ANALYSIS OF THE ROLE OF THE HISTONE METHYLTRANSFERASE G9A IN TRANSCRIPTIONAL REPRESSION OF INTRODUCED AND ENDOGENOUS RETROELEMENTS IN MURINE EMBRYONIC STEM CELLS

by

KEVIN BRIAN DONG

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2009

© Kevin Brian Dong, 2009
ABSTRACT

DNA methylation and posttranslational histone modification are epigenetic marks that influence transcription of associated DNA. DNA methylation involves the direct addition of methyl groups to cytosine bases in DNA, and is associated with transcriptional repression. Histone modification can be associated with either activation or repression, and involves covalent addition of a wide variety of chemical groups to the N-terminal tails of the core histones. A relationship between DNA methylation of repetitive elements and methylation of H3 lysine 9 (H3K9) exists in Arabidopsis thaliana, where mutants in the histone methyltransferase (HMTase) KRYPTONITE (KYP) show a defect in CpNpG DNA methylation of retrotransposons. Similarly, in Neurospora crassa the H3K9 HMTase defective in methylation 5 (DIM-5) is required for DNA methylation of repetitive elements. In mice, the major H3K9 methyltransferases (MTase) in euchromatin are G9a and GLP; G9a has been shown to influence DNA methylation of a limited number of single copy genes, however, its effect on interspersed elements, such as endogenous retroviruses (ERVs) or exogenous retroviruses (XRVs) has not been characterized.

Here, I show that G9a is responsible for H3K9 dimethylation (H3K9me2) and Kap-1 recruitment at ERVs, as well as H3K9me2 and H3K9me3 at XRVs. Moreover, G9a acted upstream of DNA methylation of both XRVs and ERVs. Mutation of G9a resulted in heightened expression of XRVs, but did not influence expression of ERVs. Analysis of the mechanism behind this histone modification-DNA methylation link revealed that recruitment of the DNA methyltransferase (DNMT) De novo methyltransferase 3a (Dnmt3a) was reduced at ERVs in the absence of G9a.
Surprisingly, this recruitment was partially rescued by catalytically null G9a.

Based on these observations, I conclude that G9a is targeted to ERVs and XRVs and marks these elements with H3K9me2. In addition, G9a acts upstream of DNA methylation and H3K9me3 at XRVs. The DNA methylation deficiency may be partly explained by recruitment of Dnmt3a independent of G9a’s catalytic activity. The differences observed between G9a activity at exogenous and ERVs suggest that developmental and/or evolutionary mechanisms place additional constraints on ERVs compared to XRVs.
TABLE OF CONTENTS

Abstract............................................................................................................................................ii
Table of Contents............................................................................................................................iv
List of Tables..................................................................................................................................vi
List of Figures................................................................................................................................vii
List of Abbreviations......................................................................................................................ix
Acknowledgements.........................................................................................................................xi
Dedication......................................................................................................................................xii

1 Introduction...................................................................................................................................1
  1.1 An Overview of Retroviruses ........................................................................................1
  1.2 Retroviral Classes .........................................................................................................2
  1.3 Repression of Retroviruses ............................................................................................4
  1.4 Endogenous Retroviruses (ERVs) .................................................................................6
  1.5 Epigenetics and the Repression of Retroelements ..........................................................7
    1.5.1 Overview of DNA methylation ..............................................................................7
    1.5.2 Dnmt1 ..................................................................................................................10
    1.5.3 Dnmt3a/3b ...........................................................................................................11
    1.5.4 Dnmt3L ...............................................................................................................12
  1.6 Overview of Histone Modification ..............................................................................12
    1.6.1 Role of histone modifications .............................................................................13
    1.6.2 Histone lysine methylation ................................................................................16
    1.6.3 The SET domain proteins ....................................................................................16
    1.6.4 SET domain H3K9 methyltransferases .................................................................17
  1.7 Statement of Thesis .....................................................................................................22

2 Materials and Methods ......................................................................................................24
  2.1 Cell Lines and ES cell culture .....................................................................................24
  2.2 Plasmids and Primers ..................................................................................................25
  2.3 Retroviral Infections ...................................................................................................26
  2.4 Flow Cytometry ...........................................................................................................27
  2.5 Realtime PCR .............................................................................................................28
  2.6 Copy Number Determination ......................................................................................29
2.7 Methylation-Sensitive Digests ................................................................................. 29
2.8 Chromatin Immunoprecipitation (ChIP) ................................................................. 29

3 Results .......................................................................................................................... 33
3.1 Experimental Design ............................................................................................... 33
3.2 Chromatin Differences Between ERVs and XRVs .................................................. 36
3.3 Reduced DNA Methylation at MSCV in G9a Mutants .............................................. 39
3.4 Defective Silencing of MSCV Retrovirus in G9a-/- ES Cells ................................. 41
3.5 Defective Silencing of MSCV in Dnmt3a/3b Mutants ................................................. 46
3.6 Expression of MSCV is Unchanged in Suv39h1/2 Mutants ...................................... 49
3.7 Reversibility of G9a-Mediated Silencing Defects ................................................... 53
3.8 Impaired Dnmt3a Recruitment at ERVs in G9a Mutants ........................................... 56
3.9 Mechanism of G9a-Independent H3K9 Trimethylation ........................................... 60

4 Discussion ..................................................................................................................... 63
4.1 A G9a-DNA Methylation Link at Retroviruses ......................................................... 65
4.2 Role of G9a in the Silencing of an XRV ................................................................. 65
4.3 Divergent Roles of G9a in XRV and ERV Repression ............................................. 68
4.4 Divergent Epigenetic Roles of G9a at XRVs and ERVs .......................................... 69
4.5 Biochemical Basis of the G9a-DNA Methylation Link ........................................... 72
4.6 Applications and Future Directions ......................................................................... 74
LIST OF TABLES

Table 1. The DNA Methyltransferases ................................................................. 8
Table 2. Histone modifications and their effects .................................................. 13
Table 3. Cell lines used in this thesis ................................................................. 25
Table 4. Primer sequences used for qPCR .......................................................... 26
Table 5. Antibodies used in this thesis ............................................................... 31
Table 6. Analysis of endogenous IAP and MLV retroelement DNA methylation status via bisulphite sequencing ......................................................... 41
Table 7. Relative expression status of ERVs ...................................................... 45
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Life cycle of a typical retrovirus</td>
<td>2</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Structure of Mo-MLV</td>
<td>4</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Domain structure of G9a</td>
<td>20</td>
</tr>
<tr>
<td>Figure 4</td>
<td>MSCV-GFP reporter construct used in this thesis</td>
<td>26</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Representative standard curve for qPCR analyses</td>
<td>28</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Outline of experimental design</td>
<td>34</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Viral copy number determination of infected ES cells</td>
<td>35</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Schematic of ChIP</td>
<td>36</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Chromatin status of exogenous MSCV retrovirus</td>
<td>37</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Chromatin status of endogenous LTR murine retroviruses</td>
<td>38</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Chromatin status of endogenous IAP and MusD LTR retroviruses</td>
<td>39</td>
</tr>
<tr>
<td>Figure 12</td>
<td>DNA methylation status of XRV</td>
<td>40</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Day 4 post-infection expression levels of integrated MSCV-GFP in WT and G9a-/- ES cells as measured by flow cytometry</td>
<td>43</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Day 24 post-infection expression levels of integrated MSCV-GFP in WT and G9a-/- ES cells as measured by flow cytometry</td>
<td>44</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Expression of exogenous MSCV-GFP retrovirus in WT and G9a-/- cells over the course of the experiment</td>
<td>45</td>
</tr>
<tr>
<td>Figure 16</td>
<td>Day 4 post-infection expression levels of integrated MSCV-GFP in WT and Dnmt3a/3b-/- ES cells as measured by flow cytometry</td>
<td>47</td>
</tr>
<tr>
<td>Figure 17</td>
<td>Day 29 post-infection expression levels of integrated MSCV-GFP in WT and Dnmt3a/3b-/- ES cells as measured by flow cytometry</td>
<td>48</td>
</tr>
<tr>
<td>Figure 18</td>
<td>MSCV-GFP retroviral expression in WT and Dnmt3a/3b-/- ES cells</td>
<td>49</td>
</tr>
<tr>
<td>Figure 19</td>
<td>Day 4 post-infection expression levels of integrated MSCV-GFP in WT and Suv39h1/2-/- ES cells as measured by flow cytometry</td>
<td>51</td>
</tr>
<tr>
<td>Figure 20</td>
<td>Day 24 post-infection expression levels of integrated MSCV-GFP in WT and Suv39h1/2-/- ES cells as measured by flow cytometry</td>
<td>52</td>
</tr>
<tr>
<td>Figure 21</td>
<td>MSCV-GFP infected Suv39h1/2-/- cells show similar silencing kinetics to WT</td>
<td>53</td>
</tr>
<tr>
<td>Figure 22</td>
<td>Day 4 post-infection expression levels of integrated MSCV-GFP in WT and G9a transgenic rescue ES cells as measured by flow cytometry</td>
<td>54</td>
</tr>
<tr>
<td>Figure 23</td>
<td>Day 18 post-infection expression levels of integrated MSCV-GFP in WT and G9a transgenic rescue ES cells as measured by flow cytometry</td>
<td>55</td>
</tr>
<tr>
<td>Figure 24</td>
<td>MSCV-GFP retroviral expression in WT and G9a Tg ES cells</td>
<td>56</td>
</tr>
<tr>
<td>Figure 25</td>
<td>Recruitment of Dnmt3a is reduced in G9a-/- ES cells</td>
<td>57</td>
</tr>
<tr>
<td>Figure 26</td>
<td>Catalytically inactive G9a partially restores Dnmt3a recruitment levels</td>
<td>59</td>
</tr>
<tr>
<td>Figure 27</td>
<td>Bisulphite analysis of DNA methylation status in G9a mutant lines</td>
<td>59</td>
</tr>
<tr>
<td>Figure 28</td>
<td>Kap-1 recruitment is reduced in the absence of G9a at endogenous IAP and MusD retroviruses</td>
<td>61</td>
</tr>
<tr>
<td>Figure 29</td>
<td>Kap-1 recruitment is reduced in the absence of G9a at endogenous IAP and MusD retroviruses</td>
<td>62</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AIDS - Acquired immunodeficiency syndrome
CMT3 - CHROMOMETHYLASE3
CpNpG - Cytosine, followed by any DNA base, followed by guanine
CtBP1 - C-terminal binding protein 1
DIM-5 - defective in methylation
DMEM - Dulbecco's Modified Eagle Media
DNMT - DNA methyltransferase
Dnmt1 - De novo methyltransferase 1
Dnmt3a - De novo methyltransferase 3a
Dnmt3b - De novo methyltransferase 3b
Dnmt3L - De novo methyltransferase 3L
ERV - Endogenous retrovirus
ES - Embryonic stem
FSC - Forward scatter
GFP - Green fluorescent protein
GLP - G9a-like protein
HIV - Human immunodeficiency virus
HMTase - Histone methyltransferase
HP1β - heterochromatin protein 1β
HTLV-1 - human T-cell leukemia virus type 1
IAP - Intracisternal A particle
ICF - Immunodeficiency, centromere instability and facial anomalies
inf. - MSCV-GFP infected
K - Lysine
KYP - KRYPTONITE
Lsh - Lymphoid specific helicase
LTR - Long terminal repeat
Mo-MLV - Moloney murine leukemia virus
mouse - Mus musculus
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSCV</td>
<td>Murine stem cell virus</td>
</tr>
<tr>
<td>MTase</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NCR</td>
<td>Negative control region</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCNA</td>
<td><em>proliferating cell nuclear antigen</em></td>
</tr>
<tr>
<td>PI</td>
<td>Propidium iodide</td>
</tr>
<tr>
<td>piRNA</td>
<td><em>piwi</em>-associated RNA</td>
</tr>
<tr>
<td>POU</td>
<td>Pit-Oct-Unc</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative, realtime polymerase chain reaction</td>
</tr>
<tr>
<td>RA</td>
<td>Retinoic acid</td>
</tr>
<tr>
<td>RBS</td>
<td>Repressor binding site</td>
</tr>
<tr>
<td>RE</td>
<td>Restriction enzyme</td>
</tr>
<tr>
<td>RT</td>
<td>Room temperature</td>
</tr>
<tr>
<td>SetDB1</td>
<td><em>SET domain protein, bifurcated 1</em></td>
</tr>
<tr>
<td>Suv39h1</td>
<td>Suppressor of Variegation 3-9 H1</td>
</tr>
<tr>
<td>Suv39h1/2</td>
<td>Suv39h1 and Suv39h2</td>
</tr>
<tr>
<td>Suv39h2</td>
<td>Suppressor of Variegation 3-9 H2</td>
</tr>
<tr>
<td>TRIM28</td>
<td><em>tripartite interaction motif 28</em></td>
</tr>
<tr>
<td>TUNEL</td>
<td>Terminal Transferase dUTP Nick End Labeling</td>
</tr>
<tr>
<td>UCRBP</td>
<td><em>upstream control region binding protein</em></td>
</tr>
<tr>
<td>Wiz</td>
<td><em>widely interspaced zinc finger motifs</em></td>
</tr>
<tr>
<td>WT</td>
<td>wildtype</td>
</tr>
<tr>
<td>XRV</td>
<td>Exogenous retrovirus</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

First and foremost, I would like to thank all the people I’ve had the opportunity to work with for their mentorship, assistance, and guidance throughout the course of this thesis. To my mentor, Dr. Matthew Lorincz, for his guidance in matters both within the field, and as a scientist as a whole. To the members of my committee, Dr. Carolyn Brown and Dr. Louis Lefebvre for keeping me on track, and for providing help in matters ranging from the abstract to the practical. I would also like to thank the wonderful members of the Lorincz Lab: Danny, Irina, Lucia, Margaret, Preeti, Ruth, Sandra, Winnie; it has truly been a team effort. I would also like to take this opportunity to thank Dr. Jacob Hodgson, whose informal hallway chats taught me a lot about the philosophy of science.

I wish to thank my friends and family for their continued support through the highs and the lows of this endeavor. Jamie, thank you for keeping things real when I get lost in the midst of my experiments. Your dry humor helped keep my feet on the ground. Gordon and Elvira, thank you for helping out with matters outside of science, and for providing me with a foundation and constant background of knowledge. It helps having parents who are both scientists; you two have been through all this long before me, and I respect your advice.

