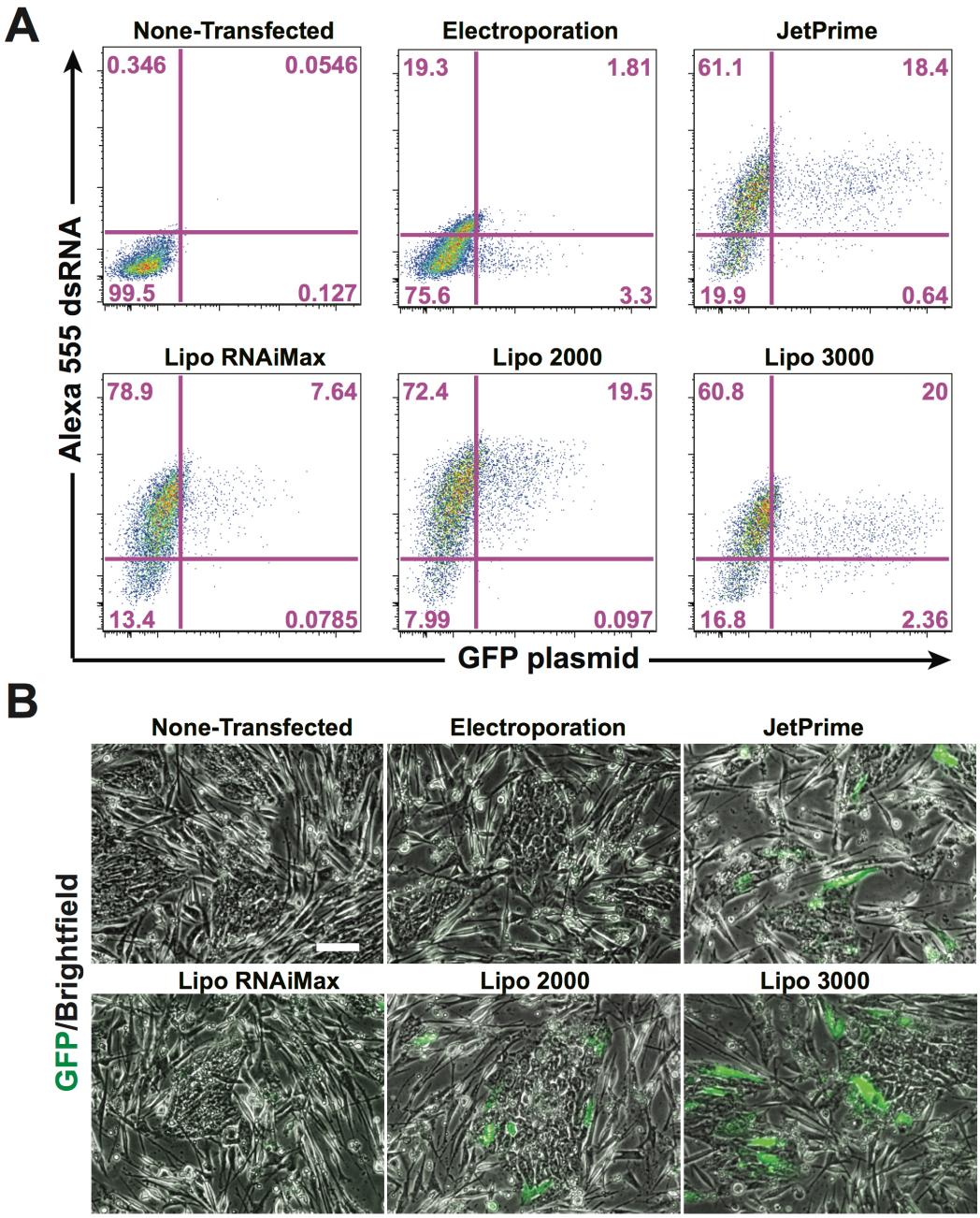

**Appendix 6. Validation of *OCT4* locus targeting using CRISPR/Cas9 and drug selection.** (A) Southern blot analysis using the external probe (WT: 4,173 bp; GFP+Puro: 6,835 bp) and the internal puromycin probe (GFP+Puro: 2,415 bp) identified eight correctly targeted clones, which are labeled in red. WT, wild-type allele; GFP+Puro, correctly targeted allele with puromycin-selection cassette. (B) Three of the correctly targeted clones (nos. 1, 4, and 7) were electroporated with Cre recombinase. Four days after electroporation, eGFP+ cells were isolated using FACS. (C) For each clone (nos. 1, 4, and 7) electroporated with CRE recombinase, two GFP+ clones were picked (e.g., C1.1 and C1.2 for clones derived from no. 1), and Southern blot analysis using the external probe (WT: 4,173 bp; GFP+Puro: 6,835 bp; GFP only: 5,015 bp) and the internal puromycin probe (GFP+Puro: 2,415 bp) showed correct removal of the puromycin-selection cassette. A clone (C1) prior to Cre electroporation was used as the Pre-Cre control. GFP+Puro, targeted allele prior to Cre-mediated excision of the PGK-Puro cassette; GFP only, targeted allele after Cre-mediated excision. (D) *OCT4-eGFP* reporter hESCs were stained for pluripotency markers OCT4, NANOG and SOX2, which overlapped with GFP expression. The GFP signal was detected using a GFP antibody. Scale bar = 100 μm.

