DNA methylation of endothelial cell-enriched genes during *in vitro* mouse ES cell differentiation and the role of TET proteins

by

Olga Vexelshtein

A thesis submitted in conformity with the requirements for the degree of Master of Science

Graduate Department of Medical Biophysics
University of Toronto

© Copyright Olga Vexelshtein 2016
DNA methylation of endothelial cell-enriched genes during in vitro mouse ES cell differentiation and the role of TET proteins

Olga Vexelshtein
Master of Science
Graduate Department of Medical Biophysics
University of Toronto
2016

Endothelial cell (EC)-enriched genes are regulated by numerous mechanisms, but especially epigenetic mechanisms such as DNA methylation. The current work examines the importance of TET1 and 2 in murine embryonic stem (ES) cells and their relevance to EC-enriched gene expression. The proximal promoters of EC-enriched genes are differentially methylated in expressing versus non-expressing cell types. To examine how this methylation pattern between EC and non-EC is established, we studied ES cells. The DNA methylation pattern was heterogeneous, and partially propagated to daughter cells upon cell division. When in vitro differentiating WT and TET1 and 2 double knockout (DKO) ES cells, the DNA methylation of EC-enriched gene promoters increases at day 4 and day 7 in both cell lines. Despite their hypermethylation at day 7, EC-enriched gene expression increases in the WT cells but it is blunted in the DKO cells. Suggesting that TET1 and 2 presence in ES cells facilitate the activation of EC-enriched genes later in development.
Acknowledgments

I would like to thank my supervisor Dr. Philip Marsden for the opportunity to work in his lab and on this project. I have learned many skills during my degree, and have grown as a scientist and a person, gaining valuable knowledge and experience.

I would also like to thank my committee members Dr. Dan Dumont, Dr. Derek van der Kooy, and Dr. Karim Mekhail for their input, advice and always making me think on my feet.

I would like to thank all previous and current Marsden lab members. For all their help, advice and making my stay in the lab not only educational but also fun. I have gained not only great colleagues, but also friends.

I have endless amount of gratitude for my dear husband, my beautiful son, and my parents, who supported me and encouraged both in the ups and the downs of my path.
Accreditation of Work

The author completed the study and experiments outlined here.

Thank you to Dr. Rudolf Jaenisch who provided us with the TET knockout ES cells\(^1\).

The list of differentially methylated EC-enriched genes in human cells were kindly provided by Apurva Shirodkar and other Marsden lab members that contributed to the manuscript\(^2\). DNA methylation of eNOS promoter and gene body figure, was kindly provided by Rosanne M. St Bernard. Gene expression data of Oct4 and EC-enriched genes during mouse *in vitro* differentiation was kindly provided by Dr. Jason Fish. DNA methylation data of VEFGR2+ and VEGFR2- sorted cells, as well as the mouse aorta was kindly provided by Anna Kop.
Table of Contents

Abstract.......................................................................................................................... ii
Acknowledgements........................................................................................................ iii
Accreditation of Work.................................................................................................... iv
Table of Contents.......................................................................................................... v
List of Tables................................................................................................................ ix
List of Figures............................................................................................................... x
List of Panels................................................................................................................ xiii
List of Abbreviations and Acronyms............................................................................ xiv

Chapter 1: Introduction.................................................................................................. 1

1.1 Epigenetics

1.1.1 Overview............................................................................................................. 2
1.1.2 Histone modification, and variants..................................................................... 6
1.1.3 RNA-based mechanisms.................................................................................... 7
1.1.4 DNA methylation.............................................................................................. 8
  1.1.4.1 DNA methyltransferases.............................................................................. 12
  1.1.4.2 TET proteins.............................................................................................. 13
  1.1.4.3 Oxymethyls............................................................................................... 17
  1.1.4.4 DNA demethylation.................................................................................. 21
  1.1.4.5 CpG islands.............................................................................................. 25
  1.1.4.6 DNA methylation during development................................................... 29
  1.1.4.7 Inheritance of epigenetic marks............................................................... 32
1.1.5 Epigenetic Heterogeneity.................................................................................. 33
1.2 Vascular Endothelium

1.2.1 Endothelium ........................................................................................................................................... 40
1.2.2 Epigenetic regulation of endothelial gene expression ............................................................................. 41

1.3 Methods for studying DNA methylation

1.3.1 Sodium bisulfite conversion .................................................................................................................. 46
1.3.2 Pyrosequencing ...................................................................................................................................... 49
1.3.3 LINE-1 – repetitive DNA sequencing .................................................................................................. 52
1.3.4 Restriction enzyme-based methods ..................................................................................................... 52
1.3.5 Antibody-based methods ..................................................................................................................... 53
1.3.6 Chromatography and mass spectrometry ........................................................................................... 54

Chapter 2: DNA methylation of endothelial cell-enriched genes during in vitro mouse ES cell differentiation and the role of TET proteins ................................................................................................................. 55

2.1 Rational, hypothesis and specific aims

2.1.1 Rational .................................................................................................................................................. 56
2.1.2 Hypothesis .............................................................................................................................................. 61
2.1.3 Aims ....................................................................................................................................................... 61

2.2 Methods

2.2.1 Cell culture ............................................................................................................................................ 62
2.2.2 ES cell differentiation ............................................................................................................................. 63
2.2.3 2i and vitamin C experiment ................................................................................................................ 63
2.2.4 Cell sorting into 96 well plate and Alkaline Phosphatase live staining ................................................. 63
2.2.5 Genomic DNA extraction ...................................................................................................................... 64
2.2.6 Sodium bisulfite DNA conversion ......................................................................................................... 64
2.2.7 Sodium bisulfite sequencing ................................................................................................................ 65
2.2.8 Pyrosequencing of bisulfite converted DNA.............................................. 65
2.2.9 RNA extraction........................................................................................................ 69
2.2.10 First strand cDNA synthesis.............................................................................. 69
2.2.11 Real-time quantitative Polymerase Chain Reaction (RT-qPCR)....................... 69
2.2.12 Statistics.............................................................................................................. 70

2.3 Results
2.3.1 DNA methylation clonality of the mouse eNOS proximal promoter in ES cells.................................................................................................................. 72
2.3.2 Regulating the DNA methylation heterogeneity of the eNOS proximal promoter in mouse ES cells.................................................................................................................. 77
2.3.3 Global DNA methylation during in vitro mouse ES cell differentiation.............. 82
2.3.4 TET and DNMT gene expression during in vitro mouse ES cell differentiation................................................................................................................................. 84
2.3.5 Characterization of expression of ES cell markers in WT and TET1 and 2 DKO mouse ES cells.................................................................................................................. 91
2.3.6 DNA methylation clonality of the mouse eNOS proximal promoter in DKO ES cells................................................................................................................................. 95
2.3.7 Global DNA methylation in mouse ES cells and TET DKO............................... 98
2.3.8 DNA methylation of EC-enriched genes in mouse ES cells and during in vitro differentiation in WT and TET DKO................................................................. 100
2.3.9 Gene expression of EC-enriched genes in mouse ES cells and during in vitro differentiation in WT and TET DKO................................................................. 104
2.3.10 eNOS gene expression in TET1 and TET2 single knockouts compared to the DKO and WT cells................................................................................................. 110

2.4 Discussion
2.4.1 ES cell DNA methylation heterogeneity and its heritability............................ 112
2.4.2 DNA methylation changes from ES cell stage to EB day 4 and the key players 114
2.4.3 TET1 and 2 DKO ES cells and their affect on DNA methylation during EB differentiation.................................................................................................................. 115
2.4.4 Gene expression changes during WT and DKO ES cell differentiation

2.5 Limitations

2.6 Conclusions

2.7 Future Directions

2.8 References
List of Tables

Introduction

Table 1.1: CGI characterization of EC-enriched gene promoters................................. 28
Table 1.2: EC-enriched genes that are not methylated in ECs and their DNA methylation status in VSMCs................................................................. 44

Methods

Table 2.1: Mouse bisulfite sequencing primers.............................................................. 67
Table 2.2: Mouse bisulfite pyrosequencing primers....................................................... 68
Table 2.3: RT-qPCR primers for mouse cDNA............................................................... 71
List of Figures

Introduction

Figure 1.1:  Comparison of key facets of genomic and epigenetic modifications... 4
Figure 1.2:  Epigenetic mechanisms.......................................................... 5
Figure 1.3:  Epigenetic-based mechanisms of gene regulation.................. 10
Figure 1.4:  Reaction products and DNA sequencing reads after bisulfite conversion.......................................................... 19
Figure 1.5:  25 possible C-G combinations........................................... 20
Figure 1.6:  Pathways of DNA methylation and demethylation.............. 23
Figure 1.7:  Base excision repair pathway............................................. 24
Figure 1.8:  DNA methylation reprogramming during development......... 31
Figure 1.9:  DNA methylation of the eNOS proximal promoter in endothelial cells and vascular smooth muscle cells............................... 38
Figure 1.10:  50% DNA methylation pattern ......................................... 39
Figure 1.11:  DNA methylation of eNOS promoter and gene body.......... 45
Figure 1.12:  Bisulfite treatment chemical reaction on a cytosine base and sequencing.......................................................... 48
Figure 1.13:  Pyrosequencing................................................................. 51

Rational

Figure 2.1:  mRNA expression levels in mouse ES cells and mouse EBs day 2 - day 13........................................................................................................ 58
Figure 2.2:  DNA methylation of the mouse promoters of eNOS and VE-cadherin... 59
Figure 2.3:  Mouse ES cell DNA methylation of the eNOS proximal promoter........ 60

Results

Figure 2.4:  Assessment of DNA methylation of the eNOS proximal promoter from 10 ES clones...................................................................................... 74
Figure 2.5:  Assessment of heritability of the DNA methylation pattern of the
mouse eNOS proximal promoter............................................................... 75

Figure 2.6: Two representative ES cell clones display both clonal and heterogeneous DNA methylation of the eNOS proximal promoter...... 76

Figure 2.7: Mouse eNOS proximal promoter methylation after treatment with 2i, vit C or both.................................................................................................................. 78

Figure 2.8: Single strand analysis of DNA methylation of the mouse eNOS proximal promoter in ES cells with treatment of 2i, vit C, or both........ 79

Figure 2.9: LINE-1 global DNA methylation assay of ES cells differentiated for 7 days................................................................................................................................. 83

Figure 2.10: mRNA expression levels of DNMT1 in mouse ES cells and EBs day 4 and day 7...................................................................................................................... 85

Figure 2.11: mRNA expression levels of DNMT3a in mouse ES cells and EBs day 4 and day 7...................................................................................................................... 86

Figure 2.12: mRNA expression levels of DNMT3b in mouse ES cells and EBs day 4 and day 7...................................................................................................................... 87

Figure 2.13: mRNA expression levels of TET1 in mouse ES cells and EBs day 4 and 7 88

Figure 2.14: mRNA expression levels of TET2 in mouse ES cells and EBs day 4 and 7 89

Figure 2.15: mRNA expression levels of TET3 in mouse ES cells and EBs day 4 and 7 90

Figure 2.16: mRNA expression levels of SOX2 in WT and DKO mouse ES cells and EBs day 4 and 7...................................................................................................................... 92

Figure 2.17: mRNA expression levels of OCT4 in WT and DKO mouse ES cells and EBs day 4 and 7...................................................................................................................... 93

Figure 2.18: mRNA expression levels of NANOG in WT and DKO mouse ES cells and EBs day 4 and 7...................................................................................................................... 94

Figure 2.19: DNA methylation of the mouse eNOS promoter in DKO ES cell clones.. 96

Figure 2.20: Single strand analysis of the DNA methylation of eNOS promoter in mouse DKO ES cells................................................................. 97

Figure 2.21: LINE-1 DNA methylation assay of mouse ES cells during differentiation in WT and TET1 and TET2 DKO cells............................................. 99

Figure 2.22: DNA methylation analysis of mouse eNOS promoter during ES cell differentiation in WT and DKO cells................................................................. 101

Figure 2.23: DNA methylation analysis of mouse CD31 promoter during ES cell differentiation in WT and DKO cells................................................................. 102
Figure 2.24: DNA methylation analysis of mouse CDH5 promoter during ES cell differentiation in WT and DKO cells.......................... 103

Figure 2.25: mRNA expression levels of eNOS in WT and DKO mouse ES cells and EBs day 4 and 7.......................................................... 106

Figure 2.26: mRNA expression levels of hnRNA of eNOS in WT and DKO mouse ES cells and EBs day 4 and 7.......................................................... 107

Figure 2.27: mRNA expression levels of CD31 in WT and DKO mouse ES cells and EBs day 4 and 7.......................................................... 108

Figure 2.28: mRNA expression levels of CDH5 in WT and DKO mouse ES cells and EBs day 4 and 7.......................................................... 109

Figure 2.29: eNOS gene expression during ES cell differentiation in WT and TET1 and TET2 individual knockouts............................................. 111
List of Panels

Introduction

Panel 1.1:  Manual calculation of observed over expected ratio for CGI..................  27
Panel 1.2:  Pyrosequencing in enzymatic reaction steps........................................  50
List of Abbreviations and Acronyms

2i - two inhibitors of MEK and GSK3
5mC - 5-methylcytosine
5hmC - 5-hydroxymethylcytosine
5fC - 5-formylcytosine
5caC - 5-carboxylcytosine
ADP - adenosine diphosphate
AID - adenine base activation-induced cytidine deaminase
ANOVA - analysis of variance
AP – alkaline phosphatase
ATP - adenosine triphosphate
C- cytosine base
CD31/PECAM1 - platelet/endothelial cell adhesion molecule 1/cluster of differentiation 31
CDH5/ VE-cadherin – cadherin 5/ vascular endothelial cell cadherin
CGI – CpG island
ChIP - chromatin immunoprecipitation
CpG - cytosine - phosphate - guanine
DKO - double knockout of TET1 and TET2
DMR - differentially methylated region
DNA - deoxyribonucleic acid
DNMT - DNA methyltransferase
EB - embryoid body
EC - endothelial cell
eNOS - endothelial nitric oxide synthase
ES cell - embryonic stem cell
G - guanine base
GSK3 - glycogen synthase kinase 3
H3K4me3 - histone H3 lysine 4 trimethylation
H3K27me3 - histone H3 lysine 27 trimethylation
HUVEC - human umbilical vein endothelial cell
IAS - inner antisense primer
iPSC - induced pluripotent stem cells
IS - inner sense primer
LIF - leukemia inhibitory factor
LINE - long interspersed nucleotide elements
IncRNA - long non-coding RNA
MBD - methyl CpG binding domain
MeCP2 – methyl CpG binding protein 2
MEK - mitogen-activated protein kinase kinase
NO - nitric oxide
OAS - outer antisense primer
OCT4 - octamer-binding transcription factor 4
OS - outer sense primer
PBS - phosphate buffered saline
PCR - polymerase chain reaction
PPI - pyrophosphate
RT-qPCR – real time quantitative PCR
RNA - ribonucleic acid
SAH - S-adenosyl homocysteine
SAM - S-adenosyl methionine
SEM - standard error of the mean
SINE - short interspersed nucleotide elements
SNP - single nucleotide polymorphism
SSRE - shear stress response element
TDG - thymine DNA glycosylase
TET - ten-eleven translocation
TF - transcription factor
TKO - triple knockout
TSS - transcriptional start site
U - uracil base
VEGFR2 - vascular endothelial growth factor receptor 2
VSMC - vascular smooth muscle cell
vWF - von Willebrand factor
Chapter 1

Introduction

Author’s note:
Portions of this chapter are in:

1.1 Epigenetics
1.1.1 Overview

The adult human body has over 250 different cell types, of which, most carry the same genomic sequence, yet each cell type has a distinct and specialized function. Epigenetics allows the specialization of the different cell types even though they are genetically identical, with the exception of disease and T and B cells’ receptor genes which undergo genomic rearrangement. From the perspective of gene expression, epigenetics can be broadly defined as chromatin-based mechanisms that regulate gene expression without changes to the DNA sequence per se. These regulatory mechanisms may or may not be inherited during mitosis and meiosis. Epigenetic modifications differ between cell types, allowing cell specialization. These modifications change through development and aging, are responsive to environmental cues, may be heritable, and may provide insight into understanding complex non-Mendelian diseases (Figure 1.1).

Epigenetics presently refers to three distinct yet interrelated pathways: DNA methylation, histone code and density, and RNA-based mechanisms (Figure 1.2).

Despite all epigenetic research, some contradictory opinions exist, Oliver Hobert and his associates do not necessarily believe in ‘epigenomics’ and instead claim that epigenetic marks are directly dependent on the DNA genomic sequence, thus epigenetics is an ‘effect’ of sequence-specific regulatory interactions, and not the ‘cause’ of cell-type-specific gene expression. Regardless of whether epigenetics is a cause or an effect, it proves to be important for the cell and organism as a whole due to the requirement of epigenetic reprogramming for proper embryonic development and cell reprogramming.

Epigenetics holds great relevance for understanding the etiology of many disorders and how the environment can affect a person and his/her future progeny. An example of that occurred during the end of World War II, when the western part of Netherlands experienced severe famine. The average daily ration distributed by the authorities was less than 700kcal, when the normal daily requirement is 2000kcal for
women, and 2500kcal for men. Prenatal famine exposure led to adverse metabolic and mental-health phenotypes in the subsequent generation. Such environmental factors may have a major impact on DNA methylation during embryogenesis where extensive epigenetic reprogramming occurs after fertilization. Studies showed DNA methylation differences in loci implicated in growth, metabolic, and cardiovascular phenotypes persisted throughout the person’s life course.