Finally, I would like to thank my colleagues in the field who have provided knowledge and support. To Dr. Yoichi Shinkai and lab: I thank you for providing me with most of the cell line stocks used in this thesis. Without your contributions to the field in general, and to our lab in particular, none of this would have happened. I would also like to thank the labs of Dr. Dirk Schübeler, Dr. Dixie Mager, and Dr. Bernard Lehnertz for their contributions to this work. For this outreach of the scientific community, I am truly grateful.
DEDICATION

To Elivra, Gordon, Rodney, Daisy, and Misty

A bicycle without someone to ride it
A mind without a heart to guide it
Lest we forget the crucial part,
The mind is but a tool of the heart.

Thank you for my inspiration
1 Introduction

1.1 An Overview of Retroviruses

Retroviruses are RNA viruses that are reverse transcribed into DNA and integrate into the genome of the host. They are of interest in the study of infectious disease and as a tool for gene delivery. A well-known disease causing XRV is human immunodeficiency virus (HIV), which is responsible for acquired immunodeficiency syndrome (AIDS) and causes roughly 2 million deaths per year worldwide (UNAIDS, 2008). Retroviruses as tools have been widely used in gene therapy, where they are employed to deliver corrective genes into patients with genetic diseases. Unfortunately, so far, they have had little success when used as vectors for gene therapy in mammals. Nevertheless, there is not much known about the factors that regulate the expression of integrated retroviruses (also known as proviruses), i.e. whether they will be active (expressed) in host cells, or inactive (silenced) in the host cell. In this thesis, I examine factors that play a role in the silencing of ERVs and XRVs using a mouse model system.

The retroviral life cycle begins as a particle termed a virion. The virion surface is studded with envelope proteins that recognize specific cell-surface proteins, and define the virus’ host range. Upon encountering a cell with the proper surface proteins, these envelope proteins bind to specific cell surface targets and trigger fusion of the virion with the host cell, leading to entry of the virus into the cell. Once in the cytoplasm, the virus reverse transcribes its RNA genome into double-stranded DNA. This DNA copy of the genome then translocates into the host cell’s nucleus and integrates into the host genome, with the help of endogenous and viral factors. The integrated ‘provirus’ then begins production of viral RNAs, which will include messenger RNAs.
that give rise to viral proteins, as well as RNA copies of its genome for packaging into new virions. These components assemble at the cellular membrane, where the maturing virion then buds out from the cell enveloped in a portion of its host membrane, thus completing the retroviral life cycle (Figure 1).

Figure 1. Life cycle of a typical retrovirus. 1) Binding of virion envelope protein to cell-surface receptor triggers fusion and cellular entry. 2) RNA is reverse transcribed into double stranded DNA. 3) Viral DNA genome translocates to the nucleus and integrates into host genome. 4) Transcription from the viral DNA genome produces the RNA genome as well as viral proteins. 5) Viral components assemble, and the new virion particles bud from the host cell. Reprinted by permission from W. H. Freeman and Company/Worth Publishers: Molecular Cell Biology (Lodish, Berk, et al, 2000), copyright 2000.

1.2 Retroviral Classes

A number of retroviral classes have been described, based on sequence conservation: the α, β, γ, δ, and ε retroviruses, as well as the lentiviruses and spumaviruses. The first retroviruses to be
discovered was Rous sarcoma virus – an alpharetrovirus – in 1911 (Rous, 1911), and the first
human retrovirus to be identified was human T-cell leukemia virus type 1 (HTLV-1), a
deltaretrovirus (Poiesz, Ruscetti, et al, 1980; Yoshida, Miyoshi, and Hinuma, 1982). The
infamous HIV retrovirus was discovered by French and American scientists in 1983 and 1984
respectively (Barre-Sinoussi, Chermann, et al, 1983; Popovic, Sarngadharan, et al, 1984), and
belongs to the lentiviral class.

One prominently studied retrovirus in mice is the Moloney murine leukemia virus (Mo-MLV),
the archetypal gammaretrovirus. Mo-MLV was originally isolated from a murine sarcoma
induced by infection with high amounts of murine leukemia virus (Moloney, 1960). Mo-MLV
possesses similar structure to other retroviruses, containing gag, pol, and env genes flanked by
direct repeats (Figure 2), but does not possess an extra region coding for oncogenic genes like in
Rous sarcoma virus or HTLV-1. The virus prefers to integrate near the transcription start sites of
active genes (Wu, Li, et al, 2003), and perhaps as a consequence of this, is still highly
pathogenic; typically causing murine T-cell lymphoma with 100% incidence and a mean latency
of 3-4 months (reviewed in (Fan, 1997)), a property that has facilitated its discovery and study.
Figure 2. Structure of Mo-MLV. (top) General DNA structure. The U3 region derives from unique sequence located at the 3' region of the original viral RNA and contains most of the transcriptional control elements of the provirus. The R region derives from a sequence that is repeated at both 5' and 3' ends of the RNA. The U5 region derives from unique sequence found at the 5' end of the viral RNA. The \textit{gag}, \textit{pol}, and \textit{env} genes code for the internal structure proteins, enzymes, and outer structural proteins of the virion respectively. Transcription is initiated at the U3-R boundary (see (Coffin, Hughes, and Varmus, 1999) for a review) (bottom) Expansion of the LTR region. The U3 region contains the viral enhancer which contains two \textasciitilde 75bp direct repeats and possesses numerous enhancer binding sites. Upstream of this is a negative control region (NCR) which binds upstream control region binding protein (UCRBP) (Flanagan, Becker, et al, 1992), termed YY1 in humans. Downstream of the transcription start site (arrow) lies an additional repressor binding site (RBS), which binds protein Kap-1 (Wolf, and Goff, 2007).

1.3 Repression of Retroviruses

In answer to the onslaught of retroviral infections, host organisms have evolved methods to impair retroviral activity. For example when undifferentiated pluripotent mouse cell lines were infected with Mo-MLV, no detectable evidence of viral replication was observed: the cells had efficiently suppressed Mo-MLV expression (Teich, Weiss, et al, 1977). Furthermore, when these lines were subsequently differentiated, they maintained retroviral silencing. Mo-MLV also fails to replicate after infection of pre-implantation embryos or embryonic carcinoma cell lines.
(Jahner, Stuhlmann, et al, 1982; Stewart, Stuhlmann, et al, 1982; Teich, Weiss, et al, 1977). In contrast to embryonic carcinoma cells, infected terminally differentiated fibroblasts showed evidence of viral replication. This suggests that once a cell is able to silence Mo-MLV, that silencing event is conferred to all its differentiated progeny; however differentiated cells appear vulnerable to new retroviral infections. One interpretation of this cell-stage-dependent expression is that stem cells, via an unknown mechanism, inhibit retroviral replication, while this repressive pathway is absent in differentiated cells.

By integrating into a host genome retroviruses become part of the host genome and are replicated during normal genomic DNA replication. However host defence mechanisms continue to act upon these proviruses in order to silence them; cells have an ability to distinguish the proviral DNA sequences from their own, and act to silence the viral sequences. Silencing is implemented by certain *cis* acting retroviral sequences that attract *trans* acting proteins that have the ability to recognize and bind these sequences. This binding then triggers downstream silencing events. For example, Mo-MLV possesses a negative control region (NCR) within the 5’ long terminal repeat (LTR), as well as a repressor binding site (RBS) located downstream of the 5’ LTR (reviewed in (Haas, Lutzko, et al, 2003)). Mutation of these sites promotes transcription of the integrated retrovirus. The repression these *cis* sequences induce is implemented by *trans*- acting proteins: the factor *upstream control region binding protein* (UCRBP) (YY1 in humans) was found to bind to the NCR of Mo-MLV (Flanagan, Krieg, et al, 1989; Flanagan, Becker, et al, 1992; Seto, Shi, and Shenk, 1991; Shi, Seto, et al, 1991; Shrivastava, and Calame, 1994), and a protein called *tripartite interaction motif 28* (TRIM28) (also termed Kap-1) was found to bind at the RBS (Wolf, and Goff, 2007). TRIM family proteins in particular have been shown to be involved in antiretroviral activity (Uchil, Quinlan, et al, 2008), and consistent with that,
decreased levels of Kap-1 leads to an attenuation of retroviral silencing (Wolf, and Goff, 2007).

1.4 Endogenous Retroviruses (ERVs)

The consequence of the interplay between retroviral infection and host defense is that organisms have come to harbour many silent copies of integrated retrovirus within their genomes, which are called endogenous retroviruses (ERVs). These ERVs are transmitted across generations through the germline: retroviruses are ‘endogenized’ when a retrovirus integrates in the germline of an individual. Should that individual survive and successfully silence the retrovirus, when the individual reproduces, the offspring created from an infected germ cell will carry the silent retrovirus in every cell in its body. This phenomenon is widespread in multicellular organisms, including humans. It is estimated that up to 46% of the human genome and 37.5% of the Mus musculus genome is made up of the remnants of transposable elements (Mouse Genome Sequencing Consortium, Waterston, et al, 2002; Weiss, 2006). Of course over the course of evolutionary time many of these silent retroviruses have become mutated, and thus are no longer capable of replicating even if host-mediated silencing were removed.

In mice, there are three classes of ERVs (reviewed in (Stocking, and Kozak, 2008)). I focus on two class II ERVs in this thesis; the intracisternal A particle (IAP) and MusD retrotransposons. For unknown reasons, some ERVs have been more successful than others, in terms of copy number expansion. IAP elements are present at 1000-2000 copies in the mouse genome (reviewed in (Kazazian, 1998)), although many of them are truncated. Full-length IAPs are ~7kb in length and contain retroviral gag and pol genes; however their potential env sequences are defective with numerous nonsense codons. MusD elements are present at ~100 copies in a common laboratory mouse strain, C57BL/6J, and similar to IAP, possesses gag and pol

### 1.5 Epigenetics and the Repression of Retroelements

Epigenetic pathways are essential for protecting against reactivation of ERVs, and are represented by DNA methylation and histone modification. These two epigenetic modifications embody cellular information over and above the genetic information within DNA bases, and instruct the cell to activate or repress the associated DNA. Formally, epigenetics is the study of stable alterations in gene expression potential that arise during development and cell proliferation that is not accompanied by a change in the DNA sequence. Epigenetics plays a crucial role in host defense against invading viral genomes and in adapting to varying environments during the lifespan of an organism (Jaenisch, and Bird, 2003), and in light of these facts I wished to explore in greater depth epigenetic aspects of retroviral regulation.

#### 1.5.1 Overview of DNA methylation

DNA methylation is accomplished by the addition of a methyl group from S-adenosyl-L-methionine to the 5' carbon position of cytosine. This reaction is catalyzed by three main MTases; Dnmt1, Dnmt3a, and Dnmt3b (see *Table 1*) and assisted by a number of isoforms of these MTases.
<table>
<thead>
<tr>
<th>DNA Methyltransferase</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dnmt1o (mouse)</td>
<td>Oocyte-specific splice variant of Dnmt1, involved in methylation of imprinted regions (Cardoso, and Leonhardt, 1999; Carlson, Page, and Bestor, 1992; Doherty, Bartolomei, and Schultz, 2002; Mertineit, Yoder, et al, 1998)</td>
</tr>
</tbody>
</table>

Together these proteins catalyze, or assist in catalyzing DNA methylation, thereby marking DNA as transcriptionally repressed. An example of the role DNA methylation plays in repression is seen in mouse germ cells that lack the chromatin remodeler lymphoid specific helicase (Lsh), a protein that associates with Dnmt3a and 3b (Zhu, Geiman, et al, 2006): these cells had reduced
DNA methylation and increased expression of IAP elements (De La Fuente, Baumann, et al, 2006). Another example showed that in mouse ES cells mutant for Dnmt1, Dnmt3a, and Dnmt3b there was reactivation of IAP and LINE1 elements (Tsumura, Hayakawa, et al, 2006).

The level of DNA methylation varies widely across organisms, but where it exists it plays a part in a variety of cellular processes. Some organisms are essentially devoid of methylation (i.e. *Saccharomyces cerevisiae, Caenorhabditis elegans*), whereas others - including humans and mice - have very high levels of methylation. A general trend is that as the complexity of the genome rises, so does the level of methylation. Organisms that possess DNA methylation utilize it in a myriad of critical processes from imprinting, to X-chromosome inactivation, to suppressing parasitic elements (Chow, and Brown, 2003; Edwards, and Ferguson-Smith, 2007; Yoder, Walsh, and Bestor, 1997). This epigenetic mark is indispensable in the development and function of many organisms, including mice (Li, Bestor, and Jaenisch, 1992; Okano, Bell, et al, 1999).

There are two mechanisms by which DNA methylation is proposed to promote repression. The first is the steric hindrance that DNA methyl groups represent, which can impair the binding of transcription factors or RNA polymerase. Watt and Molloy (1988) found evidence supporting this when methylating a binding site for the transcription factor MLTF effectively inhibited its binding and subsequent transcription *in vitro*. The second mechanism involves proteins which specifically recognize methylated DNA, and then convey the mark into a repressive cellular chromatin state. Boyes and Bird (1991) found evidence supporting this by finding that four different methylated promoters were indirectly repressed using the methyl-DNA binding protein MeCP1. Thus, DNA methylation can promote repression via both direct and indirect mechanisms.
1.5.2 Dnmt1

Dnmt1 is responsible for “maintenance” methylation in mammalian cells. *In vitro*, this protein has a roughly 30-fold higher affinity for hemimethylated DNA than for unmethylated (Bacolla, Pradhan, et al, 1999). *In vivo*, 20% of CpG-CpG dyads in human cancer cells became hemimethylated when Dnmt1 expression was removed (Chen, Hevi, et al, 2007). Thus, Dnmt1 is responsible for propagating DNA methylation patterns from one generation to the next by converting hemimethylated DNA to fully methylated DNA. Consistent with this, Chuang *et al.* showed Dnmt1 associated with *proliferating cell nuclear antigen* (PCNA) at sites of newly-replicated DNA where hemimethylated DNA is found (Chuang, Ian, et al, 1997).

Loss of Dnmt1 function is associated with a number of cellular defects; the genome becomes severely undermethylated (to a level of 5% of wildtype (WT)) in mouse ES cells, and mice with this defect die as a post-implantation embryo (Lei, Oh, et al, 1996). Consequently, there is an increase in the rate of loss of heterozygosity as a result of mitotic recombination (Chen, Pettersson, et al, 1998). This increase in mitotic recombination may be explained by the high level of reactivation of the IAP family of retrotransposons that is also observed in Dnmt1 mutants (Walsh, and Bestor, 1999): retrotransposon reactivation and insertion into new genomic locations can potentially lead to promiscuous recombination. Thus, Dnmt1 is essential for normal development and repression of repetitive elements.