## *Appendix 7.*

**Appendix 7. Targeting the *OCT4* Locus without Drug Selection.** (A) Schematics of the targeting strategy without drug selection. In the presence of the donor plasmid, HDR results in the replacement of the stop codon with 2A-mOrange. *OCT4* cr1 targets a 20-nt sequence upstream of the PAM sequence. The PCR primers (F + mOr-R) used for genotyping are indicated with red arrowheads. (B) The vector map of the piCRg Entry plasmid for expressing Cas9 and the chimeric gRNA. (C) FACS analysis for OCT4-mOrange expressing cells in HUES8 cells after electroporation of the OCT4-mOrange donor and the Cas9/gRNA plasmids. (D) Timeline for establishing hESC reporter lines using iCas9 hESCs. (E) FACS enrichment for OCT4-mOrange+ cells after transfection of the OCT4-mOrange donor plasmid and the *OCT4*-targeting gRNA into HUES8 iCas9 cells (top) and MEL1 iCas9 cells (bottom) treated with doxycycline. (F) Top: FACS analysis for OCT4-mOrange expressing cells in HUES8 iCas9 cells without doxycycline treatment after electroporation of the OCT4-mOrange donor and the Cas9/gRNA plasmids. Bottom: FACS analysis for OCT4-mOrange fluorescence in doxycycline-treated HUES8 iCas9 cells co-transfected with the OCT4-mOrange donor plasmid and the *OCT4*-targeting gRNA, compared to HUES8 iCas9 cells (not treated with doxycycline) co-transfected with the Cas9/gRNA and the donor plasmids using Lipofectamine 3000.