Epigenetics also provides great insight about genetic diseases in discordant twin studies. Twins can form from a single zygote resulting in monozygotic twins (MZ), also known as identical twins, or from two separate zygotes - dizygotic twins (DZ), also known as fraternal twins. Twin studies offer an opportunity to study the genetic contribution of disease factors, as MZ twins are genetically identical (with exceptions of their T and B cells) due to their development from a single zygote, while DZ twins share ~50% of their genomic sequence like other siblings. The power of twin studies is especially evident in MZ twins who are discordant for a phenotype or disease, which provide a unique study group where one individual develops a disorder while the other does not. The study of epigenetic pathways in the twins may provide a different perspective of understanding disease pathogenesis. The discordance may arise from skewed X-inactivation in females, asymmetric transmission of mitochondria, or genetic and epigenetic mutations that could be acquired post-zygotic separation, and may highlight the underlying biological mechanism of the disorder. Furthermore, discordant twin studies emphasize the importance of environmental influence on epigenetic modifications in particular DNA methylation, such as smoking, obesity and alcohol abuse.
Within a multicellular organism, all cells contain the same genetic code (with some exceptions). Each cell has a unique epigenetic code which make it specialized compared to other cell types. The A-T-C-G DNA code is static and conserved during cell division. The epigenetic code displays more plasticity, it is semi-conserved during cell division, can change rapidly and may respond to environment and pharmacological intervention.
Figure 1.2: Epigenetic mechanisms

Epigenetics encompasses DNA methylation and hydroxymethylation, histone code and density, and RNA-based mechanisms. These are distinct mechanisms, yet they often operate interactively during gene expression. DNA methylation occurs on a cytosine base followed by a guanine, and is correlated with gene silencing when located at proximal promoter regions. The methyl group can be oxidized to a hydroxyl group which leads to demethylation. Histone post-translational modifications create a code that can be written, read, and erased to regulate gene expression. RNA-based mechanisms are comprised of RNA that do not code for protein but instead can interact with chromatin and have gene regulatory roles.
1.1.2 Histone modifications, and variants

The field of epigenetics is predicated on the fact that DNA does not exist in a ‘naked’ state within the cell. In eukaryotes, DNA wraps around an octamer of core histone proteins (H2A, H2B, H3, and H4) to form nucleosomes. There are few histone core variants found mainly on histone H3 and H2A, which are replication and transcription dependent, and are bound by different number of DNA bases. For example, the canonical nucleosome has the histone variants H3.1/H3.2 and H2A which binds 147 DNA base pairs, while the smaller H2A variant H2A.Bbd binds 118-130 base pairs of DNA to form the nucleosome. Neighboring nucleosomes are separated by a short stretch of linker DNA as well as histone H1 which has its own variants depending on cell type and differentiation state. Post-translational modifications to the amino-terminal tail of the histone core proteins are the most highly varied of the three epigenetic pathways. In fact, more than 100 such modifications have been identified, the most prevalent of which are lysine acetylation, lysine and arginine methylation, serine and threonine phosphorylation, lysine ubiquitylation, and lysine sumoylation. The addition and removal of each modification is catalyzed by a particular family of enzymes, for instance, histone acetyl transferases (HATs) and histone methyl transferases (HMTs). Histone deacetylases (HDACs) and histone demethylases (HDMs) are responsible for catalyzing the reverse reactions, respectively.

Although some of these modifications are strictly associated with increasing transcriptional activation, such as lysine acetylation via chromatin ‘loosening,’ the effect of other modifications is context-dependent. For instance, the impact of histone
methylation is complicated, as there are 24 potential sites for this modification to exist on the four-histone core proteins, and it can be in mono-, di- or tri- methylation state\textsuperscript{21}. Furthermore, depending on the specific lysine or arginine residue methylated and on which octamer core it occurs, methylation can be associated with either transcriptional activation or repression. The modifications operate at many levels. For example, by altering the physical structure of chromatin, and histone density, they affects DNA accessibility to transcription factors and other DNA binding proteins\textsuperscript{22}. The “histone code hypothesis” is founded on the notion of growing evidence that specific combinations of these modifications cooperatively provide the cell with a particular message that is ‘read’ by one or several specific protein complexes to create a precise gene expression profile. Moreover, these messages can be ‘erased’ and ‘written’\textsuperscript{21}.

1.1.3 RNA-based mechanisms

The least well-characterized epigenetic pathway involves RNA-based mechanisms. RNA was traditionally viewed as rRNA, tRNA, and mRNA, which together play a role in the transmission of genetic information and translation into proteins. Recent advances have shown RNA to play much more active roles in gene regulation. For instance, long non-coding RNAs (lncRNAs) are a new class of functional RNAs and have been shown to play a diverse assortment of regulatory roles, including transcriptional modulation, posttranscriptional repression, RNA splicing, and microRNA sponges\textsuperscript{16,23}. Most relevant to our discussion is the ability of many lncRNAs to act as guides for epigenetic modulators, interacting with chromatin remodeling complexes and
other epigenetic machinery to regulate the transcriptional landscape\textsuperscript{24}. Many lncRNAs act along with repressive chromatin modifying complexes such as the polycomb repressive complexes, PRC1 and PRC2, while others have activating roles through the mixed lineage leukemia (MLL)\textsuperscript{25}. The classification of lncRNAs is arbitrary, many are classified based on their function on gene regulation, while others are classified based on their proximity to the genes they regulate, either on adjacent genes termed \textit{cis}, or distal genes, on other chromosomes, termed \textit{trans}. Overall, there is evidence suggesting that this novel class of RNAs plays a role in orchestrating other epigenetic processes.

1.1.4 DNA methylation

Proximal to the static DNA code, DNA methylation occurs symmetrically on both DNA strands, and it involves a covalent addition of a methyl (-CH\textsubscript{3}) on the 5-carbon position of a cytosine base creating the ‘fifth DNA base’\textsuperscript{26}. The fifth base methylcytosine (mC), was first described in 1904 by Wheeler and Johnson and purified from DNA in 1925\textsuperscript{27}. It was almost 60 years later, in 1984, when Adrian Bird and later Howard Cedar’s group who pioneered the discovery of DNA methylation and its role in gene expression\textsuperscript{28,29}. In mammals, DNA methylation occurs mostly in the context of a cytosine base followed by a guanine base on the same strand separated by a phosphate backbone, that is a CpG dinucleotide. However, in H1 human embryonic stem cells, approximately one quarter of all methylation occurs in a non-CpG context (mCpHpG and mCpHpH, where H represents adenine, cytosine or thymine) implying the presence of different methylation mechanisms or functions\textsuperscript{30,31}. Currently, it is believed that DNA
methylation of a gene promoter represses gene expression by three main mechanisms. Firstly, the methyl group acts as steric hindrance by projecting into the major groove of the DNA double helix and physically blocks recruitment of sequence specific DNA binding proteins\textsuperscript{32}. Secondly, the DNA methylation may affect the epigenetic regulation of the locus, by recruiting histone deacetylases (HDACs) and thus modifying the histone code to a heterochromatin. Lastly, DNA methylation may recruit methylcytosine-binding proteins such as MeCP2, which can block the access of activating transcription factors (TF) and recruit other chromatin remodelers\textsuperscript{33}. More recently a fourth mechanism has been suggested, implying that DNA methylation acts as a stabilizer for intrinsically labile sites that are heterogeneous between cell types, to prevent the return of TF binding after other epigenetic reprogramming\textsuperscript{34}. Unlike gene promoter methylation, intragenic DNA methylation is positively correlated with gene expression, although some suggest that it might regulate alternative splicing and alternative promoter use (Figure 1.3)\textsuperscript{35,36}. RNA polymerase II can bind at multiple sites along the genome, gene body methylation may help focus polymerase binding at the promoter, thus increasing transcription efficiency of the gene\textsuperscript{37}. The functional role of intragenic DNA methylation remains to be elucidated. Furthermore, DNA methylation has a well-established role in X-chromosome inactivation in females, and genomic imprinting where a specific locus can be methylated depending on its parental allele origin, and a wrongly imprinted allele may lead to disease such as Prader-Willi syndrome or Angelman syndrome\textsuperscript{38,39}. 
**Figure 1.3: Epigenetic-based mechanisms of gene regulation**

**A.** The epigenetic signature at the promoter and enhancer of a gene contribute to the recruitment of TF and RNA Pol II for transcription initiation regulation. Histone modifications that are observed at the proximal promoter region include H3K9Ac (pink) and H3K4me3 (orange), and common histone marks at an active enhancer include H3K27Ac (yellow) and H3K4me1 (brown). 

**B.** Intragenic epigenetic modification such as DNA methylation may affect transcription and mRNA variant expression, by recruiting TF that may promote or inhibit exon inclusion (MeCP2 and HDACs, and CTCF respectively). The H3K36me3 (purple) histone modification is associated with transcription elongation by RNA Pol II.
1.1.4.1 DNA methyltransferases

Methylation of cytosine is catalyzed by DNA methyltransferases (DNMT), it uses S-adenosyl methionine (SAM) as a methyl donor to covalently modify the 5-carbon position on the cytosine (5mC). In mammals, there are three catalytically active DNMT proteins: DNMT1, DNMT3a and DNMT3b. The de novo DNA methyltransferases, DNMT3a and DNMT3b, primarily catalyze the formation of 5mC at palindromic CpG dinucleotides\(^\text{26}\). DNMT1 is the maintenance DNA methyltransferase, and is recruited to the replication fork by UHRF1 (Set and Ring-finger-associated domain protein, also known as NP95 or ICBP90). This enables faithful propagation of CpG methylation patterns through cell division as DNMT1 methylates hemi-methylated DNA during replication\(^\text{40-42}\). Germ line DNMT1 knockout is embryonically lethal, as mouse embryo are delayed in development and do not survive past mid-gestation\(^\text{43}\). Knockout mouse embryonic stem (ES) cells are viable and show no obvious abnormalities. \(\text{Dnmt1}^{-/-}\) ES cells have 1/3 reduced DNA methylation in comparison to the wild-type cells\(^\text{43}\). The cells still have DNA methylation, which led to the discovery of DNMTs other than DNMT1\(^\text{26}\). DNMT3 knockout studies have shown that \(\text{Dnmt3a}^{-/-}\) pups are viable, although they die at about 4 weeks of after birth, and \(\text{Dnmt3b}^{-/-}\) mice are embryonically lethal, and experience multiple developmental defects including growth impairment and rostral neural tube defects with variable severity\(^\text{44}\). DNMT3a and DNMT3b double knockout embryos lack somites and do not undergo embryonic turning, indicating that their growth and morphogenesis are arrested shortly after gastrulation\(^\text{44}\). Interestingly, all ES cell lines retained undifferentiated morphology, and are high for OCT4. Triple DNMT
knockout ES cells can only contribute to extra embryonic tissues during differentiation, thus highlighting the importance of DNA methylation in cell differentiation and cell specification\textsuperscript{45}.

DNMT3L (DNMT3 Like) is a fourth member of the Dnmt family. Even though it is catalytically inactive, it has an important role in genomic imprinting and proper development of primordial germ cells (PGCs)\textsuperscript{46}. The exact function of DNMT3L is not fully understood, as it seems to play a double role in both DNA methylation and demethylation. In ES cells, DNMT3L is positively correlated with the DNA methylation of housekeeping genes’ gene body, while negatively correlated with the methylation of bivalent chromatin marked genes, those are activating and repressing histone marks at the same locus.\textsuperscript{47} A crystallography study showed that DNMT3L forms a hetero-tetrameric complex with DNMT3a. This is theorized to prevent DNMT3a from binding and localizing to heterochromatin, and thus increase its activity by binding euchromatin regions that are differentially methylated, such as imprinted regions\textsuperscript{48}. Furthermore, DNMT3L is associated with gene body DNA methylation by directing DNMT3b towards non-CpG methylation in pluripotent human cancer cells\textsuperscript{49}.

1.1.4.2 TET proteins

Ten-eleven translocation (TET) t(10;11)(q22;q23) protein was named after the fusion of \textit{TET1} gene on chromosome 10 with the mixed-lineage leukemia (\textit{MLL}) gene on chromosome 11 in some rare cases of acute myeloid leukemia (AML)\textsuperscript{50,51}. It was 2009, when the Rao group identified the potential importance of the TET protein family (TET1,}
TET2, and TET3) as mammalian homologues of Trypanosoma JBP1 and JBP2\textsuperscript{52}. The JBP (J-binding protein) proteins are iron (II)/α-ketoglutaric acid (α-KG) dependent dioxygenases that catalyze the hydroxylation of thymidine in Trypanosoma DNA which leads to the base J synthesis\textsuperscript{53}. Rao showed that TETs can oxidize 5mC into 5hmC, and 5hmC is found in surprisingly high abundance in mammals, especially in Purkinje neural cells and the brain\textsuperscript{52,54}. TETs can then iteratively oxidize 5hmC to 5fC and ScmC\textsuperscript{55}. TET1 and TET3 contain a CXXC (cysteine-X-X-cysteine) domain, which allows them to bind DNA\textsuperscript{52}. TET2 lacks CXXC DNA binding domain. It is believed that in jawed vertebrates it underwent chromosomal inversion split, splitting its CXXC domain into a separate gene, IDAX (also known as CXXC4), which is transcribed in the opposite direction\textsuperscript{56,57}. IDAX binds unmethylated CpG dinucleotides, and localizes to promoters and CpG islands where it interacts directly with the catalytic domain of TET2\textsuperscript{58}. Moreover, IDAX causes caspase activation of TET2 protein degradation, suggesting that IDAX recruits TET2 to the DNA before degradation\textsuperscript{58}. This, in turn implies that there must be other TET2 binding proteins that recruit it to methylated CpG dinucleotides. Recently, methyl-CpG-binding domain 3-like 2 (MBD3L2) protein was found to specifically modulate TET2 enzymatic activity in converting 5mC into 5hmC, by strengthening its binding to methylated DNA target\textsuperscript{59}. Therefore, it appears that MBD3L2 rather than IDAX, acts as the main DNA binding protein in TET2 activity in DNA methylation oxidation. TET proteins were also suggested to act as protein scaffolds, as functional protein partners for TET family were identified, these include chromatin-modifying proteins that are
associated with gene transcription such as OGT and HCF1, and other that are associated with gene silencing such as Sin3A$^{60}$.

Ascorbic acid, also known as vitamin-C, enhances the activity of a recombinant TET1 by interacting with its C-terminal catalytic domain, it also acts as a reducing agent for Fe(II), and thus is required as a cofactor for all three TETs oxidation and stimulates their activity$^{61-63}$. TET proteins exhibit specific developmental stage expression and tissue-specific patterns. TET1 is highly enriched at high CpG-density promoters, and is positively correlated with H3K4me3$^{64}$. It is expressed in the inner cellular mass (ICM) of a mouse blastocyst, in mouse and human ES cells and mouse E10.5-13.5 PGCs, and in the mouse brain$^{65-67}$. TET2 is broadly expressed in different mouse tissues, has been shown to regulate smooth muscle cell plasticity in humans, and like TET1 is enriched in mouse ES cells and PGCs$^{65,68}$. TET3 is downregulated after the two-cell stage and later cleavage stages when TET1 and 2 are expressed (Figure 1.8)$^{69}$. TET3 is enriched in neurons, in the oocyte and in the zygote$^{69-71}$.

As previously implied, TET proteins play a crucial role in cancer. TET proteins are often mutated in cancer, furthermore, TET2 along with DNMT3a and ASXL1 were found as the most frequently mutated genes early in the cause of hematologic malignancies$^{72,73}$. Despite TET genes being commonly mutated in cancer, TET1 knockout mice are grossly normal with the exception of variability in body size and weight$^{74}$. TET1 knockout ES cells have a slight decreased 5hmC and increased 5mC and can still form embryoid bodies (EBs)$^{74}$. The phenotype of TET2 knockout mice resembles characteristics of chronic myelomocytic leukemia (CMML) as early as 2-4 months of age,
which evolved to a wide spectrum of lethal myeloid malignancies\textsuperscript{75-78}. Creating a TET3\textsuperscript{-/-} mouse proved to be more problematic, since homozygous mutation led to neonatal lethality, a knockout was achieved by a conditional germ-line-specific deletion of TET3 in oocytes. The TET3\textsuperscript{-/-} zygotes, failed to convert the 5mC to 5hmC on the paternal pronuclei, which impede the demethylation process of the paternal Oct4 and Nanog genes. This in turn delays the activation of the paternally-derived Oct4 transgenes in early embryos. Female mice depleted of germ line TET3 show severely reduced fecundity and their heterozygous mutant offspring suffer an increased incidence of developmental failure\textsuperscript{79}.

TET1 and TET2 double knockout (DKO) ES cells are pluripotent but exhibit skewed differentiation defects toward extra embryonic lineage in a teratoma assay\textsuperscript{1}. The majority of the DKO mice died soon after birth and displayed a variety of malformations, such as exencephaly, head hemorrhaging, or profound growth retardation. The surviving DKO pups had slightly reduced body weight, compromised imprinting and reduced fertility in females\textsuperscript{1}. Triple knockout (TKO) of all three TETs is lethal, however ES cells were derived by intercrossing TET1/2/3 triple-heterozygote mice and Tet1\textsuperscript{-/-}Tet2\textsuperscript{-/-}Tet3\textsuperscript{+/--} ES cell lines. The latter cell lines were subsequently targeted to delete the wild-type (WT) Tet3 allele and obtain TET TKO ES cells. The TKO cells show a normal ES cell morphology and express pluripotency marks, however, they contributed poorly to EB differentiation and chimeras, which was rescued by TET1 ectopic expression\textsuperscript{7}. Furthermore, MEFs with a TET TKO genotype have a poor
reprogramming capacity to generate induced pluripotent cells (iPSCs)\textsuperscript{80}. The same deficiency was evident in thymine DNA glycosylase (TDG) deficient MEFs\textsuperscript{80}.