Dnmt1 has a number of different isoforms: Dnmt1b, Dnmt1o, and Dnmt1p. Dnmt1b is a splice variant of Dnmt1 that contains an extra 16 amino acids in humans (Hsu, Lin, et al, 1999) and 2 new amino acids in mice (Lin, Lee, et al, 2000), and is expressed at lower levels than Dnmt1 both at the RNA and protein level (Bonfils, Beaulieu, et al, 2000), although its function is not yet
clear. Dnmt1o is another splice variant of Dnmt1 found in mice that is present in the cytoplasm of the oocyte, and translocates into the nucleus briefly during development (Carlson, Page, and Bestor, 1992; Doherty, Bartolomei, and Schultz, 2002); it is thought to play a role in genomic imprinting. Dnmt1p is yet another splice variant of Dnmt1 that is produced in pachytene spermatocytes. This gene is somewhat unusual, as this particular variant is present stably as mRNA, but appears to be structured as to obstruct its own translation (Mertineit, Yoder, et al, 1998).

1.5.3 Dnmt3a/3b

Dnmt3a and Dnmt3b are the de novo DNMTs, capable of methylating completely unmethylated DNA. They were originally identified in a database search using bacterial cytosine-5 MTase sequences, and are present in both human and mouse (Okano, Xie, and Li, 1998a). Furthermore, while Dnmt1 showed preference for hemimethylated DNA, Dnmt3a and 3b methylated hemimethylated and unmethylated DNA with equal activity in cellular extracts, which led to their identity as de novo MTases.

Defects in Dnmt3b are responsible for the human disease Immunodeficiency, Centromere instability and Facial anomalies (ICF) syndrome (Hansen, Wijmenga, et al, 1999). This disease is an autosomal recessive disease characterized by a variable immunodeficiency, mild facial anomalies, centromeric decondensation and chromosomal instability involving chromosomes 1, 9, and 16 (Tiepolo, Maraschio, et al, 1979). Interestingly, reduced methylation at certain satellite repeats accompanies this disease (Jeanpierre, Turleau, et al, 1993), hinting at a possible role of DNA methylation in the maintenance of genome stability by maintaining these direct repeats in a condensed chromatin state.
1.5.4 Dnmt3L

Dnmt3L is involved in mediating the interaction between the de novo MTases Dnmt3a/3b and histones (Ooi, Qiu, et al, 2007). Originally identified in a database search, Dnmt3L shares 43-45% identity with human and mouse Dnmt3a and 3b in a cysteine-rich region, but lower identity with the catalytic regions of Dnmt3a and 3b, suggesting that Dnmt3L does not have DNMT activity (Aapola, Kawasaki, et al, 2000). Functionally, Dnmt3L plays an important role in development, as its absence results in imprinting-related defects in DNA methylation of the maternal mouse genome and subsequent lethality at day 9.5 post coitum (Bour'his, Xu, et al, 2001).

1.6 Overview of Histone Modification

DNA in the nucleus exists as chromatin, which includes various structural proteins that are intimately associated with DNA. The most basic set of proteins are the histones, which group together in octamers around which DNA is wound like 'beads on a string', forming the nucleosome core particle. The octamer is a grouping of 4 types of histones, each in duplicate: H2A, H2B, H3, and H4. Additional histone variants also exist, allowing for even greater diversity in structure and function. The nucleosome allows higher-order compaction of chromatin to take place, which is needed during cell division to permit proper transfer and segregation of chromosomes. The structure of the nucleosome consists of 146bp of DNA associated with each histone octamer, with the N-termini of each histone protruding like ‘tails’ into the surrounding space.

One of the most intriguing properties of histones is the myriad of post-translational modifications
that are added to them at their N-terminal 'tails'. These modifications each have specific effects on the transcription status of the associated DNA. Modifications can vary both in type, as well as location, providing a veritable language of histone modification. A summary of these is presented in Table 2.


<table>
<thead>
<tr>
<th>Type of Mark</th>
<th>Histone</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylation</td>
<td>H3</td>
<td>Activation</td>
</tr>
<tr>
<td></td>
<td>H4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H2A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H2B</td>
<td></td>
</tr>
<tr>
<td>Methylation</td>
<td>H3</td>
<td>Activation / Repression*</td>
</tr>
<tr>
<td></td>
<td>H4</td>
<td>Repression</td>
</tr>
<tr>
<td>Phosphorylation</td>
<td>H3</td>
<td>Activation</td>
</tr>
<tr>
<td></td>
<td>H2A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H2B</td>
<td></td>
</tr>
<tr>
<td>Ubiquination</td>
<td>H2A</td>
<td>Repression</td>
</tr>
<tr>
<td></td>
<td>H2B</td>
<td>Activation</td>
</tr>
<tr>
<td>Sumoylation</td>
<td>H2A</td>
<td>Repression</td>
</tr>
<tr>
<td></td>
<td>H2B</td>
<td></td>
</tr>
</tbody>
</table>

1.6.1 Role of histone modifications

Histone modifications are associated with activation or repression depending on the nature of the modification. For example, H3K9me2 is associated with transcriptional repression, while H3K9 acetylation is associated with transcriptional activation. There are a number of hypotheses as to exactly how these histone modifications enact their effects on the associated chromatin, including the charge neutralization hypothesis, and the histone code hypothesis.
1.6.1.1 Charge neutralization hypothesis

Under the charge neutralization hypothesis, histone modifications are thought to have a direct chemical effect on chromatin structure. Evidence supporting this hypothesis is the acetylation of lysines on histones. Acetylation reduces the positive charge of the histone, leading to reduction in the strength of the association between histone and DNA, and allowing for an open chromatin structure and ease of transcription. Unacetylated lysine residues possess a net positive charge, while DNA possesses a net negative charge on its sugar-phosphate backbone. These opposing charges attract, contributing to the close association between histone and DNA. If the lysine’s positive charge is neutralized by acetylation, the strength of histone-DNA association is reduced and the DNA adopts a more open, facilitative structure (Hong, Schroth, et al, 1993; Wade, Pruss, and Wolffe, 1997).

1.6.1.2 Histone code hypothesis

Although the biochemical analyses of the effects of lysine acetylation on histones may support the charge neutralization hypothesis, it does not fully explain the myriad and possibly combinatorial effects of all observed histone modifications. The histone code hypothesis acknowledges that such direct effects occur, but further proposes that a histone 'language' – composed of multiple histone modifications acting in a combinatorial or sequential fashion – may be encoded on histones that is read by other proteins to specify unique downstream functions, which in turn dictate dynamic transitions between transcriptionally active or transcriptionally silent chromatin states. This proposed ‘code’ may thus considerably extend the information potential of the genetic (DNA) code (Jenuwein, and Allis, 2001; Strahl, and Allis, 2000)
If the histone code hypothesis is true, there should be proteins which bind specific histone marks at defined residues to effect downstream functions. One such protein is murine heterochromatin protein 1β (HP1β); there are three HP1 variants in mouse: HP1α, HP1β, and HP1γ (reviewed in (Lomberk, Wallrath, and Urrutia, 2006)). HP1β localization to chromatin is implicated in transcriptional silencing of heterochromatic regions in mice (Festenstein, Sharghi-Namini, et al, 1999), and this protein specifically recognizes and binds methylated H3K9 and not a different but nearby residue H3K4 (Bannister, Zegerman, et al, 2001). Furthermore, Bannister et al. (2001) found that HP1β localization to chromatin was dependent on this binding to methylated H3K9, suggesting that H3K9 methylation recruits HP1β to mediate silencing in certain genomic regions.

Further evidence supporting the histone code hypothesis involves the protein CHROMOMETHYLASE3 (CMT3), a MTase that is responsible for CpNpG DNA methylation in *A. thaliana* (which means cytosine (C), followed by any DNA base, followed by guanine (G)) (Lindroth, Cao, et al, 2001). CMT3 possesses a chromodomain (Henikoff, and Comai, 1998), which is a methyllysine binding module, suggesting that CMT3 recognized methylated histone lysine marks. Indeed, while CMT3 could not bind methylated H3K9, Lindroth and colleagues (Lindroth, Shultis, et al, 2004) discovered that this protein was able to bind the combination of methylated K9 and K27 of histone H3. In other words, neither the K9 mark nor the K27 mark alone was sufficient to recruit this DNMT, but a combinatorial effect between the two was required. Taken together, this evidence supports the existence of a histone code whereby covalent histone marks – or a combination of them – can direct downstream effects through mediating proteins.
1.6.2 Histone lysine methylation

The previous examples highlight the role that histone lysine methylation plays in cellular function. Methylation occurs on K4, 9, 27, 36, and 79 of histone H3; as well as K20 of H4 (Berger, 2007). These residues can be mono- di- or trimethylated with the addition of one, two, or three methyl groups respectively. Lysine methylation can be either repressive or activating, depending on which histone tail residue the methylation is placed on. Of particular relevance to this thesis is H3K9 methylation, a repressive epigenetic mark. H3K9 methylation is found at heterochromatin and at repetitive elements, although its occurrence varies depending on the specific class of repetitive element (Bernstein, Meissner, and Lander, 2007; Martens, O'Sullivan, et al, 2005). Like nearly all histone lysine methylation, this mark is catalyzed by SET domain proteins.

1.6.3 The SET domain proteins

Most proteins that lay down histone lysine methyl marks possess a SET domain which serves as the catalytic domain. The SET domain was identified from the *Drosophila melanogaster* genes *Su(var)3-9, Enhancer of zeste*, and *trithorax* (Tschiersch, Hofmann, et al, 1994), and is made up of a collection of variable length β strands, some α-helical structure and an assorted collection of extended loops (reviewed in (Xiao, Wilson, and Gamblin, 2003)). The SET domain is flanked by two domains - the N and C terminal - which appear to be involved in stability and active site completion of the SET domain, and is furthermore bifurcated by a variable insert region (SET-I) which is involved in substrate specificity. Not all SET domain proteins are HMTases; SET domain proteins describe at least 50 proteins in humans (Couture, and Trievel, 2006), some of which have been shown to methylate lysines on targets such as tumor suppressor proteins.
(Chuikov, Kurash, et al, 2004), transcription factors (Kouskouti, Scheer, et al, 2004), and a kinetochore protein (Zhang, Lin, et al, 2005), although many remain undefined. The SET domain proteins thus form a diverse group which plays a role in lysine methylation and is involved in a number of cellular functions.

### 1.6.4 SET domain H3K9 methyltransferases

The known SET-domain containing H3K9 MTases are Suv39h1, Suv39h2, SetDB1, GLP, and G9a. In addition to lysine 9 of histone H3, some of these HMTases are also able to methylate other lysine targets.

**Suv39h1 and Suv39h2**

Suppressor of Variegation 3-9 H1 (Suv39h1) and Suv39h2 (Suv39h1/2) are proteins with similar (59% identity) structure and function (O'Carroll, Scherthan, et al, 2000). Together these proteins facilitate repression by catalyzing H3K9 dimethylation and trimethylation. Suv39h1/2 act mainly in the pericentric heterochromatin, where they play an important role in the maintenance of genome stability, and in turn the development of the embryos (Peters, O'Carroll, et al, 2001). Evidence of other functions have also been observed (Firestein, Cui, et al, 2000; Peters, Mermoud, et al, 2002). When both of these proteins are mutated, defects such as impaired embryo viability after day E12.5, chromosome missegregation, and nonhomologous recombination between chromatids are observed (Peters, O'Carroll, et al, 2001). In light of this, it is interesting to note that in Suv39h1/2 null ES cells there is a deficit in DNA methylation and modest increase in transcription of repetitive elements in pericentric heterochromatin (Lehnertz, Ueda, et al, 2003), which might contribute to the improper recombination phenotype. This data also suggests a mechanism that links histone modification and DNA methylation of repetitive
elements in mammalian cells.

**SetDB1/ESET**

SetDB1 and ESET are orthologous proteins in humans and mice respectively, and are euchromatic HMTases specific for H3K9 (Schultz, Ayyanathan, et al, 2002; Yang, Xia, et al, 2002), facilitating the conversion from dimethyl to trimethyl K9 (Wang, An, et al, 2003).

Similar to other H3K9 MTases, SetDB1/ESET plays a role in development (Dodge, Kang, et al, 2004), and repression. Wang *et al.* (2003) showed that ESET facilitated repression at a defined promoter *in vitro*, and Schultz *et al.* (2002) showed greater Kap-1 mediated enrichment of SetDB1, H3K9 methylation and HP1 at transgenes that were stably silenced compared to transgenes that were stably active. Furthermore, Sarraf and Stancheva (2004) showed that SetDB1 along with MBD1 was necessary for H3K9 methylation and transcriptional repression of the DNA-methylated p53-binding protein 2 gene. Taken together, this indicates that SetDB1/ESET plays a role in repression of a number of different loci.

**GLP**

G9a-Like Protein (GLP), also called Eu-HMTase1 in humans, methylates K9 and 27 of H3 (Ogawa, Ishiguro, et al, 2002; Tachibana, Ueda, et al, 2005). When mutated, mouse cells experience a reduction in H3K9me and me2, suggesting that GLP is a mono- and dimethyltransferase. Evidence in humans indicates that GLP functions to repress certain discrete sets of genes involved in proper growth and differentiation *in vivo*. Ogawa and colleagues showed that in human HeLa cells, GLP/Eu-HMTase1 is present in a complex that associates with E2F and myc-responsive genes specifically in the G0 phases of the cell cycle (Ogawa, Ishiguro, et al, 2002), and as a consequence is believed to be involved in repression of these genes. Also in humans, haploinsufficiency of Eu-HMTase1 is associated with the 9q34 subtelomeric deletion
syndrome (Kleefstra, Smidt, et al, 2005), which involves mental defects and suggests a role for this HMTase in the development of the central nervous system and proper nerve function.

In mice, Tachibana et al. (2005) showed that GLP is present in euchromatic regions, and is excluded from heterochromatic regions. It is also essential for development, as GLP null mice die around day E9.5. They went on to show that there was a reduction in H3K9me and me2 in euchromatin in GLP-/- ES cells at a level on par with G9a -/- cells. Tachibana et al. found additional similarities between G9a and GLP mutants, and found that these two proteins exist in either a homodimer or heterodimeric G9a-GLP complex in vivo and that this interaction is dependent on their SET domains. Further characterization of GLP was conducted by Ueda et al. (2006), who showed that Wiz, a nuclear protein possessing six Widely Interspaced Kruppel (C2H2)-type Zinc finger motifs that was originally identified in the mouse brain (Matsumoto, Ishii, et al, 1998), interacts with the G9a/GLP heteromer in vivo, and that this interaction is specific to G9a and GLP and requires their SET domains. Wiz was also found to interact with C-terminal Binding Protein 1 (CtBP1) and 2, which links the G9a/GLP complex to other transcriptional repression proteins that CtBPs interact with, such as Ikaros (Koipally, and Georgopoulos, 2000).