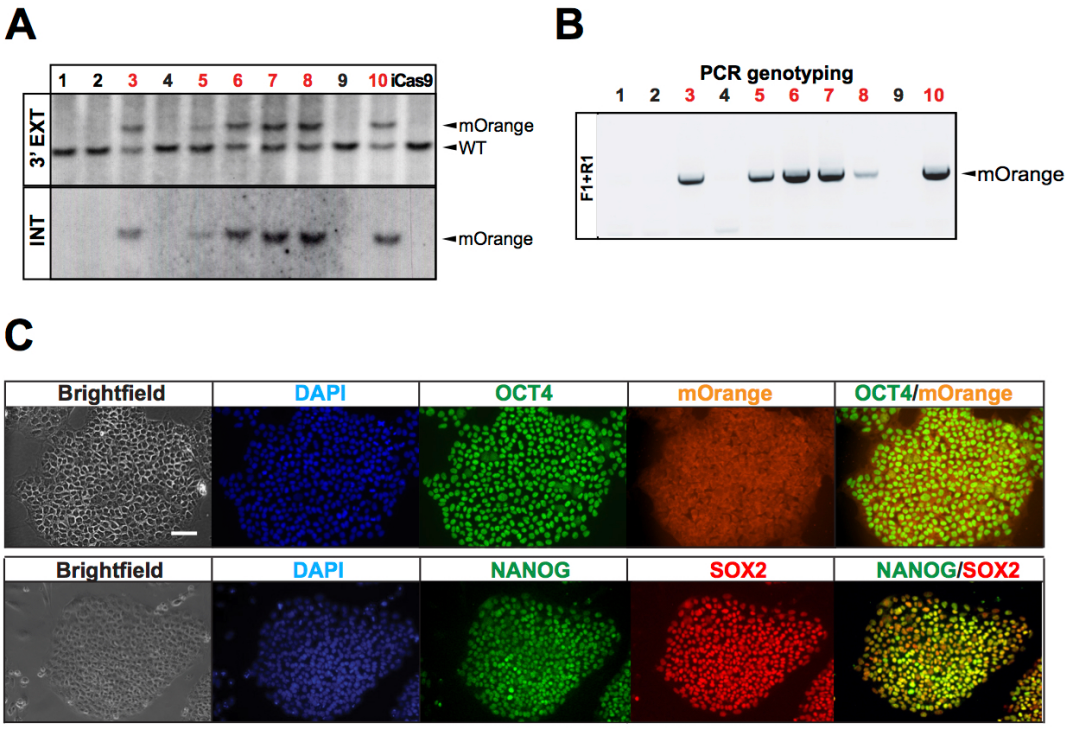

A
cr1
GGGCTCTCTCCCATGCATTCAAACTGAGGTGCC
CCCGAGAGAGGGTACGTAAGTTTGACTCCACGG
CRISPR
OCT4 Locus
Donor Plasmid
HL
2A
mOrange
HR
HDR
Targeted allele
Internal probe
3' External probe
B
attL1
hPGK
Cas9
pA
U6
gRNA
attL2
piCRg Entry
Kanr
pBR322
C
HUES8 electroporation
Donor
Donor + Cas9/gRNA
6.49e-4
1.33e-3
OCT4-mOrange
Control Fluorescence
D
iCas9 with gRNA+donor co-transfection
d-1
d0
d1
d2
d7
d14
DOX
Transfection:
gRNA+Donor Plasmid
FACS Sort
FACS Analysis
Pick colonies
E
iCas9 hESCs, Lipo 3000 transfection
+DOX, Donor
+DOX, Donor + gRNA
3.75e-4
0.46
OCT4-mOrange
Control Fluorescence
Donor only
Donor + gRNA
5.24e-3
0.241
MEL-1 iCas9 OCT4-mOrange
Control Fluorescence
F
HUES8 iCas9 -DOX, electroporation
Donor
Donor + Cas9/gRNA
0
1.31e-3
OCT4-mOrange
Control Fluorescence
iCas9 hESCs, Lipo 3000 transfection
+DOX, Donor + gRNA
-DOX, Donor + Cas9/gRNA
0.36
0.07
OCT4-mOrange
Control Fluorescence