1.1.4.3 Oxymethyls

The oxidized methylcytosine base hydroxymethylcytosine (5hmC) was first isolated from a bacteriophage in 1952, and it took 20 more years before 5hmC was detected in mammals\textsuperscript{81,82}. Unlike its precursor the 5mC, which was extensively studied, 5hmC has been mostly disregarded until recently when the TET family of enzymes were found to catalyze its conversion from 5mC\textsuperscript{52,83}. Subsequently, it was shown that all TETs can potentially catalyze a stepwise oxidation from 5hmC to 5-formylcytosine (5fC) and then 5-carboxylcytosine (5caC) (Figure 1.6). These are collectively referred as the ‘oxymethyls’, and are implicated in DNA demethylation\textsuperscript{55}. The oxymethyls have been mostly ignored and hard to study since they are enriched in specific tissues or developmental stages and represent a very small fraction of all the nucleotides in the genome. The oxymethyls were easily overlooked during DNA methylation studies, as the classic method for studying DNA methylation is bisulfite converting the DNA, which allows distinguishing between C and mC. Bisulfite treatment converts all unmodified C nucleotides into U nucleotides and thus they are read as T during PCR and sequencing. mC is protected from the conversion and remains as C during sequencing (Figure 1.4). Bisulfite treatment converts hmC into cytosine-methylsulfonate (CMS), which is read as C during sequencing, making it indistinguishable from mC. The other two oxymethyls, fC
and caC are read as T nucleotides after bisulfite conversion, making them indistinguishable from C nucleotides\textsuperscript{84,85}.

Currently, 5hmC has been found to show tissue-specific distribution, with \textasciitilde40\% as abundant as mC in Purkinje neurons in the mouse cerebellum, about 5\% of the 5mC in mouse ES cells, and 1\% or less in bone marrow hematopoietic precursor cells\textsuperscript{54,55,56}. Steady-state levels of 5fC and 5caC are present \textasciitilde100 fold less than 5hmC in mouse ES cells. The low abundance might be due to robust excision by TDG, as TDG-/- mouse ES cells have a 5-10 fold increase in 5fC and 5caC implying that they are transient intermediates, unlike the more stable modification of 5mC or even 5hmC\textsuperscript{55,57}. Also, DNA polymerases have difficulty processing nascent DNA strand synthesis across the CMS during polymerase-chain reaction (PCR) amplification, and RNA polymerase II was retarded at gene elongation at 5fC and 5caC modifications\textsuperscript{84,88}. Interestingly, 5hmC was shown to have an age-dependent increase in the cortex, hippocampus and cerebellum\textsuperscript{89}. In total, the oxymethyls can affect gene regulation as they add three more C isoforms to the already known mC and C, creating a total of five CpG methyl isoforms. Potentially, each of one of the five isoforms can pair with each of the other isoform on the opposite DNA strand. Together, this creates 25 possible C-G combinations, which may increase the potential specificity for DNA binding proteins (Figure 1.5).
Figure 1.4: Reaction products and DNA sequencing reads after bisulfite conversion

A. Cytosine (C) is deaminated into a uracil (U) and sequenced as a thymine (T).

B. mC is not changed thus sequences as C.

C. hmC is converted to CMS and sequences as C.

D. fC and E. caC are both converted to C upon bisulfite treatment, get deaminated into a U, and thus sequenced as a T.
Figure 1.5: 25 possible C-G combinations

A cytosine base has potential five isoforms: C, mC, hmC, fC and caC. Although DNA methylation mostly occurs in parallel on both DNA strands, it was suggested that the cytosine residue on the sense DNA strand could be different than on the antisense strand, especially in the oxymethyl forms. This creates the possibility of 25 different CG combinations between the two DNA strands.
1.1.4.4 DNA demethylation

DNA demethylation is believed to happen via one of two mechanisms, one being passive demethylation, and alternatively active demethylation (Figure 1.6). Conservation of DNA methylation patterns during somatic cell divisions is made possible by the activity of the maintenance methyltransferase DNMT1, which has a preference for hemi-methylated DNA and adds appropriate methylation marks to the nascent DNA strand \(^{40,41}\). Studies have shown that DNMT1 has a reduced affinity towards 5hmC, and thus has reduced recognition for 5hmC/C compared to 5mC/C. As a consequence, DNMT1 does not methylate the nascent DNA strand in the case of 5hmC/C, which will result in a diluted methylation mark upon cell division, leading to passive DNA demethylation\(^{90}\). DNA methylation could be influenced by biological activity, when local changes in pH can result in tautomerization of 5mC, from the common amino form to the imino form, which is not recognized by methyl-CpG-binding domains (MBD)\(^{91}\). Thus, DNA demethylation can also be thought of as a failure to recognize the methylated CpG site.

Many enzymes and pathways have been suggested for active DNA demethylation, the most prominent one includes the AID/APOBEC family members, a glycosylase coupled with base excision repair (BER) pathway\(^{92}\). In 2011, two papers described TDG as a potential player in TET-mediated DNA demethylation\(^{93,94}\). They showed that TDG enzyme which excises the T base from a G-T base mismatch recognizing the double-stranded template mismatch, additionally can recognize 5fC and 5caC and excise them forming an abasic site and initiating the BER pathway (Figure 1.7).
The BER pathway is employed to remove incorrect or damaged bases (mismatch repair), and consists mainly of three steps. The first includes the removal of the incorrect or modified base by an appropriate DNA N-glycosylase (TDG in the case of T-G, 5fC or 5caC), creating an AP site (apurinic/apyrimidinic). In the second step, nicking of the targeted AP site by an AP lyase occurs thereby creating a 3’OH overhang by an AP endonuclease. The last step involves replacing a new sugar to the phosphate backbone by Poly (ADP-Ribose) Polymerase (PARP). This is followed by insertion of an appropriate nitrogenous base by a DNA polymerase, and ligation of the nicked DNA backbone via a DNA ligase\textsuperscript{95,96} (Figure 1.7). PARP1 is also known to modify proteins by poly(ADP-ribosylation) (PARylation), most commonly PARP1 is known to PARylate itself, in which form it can bind DNMT1 restricting it from binding DNA and thus indirectly causing DNA demethylation\textsuperscript{97,98}. Growth arrest and DNA-damage-inducible protein 45 (GADD45) has been implicated in active DNA demethylation for a long time, but only recently it was shown to contribute to active DNA demethylation through TDG excision of 5fC and 5caC. GADD45a showed to physically interacts with TDG and contribute to its activity in DNA methylation, furthermore, when GADD45a and its sister protein GADD45b are knocked out there is hypermethylation of TDG binding sites in mouse ES cells\textsuperscript{99}. Interestingly \textit{in vitro} studies found that DNMT3a and DNMT3b have the potential to excise 5hmC depending on their protein redox state, thus raising new questions regarding the structural and functional aspects of the methyltransferases\textsuperscript{100}. However, the exact list of players and mechanisms of active DNA demethylation still remains to be elucidated.
Figure 1.6: Pathways of DNA methylation and demethylation

DNMT1/3a/3b catalyze the addition of a methyl group onto a C base. The mC can be oxidized sequentially by TET 1/2/3 into hmC, fC, and caC. hmC is not well recognized by DNMT1 the DNA methyltransferase that is active during cell replication, thus leading to passive DNA demethylation during cell replication. fC and caC can be excised by TDG creating a site lacking DNA base, which can than be repaired and substituted by the BER pathway into an unmodified C base.
Figure 1.7: Base excision repair pathway

When there is a T-G (or fC-G, or caC-G) base mismatch, TDG can recognize it and excise the base, leaving only the sugar phosphate DNA backbone (AP site). AP lyase creates a nick in the sugar, breaking the backbone, followed by an AP endonuclease, which cuts the nicked sugar leaving only the phosphates. A new sugar and a proper base are then restored by PARP, DNA polymerase, and a ligase enzyme which links and fixes the DNA backbone.
1.1.4.5 CpG islands

5mC is the most mutagenic base in the mammalian genome, as it is susceptible to deamination into thymine base, thus making the genome CpG depleted and TpG or AT-enriched\textsuperscript{101}. Interestingly, there are some regions in the genome where there is a high density of CpG dinucleotides termed CpG islands (CGI). There have been numerous definitions of CGI, based on a variety of parameters often including CpG percentage, length and the number of observed CpGs\textsuperscript{102}. Computer algorithms have been developed as biostatistical tools to identify CGIs, in particular CpGPro D program\textsuperscript{103}. Using this software CGI is defined as a stretch of DNA in a window of 500bp, with an average G+C frequency above 0.5, and a moving average high CpG content relative to the expected ratio of CpG dinucleotide to occur by chance alone (observed/expected) greater than 0.6 (Panel 1.1)\textsuperscript{103,104}. Further characterization of CGIs divides them into poor (ratio<0.45), weak (ratio 0.45-0.75), and strong (ratio>0.75)\textsuperscript{105}. It is believed that CGIs are in evolutionary transition to either loosing their CpG content or gaining CpGs towards a ratio of 1, where the abundance of CpGs is equivalent to that expected by chance alone, based on the overall G or C content in the region. This evolutionary migration of CpGs is especially evident in the differences between individuals, where there is a great variation in the edges of the CGI, namely CGI shores and shelves\textsuperscript{106}. Approximately 70% of all annotated protein coding gene promoters are associated with CGIs, making it the most common promoter regulator type in the vertebrate genome\textsuperscript{107}. Strong CGIs are usually unmethylated, and are found at promoters of housekeeping and development regulators genes, where they act as transcriptional initiation sites\textsuperscript{108,109}.
How CGIs remain unmethylated is not fully understood, however, characterization of chromatin in CGI promoters provides some clues. It was found that unmethylated CGIs are usually nucleosome depleted, are flanked by the histone variant H2A.Z and the histone modification H3K4me3, which have been shown to block de novo methyltransferases\textsuperscript{110,111}. All three TET proteins have been found to localize primarily to CGI promoter regions, which can contribute to a lack of DNA methylation\textsuperscript{64,112}. Moreover, ChIP-seq experiments have shown that there are proteins that specifically bind unmethylated CGIs such as CFP1 (also known as CXXC1), and might block DNMTs accessibility\textsuperscript{113}.

CGIs can be methylated in certain genomic regions. CGI methylation can occur in genomic imprinted regions, the inactive X-chromosome, and repressed repetitive elements\textsuperscript{105}. In cancer, there appears to be a pathological hypermethylation of CGIs, which are not methylated in healthy cells\textsuperscript{105,114}. Even though CGIs are mostly unmethylated, it was found that CGI shores could be differentially methylated. Moreover, the DNA methylation of a CGI shore located downstream of the TSS tend to have a greater association with its gene transcriptional regulation than any DNA methylation upstream of it\textsuperscript{115,116}. Promoters that have poor CGIs tend to be differentially methylated (contain a differentially methylated region - DMR) between different cell types and regulate enriched cell type specific expression\textsuperscript{2} (Table 1.1).
Panel 1.1: Manual calculation of observed over expected ratio for CGI

To manually calculate O/E ratio for CGI:

C – cytosine base, G – guanine base, N – length of the DNA examined, CpG – number of Cs followed by Gs, O – observed, E – expected.

\[ \frac{O}{E} = \frac{\# \text{ of observed CpG in the DNA stretch}}{\# \text{ of expected CpG taking into account the CG content of the length of the DNA}} \]

\[ \frac{O}{E} = \frac{(\text{CpG}/N)}{(C/N*G/N)} = \frac{\text{CpG}*N}{C*G} \]

For example: if we are looking at a 500bp DNA sequence, with a 60% CG content and observed 12 CpGs (assuming equal amount of Cs and Gs means out of 500bp, 60% of CGs would be 300 Cs or Gs, so 150 Cs and 150 Gs)

\[ \frac{O}{E} = \frac{12*500}{(150)(150)} = 0.27 \text{ Thus it would be a poor CGI} \]
Table 1.1: CGI characterization of EC-enriched gene promoters

<table>
<thead>
<tr>
<th>Gene</th>
<th>Amplicon size</th>
<th>Location</th>
<th>#CGs</th>
<th>%CG</th>
<th>O/E</th>
<th>CGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>eNOS</td>
<td>289</td>
<td>-286 to +3</td>
<td>9</td>
<td>64</td>
<td>0.23</td>
<td>Poor</td>
</tr>
<tr>
<td>CD31</td>
<td>359</td>
<td>+92 to +450</td>
<td>13</td>
<td>53</td>
<td>0.33</td>
<td>Poor</td>
</tr>
<tr>
<td>Endoglin</td>
<td>423</td>
<td>-333 to +90</td>
<td>11</td>
<td>66</td>
<td>0.42</td>
<td>Poor</td>
</tr>
<tr>
<td>ICAM-2</td>
<td>472</td>
<td>-248 to +224</td>
<td>10</td>
<td>56</td>
<td>0.19</td>
<td>Poor</td>
</tr>
<tr>
<td>P-selectin</td>
<td>451</td>
<td>-562 to -112</td>
<td>4</td>
<td>42</td>
<td>0.14</td>
<td>Poor</td>
</tr>
<tr>
<td>Tie-2</td>
<td>268</td>
<td>-295 to -28</td>
<td>5</td>
<td>44</td>
<td>0.26</td>
<td>Poor</td>
</tr>
<tr>
<td>CDH5</td>
<td>500</td>
<td>-249 to +251</td>
<td>19</td>
<td>65</td>
<td>0.27</td>
<td>Poor</td>
</tr>
<tr>
<td>vWF</td>
<td>480</td>
<td>+25 to +504</td>
<td>4</td>
<td>53</td>
<td>0.09</td>
<td>Poor</td>
</tr>
<tr>
<td>VEGFR-1</td>
<td>439</td>
<td>-301 to +138</td>
<td>49</td>
<td>70</td>
<td>0.83</td>
<td>Strong</td>
</tr>
<tr>
<td>VEGFR-2</td>
<td>206</td>
<td>-225 to -19</td>
<td>15</td>
<td>65</td>
<td>0.78</td>
<td>Strong</td>
</tr>
</tbody>
</table>
1.1.4.6 DNA methylation during development

Differentiation from one pluripotent cell into all of the different cell types of the embryo necessitates the expression of unique cell-type specific genes and silencing of others. Differentiated cells acquire epigenetic marks that distinguish them from pluripotent cells to somatic cells of other lineages. Throughout development these epigenetic marks are flexible, transient and reversible. Two major waves of epigenetic reprogramming are observed during development: those events that take place during pre-implantation development and a distinct set that occur in primordial germ cells (PGCs)

Male and female gametes are terminally differentiated and extremely specialized cells. Both sperm and oocyte are known to possess considerable levels of DNA methylation, however, the male genome is hypermethylated compared to the female genome. To facilitate efficient transport, sperm chromatin is tightly wrapped with protamines rather than histones, therefore, bears a paucity of histone-related epigenetic marks upon fertilizing an oocyte. Upon fertilization, the male pronucleus is stripped of its protamines and replaced with histones. Within six hours of fertilization, but prior to pronuclei fusion, a profound paternal genome-wide DNA demethylation event occurs. Intriguingly, this apparent DNA demethylation event has been shown to be a mass oxidation event. Maternally derived TET3 translocates to the paternal pronucleus upon fertilization, and oxidizes the methylated DNA. In contrast, the maternal genome bears histone modifications that protect its DNA methylation marks from removal. Specifically, dimethylation of H3 histone on lysine 9 (H3K9me2) attracts
developmental pluripotency-associated 3 protein (DPPA3, also known as STELLA) that protects most of the maternal genome from oxidation, with imprinting regions being an exception\textsuperscript{123,124}. Following the oxidation in the male haploid genome, DNA demethylation is achieved passively by cell division, which dilutes out the methylcytosine and oxymethyls\textsuperscript{125} (Figure 1.8).
Figure 1.8: DNA methylation reprogramming during development

There is a dynamic nature of 5mC, 5hmC, and TET enzyme levels in the pre-implantation embryo, during PGC development, and in ES cell differentiation. Intensity of blue, red, and purple shading indicates the relative levels of modified cytosine in the paternal, maternal, and zygotic genome at each developmental stage respectively. Green shading shows relative TET protein expression over the same time course.
1.1.4.7 Inheritance of Epigenetic marks

It is important to understand how epigenetic marks are being propagated, as heritable cellular memory is essential for the preservation of cell identity throughout the life of a multicellular organism (Figure 1.1). Unlike the genetic code, epigenetic marks have a certain amount of plasticity. Yet in order to maintain mature cell lineage specificity it is important that the epigenetic marks are faithfully maintained after each cell cycle. Upon DNA replication, the newly synthesized daughter strand is paired with a mother template strand that carries epigenetic marks. UHRF1 helps recruit DNMT1 to hemimethylated DNA and to the replication fork through its interaction with PCNA, a DNA polymerase delta accessory protein\(^{126,127}\). PCNA also appears to play a role in coupling chromatin-modifying enzymes with the replication machinery, as it interacts with chromatin assembly protein CAF-1, and histone modifying enzymes such as HDACs and histone methyltransferases\(^{128,129}\). After DNA replication, PCNA accumulates on newly synthesized DNA lacking histones to provide CAF-1 an opportunity for the synergistic restoration of nucleosomes and interplay with DNA mismatch repair machinery\(^{130-132}\).