**G9a**

G9a was characterized as a HMTase by Tachibana et al. (2001), who found structural similarities between it and the Suv39h family of HMTases, leading them to further characterize G9a. From N-terminal to C-terminal, G9a possesses a polyglutamic acid stretch, a cysteine-rich region, a series of 6 ankyrin repeats, followed by a SET domain flanked by cysteine-rich regions similar to Suv39h1/2 (Figure 3).
Figure 3. Domain structure of G9a. From amino to carboxy terminus, G9a possesses a polyglutamic acid stretch (E), followed by a cysteine-rich region (Cys), a series of ankyrin repeats (Ank), a cysteine-rich pre-SET domain (Pre), the SET domain (SET), and another cysteine-rich post-SET domain (Post).

With this knowledge, Tachibana et al. (2001) showed G9a to be an H3K9 and 27 MTase whose catalytic activity was dependent on its SET domain. They also showed via immunofluorescence that G9a is excluded from centromeric heterochromatin: this contrasts Suv39h1/2 which localizes almost exclusively to centromeric heterochromatin.

Expanding on this, the same group found that G9a–deficient mouse embryos and ES cells were drastically reduced in H3K9me2, indicating that G9a is a dimethyltransferase (Tachibana, Sugimoto, et al, 2002). G9a deficient mice died at day E9.5-E12.5, and exhibited gross morphology defects and drastic increases in the number of apoptotic cells. G9a mutant ES cells had one-eighth the H3K9me2 of WT via Western blot, with the remainder mainly concentrated at heterochromatic loci and thus likely catalyzed by Suv39h1/2. These results led Tachibana and colleagues to conclude that G9a was the major H3K9me2 at euchromatic loci.

G9a interacts with a number of different proteins. One such partner is GLP; in fact, the G9a-GLP heterodimer may be the native configuration of G9a in the cell (Tachibana, Ueda, et al, 2005). G9a also interacts with Wiz, which interacts with CtBP1 and CtBP2 transcriptional co-repressors (Ueda, Tachibana, et al, 2006), thus linking G9a with other repressive machinery. More recently, two groups demonstrated that G9a also interacts with itself by automethylating its

G9a has been linked with DNA methylation in some instances. Oct-3/4 is a POU (Pit-Oct-Unc) domain homeobox gene that is expressed during gametogenesis and in early embryonic cells, and normally becomes DNA methylated upon retinoic acid (RA)-mediated differentiation (Feldman, Gerson, et al, 2006). In ES cells in which G9a has been deleted, DNA methylation does not occur at Oct-3/4 upon RA mediated differentiation, indicating that G9a acts upstream of DNA methylation at this locus. Some evidence for a connection between G9a and DNA methylation in humans is that in breast cancer lines, treatment with 5-aza-2'-deoxycytidine – an inhibitor of DNA methylation – resulted in global decreases in H3K9me2. This effect was linked to 5-aza-2'-deoxycytidine’s ability to mediate dose-dependent, post-transcriptional decreases in G9a (Wozniak, Klimecki, et al, 2007).

G9a is also important in normal gametogenesis. Tachibana et al. (2007) created mice with G9a mutated in the germline and intriguingly, this resulted in infertility and a drastic loss of germ cells in adult gonads. This was caused by significant increases in apoptotic cell death as measured by Terminal Transferase dUTP Nick End Labeling (TUNEL) analysis during spermatogenesis, and by meiotic arrest at the early pachytene stage characterized by disordered progression of synaptonemal complex formation. Although Tachibana et al. showed that improper reactivation of ERVs was not seen in this case, this type of male meiotic arrest has been associated with improper reactivation of ERVs (Bour'his, and Bestor, 2004; Carmell, Girard, et al, 2007; Hata, Kusumi, et al, 2006; Webster, O'Bryan, et al, 2005).
1.7 Statement of Thesis

Given the role that epigenetics plays in the repression of endogenous and XRVs, I developed an interest in further exploring the mechanisms behind epigenetic control of retroviruses. Such research could potentially have implications for fields as diverse as infectious diseases and gene therapy. A number of reports suggested a link between H3K9 methylation and DNA methylation in non-mammalian model organisms. In *A. thaliana*, the H3K9 MTase KYP was shown to be responsible for maintenance of CpNpG DNA methylation (Jackson, Lindroth, et al, 2002). Furthermore, in the filamentous fungi *N. crassa*, a link was shown between the histone H3 MTase DIM-5 and DNA methylation (Tamaru, and Selker, 2001).

In mammalian systems, Lehnertz *et al.* (2003) showed that Suv39h1/2 H3K9 MTases were responsible for DNA methylation at major satellite elements located in pericentric heterochromatin in mice (Lehnertz, Ueda, et al, 2003). However, the authors demonstrated that in the absence of these MTases there was little or no perturbation in DNA methylation at centromeric minor satellites or at an endogenous type C retrovirus. These data suggest that either histone modification is not required for DNA methylation at minor satellites and some C type retroviruses, or that an alternative HMTase is responsible for directing DNA methylation outside of the pericentric compartment. As G9a was previously reported to act in the euchromatic compartment, I decided to investigate the role this HMTase plays in DNA methylation and repression of retroviruses in the mouse genome.

Recent studies involving G9a and DNA methylation have examined their relationship on a global scale (Ikegami, Iwatani, et al, 2007) as well as specific loci (Epsztejn-Litman, Feldman, et al, 2008), however information pertaining specifically to retroviruses is lacking. Given the insights
already obtained regarding epigenetics and retroviruses, I hypothesized that G9a plays a role in
directing both H3K9me2 and DNA methylation activity to exogenous and endogenous proviral
elements. In order to test this hypothesis, I chose to conduct experiments utilizing murine ES
cells, a cell type in which retroviruses are efficiently silenced, and investigated the epigenetic
and expression status of representative ERVs and XRVs in the presence and absence of G9a.
2 Materials and Methods

2.1 Cell Lines and ES cell culture

ES cells (Table 3) were maintained on gelatinized plates in ES media (Dulbecco's Modified Eagle Media (DMEM) (Hyclone Cat. SH30022.01) supplemented with 15% Fetal Bovine Serum, leukemia inhibitory factor, 0.1mM nonessential amino acids, sodium pyruvate, 20mM HEPES, 0.1mM 2-mercaptoethanol, 100units/ml penicillin, and 0.5mM streptomycin) in the absence of feeder cells. Passaging conditions were as follows: Every 48-72 hours cell media was aspirated from cell monolayers and 1-5ml phosphate buffered saline (PBS) was added depending on the size of the culture plate. PBS was then aspirated and 270-800µl of trypsin (Hyclone Cat. SH30042.02) was added to disperse cells. Trypsinization was carried out in a 37°C incubator under 5% CO₂ for 4-5 minutes. Trypsinization was stopped by addition of 2-5ml of quenching media (DMEM + 15% bovine calf serum) to the plates. Cells were then pipetted repeatedly up and down using a P1000 pipette tip in order to disperse cell clumps. 1/5th-1/15th of this mix was then passaged to a pregelatinized tissue culture plate with ES media. This new plate was then rocked horizontally in order to disperse the cells evenly across the plate. After short 5-20 minute room temperature (RT) incubation, the new plate was then placed at 37°C under 5% CO₂ to grow.

If the trypsinized, quenched mixture of cells were split such that this mixture would exceed ~1/6th of the total media volume of the new plate, some additional steps were performed: the entire trypsinization mix was centrifuged at 1000rpm for 4 minutes in a Heraeus Labofuge 400 to
pellet cells. The trypsinization mixture was then decanted away from the cell pellet and the pellet was resuspended in ES media. A portion of this cell mixture was then added to the new plate.

Plates were gelatinized by the addition of 1-5mls of sterile 0.1% porcine skin gelatin in water (Sigma Cat. G2500-100G) to a tissue culture plate, followed by incubation at RT for ~10 minutes. Gelatin was then aspirated from the plate before addition of ES media.

**Table 3.** Cell lines used in this thesis.

<table>
<thead>
<tr>
<th>Line Name</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT2</td>
<td>Wildtype</td>
<td>(Yagi, Tokunaga, et al, 1993)</td>
</tr>
<tr>
<td>22-10</td>
<td>G9a mutant</td>
<td>(Tachibana, Sugimoto, et al, 2002)</td>
</tr>
<tr>
<td>2-3</td>
<td>G9a mutant</td>
<td>(Tachibana, Sugimoto, et al, 2002)</td>
</tr>
<tr>
<td>15-3</td>
<td>Transgenic rescue of G9a mutant</td>
<td>(Tachibana, Sugimoto, et al, 2002)</td>
</tr>
<tr>
<td>Wt26</td>
<td>Wildtype</td>
<td>(Lehnertz, Ueda, et al, 2003)</td>
</tr>
<tr>
<td>dM57</td>
<td>Suv39h1 and Suv39h2 double mutant</td>
<td>(Lehnertz, Ueda, et al, 2003)</td>
</tr>
<tr>
<td>36c/c</td>
<td>Dnmt1 mutant</td>
<td>(Lei, Oh, et al, 1996)</td>
</tr>
<tr>
<td>J1</td>
<td>Wildtype</td>
<td>(Okano, Bell, et al, 1999)</td>
</tr>
<tr>
<td>7ab</td>
<td>Dnmt3a and Dnmt3b double mutant</td>
<td>(Okano, Bell, et al, 1999)</td>
</tr>
<tr>
<td>Tg C1168A</td>
<td>Genetic G9a null mutant rescued with catalytically inactive transgenic G9a</td>
<td>(Tachibana, Matsumura, et al, 2008)</td>
</tr>
</tbody>
</table>

**2.2 Plasmids and Primers**

The murine stem cell virus (MSCV) retroviral vector was derived from Mo-MLV, with
enhancements for expression in ES cells, specifically with mutations that reduce repression at the RBS and NCR, as well as the addition of an Sp1 site to enhance transcription (see (Hawley, Lieu, et al, 1994)). A green fluorescent protein reporter was introduced into the cassette upstream of an internal ribosomal entry site and neomycin resistance gene (Figure 4).

![MSCV-GFP reporter construct used in this thesis.](image)

Figure 4. MSCV-GFP reporter construct used in this thesis.

Primers used in this thesis were as follows:

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAP (Forward)</td>
<td>5’-CTC CAT GTG CTC TGC TTT CC-3’</td>
</tr>
<tr>
<td>IAP (Reverse)</td>
<td>5’-CCC CGT CCC TTT TTT AGG AGA-3’</td>
</tr>
<tr>
<td>MusD (Forward)</td>
<td>5’-CCC TTC CTT CAT AAC TGG-3’</td>
</tr>
<tr>
<td>MusD (Reverse)</td>
<td>5’-TAG CAT CTC TCT GCC ATT CTT CAG G-3’</td>
</tr>
<tr>
<td>MSCV LTR (Forward)</td>
<td>5’-AAC CAT CAG ATG TTT CCA GGG TG-3’</td>
</tr>
<tr>
<td>MSCV LTR (Reverse)</td>
<td>5’-TTC GGA TGC AAA CAG CAA GAG GC-3’</td>
</tr>
</tbody>
</table>

### 2.3 Retroviral Infections

Approximately 1.0×10^6 Phoenix A cells (Orbigen Cat. RVC-10001) were grown on 6cm ungelatinized tissue culture dishes in 4ml of MEL media (DMEM, 15% bovine calf serum, 100units/ml penicillin, and 0.5mM streptomycin) for 16 hours. Media was then changed with 4ml of the same, supplemented with 25µM chloroquine. Immediately after, in a 15ml conical tube 8 µg of MSCV plasmid was added along with 250 ng VSV-G and then washed down to the bottom of the tube using 430µl sterile ddH₂O. 61µl of 2M CaCl₂ was then added and lightly
finger-vortexed. To this mix, 500µl of 2× Hepes buffered saline pH7.0 was added and immediately bubbled for ~10 seconds. This transfection mixture was then added to the Phoenix A plate in a dropwise fashion with care to evenly distribute the mixture without agitation. The plates were then placed at 37°C under 5% CO₂ for 8 hours after which the media was changed to normal MEL media.

24 hours after this, the media was changed again. Also at this time, target ES cell lines were plated at 2.5×10⁴ cells per well of a 6-well gelatinized tissue culture plate in a manner consistent with the above mentioned ES cell culture protocol. 24 hours later, the virus-containing media from the transfected Phoenix A plates was harvested, centrifuged for 5min at 2000×g to pellet debris. 24h ES cell cultures were aspirated of their media, and then viral supernatant was added in defined, varied amounts. Amounts varied from 500µl-1100µl of viral supernatant. The volume of media per well of ES cells was brought to 3ml using supplemental ES media, and 4µg/ml polybrene was added. The infected ES cells were then centrifuged for 45min at 3000rpm in a Heraeus Labofuge 400 and placed at 37°C under 5% CO₂ to grow. 24 hours later, media was changed for the infected ES cells using regular ES media. Cells were then passaged as normal. Flow cytometry assays were conducted starting at day 4 post-infection.

### 2.4 Flow Cytometry

Flow cytometry was conducted as follows: trypsinized cells were pelleted as in the cell culture protocol, and then were resuspended in 300µl-500µl PBS supplemented with 2% bovine calf serum and 1µg/ml propidium iodide. Fluorescence was then analyzed using a BD FACSDiVa flow cytometer gating for live cells first using forward/sidescatter, then propidium iodide, then finally against green fluorescent protein (GFP) expression. A minimum of 10,000 events were
collected per sample.

2.5 Realtime PCR

Quantitative, realtime polymerase chain reaction (qPCR) was conducted on a Bio-Rad Opticon 2 thermal cycler with each sample in triplicate using Fermentas HotStart Taq (Cat.No. #EP0601) using the SybrGreen chemistry. Primer sequences can be found in Table 4. Program was 5 min at 95°C initial denaturing, 30 sec at 94°C subsequent denaturing, 30 sec at 59°C annealing, 30 sec at 72°C extension, 1 sec at 80°C plate read, repeated from second step 40 times.

C(t) values were converted to nanogram amounts via a standard curve run in duplicate at the same time as the experimental samples. The standard curve itself was composed of DNA purified from the actual chromatin used in the experiment, to allow for the highest possible comparability between standard and samples. A representative standard curve is depicted below in Figure 5.