*Appendix 8.*



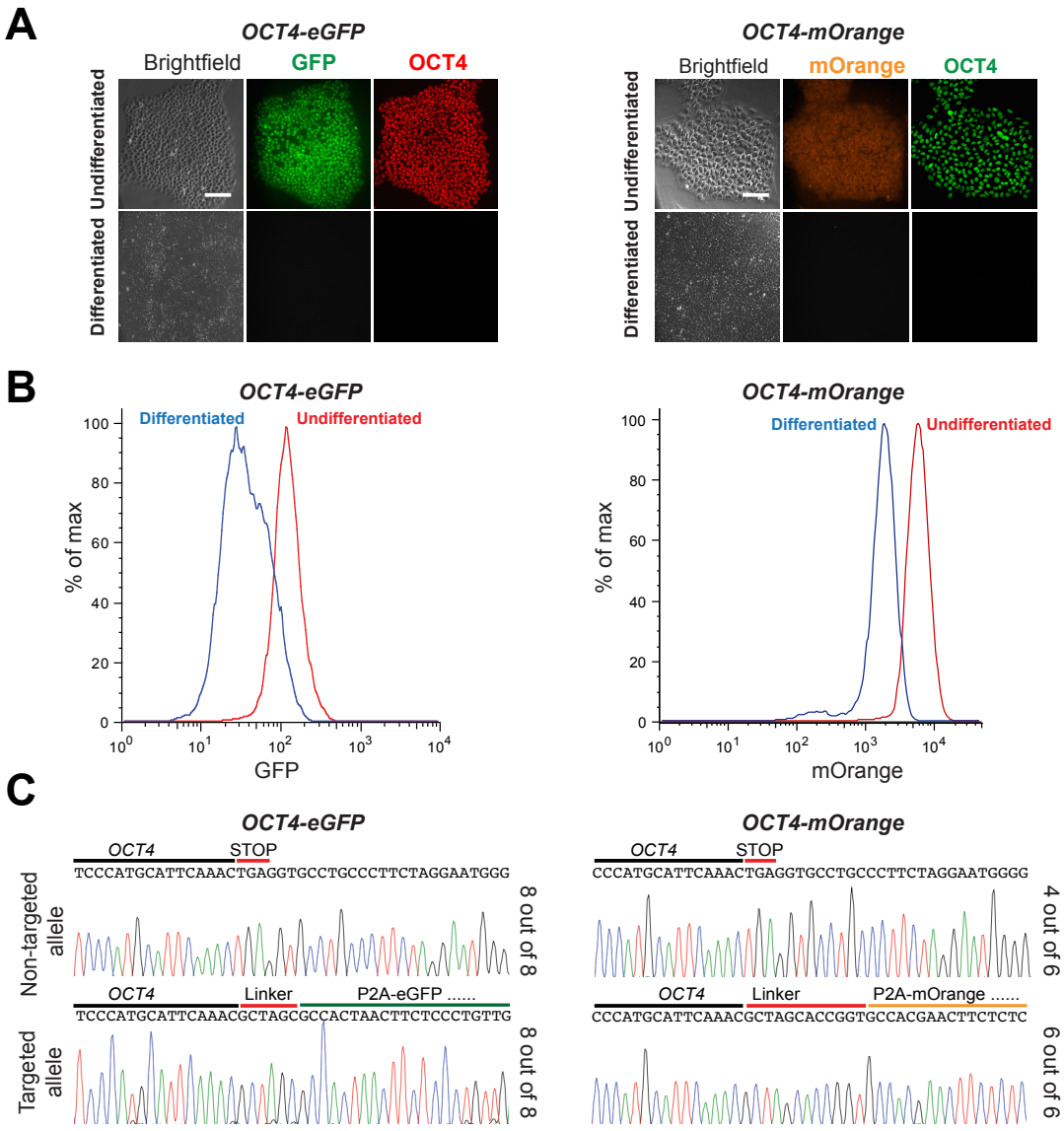

**Appendix 8. Optimizing co-transfection of DNA and RNA into hESCs.**
(A) FACS analysis of hESCs co-transfected with Alexa 555 dsRNA and a GFP-expressing plasmid (~ 10 kb) using four commonly used transfection reagents along with electroporation. Lipofectamine 3000 consistently performed better than the other conditions. Although comparable efficiencies could sometimes be achieved using JetPrime (as shown here), the outcomes were variable and appeared to relate to the amount of DNA used for transfection. (B) Stronger GFP expression was detected in hESCs transfected with Lipofectamine 3000, suggesting that increased copy number of GFP-expressing plasmid was transfected into the cell.

*Appendix 9.*




**Appendix 9. Validation of *OCT4* locus targeting using CRISPR/Cas9 without drug selection.** (A) Ten colonies were randomly picked from individual FACS-enriched mOrange+ cells. Southern blot analysis using the external probe (WT: 4,173 bp; mOrange: 4,963 bp) and the internal mOrange probe (mOrange: 4,963 bp) identified six correctly targeted clones, which are labeled in red. mOrange: correctly targeted allele. (B) PCR genotyping for OCT4-mOrange gene targeting. Correctly targeted clones are indicated in red. (C) *OCT4-mOrange* reporter hESCs were stained for pluripotency markers OCT4, NANOG and SOX2, which overlapped with mOrange expression. Scale bar = 100 μm.