Replication of the eukaryotic genome follows a systematic and temporal program, where gene density and transcriptional activity correlates with early replication, while repetitive elements and transcriptionally repressed genes correlate with late replication\(^{133}\). Differential promoter DNA methylation was also shown to play a role in replication timing, as in the case of EC-enriched genes\(^2\). The accepted dogma is that transcriptionally active genes replicate early. However, as we reported, EC-enriched
genes such as eNOS, CD31, and ICAM-2 which are regulated by DNA methylation, were found to replicate early in both expressing and non-expressing cell types, highlighting the importance of promoter DNA methylation in controlling gene expression in ECs or suppressing EC genes in other cell types\(^2\). Suggesting that late replication may provide insufficient time to replicate the epigenetic code with high fidelity. This exciting finding requires further study in ECs and other cell types. The finer details of the maintenance of the epigenetic signature in EC, are yet to be elucidated. As the relative role of epigenetic check points in the mitotic cell cycle is a poorly explored concept.

1.1.5 Epigenetic Heterogeneity

Promoter DNA methylation studies, usually are divided into hypermethylated (90%-100%) or hypomethylated (0%-10%) states. This pattern is commonly observed in mature ECs, when comparing the EC-enriched gene promoters such as eNOS in EC vs non-expressing cell type like VSMC\(^{134}\) (Figure 1.9). When the overall level of DNA methylation gives an average methylation level (~50%) various scenarios may be evident. Most current work is at the high throughput sequencing level, which only provides average reads per site. It is not possible to decipher what 50% DNA methylation means unless single strand DNA analysis is done to define the phase of differential CpG DNA methylation. Figure 1.10 depicts three scenarios for what 50% methylation could look like. Each example represents a distinct overall cell population. One case of 50% methylation can happen in imprinted loci, when one parental allele is methylated and the other is not (Figure 1.10b)\(^{135}\). In the second case a DNA methylation
pattern can be clonal within a cell population and give an average read of 50% between two epialleles (Figure 1.10c). The last case is depicted in figure 1.10a, is what we refer to as heterogeneous DNA methylation, where each cell in the population is different in its DNA methylation status of that particular DNA locus. Under standard tissue culture conditions (in fetal bovine/calf serum and LIF), ES cells receive continuous differentiation signals, however being held back from differentiating by LIF. They express high levels of DNMT3a and DNMT3b, along with TET1 and TET2 implying a continuous epigenetic reprogramming of the cells, even though they resemble somatic cells in the global level of their DNA methylation, ES cells form a heterogeneous population\textsuperscript{136-138}. Prior unpublished work in our lab has shown that in ES cells, poor CGI EC-enriched gene promoters that become differentially methylated in mature cells are heterogeneously methylated, while EC-enriched genes with strong CGI promoter that are not regulated by DNA methylation in mature cells, are not methylated in ES cells (Table 1.1, Table 1.2). This was confirmed in multiple mouse ES cell lines and replicated in human ES cells, it formed the basis for studies in this thesis.

Shipony et al. performed clonality assays in ES cells to examine inheritance of DNA methylation from mother to daughter cells\textsuperscript{139}. They found that during ES cell division daughter cells do not maintain the same DNA methylation signature as the mother cell, inferring that the DNA methylation is not clonal. On the contrary, fibroblast daughter cells had a matching DNA methylation profile to their mother cell, implying that there is a highly clonal inheritance of DNA methylation in differentiated cells. Thus, it was suggested that ES cells and differentiated cells might possess different DNA
methylation maintenance pathways from mother to daughter cell. One piece of evidence was the discovery of the four splice variants of DNMT3a, as two of them are enriched in ES cells, and the other two isoforms are expressed in differentiated cells\textsuperscript{140}. It is not known exactly how the isoforms differ functionally or what causes the cell to switch between them, but it was suggested the different DNMT3a isoforms represent distinct epigenetic regulators between ES cells and differentiated cells resulting in the epigenomic plasticity in ES cells.

Many refer to this type of heterogeneous DNA methylation as ‘noise’, however, recent work suggests that this methylation ‘noise’ might be important in cancer. Cancer cells display DNA methylation heterogeneity as well; chronic lymphocytic leukemias (CLLs) samples from patients who had increased ‘disordered’ (heterogeneous) DNA methylation are associated with adverse clinical outcome, as it might contribute to genetic instability\textsuperscript{141}. Heterogeneous DNA methylation is also associated with some healthy differentiated cells. For example, EC gene promoter of VWF was found to be associated with bimodal gene expression switching on and of in particular endothelial vascular beds to be a characteristic of healthy endothelium\textsuperscript{142}.

Histone modifications also exhibit heterogeneity in specific cases. Histone modifications are associated with chromatin density and transcription. For example, histone 3 lysine 27 trimethylation (H3K27me3) and histone 3 lysine 9 trimethylation (H3K9me3) correspond with heterochromatin and gene silencing, while H3K4me3 at the promoter region and H3K36me3 within the transcribed regions, are associated with active genes, respectively. Bivalent chromatin marks refers to the presence of repressing
(H3K27me3) and activating (H3K4me3) histone modification at the same nucleosome on one strand of DNA. Chromatin immunoprecipitation (ChIP) assay followed by re-ChIP confirmed this concept\textsuperscript{143}. This mark was first described in murine ES cells at tissue-specific genes, and is argued to mark the genes as poised for subsequent activation or repression\textsuperscript{30,143,144}. As the cells differentiate, the bivalent mark usually resolves to either an active state H3K4me3, or a repressed state H3K27me3, which premarks genes for DNA methylation\textsuperscript{145}. Interestingly, TET1 immunoprecipitation binding assays showed that its CXXC domain binds CGI promoter regions where 49\% were univalent H3K4me3, 44\% bivalent, and 2.6\% univalent H3K27me3, suggesting TET1 plays a role in the cell’s decision as to which genes are activated upon late differentiation\textsuperscript{64}. These studies are valuable especially when they are done on a single cell level, rather than on a population, as Kumar et al. have found that even gene expression can be heterogeneous in ES cell population\textsuperscript{146}.

Some groups have tried to eliminate or reduce the ES cell heterogeneity to a more ‘naïve’ or ‘ground’ state by using two differentiation inhibitor drugs referred to as 2i\textsuperscript{138,147}. The 2i are CHIR99021 which targets glycogen synthase kinase-3 (GSK3), and PD0325901 which targets mitogen-activated protein kinase kinase (MEK, also known as MAP2K or MAPKK)\textsuperscript{147}. Treating cells with 2i led to a reduced global 5mC in ES cells, and upregulation of positive regulatory domain (PR-domain) containing transcriptional regulator 14 (PRDM14). An increased PRDM14 expression was associated with the downregulation of DNMT3a and DNMT3b in 2i treated ES cells, consequently causing global reduced DNA methylation\textsuperscript{148,149}. Moreover, a recent publication found that in the
absence of LIF, PRDM14 can maintain ES cell pluripotency through TET-mediated active DNA demethylation\textsuperscript{150}. As in TET1 and 2 DKO ES cells and when administrating BER inhibitors impaired the PRDM14-induced resistance of ES cell differentiation, suggesting that PRDM14 might have a key role in both DNMT and TET activity\textsuperscript{150}. Epigenetic heterogeneity has been a very murky area, as it has been overlooked for a long time and only now some of its notions are starting to be appreciated. It is a reminder that within a population of cells, not all cells are identical, heterogeneity may exist, and it might be physiologically relevant.
Figure 1.9: DNA methylation of the eNOS proximal promoter in endothelial cells and vascular smooth muscle cells

Bisulfite sequencing of the eNOS proximal promoter in humans (A, B) and mouse (C,D). The x-axis indicates the CpG dinucleotide position from the transcriptional start site (TSS) in both human and mouse, while the y-axis indicates % methylation. The DNA methylation was measured in human umbilical vein endothelial cells (A), human aortic vascular smooth muscle cells (B), mouse endothelial cells (C), and mouse aortic smooth muscle cells (D).
Figure 1.10: 50% DNA methylation pattern

Three mechanisms of achieving 50% DNA methylation. A. Heterogeneous DNA methylation. B. Imprinted DNA methylation. C. Clonal DNA methylation in epialleles. Each DNA methylation pattern is displayed in three ways: a lollipop diagram – displayed are separate strands of DNA with the circles representing CpG sites, white circles are non methylated, black circles are methylated. CpG % methylation – histogram shows how often a particular site is methylated. Frequency histogram – histogram shows the frequency of certain amount of methylated sites per DNA strand.
1.2 Vascular Endothelium

1.2.1 Endothelium

Vascular endothelial cells (ECs) are derived from the mesodermal lineage, similar to their neighboring vascular smooth muscle cells (VSMC)\(^{151}\). ECs line the lumen of all blood vessels in mammals, and play a critical role in regulating homeostasis of the blood and tissues. The monolayer of ECs display a distinct gene expression profile, due to their interaction with the basement membrane, adjacent VSMCs and, the force of shear stress produced by blood flow\(^{152}\). ECs exhibit a great array of phenotypic vascular bed heterogeneity, which is visible at the microscope and ultrastructural level. For example, arterial ECs and venous ECs have different functions in the vasculature, as the former are elongated and align with the direction of the blood flow, and the latter appear cobblestone like due to lower overall shear stress\(^{152}\). The endothelium can also be classified as continuous, fenestrated, or discontinuous. As ECs can also be porous, or have gaps between the cells, which can also be found with complete or porous basement membrane\(^{152}\). Gene expression analysis can distinguish various ECs, such as arterial and venous ECs, and microvascular tissue ECs, implying that ECs express certain genes to serve specific functions based on their local physiology\(^{153}\).

Endothelial function, phenotype and steady-state gene expression differ in disease. ECs show a different pattern of adhesion molecules in sites of the vasculature predisposed to atherosclerotic lesion formation\(^{154}\). Moreover, this observation was made in adjacent regions of the vascular tree, therefore not necessarily due to different ECs type. A common explanation for this variability can be attributed to differences in
flow patterns that can induce differential gene expression in ECs. Notably, a unique set of highly propagating ECs has been identified and it represents a subpopulation that may play important roles in the repair of the vascular tissue. Moreover, these cells have been shown to be functionally important in therapeutics and have high proliferative potential (HPP ECs). Low proliferative potential endothelial cells (LPP-ECs) can be defined as ECs which are capable of undergoing a small amount of replication, sufficient to generate small EC outgrowth colonies (2-500 cells). HPP ECs are capable of expanding to over 2000 cells and are central to developing EC therapeutics.

1.2.2 Epigenetic regulation of endothelial gene expression

ECs are a model system for the study of gene-enriched expression, as they highly express one of three isoforms of nitric oxide synthase (NOS). Endothelial NOS (NOS3, eNOS) is responsible for the majority of endothelial cell-specific production of nitric oxide (NO), a powerful anti-atherogenic and anti-thrombic signaling molecule which acts as a potent vasodilator to maintain vessel tone. The classic cis/trans concept of gene expression is the simplest model to rationalize the cell-restricted expression of eNOS. Based on this model, cell restricted trans-acting factors are recruited and bind cis-DNA binding elements in the 5′ regulatory regions of genes. NeuroD is one such example of a master transcriptional regulator in neurons. To date, the identity of a master regulator of ECs remains unidentified, although numerous of preferentially expressed transcription factors exist. Alternatively, the cooperative activity of an ensemble of endothelial enriched transcription factors might be responsible for
constitutive expression of endothelial-enriched genes, the model gene eNOS does not contain endothelial restricted transcription factor sites\textsuperscript{161}. There are limitations to the \textit{cis/trans} model and little direct evidence to suggest it is a prominent system in eNOS gene regulation\textsuperscript{161,162}. Thus, alternative pathways such as epigenetics have been proposed largely centering on chromatin accessibility and compaction. Chromatin-based regulation of EC-enriched genes has been observed, namely eNOS, VE-cadherin (CDH5), PECAM1 (CD31), von Willebrand factor (vWF), and E-selectin, among others\textsuperscript{2,163}.

Closer observation of the eNOS promoter shows a lack of a canonical TATA box and no proximal CGI. Insights into the nuances of eNOS expression are derived largely from the epigenetic perspective at the proximal promoter. The proximal promoter is defined as a stretch of few hundred nucleotides that contain the core promoter region, direct RNA polymerase II binding, and initiation of transcription\textsuperscript{2}. Early experiments indicated greater accessibility in the chromatin structure of eNOS in ECs compared to non-ECs\textsuperscript{164}. Firstly, a differentially methylated region (DMR) in the proximal promoter of eNOS has been observed to be unmethylated in ECs, whereas it is heavily methylated in non-expressing cell types (Figure 1.9). Moreover, treating non-expressing cell types with the DNMT inhibitor 5-azacytidine resulted in increased eNOS expression\textsuperscript{134}. Even though eNOS proximal promoter is not methylated in healthy ECs, we have observed some gene body DNA methylation in both ECs and non-ECs (Figure 1.11). The gene body DNA methylation was not correlated with gene expression, and its role at the eNOS gene remains to be elucidated. Further examination of eNOS chromatin structure in human ECs revealed the presence of a myriad of histone marks associated with
transcription activation, such as the acetylation of histones H3 and H4 (H3ac and H4ac) and the di- and tri-methylation of H3 lysine4 (K4), in contrast to various non-expressing cell types\textsuperscript{162}. In addition to increased H3 acetylation at the eNOS proximal promoter, treatment of vascular smooth muscle cells (VSMCs) with a pharmacological HDAC inhibitor, trichostatin A, also resulted in increased eNOS mRNA expression levels. Moreover, recent \textit{in vitro} and \textit{in vivo} studies have found these marks to be dynamic as histone eviction and decreased histone-activating marks at the promoter of eNOS are found during acute hypoxia. In contrast, in chronic hypoxia, normal levels of the histone octamers are detected, but activating markers that are observed under normoxic conditions fail to be re-established\textsuperscript{165}. Interestingly, hypoxia also upregulates the expression of a natural antisense eNOS gene, sONE (eNOS antisense, NOS3AS) that overlaps the 3` end of the eNOS gene\textsuperscript{166}. sONE has been shown to regulate eNOS at the posttranscriptional level through destabilization of eNOS mRNA, through the loss of RNA binding proteins\textsuperscript{166,167}. Collectively showing that eNOS is regulated by DNA methylation, histone post-transcriptional modifications and density, and an RNA antisense. Our lab also shown both \textit{in vitro} and \textit{in vivo} studies of differential DNA methylation in the proximal promoter regions of other EC-enriched genes CD31/\textit{PECAM1}, Endoglin/\textit{ENG}, \textit{ICAM-2}, P-selectin/\textit{SELP}, Tie-2/\textit{TEK}, VE-cadherin/\textit{CDH5}, and \textit{von Willebrand factor (vWF)} in terminally differentiated ECs vs non-ECs in both humans and mice, and provided evidence that epigenetic modifications are functionally important for EC gene expression\textsuperscript{2}. 
Table 1.2: EC-enriched genes that are not methylated in ECs and their DNA methylation status in VSMCs²

<table>
<thead>
<tr>
<th>Gene</th>
<th>DNA methylation detected?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Endothelial cells</td>
<td>Vascular smooth muscle cells</td>
</tr>
<tr>
<td>eNOS</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>vWF</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>ICAM-2</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>P-selectin</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>VE-Cadherin</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>CD31</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>Tie2</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>Endoglin</td>
<td>X</td>
<td>+</td>
</tr>
<tr>
<td>VEGFR-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>VEGFR-2</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Figure 1.11: DNA methylation of eNOS promoter and gene body

DNA methylation status of the human eNOS gene at the proximal promoter, first exons and introns, and exon 14. The proximal promoter, exon 1, intron 1, exon 2, and intron 2 are differentially methylated between HUVECs and VSMCs. At intron 4, exon 5, intron 5, exon 6, intron 6, and exon 14 there is gene body methylation in both HUVECs and VSMCs. N=15 for each amplicon.
1.3 Methods for studying cytosine methyl modifications

Numerous methods have been developed to study DNA methylation, however, most of them are specific for one modification, most commonly 5mC. A cytosine modification is not maintained during PCR amplification, unless modified free bases are added to the reaction mix. Thus, DNA methylation assays require either to chemically modify the 5mC base by bisulfite conversion, or target the 5mC by restriction enzymes or by antibodies.

1.3.1 Sodium bisulfite conversion

Sodium bisulfite remains the gold standard method for studying DNA methylation. It is the only quantitative method that positively identifies 5mC at single base resolution. This method entails chemical treatment of the DNA using sodium bisulfite. Sodium bisulfite acts on a single stranded DNA, and requires denaturing prior to treatment. Sodium bisulfite treatment will result in deamination of cytosine residues thus converting it to uracil, mC deamination is ~100 times slower than unmodified cytosine residue deamination, thus it remains unmodified. Cytosine undergoes sulphonation and then hydrolytic deamination, followed by alkali desulphonation to yield uracil (Figure 1.12). The sodium bisulfite converted DNA can then be PCR amplified, using bisulfate-specific primers, sequenced and scored.

Sodium bisulfite treatment of the DNA presents a few limitations. It is a harsh chemical treatment of the DNA, therefore it requires a relatively large DNA input. Also, it depends on complete conversion of all cytosines, as any unconverted cytosine can be mistaken for mC. Sodium bisulfate-based methods do not distinguish between unmodified cytosine, methylated cytosine, and the oxymethyls. Sodium bisulfite converts hmC to cytosine methylsulfonate (CMS), scoring it as C during sequencing, whereas it and deaminates fC and
caC, scoring them as T during sequencing\textsuperscript{84}. The oxymethyls can also affect the efficiency of the DNA polymerase, which may lead to underrepresentation of oxymethylated DNA strands during PCR amplification\textsuperscript{84}.