![Figure 5](image.png)

**Figure 5.** Representative standard curve for qPCR analyses. Typical point values for the curve were (from greatest to smallest): 20ng, 10ng, 2.5ng, 1.25ng, 0.625ng of DNA. The equation of
this particular curve is \( y = -0.3361x + 6.89 \) with an \( r^2 \) value of 0.991, as determined by the Opticon 3 Monitor software (BioRad). Each dot represents a replicate; replicates which appeared to deviate from the curve were excluded, and are indicated by red dots.

### 2.6 Copy Number Determination

Determination of integrated MSCV-GFP provirus was accomplished via quantitative realtime PCR using primers against the GFP region of the provirus and compared against a known single-copy GFP transgene cell line to determine the number of provirus per infection.

### 2.7 Methylation-Sensitive Digests

Methylation-sensitive digestes were carried out on infected TT2 and 22-10 lines as follows: genomic DNA was isolated, and 10µl of DNA was diluted to 300µl and split equally into two volumes. Restriction buffer and bovine serum albumin was added to both, while BssHII restriction enzyme was added to one, but an equal volume of water was added to the other. The reaction was allowed to proceed overnight after which quantitative realtime PCR was carried out using the listed MSCV primers, which flank two known BssHII restriction sites in the provirus. PCR was carried out on both the digested sample, as well as the 'mock' digested sample, and the ratio of signals from both samples was calculated to determine the degree of methylation present at the two restriction sites flanked by the primers.

### 2.8 Chromatin Immunoprecipitation (ChIP)

48-hour adherent ES cells were fixed in 1% formaldehyde for 10 minutes at RT on an orbital shaker at low speed. Glycine solution was then added to a final concentration of 0.125 M to stop
the fixation reaction and further incubated on an orbital shaker for 5 minutes at RT. The liquid
was subsequently decanted from the fixed cells and cells were washed twice in PBS. Fixed cells
were then trypsinized for 10 minutes at 37°C and scraped into 50ml polypropylene conical
centrifuge tubes. Scraped plates were subsequently washed in PBS to collect remaining cells,
and this wash was added to the tube of scraped cells. Tubes of cells were centrifuged for 10
minutes at 1000rpm in a Heraeus Megafuge 1.0R at 4°C, the PBS aspirated, and the pellet
resuspended to 50ml in fresh PBS at which point a cell count was taken. Tubes were then
centrifuged again as before and the PBS aspirated. Pellets were then snap-frozen in liquid
nitrogen and stored at -85°C.

Frozen cell pellets were thawed on ice and then incubated in 1ml per 5x10^7 cells of Swelling
Buffer (0.1 M Tris-Cl pH 7.6, 10mM KOAc, 15mM MgOAc, 1x protease inhibitors (Roche Cat.
#11873580001)) for 20 minutes on ice with occasional flicking. Cells were then fractured using
15 strokes of a Dounce homogenizer on ice, transferred to 1.5ml centrifuge tubes and centrifuged
for 5 minutes at 2,500x g to pellet. Liquid was aspirated, and cells were resuspended in 1ml per
1x10^8 cells of Lysis Buffer (50mM Tris-Cl pH 8.0, 10mM EDTA, 1% SDS, 1x protease
inhibitors (Roche Cat. #11873580001)) and lysed on ice for 10 minutes. Lysates were then
transferred to 15ml polystyrene Falcon tubes and sonicated in a Diagenode Bioruptor water bath
sonicator for 10 minutes using 'HI' sonication setting, 15 second pulses with 60 second rest
intervals. Once sonicated, chromatin was briefly centrifuged to collect liquid to the bottom and a
10 µl aliquot was taken to check fragment size. The remainder was transferred to 1.5ml
centrifuge tubes and centrifuged for 10 minutes at 13,200 rpm at 4°C to pellet debris. Chromatin
was then transferred to fresh centrifuge tubes, snap-frozen in liquid nitrogen, and stored at -85
°C. For the 10 µl aliquot, in order to determine fragment size, samples were boiled at 100°C for
15 minutes in 0.2M NaCl, and then column purified and run on a 2% agarose gel.

Prepared chromatin was first pre-cleared with Protein G agarose beads at 4 °C with rotation, then diluted 1:2 in Dilution Buffer (0.01% SDS, 1.1% Triton-X-100, 1.2mM EDTA, 16.7mM Tris-Cl pH 8.0, 167mM NaCl, 1x protease inhibitors) after which antibody was added (see Table 5 for antibodies and amounts). Samples were then incubated overnight at 4 °C with rotation.

**Table 5.** Antibodies used in this thesis.

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Amount (µg)</th>
<th>Host</th>
<th>Catalog No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve IgG</td>
<td>11</td>
<td>Rabbit</td>
<td>Sigma I8140</td>
</tr>
<tr>
<td>H3</td>
<td>1.5</td>
<td>Rabbit</td>
<td>Abcam ab1791</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>2</td>
<td>Mouse</td>
<td>Abcam ab1220</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>1.6</td>
<td>Rabbit</td>
<td>Abcam ab8898</td>
</tr>
<tr>
<td>Dnmt3a</td>
<td>2.5</td>
<td>Mouse</td>
<td>Imgenex IMG-268A</td>
</tr>
<tr>
<td>Kap-1</td>
<td>5</td>
<td>Mouse</td>
<td>Abcam ab22553</td>
</tr>
</tbody>
</table>

Next day, samples were precipitated with 30µl of Protein G agarose beads for 40 minutes at 4 °C with rotation, briefly centrifuged, and the supernatant from the ‘IgG’ sample was saved in its entirety as a measure of the total input of chromatin per sample. The supernatant for the remaining samples was aspirated away from the beads and then washed 2x in Dialysis Buffer (2mM EDTA, 50mM Tris pH8.0 for monoclonal antibodies; 2mM EDTA, 50mM Tris-Cl pH8.0, 2% Sarkosyl for polyclonal antibodies) followed by 2x washes in Wash Buffer (100mM Tris-Cl pH8.0, 500mM LiCl, 1% NP40, 1% Deoxycholic acid, 1x protease inhibitors for monoclonal antibodies; 100mM Tris-Cl pH9.0, 500mM LiCl, 1% NP40, 1% Deoxycholic acid, 1x protease inhibitors for polyclonal antibodies). Samples were then released from the beads in a total of 100µl Elution Buffer (50mM NaHCO3, 1% SDS) with shaking at RT and crosslinks were reversed overnight in the presence of 0.2M NaCl at 67°C. Samples were then column purified and subject to quantitative, realtime PCR; each sample was assayed in triplicate, with 2µl of
sample being used for each replicate.

Data for ChIP experiment bar graphs was generated via qPCR. C(t) values were converted to nanogram values by comparison with a standard curve. A sample representing a fraction of the total input chromatin per immunoprecipitation was measured, and the total input was back-calculated from this. This value was used as the denominator for each sample in the bar graphs. The numerator was calculated using the value of the immunoprecipitated fraction of each sample as determined by qPCR. This final fraction was then converted into a percentage and graphed.
3 Results

3.1 Experimental Design

An investigation into the establishment of retroviral silencing as well as the maintenance of retroviral silencing (Figure 6) required that certain elements be included in the experimental design. Firstly, due to generally poor retroviral expression in ES cells (i.e. (Seliger, Kollek, et al, 1986)), a vector that expressed well initially was necessary in order to observe silencing progression. I chose murine stem cell virus (MSCV) for this investigation, as unlike Mo-MLV from which the vector is derived, MSCV expresses well in stem cells; this is a consequence of a number of mutations that were introduced into the LTR region, which alleviate the repressive effect of cis-acting sequences (Hawley, 1994). Due to these modifications, MSCV can be used to express genes of interest in ES cells. Furthermore, since MSCV is derived from Mo-MLV, which integrates in euchromatin and shows a preference for transcription start sites (Wu, Li, et al, 2003), it is an appropriate vector for the study of G9a which acts in euchromatin.
A second consideration was measurement of retroviral expression. I selected GFP as the indicator of MSCV expression, as it can be accurately measured with minimum processing via flow cytometry. Thirdly, it was necessary to be able to differentiate infection efficiency effects from G9a-related effects; if a difference in GFP expression was observed between two lines, it could be a consequence of either different genetic backgrounds, or different numbers of integrated proviruses. In order to differentiate between these two possibilities, I infected cell lines with varying amounts of retrovirus and performed copy number tests of integrants via

Figure 6. Outline of experimental design. Endogenous IAP and MusD retroviruses and exogenous MSCV retrovirus were examined for chromatin status via chromatin immunoprecipitation, DNA methylation status via restriction-sensitive digest, and expression levels via flow cytometry in murine ES cells that were either WT or G9a null. The Western blot confirming lack of G9a protein in the G9a mutant was conducted by Sandra Lee in the Lorincz Lab.
qPCR (*Figure 7*), then selected cell lines with similar numbers of integrants for further study. Finally, in order to allow for reasonable comparisons to be made between ERVs and XRVs, I selected the LTR group of retrotransposons as the ERVs to which MSCV would be compared. LTR retrotransposons possess the same basic structure as XRVs, with *gag*, *pol*, and *env* genes flanked by direct repeats, and as such are the closest match. I selected two representative ERVs, namely the IAP and MusD retroviruses (*Figure 6*), as these ERVs are among the most potentially active members of the endogenous retroviral superfamily in mice.

*Figure 7.* Viral copy number determination of infected ES cells. Genomic DNA was isolated from infected lines and subject to real-time, quantitative PCR using primers against the GFP region of the retrovirus. Copy number was determined via comparison against an infected cell line harbouring a known single-copy GFP transgene. Error bars represent standard deviation of technical replicates. A similar determination was performed by Ruth Appanah of the Lorincz Lab on the Dnmt3a/3b mutant and its matching WT, which were matched similarly to the data presented above.
3.2 Chromatin Differences Between ERVs and XRVs

Integrated MSCV retrovirus shows reduced H3K9me2 and H3K9me3 levels in G9a-/ ES cells compared to WT TT2 ES cells. To determine whether the selected retroviruses are potential targets of G9a, I performed ChIP (outlined in Figure 8) against H3K9me2 on both ERVs and XRVs. I also performed ChIP against H3K9me3. I found that levels of H3K9me2 were reduced at exogenous MSCV retrovirus in G9a-/ cells compared to WT (Figure 9), consistent with previous Western analyses showing that G9a is responsible for the majority of H3K9me2 in euchromatin (Tachibana, Sugimoto, et al, 2002). Interestingly, there was also a reduction in H3K9me3 at MSCV provirus. As H3K9me3 is not depleted in G9a-/ cells (Rice, Briggs, et al, 2003), this data suggests that an alternate HMTase activity is affected at the MSCV provirus in the absence of G9a.

Figure 8. Schematic of ChIP. Adherent live cells are fixed with 1% formaldehyde, lysed to release raw chromatin, then sonicated to an average fragment size of 200-1000bp. Antibodies to the protein of interest are then added to the chromatin where they bind to the protein of interest. These chromatin-antibody complexes are then precipitated, purified into DNA and used for subsequent analysis by PCR.
Figure 9. Chromatin status of exogenous MSCV retrovirus. (left top) H3K9me3 levels are reduced to approximately 50% of WT levels in G9a mutant ES cells at the LTR region of integrated MSCV retrovirus. Naïve rabbit immunoglobulin G was precipitated simultaneously to determine the level of nonspecific binding of chromatin to antibody. (right top) Histone H3 occupancy levels are similar at the MSCV LTR region between WT and G9a mutant cells. (bottom) H3K9me2 levels are reduced in G9a mutants at the MSCV LTR. Error bars represent standard deviation of technical replicates.

Endogenous IAP and MusD Retroviruses have reduced H3K9me2 in G9a-/- ES cells compared to WT. In order to determine if G9a influences endogenous retroviruses, levels of H3K9me2 and me3 were also determined at endogenous IAP and MusD retroviruses. Previously, data from Tachibana et al. showed that G9a is targeted to single-copy endogenous genes such as Mage-a2 (Tachibana, Sugimoto, et al, 2002), however evidence of G9a recruitment to ERVs is lacking. Although I was unable to detect G9a recruitment via ChIP using an antibody shown to identify G9a via Western blot (see the above reference for details), both IAP and MusD ERVs displayed a reduction in H3K9me2 in G9a-/- cells (Figure 10). The degree of reduction was
approximately twofold for each of the elements tested, and was further confirmed to be statistically significant in three independent experiments (Figure 11). Intriguingly, in contrast to MSCV retrovirus, the levels of H3K9me3 were not reduced at these endogenous classes of LTR retroviruses, suggesting that exogenous elements and endogenous parasitic elements are silenced by different sets of factors. I concluded that G9a plays a role in maintaining a repressive chromatin state at both exogenous and endogenous retroviruses, however the specifics of its role differs between ERVs and XRVs.

Figure 10. Chromatin status of endogenous LTR murine retroviruses. (top) IAP chromatin shows a reduction in H3K9me2, whereas H3K9me3 and H3 levels appear relatively unchanged between WT and G9a-/- lines. Naïve IgG is run as a negative control. (bottom) MusD ERVs show a similar pattern as IAP elements. Error bars represent standard deviation of technical replicates.
Figure 11. Chromatin status of endogenous IAP and MusD LTR retroviruses. Replicate of three independent experiments. IAP and MusD H3K9me2 status was reconfirmed in triplicate independent experiments showing statistically significant reduction in G9a-/- cells (P< 0.05, by Student’s t-test).

3.3 Reduced DNA Methylation at MSCV in G9a Mutants

Integrated MSCV retrovirus show DNA methylation defects at two defined sites in G9a-/- ES cells. Previous studies in model organisms, such as the plant A.thaliana and the filamentous fungi N.crassa, showed histone lysine methylation to be a prerequisite for DNA methylation (Jackson, Lindroth, et al, 2002; Tamaru, and Selker, 2001). Additionally, Walsh et al. (1998) showed DNA methylation at endogenous IAP elements was essential to maintain their silencing (Walsh, Chaillet, and Bestor, 1998b). Given these observations, I sought to determine whether the HMTase G9a directed DNA methylation and repression of MSCV. Digestion with the methylation sensitive enzyme BssHIII was conducted on genomic DNA isolated from infected lines at day 4 and day 18 post-infection, and the level of methylation was determined by the
degree of digestion measured by qPCR, using an amplicon that spans two restriction sites within
the 5’ LTR (Figure 12, MSCV primers). Consistent with data showing a DNA methylation
defect at ERVs (Table 6), the DNA methylation status at integrated MSCV is impaired in the
absence of G9a. This deficit is present both early on (day 4, approximately fourfold reduction),
and persists to day 18, with over 10-fold reduction. This data suggests that similar to ERVs,
DNA methylation of exogenous MSCV provirus is dependent on the presence of G9a.