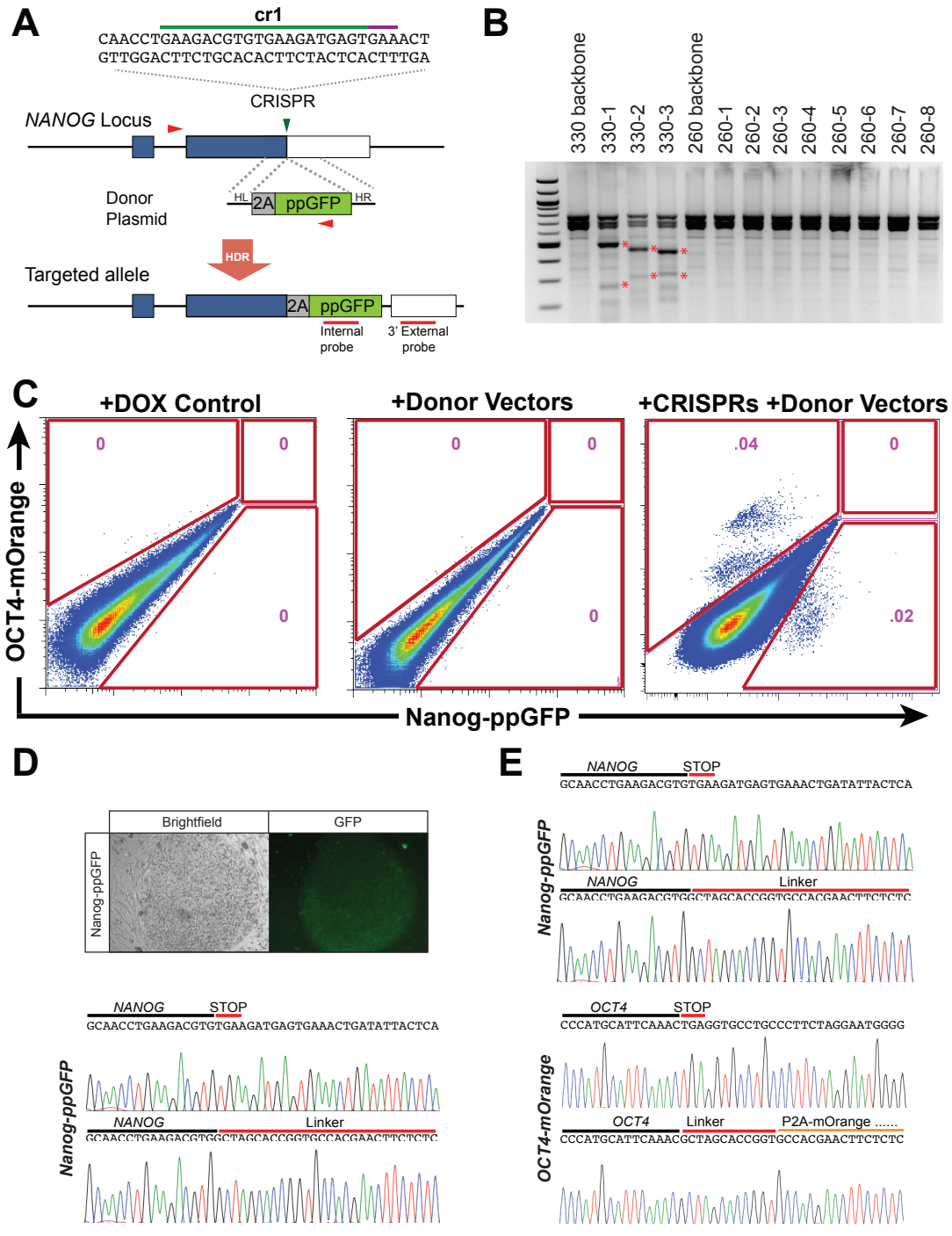
## *Appendix 10.*




**Appendix 10. Characterization of *OCT4-mOrange* Reporter Lines.** (A) *OCT4-eGFP* and *OCT4-mOrange* hESCs were treated with SB431542 and BMP4 to initiate differentiation. Three days after this treatment, the cells displayed concomitant loss of OCT4 protein expression with GFP or mOrange by immunostaining. An RFP antibody was used to detect mOrange expression, whereas the GFP expression was detected directly. The scale bar represents 100 μm. (B) Three days after SB431542 and BMP4 treatment, flow cytometry analysis showed a loss of GFP and mOrange, verifying that *OCT4-eGFP* and *OCT4-mOrange* reporter hESCs can respond to differentiation cues and that GFP and mOrange accurately reflects *OCT4* expression. (C) Sequencing results of the non-targeted allele and at the junction of correctly targeted allele in *OCT4-eGFP* and *OCT4-mOrange* reporter lines.

## *Appendix 11.*

**Appendix 11. Generation of *NANOG* reporter lines.** (A) A total of eleven crRNAs were designed to target the stop codon of the *NANOG* locus. Three crRNAs were cloned into the px330 vector and eight crRNAs were cloned into the px260 vector and transfected into 293T cells. Two days after transfection genomic DNA was collected and Surveyor analysis was used to estimate the efficiency of each CRISPR in generating indels. Asterisks indicate the cleavage products. 330 backbone: px330 vector control; 260 backbone: px260 vector control. (B) Schematics of the targeting strategy without drug selection. In the presence of the donor plasmid, HDR results in the replacement of the stop codon with 2A-ppGFP. *NANOG* cr1 targets a 20-nt sequence upstream of the PAM sequence. The PCR primers used for genotyping are indicated with red arrowheads. (C) FACS enrichment for OCT4-mOrange+ or NANOG-ppGFP+ cells after transfection of the OCT4-mOrange, NANOG-ppGFP donor plasmids and the *OCT4*-targeting and *NANOG*-targeting gRNAs into HUES8 iCas9 cells treated with doxycycline. GFP positive cells were sorted and expanded for further analysis and targeting. (D) Top: Brightfield and GFP fluorescence for one of the GFP positive sorted and expanded clones. Bottom: sequencing at the *NANOG* locus showed correct integration of the 2A-ppGFP donor. The non-targeted allele of *NANOG* was unchanged. (E) Targeting of the GFP positive targeted clone from (D) was targeted with *OCT4* gRNA and the 2A-mOrange donor vector. GFP and mOrange double positive cells were sorted and expanded. Sequencing of one such clone showed correct integration of the 2A-mOrange donor into the *OCT4* locus. The non-targeted allele of *OCT4* was unchanged.


A
cr1
CAACCTGAAGACGTGTGAAGATGAGTGAAACT
GTTGGACTTCTGCACACTTCTACTCACTTTGA
CRISPR
NANOG Locus
Donor Plasmid
HL
2A
ppGFP
HR
HDR
Targeted allele
Internal probe
3' External probe
B
330 backbone
330-1
330-2
330-3
260 backbone
260-1
260-2
260-3
260-4
260-5
260-6
260-7
260-8
C
+DOX Control
+Donor Vectors
+CRISPRs +Donor Vectors
OCT4-mOrange
Nanog-ppGFP
0
.04
.02
D
Brightfield
GFP
Nanog-ppGFP
NANOG
STOP
GCAACCTGAAGACGTGTGAAGATGAGTGAAACTGATATTACTCA
Linker
E
NANOG
STOP
GCAACCTGAAGACGTGTGAAGATGAGTGAAACTGATATTACTCA
Nanog-ppGFP
Linker
OCT4
CCCATGCATTCAAACTGAGGTGCCTGCCCTTCTAGGAATGGGG
OCT4-mOrange
P2A-mOrange .....
CCCATGCATTCAAACGCTAGCACCGGTGCCACGAACTTCTCTC

## *5.1 REFERENCES*


Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R.*, et al.* (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144-147.

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L.*, et al.* (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS genetics *8*, e1002781.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Anderson, T.R., Hedlund, E., and Carpenter, E.M. (2002). Differential Pax6 promoter activity and transcript expression during forebrain development. Mechanisms of development *114*, 171-175.

Balasubramanian, D., Akhtar-Zaidi, B., Song, L., Bartels, C.F., Veigl, M., Beard, L., Myeroff, L., Guda, K., Lutterbaugh, J., Willis, J.*, et al.* (2012). H3K4me3 inversely correlates with DNA methylation at a large class of non-CpG-island-containing start sites. Genome medicine *4*, 47.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470-484.