There are other variable sequencing techniques that can be applied to bisulfite converted DNA. After PCR amplification the product can be ligated into prokaryotic plasmids and subcloned prior to sequencing. Since bisulfite conversion of the sense DNA strand causes it to be non-complementary to its antisense, each cloning will represent single ‘original’ DNA strand methylation. Other methods such as pyrosequencing, and high-throughput sequencing gives an average read of methylation level per site, yet it is more quantitative.
Figure 1.12 Bisulfite treatment chemical reaction on a cytosine base and sequencing

A. Sulphonation occurs as bisulfite reacts with the double bond between carbons 5 and 6 of C to form the cytosine sulphonate adduct. Next, hydrolytic deamination occurs as cytosine sulphonate is deaminated at carbon 4 to form the uracil sulphonate adduct. Finally, alkali desulphonation occurs as uracil sulphonate is converted to uracil. B. After bisulfite treatment, the two DNA strands will not be complimentary any more as all Cs will be sequenced as Ts, and all mCs will be sequenced as Cs.
1.3.2 Pyrosequencing

Pyrosequencing is a four-enzyme step DNA sequencing method that detects the sequential incorporation of nucleotides using bioluminescence\textsuperscript{171}. DNA is denatured and only one strand is sequenced using a single primer (Panel 1.2). Individual nucleotides, in the form of dNTP, are sequentially added to the DNA template. When a nucleotide is incorporated by the DNA polymerase, there is a release of PPI. In the second step, adenosine phosphosulfate (APS) is added with ATP sulfurylase to convert the PPI into ATP. In the third step, Luciferin and Luciferase enzyme are added. In the presence of ATP, there will be an oxidation reaction producing light. The byproducts are oxyluciferin, AMP, and PPI. The last step involves the enzyme Apyrase; which will degrade any unincorporated nucleoside triphosphates, allowing nucleotides to be added sequentially in one tube\textsuperscript{172}. The emitted light from step 3 is detected by a charged-coupled device (CCD) camera which converts the light intensity into an electronic signal. The amount of light emitted is proportional to the number of incorporated nucleotides (Figure 1.13). Pyrosequencing can be done on specific genes using specific primers, or can be used in a global methylation assay such as looking at the methylation level of the repetitive LINE-1 elements.
Panel 1.2: pyrosequencing in enzymatic reaction steps

(DNA)_n + dNTP → (DNA)_{n+1} + PPI

PPI + APS → ATP + SO_4^{2-}

ATP + Luciferin + O_2 → AMP + PPI + oxyluciferin + CO_2 + light

(DNA)_n — sequence of DNA of length n, dNTP — deoxynucleotide triphosphate, PPI — inorganic pyrophosphate, APS — adenosine phosphosulfate, ATP — adenosine triphosphate, AMP — adenosine monophosphate.
Figure 1.13: Pyrosequencing

Pyrosequencing couples incorporation of a nucleotide with the production of light. When a nucleotide is incorporated, it releases an inorganic phosphate (PPI) as a byproduct, which is then converted to ATP by ATP-sulfurylase. The ATP is used up in the oxidation of luciferin by luciferase, producing light that is recorded and converted to an electrical signal. Apyrase enzyme is required to digest any unincorporated nucleotides. The light intensity is proportional to the amount of bases incorporated. This method could be used to sequence bisulfite DNA and LINE-1.
1.3.3 LINE-1 – repetitive DNA sequencing

Long interspersed nucleotide elements (LINE-1) are active repetitive elements in a subclass of retrotransposons. There are ~500,000 LINE-1 (L1) copies in the human and mouse genome, which span across multiple chromosomes and comprise ~19% of the total genome\textsuperscript{173}. The full length of an L1 is about 7Kb; however, most of them are inactive due to various mutations, truncations and transcriptional inactivation by DNA methylation\textsuperscript{174,175}. Since L1 are highly abundant and are heavily methylated in both human and mouse genome, they have been used as a tool to estimate global methylation in cells\textsuperscript{136}.

1.3.4 Restriction Enzyme-based methods

Restriction enzyme methods assay genome-wide DNA methylation level by utilizing two isoschizomer restriction enzymes, one of which is DNA methylation sensitive and the other insensitive. The most commonly used pair is \textit{MspI} and \textit{HpaII}, which both recognize and cut 5'-CCGG-3'. \textit{MspI} is an efficient cutter and is insensitive to CpG methylation. \textit{HpaII} on the other hand, only cuts unmethylated CpG sites, thus is DNA methylation sensitive. \textit{HpaII} is not an efficient cutter except at multiple adjacent \textit{HpaII} sites, thus to increase its cutting efficiency double stranded oligonucleotides containing the cutting site should be spiked into the digestion mixture\textsuperscript{134,176}. The CCGG sequence constitutes for ~1.5 million CpG sites out of a total of 21 million CpGs in the mouse genome\textsuperscript{177}. After digestion, DNA methylation can be assayed by a variety of methods, most common of which are reduced representation bisulfite sequencing (RRBS), luminometric assay (LUMA), and the EpiMark kit (New England BioLabs).
RRBS decreases the complexity of whole genome high-throughput sequencing by digesting the DNA, which greatly reduces the number of CpG sites to be analyzed, making it more CG-rich method\textsuperscript{178}.

LUMA assay includes a third enzyme digest. \textit{EcoRI} cuts at 5'-'GAATTC-3' and is unaffected by CpG methylation. It is added as an internal control for the mass input of the genomic DNA. The DNA is digested with either: \textit{MspI} + \textit{EcoRI} or \textit{HpaII} + \textit{EcoRI} and then pyrosequenced\textsuperscript{179}. The overall percent methylation is calculated as the ratio between the two digestions:

\[
\text{% Methylation} = 1 - \left( \frac{HpaII}{MspI} \right) \times 100\%
\]

The most recently introduced EpiMark kit from NEB, allows to differentiate between mC and hmC by adding T4 Phage β-glucosyltransferase, which adds a glucosal moiety onto hmC. This prevents the digestion of the ChmCGG site by both \textit{MspI} and \textit{HpaII} providing a tool to quantify and detect hmC\textsuperscript{180}.

1.3.5 Antibody based methods

Affinity-based methods use antibody to target a specific C modification. Antibodies against mC and all three oxymethyls are commercially available. They are DNA immunoprecipitation sequencing (DIP-seq)-based methods\textsuperscript{87,136,181}. However, these methods are limited to the antibody specificity, as they may suffer from high background and bias towards densely modified genomic regions\textsuperscript{85,182}. Methyl-DNA immunoprecipitation (MeDIP) can provide a whole genome profile of DNA methylation levels\textsuperscript{181}. MeDIP involves the fragmentation of DNA, denaturation, and immunoprecipitation (IP) with an antibody
targeting 5mC. The precipitated DNA can then be sequenced by quantitative PCR or used in a DNA microarray analyses\textsuperscript{183}. Hydroxymethyl-DNA IP (OH-MeDIP) has been used to measure hmC levels in mouse ES cells\textsuperscript{184}. Other antibody-based assays developed to enrich for hmC, modify the hydroxyl group by incorporating a glucosal moiety, a radioisotope label, or using a chemical affinity tag such as biotin\textsuperscript{185,186}.

1.3.6 Chromatography and mass spectrometry

Chromatography alone or in combination with mass spectrometry can be used to resolve different nucleotides. Since each C modification has a different mass, it can be separated and quantified, providing the % of each modification in the total DNA genomic input\textsuperscript{55,187}. High-performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) has been used to separate, distinguish and quantify nucleic acids\textsuperscript{188}. It has been used and adapted to profile and characterize the different oxymethyl forms\textsuperscript{187,189}. HPLC paired with mass spectrometry can be used to resolve distinct compounds in solution, based on polarity and mass. Reversed-phase HPLC is based on utilizing a column with a nonpolar, hydrocarbon solid phase, and individual components are eluted based on their solubility in a polar solution\textsuperscript{190}. Discrete elution times allow for accurate detection and quantification of multiple compounds. The solid phase, mobile phase, pH and temperature may affect the level of separation, all of which can be optimized when designing a specific assay.
Chapter 2
DNA methylation of endothelial cell-enriched genes during *in vitro* mouse ES cell differentiation and the role of TET proteins
2.1 Rational, Hypothesis and Specific Aims

2.1.1 Rational

The ECs reside at the interface between the blood and vascular wall. An understanding of the regulatory mechanisms that control EC-enriched genes is fundamental to the understanding of EC physiology and function. Our lab has previously reported that EC-enriched genes are regulated by epigenetic mechanisms including DNA methylation of the eNOS proximal promoter, which is not methylated in ECs\textsuperscript{134}. In contrast, it is fully methylated in non-expressing cells such as VSMCs (vascular smooth muscle cells) (Figure 1.9). The differential DNA methylation between ECs and VSMCs was observed in other EC-enriched genes, in both human and mouse\textsuperscript{2,134}. The excepted dogma for DNA replication timing is that tissue-specific genes that are being expressed in a specific cell type will replicate early in those cells, and will replicate late in non-expressing cells. Our lab has found that EC-enriched genes that are regulated by DNA methylation, replicate early in both expressing and non-expressing cell types (EC and VSMC respectively)\textsuperscript{2}. In contrast, EC-enriched genes that are not regulated by DNA methylation such VEFGR1 and VEGFR2 (also known as FLT1 and FLK1, respectively) replicate early in expressing ECs, and replicate late in VSMCs, following the excepted dogma\textsuperscript{2}. This highlights the potential importance of the maintenance of the epigenetic code during cell division. As early replication in genes that are epigenetically regulated, may provide enough time to re-establish the proper epigenetic code in all cell types. If these genes replicate late in non-expressing cell types, there would be insufficient time for the maintenance of the epigenetic signature. This concept requires further study. As
the finer details remain to be elucidated, this piece of evidence further signifies the importance of maintaining proper epigenetic signature in EC-enriched genes.

To learn how the DNA methylation signature in ECs is established in development or during differentiation, we examined the proximal promoters of EC-enriched genes in murine ES cells. Using the EB method the expression of early EC markers was surveyed, and it was found that the steady-state mRNA of VEGFR2 goes up at EB day 4, while other markers are upregulated later in differentiation (Figure 2.1). Sorting day 4 EBs, for positive or negative cell surface VEGFR2 expression, allowed us to determine if there is a different DNA methylation signature in EC-enriched genes between EC-progenitor and non-EC-progenitor cells. Examination of the sorted populations showed that both are fully methylated at the proximal promoters of EC-enriched genes such as eNOS and CDH5 (Figure 2.2a). The most surprising finding was that ES cells in the pluripotent state were heterogeneously DNA methylated at the EC-enriched gene promoters. We found this heterogeneous DNA methylation in several independent mouse ES cell lines and in human ES cells (Figure 2.3). These data suggest that something happens between ES cell stage and EB day 4, which causes an increase in DNA methylation at the proximal promoter regions of these genes. We posit that heterogeneous DNA methylation at baseline is a key component of establishing future cell-specific gene expression. Though this concept has been described for the histone code, and termed bivalent histone marks, the role of differential and heterogeneous DNA methylation has not been studied.
Figure 2.1: mRNA expression levels in mouse ES cells and mouse EBs day 2 – day 13

The x-axis shows time points starting from ES cells all the way to day 13 after EB differentiation. The y-axis shows gene expression relative to GAPDH and TBP. As ES cells differentiate, the mRNA for Oct4, a marker of pluripotency decreases. EC-enriched genes go up, VEGFR2 being the first to increase at day 4, followed by CD31 at day 7 (data provided by Dr. Jason Fish).
Figure 2.2: DNA methylation of the mouse promoters of eNOS and VE-cadherin

The x-axis shows percent DNA methylation. The y-axis shows CG dinucleotide position from the TSS of the gene. 

A. Day 4 EBs were FACS sorted for top 5% VEGFR2+ (early ECs) and bottom 5% VEGFR2- cell populations. There was no DNA methylation difference between the populations of both genes, using a two way ANOVA. Data is presented as the mean ±SEM (n=3-4).

B. EC and VSMC were isolated from the mouse thoracic aorta, and both eNOS and VE-cadherin proximal promoters are significantly less methylated in ECs than in VSMC. Significance test was performed by a two way ANOVA (p < 0.001). Data is shown as the mean ±SEM (n=3-7).
Figure 2.3: Mouse ES cell DNA methylation of the eNOS proximal promoter

A. Pyrosequencing shows an average of ~50% methylation overall, data presented as mean ±SEM (n=3). X-axis shows GC position from TSS, and y-axis shows percent methylation. 

B. Lollipop diagram representing single strand analysis of DNA methylation. Closed circles represent methylated CG sites, open circles represent non-methylated CG sites. There is heterogeneous methylation at the ES cell stage, as each allele has a different methylation pattern.

C. Methylation histogram of panel B, showing the % methylation of each site. Since the distribution of the histogram is rectangular and not normal or bimodal, it indicates uniform distribution and thus heterogeneous population.

D. Frequency histogram of panel B shows frequency of methylated alleles. Since the distribution of the histogram is in a rectangular shape, it indicates uniform distribution and thus heterogeneous population.
2.1.2 Hypothesis

We hypothesize that TET protein function is responsible for maintaining low and heterogeneous levels of DNA methylation in pluripotent mouse ES cells. This heterogeneous DNA methylation pattern marks EC-lineage restricted genes prior to EC-lineage specification, and is functionally relevant for their subsequent expression.

2.1.3 Aims

To define the functional relevance of heterogeneous DNA methylation of EC-enriched genes in murine ES cells.

Examine TET gene expression during *in vitro* mouse EB differentiation.

To define the functional contribution of TET1 and TET2 proteins to DNA methylation and gene expression of EC-enriched genes in murine ES cells.
2.2 Methods

2.2.1 Cell culture

Mouse ES cell lines E14 and R1 were maintained at 37°C in humidified air with 5% CO2. The cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) with high glucose (4,500 mg/L) and without L-glutamine or sodium pyruvate. The ES cell maintenance medium was supplemented with 15% fetal bovine serum (Gibco, ESC qualified), 0.1mM MEM non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 50 U/ml penicillin and 50 ug/ml streptomycin (Gibco), 0.1 mM beta-mercaptoethanol (Gibco), 2 mM glutamine (GlutaMAX, Gibco), 1000 U/ml leukemia inhibitory factor (LIF; Gibco). The culture medium was replaced daily and ES cells were passaged every two days or as needed with 0.05% trypsin-EDTA onto 0.1% gelatin (Sigma) coated tissue culture plates.

Wild type mouse ES cell lines V6.5 and TET knockout line (TET1 and TET2 DKO) were generously provided by Dr. Rudolf Jaenisch, were seeded on mito-C treated mouse embryonic fibroblast (MEF) cells (Millipore), and cultured under the same conditions as E14 and R1 cells. To separate ES cells from MEFs, the cells were treated with 0.05% trypsin for 5-7 min at 37°C, neutralized with maintenance media and plated on a freshly gelatinized plate for 1hr. MEFs are almost double the size of ES cells, they are ~16μm in diameter vs ~8μm diameter of ES cells, and MEFs attach first to the plate. After 1hr of incubation the supernatant media containing the ES cells can be utilized for cell differentiation or DNA/RNA extraction.
2.2.2 ES cell differentiation

ES cells were induced to differentiate upon LIF removal and formed embryoid bodies (EB) using the hanging drop method. Cells were counted on a Beckman Coulter counter or haemocytometer and ~500 ES cells were suspended in 20 ul drops on the lids of 150mm bacteriological grade dishes and left inverted for two days. At day 2 of differentiation EBs were transferred to 60mm bacteriological grade dishes (STARDISH, Phoenix Biomedical) and cultured in Differentiation media (identical to ES cell Maintenance media except without LIF). Media was replaced daily.

2.2.3 2i and Vitamin C experiment

Two inhibitors of differentiation, CHIR99021 (3uM of GSK3 inhibitor) and PD0325901 (1uM of MEK inhibitor) (TOCRIS bioscience) were added to the ES cell media. Vitamin C 100ug/mL (Sigma) was added daily to the ES cell’s media. Cells were maintained under test conditions for 7 days before DNA isolation.

2.2.4 Cell sorting into 96 well plate and Alkaline Phosphatase Live staining

Mouse ES cells were sorted by a BD FACS Aria III SORP into a 96-flat bottom well plate (Starstdet) that was already seeded with MEFs, and contained 50ul of ES cell media. After sorting, each well was topped up with additional 50uL of media. Media was changed daily for 7 days, and the cells were monitored under microscope to confirm the presence of only one colony per well. Wells with multiple colonies were not used. To separate the ES cells from the MEFs they were seeded on, 10 wells with only one ES colony were selected and than stained
with Alkaline Phosphatase (Thermo Scientific) following the factory protocol. After staining, the 10 wells were washed with 100ul PBS -/-, than trypsinized for 5min, and neutralized with 150uL of ES cell media. The single cell suspension within each well was sorted using the same cell sorter for AP positive into individual 1.5mL tubes containing 50uL of AL lysis buffer (Qiagen DNA blood minikit). 0.1uL of 10mg/mL RNase A (Thermo Scientific) was added to each tube and incubated for 1hr at 37°C, and 0.5uL of 20mg/mL of Proteinase K (Thermo Scientific) was added and incubated overnight at 56°C. The DNA isolation was completed using the manufacture’s protocol.

2.2.5 Genomic DNA extraction

Cells were lysed with a lysis buffer containing final concentrations of 0.1M Tris pH8, 5mM EDTA pH8, 0.2% SDS, 0.2M NaCl, and 10mg/mL RNase A (Thermo Scientific). The lysates were incubated at 37°C water bath for 2hrs, than 20mg/mL of Proteinase K (Thermo Scientific) was added, and the samples were incubated overnight at 56°C. DNA was extracted using a standard phenol/chloroform method with salt-ethanol precipitation.