Figure 12. DNA methylation status of XRV. A) Schematic of methylation-sensitive digests.
Equal amounts of purified genomic DNA in restriction enzyme (RE) buffer were subject to real
and mock digests under identical conditions. The ratio of cut over mock was determined via
quantitative PCR. B) MSCV-GFP retrovirus depicting BssHII cut sites in the 5’ LTR. Cut sites
are indicated via red arrows, amplicon by blue arrows. C) Quantitative PCR results for WT and
G9a−/− lines. Bars represent amplification of BssHII cut DNA over mock/uncut DNA, and thus
the degree of DNA methylation at BssHII sites at MSCV retrovirus. Error bars represent
standard deviation of technical replicates.
### Table 6. Analysis of endogenous IAP and MLV retroelement DNA methylation status via bisulphite sequencing. Mean numbers of methylated CpGs per sequenced molecule, as well as levels of methylation relative to WT are shown. Adapted by permission from Macmillan Publishers Ltd: The EMBO Journal (Dong, Maksakova, et al, 2008), copyright 2008.

<table>
<thead>
<tr>
<th></th>
<th>IAP Mean CpG</th>
<th>MLV Mean CpG</th>
<th>IAP Relative CpG</th>
<th>MLV Relative CpG</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT2</td>
<td>15.5</td>
<td>19.8</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>G9a-/-</td>
<td>6.1</td>
<td>7.6</td>
<td>39.3</td>
<td>38.4</td>
</tr>
<tr>
<td>G9a-/- Tg</td>
<td>15.8</td>
<td>15.6</td>
<td>101.9</td>
<td>78.8</td>
</tr>
<tr>
<td>Dnmt1-/-</td>
<td>4.1</td>
<td>5.5</td>
<td>26.3</td>
<td>27.8</td>
</tr>
<tr>
<td>Dnmt3a/3b-/-</td>
<td>13.2</td>
<td>1.6</td>
<td>85.0</td>
<td>8.3</td>
</tr>
</tbody>
</table>

### 3.4 Defective Silencing of MSCV Retrovirus in G9a-/- ES Cells

_Elevated MSCV-GFP Expression in G9a-/- ES cells Relative to WT._ I was intrigued by the discovery that G9a was responsible for H3K9me2 and sometimes H3K9me3 activity at retroviruses, and acts upstream of DNA methylation. A prediction from this data is that G9a mutant ES cells may have an impaired ability to silence retrovirus. While Walsh _et al._ (1998) showed that mutation of _Dnmt1_ resulted in IAP reactivation in mouse embryos (Walsh, Chaillet, and Bestor, 1998a), the role that H3K9 MTases play in proviral silencing in ES cells has not been addressed. I sought to answer this question initially using flow cytometry to assay GFP expression in ES cells expressing the MSCV-GFP retrovirus. Similar levels of GFP expression
between WT and $G9a^{-/-}$ cells matched for retroviral copy number (Figure 13) were detected at early time points, with WT and $G9a^{-/-}$ ES cells populations being 42.9% and 46% GFP positive respectively 4 days post-infection. However by day 24 post-infection (Figure 14), a marked difference appeared in the $G9a^{-/-}$ line. While the WT line showed considerable attenuation of GFP expression (roughly 1/2 of early-infection expression), the $G9a^{-/-}$ cell population maintained essentially the same proportion GFP-positive cells, indicating an impaired ability to silence the retrovirus (progression shown in Figure 15). This result was confirmed at the RNA level, showing significantly higher GFP transcript levels in $G9a^{-/-}$ infected cells compared to WT (Table 7) (Dong, Maksakova, et al, 2008), and is consistent with previously published data in the model plant *A.thaliana* demonstrating that H3K9 methylation by KYP acts upstream of DNA methylation and retrotransposon silencing (Jackson, Lindroth, et al, 2002). This result is in contrast to data showing no silencing defect in $G9a^{-/-}$ cells at ERVs (Table 7) (Dong, Maksakova, et al, 2008), suggesting that while G9a is necessary for efficient silencing of exogenous MSCV retrovirus, additional mechanisms act to suppress ERVs.
Figure 13. Day 4 post-infection expression levels of integrated MSCV-GFP in WT and G9a-/- ES cells as measured by flow cytometry. Contour plots are of propidium iodide (PI) versus GFP protein expression of MSCV-GFP. Infected WT and G9a mutant ES cells were compared against identical but uninfected ES cells by gating to exclude approximately 99% of the cells along the GFP expression axis on the uninfected cells and then applying the same gates to the infected lines. At this timepoint, GFP expression levels are similar between the WT and G9a mutant (42.9% and 46% respectively).
Figure 14. Day 24 post-infection expression levels of integrated MSCV-GFP in WT and G9a-/- ES cells as measured by flow cytometry. As in the previous figure, contour plots of MSCV-GFP infected WT and G9a mutant ES cells were compared against identical but uninfected ES cells and gating to exclude approximately 99% of the cells along the GFP expression axis of uninfected cells and then applying the same gates to the infected lines. At this timepoint, GFP expression levels in the WT cells have dropped to approximately one-half of the G9a mutant (21.9% versus 44% respectively).
Figure 15. Expression of exogenous MSCV-GFP retrovirus in WT and G9a-/- cells over the course of the experiment. GFP expression shows delayed silencing kinetics over time in G9a-/- ES cells (top solid line) compared to WT (top dotted line) as measured by flow cytometry. Uninfected cells of the same lines were run in parallel and were gated to exclude 99% of the cells (bottom two lines).

Table 7. Relative expression status of ERVs. In G9a-/- ES cells, levels of IAP transcripts are comparable to that of WT cells as measured by RT-PCR. Mutants of the DNMTs Dnmt1 and Dnmt3a & 3b in contrast show elevated levels of transcript. Error is displayed as standard deviation. Adapted by permission from Macmillan Publishers Ltd: The EMBO Journal (Dong, Maksakova, et al, 2008), copyright 2008.

<table>
<thead>
<tr>
<th></th>
<th>+RT</th>
<th>Error</th>
<th>-RT</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>1.00</td>
<td>0.11</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>G9a-/-</td>
<td>0.45</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>G9a Tg</td>
<td>1.19</td>
<td>0.07</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Dnmt1-/-</td>
<td>5.21</td>
<td>0.06</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>H2O</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.5 Defective Silencing of MSCV in Dnmt3a/3b Mutants

Given the importance of DNA methylation in suppression of endogenous IAP elements (Walsh, Chailliet, and Bestor, 1998a), I next addressed the question as to whether G9a-/- cells displayed a similar phenotype to DNA methyltransferases mutants. Since both histone modifications and DNA methylation were perturbed in G9a-/- cells, I sought to discover the relative significance that DNA methylation plays in the silencing of MSCV retrovirus. DNA methylation is enacted by three main MTases: Dnmt3a, Dnmt3b, and Dnmt1. Unlike the IAP ERVs studied by Walsh et al. (1998), the MSCV XRV is initially completely unmethylated upon integration into its host genome. Therefore, the most likely candidates responsible for DNA methylation of MSCV proviral elements are the de novo DNMTs Dnmt3a and/or Dnmt3b. I began this investigation by examining expression of MSCV-GFP in Dnmt3a/3b double mutant cells (7ab cells) via flow cytometry just as I did for G9a-/- cells. Similar to G9a-/- cells, 7ab cells also have initially high levels of GFP expression (Figure 16), which persists throughout the course of the experiment (Figure 17, and summarized in Figure 18). In contrast, the corresponding WT line behaved similarly to the G9a WT line. In summary, despite possessing mutations in different genes, the two mutant lines – G9a and Dnmt3a/3b – produced similar phenotypes. Furthermore, given that different WT lines silence MSCV retrovirus similarly, the phenotype observed in the mutants is likely attributable to the mutation itself rather than different genomic backgrounds. Taken together with the previous data, this indicates that DNA methylation contributes a significant role in the silencing of MSCV retrovirus, and that G9a plays a role in directing this methylation.
Figure 16. Day 4 post-infection expression levels of integrated MSCV-GFP in WT and Dnmt3a/3b-/− ES cells as measured by flow cytometry. Plots of MSCV-GFP infected WT and Dnmt3a/3b-/− ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels are similar between the WT and Dnmt3a/3b double mutant (40.7% and 44.5% respectively).
Figure 17. Day 29 post-infection expression levels of integrated MSCV-GFP in WT and Dnmt3a/3b-/- ES cells as measured by flow cytometry. Plots of MSCV-GFP infected WT and Dnmt3a/3b ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels in the WT have dropped to less than one-half that of the Dnmt3a/3b double mutant (14.4% versus 36.3% respectively).
3.6 Expression of MSCV is Unchanged in Suv39h1/2

**Mutants**

Suvar39h1 and Suv39h2 are HMTases that are primarily located in constitutive heterochromatin, (although they have been found at other locations (Nielsen, Schneider, et al, 2001)), and as such may not overlap G9a target areas in euchromatin (Peters, O'Carroll, et al, 2001). To investigate whether levels of H3K9me2 in pericentric heterochromatin might explain the G9a mutant phenotype, and whether another K9 HMTase might have an effect on retroviral silencing, I infected a *Suv39h1/2* ES line with MSCV-GFP retrovirus, and monitored GFP expression. In contrast to the G9a and *Dnmt3a/3b* mutant lines, *Suv39h1/2* mutant lines did not show a defect in

---

**Figure 18.** MSCV-GFP retroviral expression in WT and *Dnmt3a/3b* mutant ES cells. Flow cytometry shows consistently higher levels of GFP expression in *Dnmt3a/3b*-/− cells (solid line) over time compared to WT (dotted line).
silencing, as initial levels of expression of the mutant (*Figure 19*) decrease with similar kinetics to WT (*Figure 20*), indicating that Suv39h1/2 do not play as significant a role in the silencing of MSCV retrovirus. The silencing progression is summarized in *Figure 21*. Furthermore, this data indicates that global levels of K9 methylation may not be the only factor involved in efficient retroviral silencing, as *Suv39h1/2* mutants do have reduced levels of H3K9 methylation (Lehnertz, Ueda, et al, 2003), and that either location of K9 methyl marks or protein-protein interactions specific to G9a may play a more significant role in silencing.
Figure 19. Day 4 post-infection expression levels of integrated MSCV-GFP in WT and Suv39h1/2-/- ES cells as measured by flow cytometry. Plots of MSCV-GFP infected WT and Suv39h1/2-/- ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels are similar between the WT and Suv39h1/2 double mutant (47.5% and 51.6% respectively).
Figure 20. Day 24 post-infection expression levels of integrated MSCV-GFP in WT and Suv39h1/2-/ ES cells as measured by flow cytometry. Plots of MSCV-GFP infected WT and Suv39h1/2-/ ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels remain similar between WT and the Suv39h1/2 double mutant (27.4% versus 35.1% respectively).
3.7 Reversibility of G9a-Mediated Silencing Defects

Reintroduction of G9a into G9a-/- Cells Rescues the Silencing Defect at MSCV Retrovirus.

There are certain risks when interpreting the phenotype of a mutant cell line: the presence of extraneous additional sequences such as antibiotic resistance genes, as well as the fact that spontaneous mutations, which could influence the experimental phenotype apart from the intended mutation, may arise during derivation. The G9a mutant line possesses a pgk-neomycin resistance cassette that replaces the glutamic acid stretch, adjacent cysteine-rich region, and ankyrin repeats (Tachibana, Sugimoto, et al, 2002); and so in order to address potential unforeseen mutations created during the generation of the G9a mutants, and to determine whether the silencing defect observed in G9a-/- cells was reversible, I examined the expression

Figure 21. MSCV-GFP infected Suv39h1/2-/- cells show similar silencing kinetics to WT. Flow cytometry data collected over time displays reductions in retroviral expression in both infected lines (top solid and dashed lines).
status of MSCV retrovirus in the G9a-/- line rescued with transgenic G9a. The results indicate that similar to WT cells, initial MSCV-GFP levels in G9a Tg cells demonstrated robust expression (Figure 22), which over time became progressively silenced (Figure 23, summarized in Figure 24). This shows that the silencing defect observed in G9a-/- cells can be rescued with the introduction of exogenous G9a, and that the defect is not a consequence of non-G9a related sequences or mutations introduced in the creation of the G9a-/- cell line.

**Figure 22.** Day 4 post-infection expression levels of integrated MSCV-GFP in WT and G9a transgenic rescue ES cells as measured by flow cytometry. Contour plots are of forward scatter (FSC) versus GFP protein. Plots of MSCV-GFP infected WT and G9a Tg ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels are similar between the WT and G9a Tg lines (38.4% and 42.6% respectively).
Figure 23. Day 18 post-infection expression levels of integrated MSCV-GFP in WT and $G9a$ transgenic rescue ES cells as measured by flow cytometry. Plots of MSCV-GFP infected WT and $G9a$ Tg ES cells compared against their uninfected counterparts show that at this timepoint, GFP expression levels are still similar between the two lines, however expression has decreased in comparison with Day 4 levels.
3.8 Impaired Dnmt3a Recruitment at ERVs in G9a Mutants

Given the defects in H3K9me2 and H3K9me3, DNA methylation status, and expression status I observed in the absence of G9a at various classes of retroviruses, I sought to clarify the mechanism(s) by which they occur. Reductions in H3K9me2 may be a direct consequence of reduced H3K9me2 catalytic activity in the absence of G9a; however mechanisms resulting in reductions in DNA methylation are less clear. DNA methylation is enacted by three main DNMTs: Dnmt3a, Dnmt3b, and Dnmt1; and DNA methylation defects could be a result of reduced expression, recruitment, or activity of any of these DNMTs. Our lab has discovered that despite the absence of G9a, expression levels of the DNMTs do not change to any great extent (Sandra Lee, in (Dong, Maksakova, et al, 2008)). To elucidate the mechanism by which the
methylation defect occurred, I performed ChIP against each of the DNMTs. The results indicate that there is an approximately fourfold reduction in Dnmt3a localization at ERVs in G9a mutants (Figure 25) once background IgG is subtracted, suggesting that altered DNA methylation at these elements may be a consequence of reduced recruitment of Dnmt3a. Unfortunately, I was not able to obtain successful immunoprecipitations against Dnmt1 or Dnmt3b despite trying at least two different antibodies against each (data not shown). Despite this, the data provides evidence of a recruitment defect in Dnmt3a in the absence of G9a, and at face value, suggests that G9a is required for efficient recruitment of Dnmt3a-mediated DNA methylation of IAP and MusD LTR retroviruses.