Bartolomei, M.S. (2009). Genomic imprinting: employing and avoiding epigenetic processes. Genes & development *23*, 2124-2133.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-247.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Bogdanovic, O. (2017). Tet proteins: master regulators of vertebrate body plan formation? Epigenomics *9*, 93-96.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nature reviews Genetics *17*, 661-678.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934-937.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B., and Bestor, T.H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science *294*, 2536-2539.

Byrne, S.M., Ortiz, L., Mali, P., Aach, J., and Church, G.M. (2015). Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. Nucleic acids research *43*, e21.

Cantone, I., and Fisher, A.G. (2013). Epigenetic programming and reprogramming during development. Nature structural & molecular biology *20*, 282-289.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology *27*, 275-280.

Chen, C.C., Wang, K.Y., and Shen, C.K. (2012). The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. The Journal of biological chemistry *287*, 33116-33121.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. Nature *493*, 561-564.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Molecular and cellular biology *23*, 5594-5605.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome research *24*, 132-141.

Choudhury, S.R., Cui, Y., Lubecka, K., Stefanska, B., and Irudayaraj, J. (2016). CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter. Oncotarget.

Collinson, A., Collier, A.J., Morgan, N.P., Sienerth, A.R., Chandra, T., Andrews, S., and Rugg-Gunn, P.J. (2016). Deletion of the Polycomb-Group Protein EZH2 Leads to Compromised Self-Renewal and Differentiation Defects in Human Embryonic Stem Cells. Cell reports *17*, 2700-2714.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011).

Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D*., et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S*., et al.* (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370-374.

Dai, H.Q., Wang, B.A., Yang, L., Chen, J.J., Zhu, G.C., Sun, M.L., Ge, H., Wang, R., Chapman, D.L., Tang, F*., et al.* (2016). TET-mediated DNA demethylation controls gastrulation by regulating Lefty-Nodal signalling. Nature *538*, 528-532.

Davis, R.P., Costa, M., Grandela, C., Holland, A.M., Hatzistavrou, T., Micallef, S.J., Li, X., Goulburn, A.L., Azzola, L., Elefanty, A.G*., et al.* (2008). A protocol for removal of antibiotic resistance cassettes from human embryonic stem cells genetically modified by homologous recombination or transgenesis. Nature protocols *3*, 1550-1558.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F*., et al.* (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Developmental cell *29*, 102-111.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M.*, et al.* (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Developmental cell *24*, 310-323.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C.*, et al.* (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell stem cell *9*, 166-175.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E.*, et al.* (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. The EMBO journal *32*, 645-655.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P.*, et al.* (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371-376.

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T.*, et al.* (2013a). A TALEN genome-editing system for generating human stem cell-based disease models. Cell stem cell *12*, 238-251.

Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013b). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell stem cell *12*, 393-394.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A.*, et al.* (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652-655.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *102*, 15785-15790.

Farthing, C.R., Ficz, G., Ng, R.K., Chan, C.F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. PLoS genetics *4*, e1000116.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Ferguson-Smith, A.C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. Nature reviews Genetics *12*, 565-575.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology *31*, 822-826.

Gal-Yam, E.N., Egger, G., Iniguez, L., Holster, H., Einarsson, S., Zhang, X., Lin, J.C., Liang, G., Jones, P.A., and Tanay, A. (2008). Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proceedings of the National Academy of Sciences of the United States of America *105*, 12979-12984.

Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y*., et al.* (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell stem cell *12*, 453-469.

Giannopoulou, E.G., and Elemento, O. (2011). An integrated ChIP-seq analysis platform with customizable workflows. BMC bioinformatics *12*,

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J.*, et al.* (2004). Differences between human and mouse embryonic stem cells. Developmental biology *269*, 360-380.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annual review of biochemistry *74*, 481-514.

Gonzalez, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell stem cell *15*, 215-226.

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011a). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nature protocols *6*, 468-481.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011b). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J.*, et al.* (2014). The DNA methylation landscape of human early embryos. Nature *511*, 606-610.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and

imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448-452.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science *329*, 78-82.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic acids research *40*, 4841-4849.

Hayashi, K., de Sousa Lopes, S.M., and Surani, M.A. (2007). Germ cell specification in mice. Science *316*, 394-396.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141-1143.

Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B.*, et al.* (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. Nature biotechnology *27*, 851-857.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C.*, et al.* (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nature biotechnology *29*, 731-734.

Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J., and Thomson, J.A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences of the United States of America *110*, 15644-15649.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology *31*, 827-832.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W.*, et al.* (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell stem cell *14*, 512-522.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nature biotechnology *26*, 1269-1275.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS letters *583*, 1713-1720.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M.*, et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature genetics *41*, 178-186.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. International journal for parasitology *38*, 1-31.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nature reviews Genetics *10*, 161-172.