2.2.6 Sodium bisulfite DNA conversion

Up to 500ng of genomic DNA was subjected to sodium bisulfite treatment using the EZ DNA Methylation Kit (Zymo Research) following the manufacturer’s protocol with some modifications. During the DNA denaturation step, the DNA was denatured for 30 minutes at 42°C instead of 15 minutes at 37°C. The conversion of C to T in a non-CG context was 99% or higher, as detected after PCR amplification and sequencing.
2.2.7 Sodium bisulfite sequencing

An aliquot of the sodium bisulfite treated DNA (~10ng, 1-3ul of the 10ul elution) was subjected to nested PCR amplification in a 50ul reaction: 1x PCR buffer, 1.5mM MgCl₂, 200uM of each dNTP, 10pmol of each forward and reverse primers, and 1 unit of Platinum Taq DNA polymerase (Invitrogen). First PCR was completed using outer primers (OS – outer sense, and OAS – outer antisense) and 2ul of the first PCR product (was used as a template input for the second PCR using inner primers (IS – inner sense, IAS – inner antisense) (Table 2.1). For single strand analysis the second PCR product was ligated into the PCR II cloning vector (Invitrogen), ethanol precipitated, and electroporated into electrocompetent DH10β E.coli cells (Invitrogen). Using antibiotic and β-gal selection, individual clones were picked and mini-prepped using the QuickLyse Miniprep Kit (Qiagen) following the manufacturer’s protocol. The plasmid was than subjected to Sanger sequencing for base resolution analysis, as previously discussed¹³⁴,¹⁹¹.

2.2.8 Pyrosequencing of bisulfite converted DNA

Pyro-PCR was performed using 2ul of the first PCR product from the nested PCR as an input template. The pyro-PCR biotinylated primers were designed by EpigenDx Inc. and are specific for bisulfite converted DNA (Table 2.2). Using the company protocol, the samples were amplified, and then subjected to pyrosequencing (EpigenDx Inc). The biotinylated PCR products are bound to sepharose beads, purified, washed and denatured in 0.2M NaOH and then washed with Pyrosequencing Vacuum Prep Tool (Pyrosequencing Inc.). The sequencing primer (0.3uM) was used with the single stranded template and analyzed with a PSQ96HS
Instrument (EpigenDx Inc) in a SNP (C/T) analysis. The relative amount of 5mC was calculated as the percentage of methylated cytosines divided by the sum of methylated and unmethylated cytosines at a particular CG site:

\[ 5\text{mC}\% = \frac{5\text{mC}}{5\text{mC} + \text{unmethylated C}}. \]
### Table 2.1 Mouse bisulfite sequencing primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Amplication Location</th>
<th>Amplicon size (bp)</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>eNOS</strong></td>
<td>OS 5' AGATAGGAGAGGAGTAAGGGTGAATT 3'</td>
<td>-409 to +73</td>
<td>482</td>
<td>50.4</td>
</tr>
<tr>
<td></td>
<td>OAS 5' CCCTAAACCACAAAAATAACCCAAACTC 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IS 5' GGGTTTTATTTATAGTTTTAGTTTTT 3'</td>
<td>-230 to +57</td>
<td>287</td>
<td>48.3</td>
</tr>
<tr>
<td></td>
<td>IAS 5' AACCCAAAACCTCTAACCCACACTCTTC 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD31</td>
<td>OS 5' TAGGAGGTTTAGATTATTTATTTATG 3'</td>
<td>-27 to +462</td>
<td>490</td>
<td>48.5</td>
</tr>
<tr>
<td></td>
<td>OAS 5' TAAAACACTCAACTATCTATCTT 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IS 5' AGGAGAAGAAGTTTTTTAAATTTTTGA 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IAS 5' CACAAACCAAAAACCTACTATATA 3'</td>
<td>+101 to +368</td>
<td>293</td>
<td>46.9</td>
</tr>
<tr>
<td>VE-cadherin</td>
<td>OS 5' AGGGATGGTGAATGTTTG 3'</td>
<td>-312 to +306</td>
<td>619</td>
<td>49.2</td>
</tr>
<tr>
<td></td>
<td>OAS 5' GACAATCTACTTACAAACTACTAA 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IS 5' TTGAGGTATGAGTTGAATATTTTA 3'</td>
<td>-230 to +249</td>
<td>480</td>
<td>49.2</td>
</tr>
<tr>
<td></td>
<td>IAS 5' TTATATCCTTACACACTAAAAATATA 3'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OS = outer sense  
OAS = outer antisense  
IS = inner sense  
IAS = inner antisense
### Table 2.2 Mouse bisulfite pyrosequencing primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers</th>
<th>Amplicon location</th>
<th>Amplicon size (bp)</th>
<th>Ta (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eNOS</td>
<td>F – ADS251FP R-b – ADS251RPB</td>
<td>-235 to +29</td>
<td>256</td>
<td>56</td>
</tr>
<tr>
<td>CD31</td>
<td>F – ADS1723FP R-b – ADS1723RPB</td>
<td>+102 to +366</td>
<td>266</td>
<td>61</td>
</tr>
<tr>
<td>VE-cadherin</td>
<td>F – ADS1886FP R – ADS1886RPT UPB</td>
<td>-241 to -74</td>
<td>170</td>
<td>61</td>
</tr>
<tr>
<td>LINE-1</td>
<td>F-b – ADS685-300FPB R – ADS685-300RP</td>
<td>abundant</td>
<td>113</td>
<td>50.5</td>
</tr>
</tbody>
</table>

F = forward, R = reverse, b = biotin label  
FP = forward primer, RPB = reverse primer biotinylated  
RPT = reverse primer tailed, UPB = universal reverse biotinylated primer  
FPB = forward primer biotinylated
2.2.9 RNA extraction

RNA was extracted from ES cells and EBs using TRIzol reagent (Ambion life Technologies) following the manufacturer’s protocol. Exogenously synthesized, capped and polyadenylated 0.025ng firefly RNA for luciferase was added to the RNA cell lysates as a control for RNA yield and first-strand synthesis efficiency. DNase I (Fermentes) treatment was performed prior to RNA precipitation. RNA was eluted in 40uL of RNase-free water and stored at -80°C.

2.2.10 First strand cDNA synthesis

1ug of total cellular RNA was used to synthesize first strand cDNA with Superscript II kit using random hexamer priming (Invitrogen). Final volume of cDNA was adjusted to 100uL and stored at -20°C.

2.2.11 Real-time quantitative Polymerase Chain Reaction (RT-qPCR)

The amount of target cDNA was quantified using an ABI 7900HT sequence Detection system (Applied Biosystems) using SYBR green (Applied Biosystems, Life Technologies). Quantification was done in triplicate on 2uL of cDNA in a 10uL volume of reaction mix. Copy number of transcripts was determined using a standard curve generated by amplicon-containing plasmids. Primers were designed using Oligo 7 software to cross exon-exon boundaries with an intron of at least 500bp in between them and their specificity was tested and optimized by using different concentration of primers (900ng, 450ng, 300ng, 150ng). The PCR product was electrophorased on a 1% agarose gel to verify a presence of a single amplicon at the predicted size (Table 2.3). Resulting copy numbers were normalized to
luciferase efficiency. Since 0.025ng of luciferase mRNA (2.73 x 10^7 copies of luciferase RNA) was added during the RNA extraction, and was measured copy number per sample, it can be used to estimate the original amount of a mRNA in the RNA cell lysate and thus account for RNA degradation during RNA isolation and first strand synthesis efficiency. Measured copy numbers were then normalized to GAPDH as a house-keeping gene.

2.2.12 Statistics

Unless otherwise stated, all experiments were performed with a minimum of three replicates. Data represents the mean ± SEM. Unless stated otherwise, ANOVA statistical test with a 95% confidence interval was used when we were comparing more than two variable groups such as multiple days of differentiation, and a dependent variable such as % methylation. One-way ANOVA was used when there was one independent variable with 2 or more treatments, while a two-way ANOVA was used when there were two independent variables with multiple conditions.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence</th>
<th>Location</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eNOS</td>
<td>F 5’ CGCCAACGTGGAGATCACT3’ R 5’ ATCAAAGCGGCCATTTCCCT3’</td>
<td>Exon 6 Exon 7</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>CD31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F 5’TCAAGACCACATCCAGGAGTGAATACG3’ R 5’TGCTTTGGAGGTGGTCAATA3’</td>
<td>Exon 4 Exon 5</td>
<td>197</td>
</tr>
<tr>
<td>VE-cadherin</td>
<td>F 5’TGCCCACCATCGCAAAAAGAGAGAC3’ R 5’CTGGCGGTTCACGGGACTTG3’</td>
<td>Exon 2 Exon 3</td>
<td>115</td>
</tr>
<tr>
<td>TET1</td>
<td>F 5’ AGCTGGATTGAAGGAACGGAAG3’ R 5’TCTTGCCAAAAGGTGTGTATA3’</td>
<td>Exon 3 Exon 4</td>
<td>94</td>
</tr>
<tr>
<td>TET2</td>
<td>F 5’CGAGGCTGAGGGACAGAAC3’ R 5’AGCGGAACAGGAACAGGAAC3’</td>
<td>Exon 1 Exon 2</td>
<td>86</td>
</tr>
<tr>
<td>TET3</td>
<td>F 5’TCTGGCCCCACAGTAGCTT3’ R 5’CCTCCTTGGCCTGATAG3’</td>
<td>Exon 4 Exon 5</td>
<td>103</td>
</tr>
<tr>
<td>DNMT1</td>
<td>F 5’AGCTGGTCTATCGATCAGGACA3’ R 5’CCACCAAATCTCATATATCTT3’</td>
<td>Exon 21 Exon 22</td>
<td>169</td>
</tr>
<tr>
<td>DNMT3a</td>
<td>F 5’GCACCGAGGAAATGTATGT3’ R 5’AGAGGTCAATGCGAGGACTG3’</td>
<td>Exon 17 Exon 18</td>
<td>112</td>
</tr>
<tr>
<td>DNMT3b</td>
<td>F 5’CAATCTGCACAGAGCCAGTC3’ R 5’GGCTGGGAGACCTCCCTTTA3’</td>
<td>Exon 3 Exon 4</td>
<td>84</td>
</tr>
<tr>
<td>hnRNA eNOS</td>
<td>F 5’GCTGCCCTTGGAAAAGTGAT3’ R 5’TGCTGGCCCCACCTTCCATTTA3’</td>
<td>Intron 1 Exon 2</td>
<td>149</td>
</tr>
<tr>
<td>SOX2</td>
<td>F 5’TTGGGGGAGATTTTTTTGACAGAG3’ R 5’AGCGCCTAACGTACACTAGA5’</td>
<td>Exon 1 Exon 1</td>
<td>132</td>
</tr>
<tr>
<td>OCT4</td>
<td>F 5’AGCCGACAACATGAGAAACC3’ R 5’CTTCGGGCAACTTCAAAACAT5’</td>
<td>Exon 3 Exon 4</td>
<td>133</td>
</tr>
<tr>
<td>NANOG</td>
<td>F 5’CAAGCGGTGGGAGAGAAAAAAAAAC3’ R 5’TGGTGCTGAGCCCTTCTGAAT3’</td>
<td>Exon 4 Exon 4-5</td>
<td>67</td>
</tr>
<tr>
<td>GAPDH</td>
<td>F 5’TTCACCACCATGGAGAGG3’ R 5’CTCGTGGTTCACAcCACCAC3’</td>
<td>Exon 3 Exon 4</td>
<td>111</td>
</tr>
<tr>
<td>Luciferase</td>
<td>F 5’ACTCCTGAGTCTACTGGTC3’ R 5’GTAATCCCTGAGGCTCCTCA3’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3 Results

2.3.1 DNA methylation clonality of the mouse eNOS proximal promoter in ES cells

The DNA methylation pattern of EC-enriched gene promoters is mostly homogeneous in mature and healthy cells. It is either non-methylated in ECs, or methylated in non-expressing cell types such as VSMC (Figure 1.9). Interestingly, in ES cells the DNA methylation of EC-enriched gene promoters is heterogeneous, as shown for eNOS (Figure 2.3). Our objective was to determine whether this DNA methylation pattern of a single ES cell at EC-enriched gene promoters is heritable. We chose to use the mouse eNOS proximal promoter as an example of an EC-enriched gene promoter. Single ES cells were sorted per well of a 96 well plate, allowed to expand into a colony, and DNA methylation pattern of 10 individual clones was examined. We predicted two possible outcomes: 1) a unique pattern of DNA methylation within a single clone, or 2) the DNA methylation pattern within a clone is going to be heterogeneous, as with the parental clone, indicating that DNA methylation is not heritable from mother to daughter cell and thus heterogeneity is being reestablished. Pyrosequencing was performed to analyze the total methylation of the individual clones. Some clones possessed ~50% methylation at the CpG sites implying heterogeneity, while other clones had a mixture of ~100% and ~0% methylated CpG sites, implying a certain pattern (Figure 2.4). Single strand analysis of 9 clones, showed the existence of two types of clones: heterogeneous clones, and clonal clones. 6/9 clones had a heterogeneous DNA methylation pattern, while 3/9 clones displayed a clonal DNA methylation pattern (Figure
2.5). This was further confirmed by examining the frequency histograms. Representative frequency histograms for one heterogeneous clone and one clonal clone further confirm the two different DNA methylation inheritance patterns within the ES cells (Figure 2.6).
Figure 2.4: Assessment of DNA methylation of the eNOS proximal promoter from 10 ES clones

DNA was isolated from 10 different ES cell clones. After bisulfite treatment and nested PCR amplification, the eNOS amplicon was pyrosequenced. The x-axis represents the CpG sites present at the mouse eNOS promoter relative to the transcriptional start site of the gene. The y-axis shows percent methylation. Each coloured row represents a distinct ES line established from a single murine ES cell.
Figure 2.5: Assessment of heritability of the DNA methylation pattern of the mouse eNOS proximal promoter

A schema of the eNOS mouse promoter is displayed at the top of the figure, with each circle and number indicating the position of a CpG site, relative to the TSS (indicated by the arrow). Outlined also are binding sites of different transcription factors in that locus. At the bottom half of the figure, lollipop diagrams of 9 different ES cell clones are being shown. All clones have variable levels of DNA methylation from 29% to 89%. There are two types of clones: heterogeneous, and clonal. 6/9 clones have heterogeneous DNA methylation, while 3/9 clones exhibit clonal DNA methylation. 13-23 DNA strands were sequenced per clone.
Figure 2.6: Two representative ES cell clones display both clonal and heterogeneous DNA methylation of the eNOS proximal promoter

The frequency and methylation histograms of two representative clones from figure 2.5 are being shown. The clone on the left has a clonal DNA methylation pattern, as the frequency histogram has two major peaks, depicting two populations or epi-alleles. The clone on the right is heterogeneously methylated, as there is no pattern to the DNA methylation in the lollipop diagram, and the frequency histogram shows a rectangular uniform distribution.
2.3.2 Regulating the DNA methylation heterogeneity of the eNOS proximal promoter in mouse ES cells.

The recent assertion about the ES cell ‘ground state’ when using 2i culture conditions, is that global DNA methylation is reduced through indirect inhibition of DNMT3s via PRDM14, and some suggest active demethylation through TET activity. We wanted to examine whether the heterogeneity of DNA methylation at the eNOS promoter is retained at the ‘ground state’. In parallel, we decided to treat mouse ES cells with Vitamin C (vit C), to determine if augmenting TET activity and enhanced active DNA demethylation will retain the heterogeneous eNOS promoter methylation. Mouse ES cells were treated for 7 days with 2i, vit C, or both. Extracted genomic DNA was bisulfite converted, and pyrosequenced to assess overall methylation of the mouse eNOS proximal promoter. In all three treatments there was a significant decrease in DNA methylation (Figure 2.7). When treating with both 2i and vit C, it appears as if the effect is additive since it has the lowest methylation level (Figure 2.7c). Furthermore, single strand analysis revealed that upon either treatment, the heterogeneity of the DNA methylation is less apparent, but that might also be since the DNA methylation is reduced to almost non-existent (Figure 2.8).
Figure 2.7: Mouse eNOS proximal promoter methylation after treatment with 2i, vit C or both

WT R1 ES cells were treated for 7 days with 2i (A), vit C (B), or both (C). Genomic DNA was bisulfite converted and pyrosequenced.
A. 2i treatment shows a decrease of the average DNA methylation. Data is presented as the mean ± SEM (n=3). Two-tailed t-test between the treatment and control indicated a significant difference (p<0.0001).
B. Vitamin C treatment also showed a decrease of the average DNA methylation. Data is presented as the mean ± SEM (n=3). Two-tailed t-test between the treatment and control indicated a significant difference (p<0.05).
C. Treatment with both 2i and Vitamin C had a further decrease of the average DNA methylation than the individual treatment. Data is presented as the mean ± SEM (n=3). Two-tailed t-test between the treatment and control indicated a significant difference (p<0.0001).
Control
N=15
43% methylation

2i
N=23
8.3% methylation
Figure 2.8: Single strand analysis of DNA methylation of the mouse eNOS proximal promoter in ES cells with treatment of 2i, vit C, or both

R1 mouse ES cells were subjected to 7 days treatment of 2i, vitamin C, or both. Fresh media with drugs, and vitamin C were added daily. DNA methylation of the eNOS proximal promoter was examined through Sanger sequencing of plasmid subclones. The data for each treatment is presented in a lollipop diagram representing single strand analysis of DNA methylation. Filled circles represent methylated sites, open circles represent non methylated CG sites, and vertical lines indicated a non-converted cytosine by the bisulfite. Methylation histogram shows the % methylation of each site, and a frequency histogram shows how frequent certain amount of methylated sites occurs per DNA strand are displayed for each treatment: a. control, b. 2i, c. vit C, d. 2i and vit C.
2.3.3 Global DNA methylation during *in vitro* mouse ES cell differentiation