Figure 25. Recruitment of Dnmt3a is reduced in G9a-/- ES cells. Dnmt3a recruitment to endogenous IAP, and MusD LTR retroviruses is reduced approximately twofold in ChIP assays. H3 occupancy, however, remains similar at these elements between WT and G9a mutants. Error bars represent standard deviation of technical replicates. Adapted by permission from Macmillan Publishers Ltd: The EMBO Journal (Dong, Maksakova, et al, 2008), copyright 2008.
A G9a catalytic mutant partially rescues Dnmt3a recruitment at ERVs. In order to further define the biochemical events that led to the G9a mutant phenotype, I attempted to differentiate between effects of G9a catalytic activity and effects of G9a protein interactions. My previous observations showed a reduction in H3K9me2, but not H3K9me3 at ERVs, indicating that the DNA methylation defect coexists with significant levels of K9 methylation. This raises the question as to whether K9 methylation or the G9a protein itself is responsible for directing DNA methylation. In order to answer this question, we utilized G9a-/ ES cells that had been rescued with G9a possessing a C1168A mutation that Tachibana et al. (2008) showed to have essentially no catalytic activity (Tachibana, Matsumura, et al, 2008), and examined the chromatin status of these cells. Our data indicate that while H3K9me2 is reduced in catalytic mutants to a level comparable with the genetic null mutant, recruitment of Dnmt3a is partially rescued (Figure 26). This suggests that the presence of G9a protein, aside from its catalytic activity, plays a role in the efficient recruitment of Dnmt3a to retroviruses, and by extension, in the efficient DNA-methylation mediated silencing of these elements. This result regarding DNA methylation was confirmed by bisulphite sequencing showing that catalytically null G9a is able to partially rescue the DNA methylation defect at ERVs (Figure 27) (Dong, Maksakova, et al, 2008).
**Figure 26.** Catalytically inactive G9a partially restores Dnmt3a recruitment levels. (top) Levels of H3K9me2 between G9a/- and Tg C1168A catalytic G9a mutant are similar, showing minimal levels of H3K9me2 in the catalytic mutant. (bottom) Dnmt3a recruitment is partially restored in the Tg C1168A line (53% of Tg wt compared to G9a/- at 11% of TT2 wt). Error bars reflect standard deviation of technical replicates. Adapted by permission from Macmillan Publishers Ltd: The EMBO Journal (Dong, Maksakova, et al, 2008), copyright 2008.

**Figure 27.** Bisulphite analysis of DNA methylation status in G9a mutant lines. TT2 WT, G9a genetic mutant, G9a transgenic rescue, and G9a catalytic mutant rescue ES cells DNA methylation levels at the LTR region of IAP endogenous elements. Adapted by permission from Macmillan Publishers Ltd: The EMBO Journal (Dong, Maksakova, et al, 2008), copyright 2008.
3.9 Mechanism of G9a-Independent H3K9 Trimethylation

These results indicated that while both H3K9me2 and me3 were reduced at XRVs, only dimethylation was reduced at ERVs in the absence of G9a. Given this observation, I reasoned that an alternate H3K9 MTase that is capable of trimethylation is targeted specifically to ERVs independently of G9a recruitment. Previously, Wolf and Goff showed that Kap-1 is recruited to the RBS of Mo-MLV (Wolf, and Goff, 2007). Furthermore, Kap-1 recruits the histone H3K9 trimethyltransferase SetDB1 (Schultz, Ayyanathan, et al, 2002). Taken together, these data suggested that reduced recruitment of SetDB1 – via reduced recruitment of Kap-1 – to the RBS of MSCV was a possible cause for the reduced H3K9me3 seen at MSCV retrovirus. The primer binding site of MSCV has been changed to prime from tRNA\textsubscript{Gln} instead of the usual tRNA\textsubscript{Pro} (Hawley, 1994) which relieves RBS-mediated repression, thus reduced recruitment of Kap-1 and SetDB1 seemed reasonable as an explanation for reduced H3K9me3 at MSCV in G9a\textsubscript{-/-} cells. Under this premise, I chose to test for the presence of Kap-1 and SetDB1 at retroviruses. Although I was not able to detect enrichment of SetDB1 (data not shown), I was able to partially determine the status of Kap-1 recruitment. In contrast to what was expected, Kap-1 recruitment was reduced at ERVs in the absence of G9a at both IAP and MusD endogenous loci (Figure 28). This result was confirmed in replicate, and utilizing two different G9a mutant cell lines to control for possible confounds due to the effects of particular genomic backgrounds (Figure 29).

Unfortunately, I was unable to ascertain the recruitment status of Kap-1 for exogenous MSCV retrovirus, possibly due to difficulties arising from the fact that there are much fewer copies of MSCV in the cells as opposed to ERVs, or due to difficulties with changing antibody batches. These results were unexpected, as under the current hypothesis, reduced levels of H3K9me3 at XRVs and unchanged levels of H3K9me3 at ERVs would suggest reduced levels of SetDB1 (and
thus Kap-1) at XRVs and unchanged levels of Kap-1 at ERVs. A possible explanation for this result is that SetDB1 is being recruited to ERVs independently of Kap-1.

Figure 28. Kap-1 recruitment is reduced in the absence of G9a at endogenous IAP and MusD retroviruses. (top) Recruitment of Kap-1 is reduced over tenfold at IAP elements, however H3 occupancy remains similar. (bottom) MusD elements display a similar reduction in Kap-1 recruitment. Error bars represent standard deviation of technical replicates.
Figure 29. Kap-1 recruitment is reduced in the absence of G9a at endogenous IAP and MusD retroviruses. (top) Recruitment of Kap-1 is reduced over twofold at IAP elements in two different cell lines mutant for G9a, however H3 occupancy remains similar. (bottom) MusD elements display a similar reduction in Kap-1 recruitment. Error bars represent standard deviation of technical replicates.
4 Discussion

While a connection between histone modification, DNA methylation, and silencing of retrotransposons was shown previously in non-mammalian systems (Jackson, Lindroth, et al, 2002), evidence of such a connection in mammals is lacking. In mice H3K9 methylation, DNA methylation, and repression of direct repeat pericentromeric repetitive elements was dramatically reduced in Suv39h1/2 mutants (Lehnertz, Ueda, et al, 2003), but ERVs were unaffected. This raises the question as to whether a similar mechanism exists with ERVs in the euchromatic compartment, and/or XRVs. Since G9a is an H3K9 MTase responsible for the majority of K9 dimethylation in euchromatin (Tachibana, Sugimoto, et al, 2001), I chose to focus on this enzyme in my investigation.

In this thesis, I examined the effect of G9a on exogenous MSCV retrovirus and endogenous IAP and MusD retroviruses. Similar to our lab’s data showing that DNA methylation at ERVs is reduced in G9a-/- ES cells (Dong, Maksakova, et al, 2008), DNA methylation at exogenous MSCV was reduced in G9a-/- cells compared to WT, and this reduction persisted throughout the course of the infection. Examining the chromatin status of ERVs and XRVs, I observed a reduction in H3K9me2 at both classes of retroviruses in G9a-/- ES cells; however, I found that while H3K9me3 was reduced at exogenous elements, this was not the case with ERVs, which retained H3K9me3 levels nearly identical to WT cells in the G9a mutants. Thus, while repressive epigenetic marks of both exogenous and endogenous retroviruses are perturbed in the absence of G9a, exogenous elements are more severely perturbed.

The trend of XRVs being more greatly affected by G9a mutation than ERVs was mirrored in
their expression status. In contrast to ERVs, where silencing was not perturbed in G9a mutant ES cells (Dong, Maksakova, et al, 2008), I detected a silencing defect at MSCV retrovirus using flow cytometry. Initial levels of expression were high in WT and G9a-/− cells, however while WT cells progressively and efficiently silenced the retrovirus, the G9a mutant line showed a defect wherein GFP silencing occurred less efficiently and to a lesser degree. I observed a similar defect in ES cells mutant for the de novo DNMTs Dnmt3a and Dnmt3b. In contrast, mutants of Suv39h1/2 showed no silencing defect, revealing that the MSCV retrovirus is affected by G9a and/or DNA methylation, but not global H3K9 methylation levels or Suv39h1/2. This result is consistent with MSCV’s preference for integrating near transcription start sites (Wu, Li, et al, 2003).

Based on these results, I decided to further explore the mechanism behind the DNA methylation defect seen in G9a mutant cells. To this end, I conducted ChIP experiments and determined that there was a reduction in Dnmt3a binding in G9a mutant ES cells – a defect that was partially rescued by reintroduction of a catalytically null form of G9a. This result indicates that Dnmt3a may play a role in the DNA methylation of retroviruses, consistent with previous publications (Okano, Bell, et al, 1999), and that this role may function independently of G9a activity. Finally, I turned my attention towards elucidating the mechanism behind the difference in H3K9me3 observed in G9a mutants between XRVs and ERVs. Although I was unable to ascertain recruitment of H3K9 MTase SetDB1 to ERVs via ChIP, I was able to measure recruitment of Kap-1, a protein that recruits SetDB1. Contrary to expectations, recruitment of Kap-1 to ERVs was reduced in G9a-/− ES cells, suggesting that if SetDB1 is responsible for maintenance of the trimethyl mark at ERVs is not dependent on Kap-1 recruitment.
4.1 A G9a-DNA Methylation Link at Retroviruses

The finding that DNA methylation was significantly reduced at all surveyed elements in G9a/- cells suggests a general role of G9a in the DNA methylation of euchromatic repetitive elements in ES cells. Consistent with this, data from A. thaliana shows that when the H3K9 HMTase KYP is mutated, DNA demethylation occurs that phenocopies mutants of the DNMT CMT3, and shows loss of cytosine methylation at sites of CpNpG trinucleotides and importantly, reactivation of endogenous retrotransposon sequences (Jackson, Lindroth, et al, 2002). This data is also consistent with N. crassa mutants of H3K9 MTase DIM-5 which also showed reductions in DNA methylation (Tamaru, and Selker, 2001). Furthermore, it is consistent in mammalian systems where Tachibana et al. showed depletion of H3K9me2 at a defined locus, Mage-a2, in G9a/- cells (Tachibana, Sugimoto, et al, 2002); and with Lehnertz et al. (2003) who showed that HMTases Suv39h1/2 were linked to DNA methylation and transcriptional repression of pericentromeric repetitive elements (Lehnertz, Ueda, et al, 2003). In summary, in light of the information presented here and in prior publications, the data supports the idea that G9a is necessary for efficient DNA methylation in the euchromatic compartment.

4.2 Role of G9a in the Silencing of an XRV

What are the molecular events that could explain the impaired silencing of exogenous MSCV retrovirus that occurred in G9a/- ES cells? One possibility is that the reduction in the repressive marks of H3K9me2 and DNA methylation resulting from the G9a mutation results in increased recruitment of activating complexes and the transcription machinery. In other words, consistent with the histone code hypothesis, reduction of repressive histone marks leads to reduction of
downstream repressive effectors and subsequent relief of silencing. Furthermore, there may be antagonistic effects that act to reverse repressive marks due to transcription complexes recruited to MSCV. Coupled with the impairment of repressive complexes in G9a-/- cells, this may explain the silencing defect. Consistent with this, while HP1 was shown to recognize H3K9me2 via its chromodomain (Jacobs, and Khorasanizadeh, 2002; Nielsen, Nietlispach, et al, 2002) and recruit Dnmt1 (Smallwood, Esteve, et al, 2007) to enact DNA methylation and repression; Rea et al. (2000) showed that H3K9 methylation and serine 10 phosphorylation are antagonistic with each other, thus in the absence of H3K9me2, not only is HP1 less efficiently recruited, but also serine 10 phosphorylation (an active mark) may accumulate in its absence.

Examining the data more closely reveals intriguing patterns in the silencing kinetics between WT and G9a-/- lines. In WT ES cells, there is a progressive decrease in MSCV expression which is not present to the same extent in G9a-/- cells, indicating that the MSCV vector is still prone to silencing despite possessing a number of expression-enhancing mutations in its LTR region. Additionally, if one examines GFP expression in G9a mutants over the course of the experiment it appears that alternate, albeit less efficient silencing may take place. There is a marginal increase in expression in G9a-/- cells initially, followed by a small decrease over the remainder of the experiment. WT cells, in contrast, demonstrate significantly more efficient silencing progression to a steady-state level. This difference in silencing kinetics suggests two distinct but related defects in retroviral silencing: first, a delay in the repressive response where the reporter GFP protein is expressed well and protein levels accumulate; and second, a reduction in the effectiveness of the silencing response as shown by the much more modest rate of silencing compared to WT cells. What might explain this pattern? Under the aforementioned model of antagonism between activating and repressive complexes, if repressive events are delayed, viral
protein levels are able to increase in the short term, then as silencing occurs, expression decreases.

Despite silencing differences between WT and G9α mutant cells, neither line is able to completely silence MSCV; some GFP expression (roughly 15% of cells) is present throughout the course of the experiment. Observations from Pannell and colleagues (Pannell, Osborne, et al, 2000) suggest that silencing occurs within a relatively short time frame in ES cells, indicating that there may be a subset of integration sties within the pool of retroviral integrants that is not susceptible to silencing. Given that MSCV retrovirus prefers to integrate near the transcription start sites of genes, this persistent level of expression may be a consequence of integrants that have inserted into exceptionally permissive locations in the genome, and thus whose expression is assured via close proximity to transcriptionally active loci. This positional effect phenomenon has already been demonstrated for Mo-MLV vectors in murine fibroblasts (Hoeben, Migchielsen, et al, 1991), and given that MSCV is based on Mo-MLV, is likely to hold for this vector as well.

Taken together, these data indicate that G9α is required for efficient silencing of MSCV retrovirus, and may be required for other classes of XRVs as well. The requirement for an HMTase in efficient suppression of retroviruses is consistent with findings in A.thaliana where an HMTase is responsible for directing DNA methylation and efficient silencing at Athila and Ta3 retrotransposons (Jackson, Lindroth, et al, 2002), since in the absence of the H3K9 HMTase KYP, these transposons were reactivated.
4.3 Divergent Roles of G9a in XRV and ERV Repression

In contrast to endogenous elements, which our lab showed to remain silent in the absence of G9a (Dong, Maksakova, et al, 2008), MSCV retrovirus showed a silencing defect. What can explain this difference? One explanation is the amount of time each has been exposed to the cellular silencing machinery; given that ERVs have been exposed to silencing much longer than MSCV, they may be more strongly silenced. However, because previous evidence suggests that silencing occurs within a relatively short time frame (Pannell, Osborne, et al, 2000), this may not be the case. Furthermore, after the initial silencing events, a portion of cells demonstrate sustained expression from exogenous virus, which is stable over time (Swindle, Kim, and Klug, 2003). Thus, time is unlikely to be the only factor that plays a role in the silencing of retroviruses.