Jin, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W., and Kusch, T. (2005). In and out: histone variant exchange in chromatin. Trends in biochemical sciences *30*, 680-687.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nature reviews Genetics *13*, 484-492.

Jones, P.A., Wolkowicz, M.J., Rideout, W.M., 3rd, Gonzales, F.A., Marziasz, C.M., Coetzee, G.A., and Tapscott, S.J. (1990). De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. Proceedings of the National Academy of Sciences of the United States of America *87*, 6117-6121.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. The EMBO journal *32*, 340-353.

Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? Genes & development *19*, 295-310.

Kammandel, B., Chowdhury, K., Stoykova, A., Aparicio, S., Brenner, S., and Gruss, P. (1999). Distinct cis-essential modules direct the time-space pattern of the Pax6 gene activity. Developmental biology *205*, 79-97.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kelly, T.K., Miranda, T.B., Liang, G., Berman, B.P., Lin, J.C., Tanay, A., and Jones, P.A. (2010). H2A.Z maintenance during mitosis reveals

nucleosome shifting on mitotically silenced genes. Molecular cell *39*, 901-911.

Khoueiry, R., Sohni, A., Thienpont, B., Luo, X., Velde, J.V., Bartoccetti, M., Boeckx, B., Zwijsen, A., Rao, A., Lambrechts, D.*, et al.* (2017). Lineage-specific functions of TET1 in the postimplantation mouse embryo. Nature genetics.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nature reviews Genetics *15*, 321-334.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571-1572.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R.*, et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.

Kowalczyk, M.S., Higgs, D.R., and Gingeras, T.R. (2012). Molecular biology: RNA discrimination. Nature *482*, 310-311.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial

sequencing and analysis of the human genome. Nature *409*, 860-921.

Lee, D.S., Shin, J.Y., Tonge, P.D., Puri, M.C., Lee, S., Park, H., Lee, W.C., Hussein, S.M., Bleazard, T., Yun, J.Y*., et al.* (2014). An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nature communications *5*, 5619.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, X.J., Du, Z.W., Zarnowska, E.D., Pankratz, M., Hansen, L.O., Pearce, R.A., and Zhang, S.C. (2005). Specification of motoneurons from human embryonic stem cells. Nature biotechnology *23*, 215-221.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509-4518.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Molecular and cellular biology *22*, 480-491.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C.*, et al.* (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nature genetics *47*, 469-478.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233-247 e217.

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. Nature chemical biology *5*, 400-402.

Lomberk, G., Wallrath, L., and Urrutia, R. (2006). The Heterochromatin Protein 1 family. Genome biology *7*, 228.

Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia *17*, 637-641.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology *15*, 550.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes & development *28*, 2103-2119.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McLaren, A., and Lawson, K.A. (2005). How is the mouse germ-cell lineage established? Differentiation; research in biological diversity *73*, 435-437.

Meier, I.D., Bernreuther, C., Tilling, T., Neidhardt, J., Wong, Y.W., Schulze, C., Streichert, T., and Schachner, M. (2010). Short DNA sequences inserted for gene targeting can accidentally interfere with off-target gene expression. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *24*, 1714-1724.

Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell *151*, 1417-1430.

Merkert, S., Wunderlich, S., Bednarski, C., Beier, J., Haase, A., Dreyer, A.K., Schwanke, K., Meyer, J., Gohring, G., Cathomen, T.*, et al.* (2014). Efficient designer nuclease-based homologous recombination enables direct PCR screening for footprintless targeted human pluripotent stem cells. Stem cell reports *2*, 107-118.

Messerschmidt, D.M., Knowles, B.B., and Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes & development *28*, 812-828.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.

Molina-Serrano, D., Schiza, V., and Kirmizis, A. (2013). Cross-talk among epigenetic modifications: lessons from histone arginine methylation. Biochemical Society transactions *41*, 751-759.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer cell *20*, 11-24.

Morita, S., Noguchi, H., Horii, T., Nakabayashi, K., Kimura, M., Okamura, K., Sakai, A., Nakashima, H., Hata, K., Nakashima, K.*, et al.* (2016). Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. Nature biotechnology.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nature chemical biology *8*, 751-758.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M.*, et al.* (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature cell biology *9*, 64-71.

Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., Parlato, C., and Oliviero, S. (2013a). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. Genome biology *14*, R91.

Neri, F., Krepelova, A., Incarnato, D., Maldotti, M., Parlato, C., Galvagni, F., Matarese, F., Stunnenberg, H.G., and Oliviero, S. (2013b). Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell *155*, 121-134.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W.*, et al.* (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nature genetics *39*, 237-242.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L

connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Pan, G., Tian, S., Nie, J., Yang, C., Ruotti, V., Wei, H., Jonsdottir, G.A., Stewart, R., and Thomson, J.A. (2007). Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. Cell stem cell *1*, 299-312.