LINE-1 (L1) is the most abundant mammalian transposable element. Repetitive DNA is known to be highly methylated. Alteration in the DNA methylation status may lead to its expression and is associated with numerous diseases\(^\text{192}\). As L1 is the most abundant repetitive element comprising almost 20% of the genome, and is heavily methylated in normal somatic cells, the methylation status of L1 elements can be taken as a reflectance of global repetitive DNA methylation status, as described by others\(^\text{136}\). Using L1 methylation as a global DNA methylation assay, we decided to observe the changes in DNA methylation in our time course of interest. Mouse ES cells were *in vitro* differentiated for 7 days, and L1 methylation was assayed. A significant increase was observed as early as EB day 4 compared to EB day 0 (undifferentiated ES cell) (Figure 2.9). This is consistent with the DNA methylation status of EC-enriched genes eNOS and VE-cadherin at day 4, showing that it might be a global DNA methylation increase effect (Figure 2.2).
Global DNA methylation was assayed during murine ES cell *in vitro* differentiation into EBs. The x-axis shows the days of differentiation. The y-axis shows the overall %DNA methylation assayed by pyrosequencing. Data is presented in a box plot, showing the median, min and max (n=3). ANOVA statistical test showed a significant increase in DNA methylation from ES cell stage to EB day 4 (p < 0.0001) and day 7 (p < 0.0001).
2.3.4 TET and DNMT gene expression during *in vitro* mouse ES cell differentiation

Numerous enzymes and cofactors play a role in DNA methylation placement, maintenance and erasure; however, the DNMTs and TETs appear to have the most central roles in this process\(^9\) (Figure 1.6). In order to see which changes in the total cellular level of these regulators could contribute to the increased DNA methylation from ES cell stage to EB day 4, we looked at the mRNA expression levels of the three catalytically active DNMTs, and all three TET proteins. The expression of DNMT1 was consistent throughout ES cell differentiation for 7 days, as no significant difference was observed (Figure 2.10). DNMT3a showed an increased expression at EB day 7 compared to the ES cell stage (Figure 2.11). This increased expression might contribute to the further increased DNA methylation at day 7, however, it does not match our time point of interest which is the transition from ES cell stage to EB day 4. DNMT3b gene expression displayed no significant changes during ES cell differentiation for 7 days (Figure 2.12). TET1 is highly expressed in ES cells\(^136,138\). We found that TET1 was significantly down regulated by EB day 4 and 7 (Figure 2.13). TET2 showed a similar trend to TET1, where it was highly expressed in ES cells, and then being repressed by EB day 4 which continued to day 7 (Figure 2.14). TET3 on the other hand was not present in ES cells, and only by EB day 7 it had a significant increase in expression (Figure 2.15). This implies that since TET1 and TET2, who play a role in DNA demethylation, are present in ES cells and are gone by EB day 4, they may be important in the switch in the global and local DNA methylation at those time points, from low heterogeneous DNA methylation, to highly methylated.
Expression of mouse DNMT1 during ES cell differentiation

Figure 2.10: mRNA expression levels of DNMT1 in mouse ES cells and EBs day 4 and day 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. ANOVA statistical test did not find any significant change in expression during ES cell differentiation into EBs. Data is shown as the mean±SEM (n=3).
Expression of mouse DNMT3a during ES cell differentiation

**Figure 2.11: mRNA expression levels of DNMT3a in mouse ES cells and EBs day 4 and day 7**

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. A significant increase at EB day 7 relative to EB day 0 was found using an ANOVA (p<0.05). Data is shown as the mean±SEM (n=3).
Figure 2.12: mRNA expression levels of DNMT3b in mouse ES cells and EBs day 4 and day 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. ANOVA statistical test did not find any significant change in expression during ES cell differentiation into EBs. Data is shown as the mean±SEM (n=3).
Figure 2.13: mRNA expression levels of TET1 in mouse ES cells and EBs day 4 and 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. A significant decrease at EB day 4 and day 7 relative to EB day 0 was found using an ANOVA (p<0.01 for both days). Data is shown as the mean±SEM (n=3).
Figure 2.14: mRNA expression levels of TET2 in mouse ES cells and EBs day 4 and 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. A significant decrease at EB day 4 and day 7 relative to EB day 0 was found using an ANOVA (p<0.001 for both days). Data is shown as the mean±SEM (n=3).
Expression of mouse TET3 during ES cell differentiation

Figure 2.15: mRNA expression levels of TET3 in mouse ES cells and EBs day 4 and 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cells. A significant increase at EB day 7 relative to EB day 0 was found using an ANOVA statistical test (p<0.05). Data is shown as the mean±SEM (n=3).
2.3.5 Characterization of expression of ES cell markers in WT and TET1 and 2 DKO mouse ES cells

As it was observed that both TET1 and TET2 were associated with the time point at which DNA methylation increases from ES cell stage to EB day 4, double knockout of TET1 and 2 (DKO) cell line was obtained from Dr. R. Jaenisch. Characterization of the ES cells was performed by observing common ES cell markers. It was found that just like their equivalent WT mouse ES cell line V6.5, the DKO ES cells had high expression of SOX2, OCT4, and NANOG (Figure 2.16, 2.17, 2.18, respectively). Also, similar to the WT cells, the DKO showed a significant decrease in those markers with cell differentiation at EB day 4 and 7.
Figure 2.16: mRNA expression levels of SOX2 in WT and DKO mouse ES cells and EBs day 4 and 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. A significant decrease at EB day 4 and day 7 relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.001 both day 4 and 7 in WT, and p<0.05 both day 4 and 7 in DKO). Data is shown as the mean±SEM (n=3).
Figure 2.17: mRNA expression levels of OCT4 in WT and DKO mouse ES cells and EBs day 4 and 7

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. A significant decrease relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.05 and p<0.001 for day 4 and 7 respectively in WT cells, and p<0.05 at day 7 in DKO cells). Data is shown as the mean±SEM (n=3).
**Figure 2.18: mRNA expression levels of NANOG in WT and DKO mouse ES cells and EBs day 4 and 7**

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. A significant decrease at EB day 4 and day 7 relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.05 and p<0.001 respectively in WT, and p<0.05 in DKO for both days). Data is shown as the mean±SEM (n=3).
2.3.6 DNA methylation clonality of the mouse eNOS proximal promoter in DKO ES cells

As TET1 and 2 are enriched in ES cells, we decided to test whether they contribute to the heterogeneous DNA methylation of EC-enriched promoters, and whether they would affect the inheritance pattern of the DNA methylation. A clonality assay was performed on the DKO mouse ES cells. Single DKO ES cell was sorted per well of a 96 well plate onto pre-seeded MEFs and expanded into a colony. Since the knockout cells are feeder dependent, the DKO ES cells where stained with an Alkaline Phosphatase (AP) live stain prior to DNA isolation. The DNA methylation of the eNOS promoter was examined using pyrosequencing to give an average DNA methylation of all the cells that originated from one clone. We assayed a total of 6 clones by pyrosequencing. We found that some clones had very high methylation in some CpG sites and very low methylation in others, indicating a patterned and potentially inherited methylation (Figure 2.19). Since heterogeneous DNA methylation would appear around the 50% methylation level, it seems that the examined clones had clonal DNA methylation propagation of the eNOS promoter. Single strand analysis of 4 clones, confirmed this DNA methylation pattern (Figure 2.20). As predicted from the pyrosequencing, DKO ES cells exhibited clonality of DNA methylation of the eNOS promoter. 3 of the 4 examined clones had patterned DNA methylation, and one clone was fully methylated at the eNOS promoter (Figure 2.20). Although the overall methylation in DKO cells was higher than the WT cells, and the average methylation of each clone is slightly different, there was observed higher clonality of the eNOS promoter methylation in the absence of TET1 and 2.
The DNA methylation of the eNOS promoter of 6 DKO ES cell clones was assayed using pyrosequencing. The clones examined high methylation at some CpG sites and low methylation in other CpG sites predicting a clonal inheritance since it was not 50% methylation. The x-axis represents the CpG sites at the mouse eNOS promoter relative to the transcriptional start site of the gene. The y-axis shows percent methylation. Each coloured row represents a distinct ES line established from a single DKO ES cell.

Figure 2.19: DNA methylation of the mouse eNOS promoter in DKO ES cell clones
Figure 2.20: Single strand analysis of the DNA methylation of eNOS promoter in mouse DKO ES cells

The eNOS mouse amplicon is being displayed in a lollipop format, with each circle indicating the position of a CpG site from the TSS as indicated by the arrow. The DNA methylation of each clone is being viewed as single strands of DNA. All clones have variable levels of DNA methylation ranging from 33% to 100%, 8-13 DNA strands were sequenced per clone.
2.3.7 Global DNA methylation in mouse ES cells and TET DKO

Global DNA methylation in DKO ES cells was examined using LINE-1 DNA methylation assay. The DKO ES cells were compared to its WT cell line V6.5. Both cell lines showed increased L1 DNA methylation with cell differentiation at EB day 4 and day 7 (Figure 2.21). The WT V6.5 ES cells had slightly higher DNA methylation than what was found in other mouse ES cell lines (Figure 2.9). The L1 elements DNA methylation in DKO ES cells was significantly higher than in the WT V6.5 ES cells (Figure 2.21).
Figure 2.21: LINE-1 DNA methylation assay of mouse ES cells during differentiation in WT and TET1 and TET2 DKO cells.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows % methylation. A significant increase of DNA methylation from EB day 0 to EB day 4 (P < 0.0001) and EB day 7 (P < 0.0001) in the WT cells, and a significant increase from EB day 0 to EB day 4 (P < 0.05) and EB day 7 (P < 0.05) in the DKO cells was found using a two way ANOVA. Second ANOVA test showed a significant increase in DNA methylation in EB day 0 between the WT and the DKO cells (P < 0.0001). Data is presented in a box plot, showing the median, min and max (n=3-4).
2.3.8 DNA methylation of EC-enriched genes in mouse ES cells and during \textit{in vitro} differentiation in WT and TET DKO

Using TET1 and 2 knockout murine ES cells, we observed an increase in global DNA methylation (Figure 2.21), and furthermore we observed an increased tendency of DNA methylation marks at the eNOS promoter to be propagated from mother to daughter cells (Figure 2.20). Next we decided to look at the DNA methylation marks in EC-enriched genes during ES cell differentiation. We found that overall eNOS methylation state was 34% in WT ES cells, versus 49% in DKO ES cells (Figure 2.22). The DNA methylation of CD31 and CDH5 promoters was found to be significantly higher in DKO ES cells than in the WT (Figure 2.23, and 2.24 respectively). In WT ES cells the overall CD31 promoter methylation was 26% versus 54% in the DKO ES cells, and the CDH5 promoter in WT ES cells was 41% methylated versus 60% in the DKO ES cells. Similar to the increased global DNA methylation during ES cell differentiation, all three examined EC-enriched genes had increased DNA methylation in both the WT and DKO cells during differentiation at their promoters (Figures 2.22-2.24).
Figure 2.22: DNA methylation analysis of mouse eNOS promoter (CGs at position -199, -188, -184, -161, -149, -127, -90, -85, and -4 from TSS) during ES cell differentiation in WT and DKO cells.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows % methylation. A significant increase in methylation at EB day 4 and day 7 relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.001 and p<0.0001 respectively in WT, p<0.001 and p<0.01 in DKO cells). Data is shown as the mean±SEM (n=3).
Figure 2.23: DNA methylation analysis of mouse CD31 promoter (CGs at position +233, +306, +341, and +343 after TSS) during ES cell differentiation in WT and DKO cells.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows % methylation. A significant increase in methylation at EB day 4 and day 7 relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.001 and p<0.0001 respectively in WT, p<0.01 and p<0.05 in DKO cells). A second ANOVA test showed a significant increase in methylation between the WT and DKO cell lines at the ES cell stage (p < 0.01). Data is shown as the mean±SEM (n=3).
Figure 2.24: DNA methylation analysis of mouse CDH5 promoter (CGs at position -187, -162, -155, and -98 from TSS) during ES cell differentiation in WT and DKO cells.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows % methylation. A significant increase in methylation at EB day 4 and day 7 relative to EB day 0 was found in both cell lines using a two way ANOVA (p<0.0001 and p<0.001 respectively in WT, p<0.01 and p<0.05 in DKO cells). A second ANOVA test showed a significant increase in methylation between the WT and DKO cell lines at the ES cell stage (p < 0.05). Data is shown as the mean±SEM (n=3).
2.3.9 Gene expression of EC-enriched genes in mouse ES cells during in vitro differentiation in WT and TET DKO

DNA methylation is a known transcriptional repressive mark\textsuperscript{28}. TETs are part of the DNA demethylation pathway. We looked at the expression of EC-enriched genes in WT V6.5 cells and in the TET DKO at the ES cell stage and during differentiation at days 4 and 7. In WT ES cells, eNOS has low expression, it remains low at the beginning of differentiation but shows an increase by day 7 (Figure 2.25). This is counterintuitive, since we have observed that the eNOS promoter is heavily methylated at day 7 of differentiation (Figure 2.22). Similar patterns were observed in CD31 and CDH5 genes (Figure 2.27, and 2.28). It is important to note that unlike eNOS and CDH5, CD31 has high basal expression levels in ES cells and the ICM as has been previously noted\textsuperscript{193}. However, its expression goes down by day 4, and has a significant increase at day 7 (Figure 2.27). In order to confirm that the increased mRNA levels is due to an increase in transcriptional and not due to other factors such as changes in mRNA half-life, we looked at the expression levels of the heterogeneous RNA (hnRNA) of eNOS. We found that similar to mature RNA, the hnRNA of eNOS is low in ES cell, and is increased at day 7 (Figure 2.26).

In contrast to WT cells, EC-enriched genes in DKO cells did not increase at day 7 (Figure 2.25-2.28). Again we looked at hnRNA levels of eNOS in WT and DKO cells. When compared to WT cells, we observed markedly reduced eNOS hnRNA levels in DKO cells at day 7 (Figure 2.26). Interestingly similar to the WT cells, DKO ES cells express high levels of CD31. However, after its reduction at EB day 4, CD31 expression does not increase at day 7 in the DKO cells as it did in the WT cells (Figure 2.27). At EB day 7, eNOS is only partially activated. This modest effect can be because eNOS is a flow-regulated gene, and is not highly expressed until E9.5 in the mouse embryo when blood flow and other factors help to augment its expression\textsuperscript{194}. CDH5 expression, on the other hand, increases approximately 100 fold at day 7 in WT cells in comparison to day 0 or ES cell stage (Figure 2.28). Thus, it is not fully surprising that at day 7 the DKO cells have
some increased expression of CDH5, although it is not statistically significant compared to day 0 (Figure 2.28).
Expression of mouse eNOS during ES cell differentiation

Figure 2.25: mRNA expression levels of eNOS in WT and DKO mouse ES cells and EBs day 4 and 7.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. A significant increase at EB day 7 relative to EB day 0 was found in the WT cell line using a two way ANOVA (p<0.0001). Second ANOVA was done to compare cell lines, and it was found that at EB d7 there is a significant change in expression between WT and DKO cells, (p<0.0001). Data is shown as the mean±SEM (n=3).
Figure 2.26: mRNA expression levels of hnRNA of eNOS in WT and DKO mouse ES cells and EBs day 4 and 7.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. A significant increase at EB day 4 and day 7 relative to EB day 0 was found in the WT cell line using a two way ANOVA (p<0.05, and p<0.0001 respectively). A second ANOVA was done to compare cell lines, and it was found that at EB d7 there is a significant change in expression between WT and DKO cells, (p<0.0001). Data is shown as the mean±SEM (n=3).
Figure 2.27: mRNA expression levels of CD31 in WT and DKO mouse ES cells and EBs day 4 and 7.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. Using a two way ANOVA, a significant increase at EB day 7 relative to EB day 4 was found in the WT cell line (p<0.01), and a significant decrease of gene expression was found at EB day 4 and day 7 relative to EB day 0 in DKO cell line (p<0.01 for both). A second ANOVA was done to compare cell lines, and it was found that at EB d7 there is a significant change in expression between WT and DKO cells, (p<0.05). Data is shown as the mean±SEM (n=3).
Expression of mouse CDH5 during ES cell differentiation

Figure 2.28: mRNA expression levels of CDH5 in WT and DKO mouse ES cells and EBs day 4 and 7.