What then, can explain the difference in expression observed in MSCV compared to ERVs? An alternative explanation is that there is less functional overlap in proteins that contribute to silencing of XRVs. While a number of HMTases may contribute to silencing at ERVs, G9a may be the principal HMTase responsible for silencing at exogenous elements. Thus, a difference in expression may be a consequence of a reduction in pathway redundancy at XRVs compared to ERVs. A final explanation is that ERVs, being already DNA methylated, require only the maintenance DNMT Dnmt1 for maintenance of silencing; this is supported by data showing that IAP elements are reactivated in Dnmt1 mutants (Walsh, Chaillot, and Bestor, 1998a). In contrast to ERVs, XRVs such as MSCV are initially unmethylated and require the activity of de novo MTases; the data presented here suggests that this MTase activity is dependent on G9a-mediated recruitment of Dnmt3a.
4.4 Divergent Epigenetic Roles of G9a at XRVs and ERVs

Consistent with data published by Tachibana et al. (2002) which showed that G9a is responsible for the majority of cellular H3K9me2, this mark is reduced at both exogenous MSCV and endogenous IAP and MusD retroviruses in G9a mutants. However, while Tachibana and colleagues found near absence of the mark at the Mage-a2 promoter, the current results show only a decrease of H3K9me2 at integrated retroviruses – not an absence. A possible explanation is that different targets in the genome are regulated by different sets of histone modifying enzymes; in this case Mage-a2 would be targeted by G9a, while integrated retroviruses are targeted by G9a and other HMTases. This coordinated recruitment of histone modifying enzymes has already been observed in a C-terminal binding protein (CtBP) complex, which contained both G9a and HDAC2, among other proteins (Shi, Sawada, et al, 2003). Another explanation is that since retroviruses are typically present in multiple copies per genome, some copies may be solely dependent on G9a for H3K9me2, while others may be methylated independently of G9a; thus ChIP would report an average of these two levels of H3K9me2.

Why would different copies of a single ERV class be targeted differently? Each ERV copy has accumulated mutations, and differences in mutations may be sufficient to account for the differences in HMTase targeting. Furthermore, the nature of the sequences surrounding an ERV may also influence what factors are recruited to it; since retroviral sequences are able to affect nearby host sequences (i.e. (Morgan, Sutherland, et al, 1999)), the reverse is also plausible.

The observation that G9a influences H3K9me3 at MSCV poses an intriguing question: does G9a catalyze H3K9me3, or does G9a indirectly contribute to the H3K9me3 levels? Since no strong evidence exists in vivo supporting G9a trimethylase activity, one alternative explanation is that
H3K9me2 is a required substrate for a different HMTase that catalyzes the H3K9me3 mark, then if there is less H3K9me2, there would be correspondingly less H3K9me3. A second possibility is that G9a recruits trimethyltransferase activity via protein-protein interactions independent of histone methylation by interacting directly or indirectly with another HMTase. Possible HMTases responsible for H3K9me3 at XRVs include Suv39h1/2 (Lehnertz, Ueda, et al, 2003), SetDB1 (Wang, An, et al, 2003), or a yet-unidentified trimethyltransferase such as the candidate SetDB2. Given that Suv39h1/2 appears to be mainly confined to the pericentromeric heterochromatin and not in the euchromatic compartment where MSCV retroviruses reside (Lehnertz, Ueda, et al, 2003; Peters, O'Carroll, et al, 2001), and given that the data presented here shows that mutating Suv39h1/2 does not affect retroviral silencing, SetDB1 is the most likely candidate for this activity.

While SetDB1 may explain the H3K9me3 deficit at exogenous MSCV retrovirus, it does not explain the WT levels of H3K9me3 observed at ERVs in the absence of G9a, given that I saw reduced Kap-1 binding at these ERVs. The protein Kap-1 was recently shown to bind to the RBS of Mo-MLV (Wolf, and Goff, 2007), and is a known transcriptional corepressor, acting as a bridge between KRAB-domain zinc finger DNA binding factors and other transcriptional repressors such as the NuRD deacetylase complex, HP1, and SetDB1 (Le Douarin, Nielsen, et al, 1996; Schultz, Friedman, and Rauscher, 2001; Schultz, Ayyanathan, et al, 2002). Taken together, one can envision a model whereby Kap-1 recruits SetDB1 which enacts H3K9me3 at retroviruses, and that mutation of the RBS in MSCV impairs this recruitment thus resulting in MSCV having lower H3K9me3 than ERVs. However, the data presented in this thesis suggests that Kap-1 recruitment is impaired at ERVs in G9a-/- cells, which is inconsistent with the aforementioned model, as one would expect that the maintenance of H3K9me3 at endogenous
elements in the absence of G9a would also be accompanied by maintenance of Kap-1 recruitment. One possible explanation is that SetDB1 is recruited similarly to ERVs and XRVs; however its recruitment is dependent on G9a and not Kap-1. Another possible explanation is that an alternate trimethyltransferase other than SetDB1 is recruited to ERVs.

Another factor to take into consideration, when asking why MSCV and ERV chromatin packaging differs in the absence of G9a, is the nature of the chromatin in which each is found. Mo-MLV, and by extension MSCV, integrates near transcription start sites of actively transcribed genes (Wu, Li, et al, 2003). The chromatin ‘neighborhood’ at these sites would be euchromatic with an open, permissive structure and marked with active histone marks. In contrast, the chromatin that ERVs reside in may not be associated with transcriptional activity and adopted a repressive epigenetic state. Under this hypothesis, the differences observed between exogenous and endogenous retroviruses are a result of being influenced, and adopting the state of the surrounding chromatin in which they find themselves. This is supported by the fact that at least one class of endogenous LTR retroviruses (the MaLR superfamily) is underrepresented at gene promoter regions (Tomilin, 2008).

A final explanation for the observed differences between exogenous and endogenous retroviruses is the possibility that certain silencing mechanisms are active only at specific developmental stages. For example, XRVs tend to be silenced in ES cells (as demonstrated in this thesis), but express well in differentiated cells; however if an ES cell is induced to differentiate, any silent XRVs remain silent (Niwa, Yokota, et al, 1983). These observations suggest that certain silencing mechanisms are active at certain phases of development, and that their effects are heritable even after the mechanism is no longer active. ERVs have gone through all stages of the murine life cycle, whereas under our experimental scheme, the XRV was introduced in ES cells.
Since ES cells are isolated from the inner cell mass of a preimplantation blastocyst (Evans, and Kaufman, 1981), by the stem cell stage ERVs have gone through germ cell development through to growth of the embryo. In other words, in comparison to XRVs, ERVs were present when alternative developmentally-regulated silencing mechanisms may have been active, and even though it is no longer active in ES cells, the silencing is maintained.

If developmentally-related silencing mechanisms exist, what specific mechanism could account for the silencing observed at ERVs but not XRVs? One possibility includes a recently-discovered class of small interfering RNAs called piwi-associated RNAs (piRNAs). piRNAs are small interfering RNAs that associate with a subfamily of argonaute proteins that consists of MIWI, MILI/PIWIL2, and MIWI2/PIWIL4. These small RNAs play an important role in spermatogenesis (reviewed in (Kim, 2006)). piRNAs have been isolated from testes (Aravin, Gaidatzis, et al, 2006; Girard, Sachidanandam, et al, 2006; Grivna, Beyret, et al, 2006; Watanabe, Takeda, et al, 2006) and also oocytes (Tam, Aravin, et al, 2008; Watanabe, Totoki, et al, 2008). In general, small interfering RNAs play an important role in repression of retroviruses (reviewed in (Girard, and Hannon, 2008)), and in particular in MILI mutant oocytes levels of IAP transcripts were increased ~3.5 fold (Watanabe, Totoki, et al, 2008), thus implicating piRNAs in the suppression of ERVs in the germline. Thus, piRNAs represent a specific, developmentally timed silencing mechanism that influences ERVs and may help account for the differences observed between XRVs and ERVs.

4.5 Biochemical Basis of the G9a-DNA Methylation Link

ChIP results indicated that G9a recruited Dnmt3a independent of its catalytic activity suggesting that G9a promotes de novo DNA methylation via protein-protein interactions. There are two
likely candidate domains that mediate this interaction: the G9a SET domain and the ankyrin domain (Collins, Northrop, et al, 2008; Tachibana, Sugimoto, et al, 2001; Tachibana, Ueda, et al, 2005). The SET domain is necessary and sufficient for G9a binding to GLP, and the ankyrin domain is a common motif for mediating protein-protein interactions. Recently, an interaction between G9a’s ankyrin domain with the de novo DNMTs was shown by Epsztejn-Litman et al. (2008); this study demonstrated G9a catalytic activity-independent DNA methylation at the Oct3/4 promoter, consistent with our results. Furthermore, Epsztejn-Litman et al. showed via deletion mapping, that recruitment of Dnmt3a/3b by G9a requires the ankyrin domain. Further research is needed, however, to fully characterize and confirm the binding partners of each of these domains in G9a. Interestingly, Dnmt3a was reported to interact directly with another H3K9 HMTase, namely SetDB1 (Li, Rauch, et al, 2006). These two interact via the N terminal domain of SetDB1 and the plant homeodomain of Dnmt3a, thus providing another example of association between an HMTase and a DNMT, and inviting the question of whether a G9a-Dnmt3a-SetDB1 complex might be responsible for H3K9me3 at ERVs. A link between a DNMT and G9a is also supported by previous data published by Esteve at al. (2006), which demonstrated a direct interaction between the maintenance DNMT Dnmt1 and G9a. Furthermore, Lei et al. (1996) showed that even in the absence of Dnmt1 ES cells can still DNA methylate an XRV, indicating that other DNMTs are recruited to retroviruses and raising the possibility that G9a could interact with the de novo DNMTs Dnmt3a and/or Dnmt3b. Taken together, these data suggest that G9a recruits Dnmt3a to ERVs possibly through a direct protein-protein interaction that does not depend on its catalytic function.

While catalytically null G9a was able to recruit Dnmt3a, the level of recruitment was lower than that observed with catalytically active G9a. This suggests that there also exists an aspect of
Dnmt3a recruitment that is dependent on G9a catalytic activity. There is no current evidence that Dnmt3a itself is able to recognize and bind dimethyl H3K9, however there are a number of proteins which do, most notably chromodomain-containing proteins such as HP1 (Jacobs, and Khorasanizadeh, 2002), as well as plant homeodomain proteins (Karagianni, Amazit, et al, 2008); any of these methyllysine binding proteins could contribute to recruitment of Dnmt3a. In addition, G9a itself is a recognition protein for dimethyl lysine: Collins et al. showed that the ankyrin domain to be a dimethyl lysine binding module (Collins, Northrop, et al, 2008) as well as the SET domain which enacts the dimethyl mark itself. It may be that by removing G9a catalytic activity, recruitment of repressive partners by G9a recognition of H3K9me2 is also reduced. Taken together, these data suggest that there also exists an H3K9me2-dependent component of Dnmt3a recruitment.

4.6 Applications and Future Directions

The mechanisms which control retroviral expression and silencing are of interest to many facets of biology and medicine. Researchers of disease processes, such as HIV, may be interested in knowing the mechanisms whereby retroviral silencing occurs, and the interplay between silencing pathways. The knowledge that G9a plays a role in not only laying down H3K9me2 at retroviruses but also influences DNA methylation, provides insight into and could potentially facilitate treatment of retrovirally-related diseases. For example, in many cancers there is global hypomethylation, particularly in intergenic regions wherein lie the bulk of ERVs, and this is accompanied by hypermethylation of selected genetic loci (reviewed in (Wilson, Power, and Molloy, 2007)). The hypomethylation may be linked to transposon reactivation and genomic instability related to actively transposing elements. This altered methylation pattern is
suggestive of altered targeting of the DNMTs in cancer cells. The data presented in this thesis suggests that G9a targets DNMT activity to retroviruses – in particular via recruitment of Dnmt3a. Thus, researchers concerned with restoring DNA methylation patterns in affected cells may find it productive to examine G9a and its expression status to determine if the altered DNA methylation patterns that accompany many cancers are in part a function of altered G9a expression.

Gene therapy is a method of treating disease via the insertion of a gene whose product is intended to correct or compensate for the mechanism underlying the disease process. Gene therapy often involves the use of retroviruses to deliver the therapeutic payloads into individuals. A major obstacle to effective gene therapy has been silencing of the introduced therapeutic gene. Here, I show that the HMTase G9a plays an important role in silencing an exogenous MSCV retrovirus. A possible solution to the silencing difficulty observed in gene therapy applications could involve inhibiting G9a expression, thereby relieving repression of the exogenous gene therapy construct. This would have the additional benefit of not affecting repression of ERVs as well, as expression of ERVs was not observed in the G9a deficient ES line utilized here (Dong, Maksakova, et al, 2008).

In closing, I would like to highlight some lines of inquiry in which research could be directed. Firstly, while I was able to detect Kap-1 enrichment at ERVs, I was unable to do so at MSCV. While this may reflect the true biological state of the cell, some technical difficulties with chromatin immunoprecipitation at the time may make a reexamination worthwhile. Secondly, in order to more directly determine the mechanisms involved in retroviral silencing, instead of measuring H3K9 methylation, examination of the G9a protein as well as the known DNMTs at both MSCV and ERVs via ChIP would more clearly explain how G9a recruits DNMT activity to
retroviruses. ChIP on candidate H3K9 HMTases would yield insight as to why the trimethyl mark is reduced at MSCV, but not at endogenous loci. Thirdly, this data opens possibilities in medicine: targeted knockdown of G9a at different stages of development may affect the stable expression levels of retrovirally-introduced transgenes. One could temporarily knockdown G9a in ES cells carrying a virally-introduced transgene and then immediately differentiate those cells: when compared to infected ES cells that did not have G9a knockdown prior to differentiation, is there a difference in transgene expression in the differentiated cells? If so, it would provide potential solutions to unwanted silencing in gene therapy: if silencing could be inhibited at the stem cell stage, it would allow stable transgene expression upon differentiation and reintroduction into the patient. Lastly, given that G9a was responsible for silencing of MSCV retrovirus in mice, it may be fruitful to explore its role in the silencing of retroviruses relevant in human disease. For example, what role might G9a play in the silencing of HIV in infected cells? As the field of epigenetics expands, it is my hope that insights and applications of this knowledge will work to the benefit of humankind.
Reference List


fibroblast cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. 2, 904-912.


in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. Genes Dev 14, 1779-91.