Pan, H., Jiang, Y., Boi, M., Tabbo, F., Redmond, D., Nie, K., Ladetto, M., Chiappella, A., Cerchietti, L., Shaknovich, R.*, et al.* (2015). Epigenomic evolution in diffuse large B-cell lymphomas. Nature communications *6*, 6921.

Pankratz, M.T., Li, X.J., Lavaute, T.M., Lyons, E.A., Chen, X., and Zhang, S.C. (2007). Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. Stem cells *25*, 1511-1520.

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nature reviews Molecular cell biology *14*, 341-356.

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P.*, et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals

RNA-programmed Cas9 nuclease specificity. Nature biotechnology *31*, 839-843.

Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. Annual review of genetics *36*, 233-278.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M.*, et al.* (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440-455.

Plaza, S., Saule, S., and Dozier, C. (1999). High conservation of cis-regulatory elements between quail and human for the Pax-6 gene. Development genes and evolution *209*, 165-173.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H.*, et al.* (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer cell *20*, 25-38.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Genes & development *30*, 733-750.

Ratnam, S., Mertineit, C., Ding, F., Howell, C.Y., Clarke, H.J., Bestor, T.H., Chaillet, J.R., and Trasler, J.M. (2002). Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Developmental biology *245*, 304-314.

Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J.I., Matallanas, D., Von Kriegsheim, A.*, et al.* (2016). Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. Cell stem cell *19*, 491-501.

Rostovskaya, M., Fu, J., Obst, M., Baer, I., Weidlich, S., Wang, H., Smith, A.J., Anastassiadis, K., and Stewart, A.F. (2012). Transposon-mediated BAC transgenesis in human ES cells. Nucleic acids research *40*, e150.

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G.*, et al.* (2011). Lineage-specific distribution of high levels of genomic 5-

hydroxymethylcytosine in mammalian development. Cell research *21*, 1332-1342.

Schiesser, S., Hackner, B., Pfaffeneder, T., Muller, M., Hagemeier, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angewandte Chemie *51*, 6516-6520.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome research *19*, 1165-1174.

Schnerch, A., Cerdan, C., and Bhatia, M. (2010). Distinguishing between mouse and human pluripotent stem cell regulation: the best laid plans of mice and men. Stem cells *28*, 419-430.

Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L.*, et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic acids research *40*, 8255-8265.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692-706.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nature reviews Genetics *15*, 272-286.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74-79.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews Genetics *14*, 204-220.

Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X*., et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature biotechnology *29*, 68-72.

Song, C.X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. Nature biotechnology *30*, 1107-1116.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D*., et al.*

(2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490-495.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome biology *12*, R54.

Stroud H, F.S., Morey Kinney S, Pradhan S, Jacobsen SE. (2011). Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. . Genome biology, R54.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K.*, et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS genetics *7*, e1002154.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem

cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes to cells : devoted to molecular & cellular mechanisms *11*, 805-814.

Valinluck, V., and Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer research *67*, 946-950.

Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell stem cell *15*, 27-30.

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes & development *27*, 1318-1338.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nature genetics *20*, 116-117.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. Mechanisms of development *118*, 187-190.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature genetics *37*, 853-862.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I.*, et al.* (2007). Epigenetic stem cell signature in cancer. Nature genetics *39*, 157-158.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wolf, S.F., Jolly, D.J., Lunnen, K.D., Friedmann, T., and Migeon, B.R. (1984). Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proceedings of the National Academy of Sciences of the United States of America *81*, 2806-2810.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature communications *2*, 241.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes & development *25*, 2436-2452.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., Ruan, K., Wang, F., Xu, G.L., and Hu, R. (2016). A CRISPR-based approach for targeted DNA demethylation. Cell discovery *2*, 16009.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S.*, et al.* (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular cell *42*, 451-464.

Xu, Z.P., and Saunders, G.F. (1998). PAX6 intronic sequence targets expression to the spinal cord. Developmental genetics *23*, 259-263.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443-447.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498-1510.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B.*, et al.* (2012). Base-resolution analysis of

5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368-1380.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nature chemical biology *8*, 328-330.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M.*, et al.* (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell stem cell *7*, 90-100.

Zhang, X., Su, J., Jeong, M., Ko, M., Huang, Y., Park, H.J., Guzman, A., Lei, Y., Huang, Y.H., Rao, A.*, et al.* (2016). DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. Nature genetics *48*, 1014-1023.

Zheng, J.B., Zhou, Y.H., Maity, T., Liao, W.S., and Saunders, G.F. (2001). Activation of the human PAX6 gene through the exon 1 enhancer by transcription factors SEF and Sp1. Nucleic acids research *29*, 4070-4078.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nature reviews Genetics *12*, 7-18.

Zhu, Z., and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. Development *140*, 705-717.

Zhu, Z., Gonzalez, F., and Huangfu, D. (2014). The iCRISPR platform for rapid genome editing in human pluripotent stem cells. Methods in enzymology *546*, 215-250.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.

Zlatanova, J., and Thakar, A. (2008). H2A.Z: view from the top. Structure *16*, 166-179.