The x-axis shows time points starting from ES cells at EB day 0, to EB day 4 and day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in WT ES cells. Using a two way ANOVA, a significant increase at EB day 7 relative to EB day 0 was found in the WT cell line (p<0.05), no significant increase was found in the DKO cells. Second ANOVA was done to compare cell lines, and there is no significant difference between the cell lines. Data is shown as the mean±SEM (n=3).
2.3.10 eNOS gene expression in TET1 and TET2 single knockouts compared to the DKO and WT cells

Single TET1 and TET2 knockout mouse ES cell lines were obtained from Dr. R. Jaenisch. We examined the steady state expression of eNOS RNA from the individual knockouts at ES cell stage and EB day 7 (Figure 2.29). At ES cell stage of EB day 0, all four cell lines had similar and low expression of eNOS. The WT and TET1-/- EB day 7 cells, exhibited a significant increase of eNOS expression relative to their basal levels at day 0. There was no significant increase at EB day 7 relative to day 0 in the TET2-/- and DKO cells. Also, when comparing cell lines at day 7 TET2-/- and DKO cells have significantly less eNOS than in the WT cells. This suggests that TET2 is the major player that helps eNOS upregulation when the ES cells differentiate, as in its absence eNOS mRNA does not increase with differentiation.
Figure 2.29: eNOS gene expression during ES cell differentiation in WT and TET1 and TET2 individual knockouts

The x-axis shows time points starting from ES cells at EB day 0, and EB day 7. The y-axis shows gene expression relative to GAPDH, and is normalized to expression in ES cell stage at each cell line respectively. A significant increase at EB day 7 relative to EB day 0 was found in the WT cell line and TET1-/- using a two way ANOVA (p<0.0001 for both cell lines), no significant increase was found in the TET2-/- and DKO cells. Second ANOVA was done to compare the knockout cell lines to the WT, there is a significant decrease in TET2-/- and DKO EB day 7 when compared to the WT day 7 (p<0.001, and p<0.01 in TET2-/- and DKO cell lines respectively). Data is shown as the mean±SEM (n=3).
2.4 Discussion

The main goal of this study was to determine when the contrasting DNA methylation status was established between ECs and non-ECs at EC-enriched gene promoters. In unpublished work in our lab, we have discovered that EC-enriched genes that are regulated by DNA methylation are heterogeneously methylated at the ES cell stage. These gene promoters are fully methylated at day 4 of ES cell differentiation. We wanted to determine what contributes to this switch and whether it is functionally important for further cell lineage determination. This is an important question for basic research and for potential application in therapies targeting EC dysfunction. The major findings of this thesis are: \( i \) the DNA methylation in ES cells of the eNOS promoter sometimes propagates to the daughter cells and sometimes it retains heterogeneity, but this heterogeneity was not observed in TET 1 and 2 double knockout ES cells. \( ii \) LINE-1 elements are not methylated in ES cells, however they get fully methylated at day 4 of differentiation, matching the time of methylation of EC-enriched promoter genes. \( iii \) TET1 and 2 are highly expressed at ES cell stage, but are gone by day 4 of differentiation, matching the time course of gene hypermethylation. \( iv \) When knocking out TET1 and 2 in ES cells, EC-enriched genes that are regulated by DNA methylation have delayed activation of transcription when compared to the WT cells. Examination of eNOS expression in TET1 and TET2 individual knockouts showed that TET2 is the key regulator in eNOS gene induction, rather than TET1 or a complimentary activity of both TETs.

2.4.1 ES cell heterogeneity and its heritability

Waddington’s famous metaphor from 1957, of a ball rolling down a landscape of hills and valleys has been often interpreted as a model for cell differentiation\(^{195}\). It is predicated on the idea that all ES cells start off the same at the top of the hill, and then as they start to differentiate they are forced towards one lineage or the other. The ability to create induced pluripotent stem cells, raised the possibility that cells can be pushed uphill by some key TFs, where they will reach a point equivalent to the ES cell\(^ {196}\).
However, it is important to remember that not all ES cells are the same, as they form a heterogeneous population, thus the differentiation path one cell takes can be completely different from another ES cells. My project was initiated to better understand this ES cell heterogeneity, its heritability and how it can contribute to cellular differentiation towards the EC lineage.

ES cell heterogeneity is evident in the epigenetic signature and can be somewhat translated to gene expression as well. ES cells are enriched with epigenetic reprogrammers, some of which write repressive epigenetic marks such as DNA methylation, and others than counteract it. Thus it is suggested that ES cells continuously receive differentiation signals from its environment, but are contravened from differentiation\textsuperscript{136,138}. Epigenetic reprogramming seems to contribute to this process. Another perspective suggests that remaining pluripotent does not necessarily means that the pluripotency factors block differentiation, but that instead each pluripotency factor on its own directs ES cell differentiation to a specific lineage, while prohibiting commitment to other mutually exclusive lineages. So, individual pluripotency factors are continuously attempting to specify ES cell differentiation to their own lineage, but collectively they cancel each other, and maintain a diverse range of lineage commitment options and differentiation opportunities\textsuperscript{197}. Austin Smith popularized this concept as “the battlefield of pluripotency”\textsuperscript{198}.

Epigenetic heterogeneity in ES cells might also result from lower DNA methylation fidelity during mitosis. While somatic cells are observed to have a degree of clonal methylation inheritance, ES cells displayed plasticity in epigenetic inheritance\textsuperscript{139}.

Many EC-enriched genes are epigenetically regulated; their promoters are not methylated in ECs, and are fully methylated in other cell types like VSMCs\textsuperscript{2}. These promoter genes are heterogeneously methylated in ES cells. However, some EC-enriched genes such as VEGFR2, are not regulated by DNA methylation and are completely unmethylated in ES cells (Thesis by Anna Kop 2011). This showed that EC-enriched genes that are regulated by DNA methylation, have DNA methylation marks as
early as ES cell stage, while genes that are not regulated by DNA methylation, also
display no regulation at the ES cell stage. Thus, we decided to study the propagation of
the DNA methylation marks in the EC-enriched genes that are regulated by DNA
methylation. It was surprising to find two types of ES cell clones, where in one case the
alleles were acting independently and heterogeneity was reestablished during cell
division, while in the other type there was allele cross talk where some ES cells
propagate their DNA methylation marks at the eNOS promoter (Figure 2.5). Treating the
cells with 2i and or vitamin C, showed reduced DNA methylation at the eNOS promoter,
but retaining the heterogeneity, since the few remained methylated CpGs were
different from one another (Figure 2.8). It will be important to perform ES clonality
assays in vitamin C and 2i treated ES cells to define whether the clonality of DNA
methylation marks is affected. We anticipate that heterogeneity of clones will be
augmented.

On the other hand, knocking out TET1 and 2, which are enriched in ES cells,
showed increased methylation of the eNOS promoter, and increased clonality in the
propagation of the DNA methylation profile (Figures 2.19 and 2.20). This might suggest
that TET1 and 2 play a role in maintaining the epigenetic heterogeneity of the eNOS
promoter, since when adding vitamin C there is less methylation but still some
heterogeneity, and when knocking them down there is increased methylation and
clonality. Thus the initial finding of a mixture of heterogeneity and clonal DNA
methylation mark propagation at the eNOS promoter, may be due to different levels of
TET expression in each ES cell. An ES cell that expresses high levels of TET can have
lower and heterogeneous DNA methylation at the eNOS promoter, while an ES cell that
has lower expression levels of TETs can be more methylated and clonal in nature. This
remains to be tested.

2.4.2 DNA methylation changes from ES cell stage to EB day 4, and the key
players

It is a common misconception that ES cells are hypomethylated, since research
has shown that ES cell resemble somatic cells in their global DNA methylation\textsuperscript{137}. Instead, only some elements and specific genes show hypomethylation in ES cells when compared to somatic cells\textsuperscript{137}. LINE-1 repetitive element is often used to estimate global methylation, as it is abundant and distributed throughout the genome, allowing “random sampling”. However, it is not methylated in ES cells and thus is a poor representative of global DNA methylation in ES cells. Although it is not methylated in ES cells, LINE-1 gets fully methylated by EB day 4, which matches the same time course as the EC-enriched gene promoters that switch from being low heterogeneously methylated at ES cell stage, to fully methylated at day 4 (Figure 2.2). To determine the key players in this switch, we examined gene expression of the three catalytically active DNMTs and the three TETs. We found that DNMT1 and DNMT3b had no significant change in expression between ES cell stage and EB differentiation for 7 days (Figures 2.10 and 2.12). DNMT3a showed an increased expression at day 7 of differentiation, but not at day 4, thus not matching the time course of the observed hypermethylation (Figure 2.11). Looking at TET1 and 2, both are enriched at the ES cell stage and are downregulated by day 4 and day 7 of differentiation. Since TETs are demethylases, their absence causes increased DNA methylation, which matches our time course. As TET2 has a known accessory CXXC binding proteins IDAX and MBD3L2, it would be important to determine whether they may play a role in ES cell DNA methylation\textsuperscript{58,59}. TET3 is not expressed in ES cells, and is only upregulated towards day 7 of differentiation, thus it is only TET1 and 2 that are potential regulators (Figure 2.15). It will be important to do Western blots of the TET and DNMT proteins, to assess whether the protein levels follow a similar pattern to the steady state mRNA levels.

2.4.3 TET1 and 2 DKO ES cells and their affect on DNA methylation during EB differentiation

DKO mouse ES cells expressed comparable levels of the OCT4, NANOG and SOX2 pluripotency markers compared to WT ES cells. Expression was downregulated during EB differentiation in both cell line (Figures 2.16-2.18). LINE-1 is significantly more...
methylated in DKO ES cells than in the WT cells. It gets hypermethylated with differentiation in both WT and DKO cell lines (Figure 2.21). Next we decided to see whether knocking out TET1 and 2 would affect the DNA methylation at EC-enriched genes. The DNA methylation of the eNOS promoter is increased in the DKO ES cells compared to the WT cells although non-significant, and both cell lines had increased methylation with differentiation (Figure 2.22). Both CD31 and CDH5 promoters were more significantly methylated in DKO ES cells than in the WT, and they further increased in methylation with differentiation (Figures 2.23 and 2.24). Collectively this shows that DKO ES cells are more methylated than WT ES cells at LINE-1 and EC-enriched gene promoters. When differentiating DKO ES cells there is further methylation matching the hypermethylation that occurs in the WT cells.

2.4.4 Gene expression changes during WT and DKO ES cell differentiation

EC-enriched genes have mostly low expression in ES cells, and they go up with differentiation. eNOS expression starts to increase at day 7 of differentiation in WT cells, however, in DKO cells at day 7 there is no such increase (Figure 2.25). Although both WT and DKO EB day 7 are hypermethylated at the eNOS promoter suggesting that its expression should be repressed, there is an increased gene expression at EB day 7 in the WT cells. eNOS has a placental promoter, which regulates eNOS expression in the synctiotrophoblast\(^\text{199}\). Unpublished data from our lab showed that the eNOS placental promoter is also regulated by DNA methylation, where it starts-off heterogeneous and than becomes hypermethylated at term (Thesis by Megan Thompson 2012). Thus, the increase in eNOS mRNA levels at EB day 7 despite the hypermethylation of the proximal promoter, might be driven from the alternative placental promoter. Further examination of the eNOS placental promoter during ES cell differentiation must be done to confirm it. To determine whether the increase in mRNA detection was due to increased transcription or due to mRNA stability changes, we looked at the hnRNA expression of eNOS, and found that it mirrored the mature mRNA, where there is increased expression at WT EB day 7 compared to the ES cell and EB day 4, and no such
increased expression occurs in the DKO cells (Figure 2.26).

CD31 is basally expressed in murine ES cells, thus compared to other EC-enriched genes, there are high levels of basal CD31 in ES cells. There is a sharp decline in CD31 mRNA at EB day 4, which might be associated with the increased DNA methylation at EB day 4. At EB day 7, the mRNA levels of CD31 go up in WT cells but not in the DKO (Figure 2.27). Confirming that like eNOS, CD31 mRNA does not get upregulated in DKO cells after it was silenced at the start of EB differentiation. CDH5 is not highly expressed in ES cells, however, unlike eNOS that goes up ~10 fold at EB day 7 in the WT cells, CDH5 goes up ~100 fold at EB day 7 of WT cells when compared to ES cells. Although DKO cells show some increase in CDH5 expression at EB day 7, it is not significant compare to DKO ES cells (Figure 2.28). Taken together it seems that the absence of TET1 and 2 at the ES cell stage (since they are not present at EB ay 4 or day 7), inhibits or retards EC-enriched gene activation at EB day 7. As the deletion of TET1 and 2 is not lethal to the mouse, it can be assumed that these EC-enriched genes will turn on later in development, but are just being stalled by the absence of the TETs. Since in both the WT and DKO cells the EC-enriched promoter genes are hypermethylated at EB day 7, and only the WT cells have increased gene expression at EB day 7, it appears that the change in expression is DNA methylation independent. eNOS expression in individual TET1 and TET2 knockout mouse ES cell line was examined, and showed that similar to the WT cells, TET1-/- EBs at day 7 had increased eNOS expression, however, TET2-/- EBs at day 7 did not have eNOS expression (Figure 2.29). Suggesting that TET2 is the key protein in upregulation of eNOS gene during ES cell differentiation. Further experiments have to be done on the expression of other EC-enriched genes using the individual TET1 and TET2 knockouts. Thus, so far we can conclude that the presence of TET1, 2 or both is important in ES cells to help activate EC-enriched gene expression, which is not dependent on their role as demethylases.

The literature shows that genes that contain 5hmC in ES cells are more likely to contain bivalent chromatin domains at their promoters, those are activating and repressing histone marks on the same nucleosome that poise a gene for activation or
repression\textsuperscript{145}. Also, 5hmC was less likely to be found at genes with only the activating H3K4me3 mark. Moreover, genes with 5hmC at their start sites showed lower expression in mouse ES cells than other genes, and were more likely to be upregulated upon differentiation. Although 5mC at the TSS also correlates with lower gene expression in murine ES cells, 5mC is not associated with bivalent marks\textsuperscript{144}. TET1 binding to DNA has been positively correlated with H3K4me3, and bivalent genes are mostly unmethylated and present in low CpG promoters\textsuperscript{64}. Moreover, TET1 also recruits the PRC2 complex, which in turn catalyzes the repressive H3K27me3 mark\textsuperscript{200}. This suggests that TET1 might facilitate the establishment of histone bivalent marks in ES cells, thus poising the genes for activation. When knocking out TET1, the bivalent code is not set properly which might cause in delayed gene activation upon cell differentiation.

TET2 was shown to form a complex with O-linked N-acetylglucosamine (GlcNAc) transferase (OGT), which might regulate gene transcription\textsuperscript{201}. TET2 targets OGT to the chromatin and regulates histone H2B O-GlcNAcylation in ES cells\textsuperscript{201}. H2B O-GlcNAcylation is associated with active TSS and positively regulates transcription\textsuperscript{202}. Furthermore, TET2 and OGT together can act as a scaffold and recruit other chromatin modifying proteins that may affect transcription of genes\textsuperscript{60}. Thus it is possible that in the absence of TET2, there is no OGT activation of gene transcription, and thus a delayed activation of EC-enriched genes expression in DKO EB day 7 cells.

2.5 Limitations

The findings described in this study are based on \textit{in vitro} studies, and thus is not a complete representation of what occurs \textit{in vivo} during embryonic development. Although we discussed the importance of cell heterogeneity, we did not account for cell-cell interactions and how they can affect ES cell heterogeneity or behavior. Our ES cell differentiation model did not included directed differentiation towards an EC lineage, and thus allowed the cells in the EB to differentiate to all lineages. This might contribute to EB variability and greater EC-enriched gene expression variability, as not all the differentiated cells were EC progenitors.
Our study focuses on promoter DNA methylation, and on estimating CpG methylation by assaying a small region of the gene. There are obvious limitations to this assay, which can be extended and confirmed by using genomic arrays that cover larger regions, and take into account intragenic DNA methylation.

All studies have been done in mouse cells, and although the examined proteins, and the DNA methylation marks are conserved between mouse and human, it should still be confirmed to also be relevant to humans.

2.6 Conclusion

This study provided evidence that the presence of TET1 and TET2 in ES cells contribute to the maintenance of the heterogeneous DNA methylation in EC-enriched gene promoters, and their plasticity during cell division. We have also shown that in the absence of TET1 and 2 there is increased LINE-1 and EC-enriched gene promoter methylation in ES cells and during EB differentiation. Despite, the hypermethylation at EB day 7 EC-enriched gene promoters, there is increased expression of those genes, which does not occur if TET1 and 2 are not expressed at the ES cell stage.

2.7 Future Directions

Who are the key players: TET1, TET2 or both?

Gene expression of other EC-enriched genes besides eNOS should be assayed for the individual TET1/- cells and TET2/- cells. This will help determine which one of them regulates the EC-enriched genes expression, or if they are complementing each other, and thus both have to be deleted for the gene expression retardation to occur.

Are TET1, TET2 or both are directly binding to the EC-enriched gene promoters?

ChIP assay should be done, to determine if the TETs directly bind to EC-enriched gene
promoters, or if there are other DNA binding proteins that facilitate this interaction. The precipitated chromatin could be tested for histone bivalent marks. As it is not documented in the literature whether TET knockout ES cells have less histone bivalent marks. This will help to determine whether the retarded gene activation in the DKO cells is because the genes were not poised for activation at the ES cell stage.

Is the retarded EC-enriched gene activation is due to hindered RNA POL II binding?

RNA POL II ChIP assay can be done to determine if absence of TETs inhibits binding, or whether the reduced mRNA levels of EC-enriched genes at EB day 7 might be due to post-transcriptional modification. Also, POL II ChIP can be done on WT ES cells, and the DNA methylation of the precipitated chromatin can be determined to examine whether the DNA methylation inhibits direct POL II binding to the promoter.

Are there other important players in EC progenitor cells that might contribute to EC-enriched gene activation?

Sort EB day 7 for EC progenitors CD31+ and non-progenitors CD31- cells. Since both TET3 and DNMT3a expression goes up at EB day 7, it will help determine if there is a difference in their expression between the two-sorted populations which might serve as another explanation to the change in EC-enriched gene expression upregulation.

Does IDAX (CXXC4) or MBD3L2 interact with TET2 in ES cells at the EC-enriched gene promoters?

Immunoprecipitate IDAX and MBD3L2 from ES cells protein extract, and on a Western blot probe with a TET2 antibody to determine if the two proteins physically interact in ES cells. ChIP of IDAX and MBD3L2 can be done to determine if they bind to EC-enriched gene promoters. If confirmed, ChIP can be done in TET2 knockout ES cells, to determine whether IDAX or MBD3L2 are guided by TET2 to EC-gene enriched promoters.
2.8 References

18. Terme, J. M. et al. Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. The


Lorsbach, R. B. *et al.* 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, doi:10.1038/sj.leu.2402834 (2003).


van der Heijden, G. W. *et al.* Asymmetry in histone H3 variants and lysine methylation between paternal and maternal chromatin of the early mouse zygote.


