WHO’S THE BOSS?
AN INVESTIGATION INTO THE COMPLEX RELATIONSHIP BETWEEN ENDOGENOUS RETROVIRUSES AND NEARBY GENES

by

KATHARINE ELIZABETH MICELI

B.Sc., York College of Pennsylvania, 2008

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)

October 2011

© Katharine Elizabeth Miceli, 2011
Abstract

Nearly half of the human and mouse genomes is composed of transposable elements (TEs) which are pieces of DNA capable of or once capable of moving to different locations in the genome. These mobile elements can potentially impact host gene expression in a variety of ways either directly through insertional mutagenesis or indirectly by serving as an alternative promoter/enhancer or through modification of splice signals. In order to protect the integrity of the genome, various mechanisms have evolved to silence TEs. Chromatin modifications, namely DNA methylation and histone modifications, represent two methods of transcriptional silencing of transposable elements. In previous studies it has been shown that methylation of short interspersed nucleotide element (SINE) TEs in the plant and mouse genome can “spread” into flanking sequences.

I chose to investigate this “spreading” phenomenon with respect to endogenous retroviruses (ERVs) which are highly active in the mouse genome. Direct evidence has been provided for ERVs in the mouse genome inducing the spread of repressive chromatin. However, the ability of ERV-induced “spreading” to impact nearby gene expression has not yet been characterized. In my thesis, I provide evidence for repressive chromatin spreading as an infrequent occurrence with the exception of the B3gal1 gene where DNA methylation and repressive histone marks spread from a solitary ERV long terminal repeat (LTR) into the CpG island promoter of the gene, which correlates with reduced gene transcription. In most cases, however, I show that CpG islands are unmethylated despite the presence of a nearby methylated ERV. Furthermore, I show examples of differential epigenetic marking of insertions with the 3’LTR of the insertion located closest to the nearby gene exhibiting hypomethylation. In one case in particular, I show an interesting trend between expression of the CdgAP gene and hypomethylation of the nearby 3’LTR.

These results lead to the conclusion that susceptibility to or protection from “spreading” of repressive chromatin marks is a locus-specific event that may be mediated by various factors such as the ERV-gene distance, the presence of a CpG island gene promoter, the chromatin state of the gene promoter, and the expression of the corresponding gene.
Preface

Some of the data in Chapter 3 (3.1) is in press. Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., Lorincz, M.C., and Mager, D.L. (2011). Variation in retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genetics, in press. Bisulfite sequencing in F1 hybrid ESCs (Figure 3.3) and tissues (Figure 3.6) was completed by L. Gagnier. Bisulfite sequencing (Figure 3.3) and chromatin immunoprecipitation (Figure 3.4) in J1 and TT2 ESCs was completed by R. Rebollo. Allelic gene expression analysis in F1 hybrid ESCs (Figure 3.5) and tissues (Data not shown) was completed by L. Gagnier. Quantitative reverse transcriptase PCR (qRT-PCR) for gene expression in J1 and TT2 ESCs (Figure 3.5) was completed R. Rebollo. Knockdown experiments in J1 and TT2 ESCs (pertaining to figures 3.7 and 3.8) were performed by Irina Maksakova of Matt Lorincz’s lab at the University of British Columbia. J1 and TT2 ESCs knockdown samples were obtained from Irina as cell pellets. I completed the qRT-PCR and methylated DNA immunoprecipitation (meDIP) experiments to analyze knockdown efficiency and measure B3galt1 expression (Figures 3.7 and 3.8).

Chapter 3 (3.2) is based on work completed by me and Liane Gagnier. I completed all of the bisulfite sequencing and allelic expression analysis with respect to Hus1 (Figure 3.9 and Table 3.1). I completed all of the bisulfite sequencing for Pnpt1 except for data corresponding to the 3’LTR in heart tissue which was completed by Liane Gagnier (Table 3.2). Liane Gagnier completed all of the bisulfite sequencing and COBRA for Mthfd2l and Atxn1l (Tables 3.3 and 3.4). For Catsper3, I completed all of the bisulfite sequencing (Table 3.5) and Liane Gagnier completed qRT-PCR for expression (Data not shown).

Chapter 4 is based on work completed by me and Liane Gagnier. In section 4.1, Liane Gagnier completed the bisulfite sequencing, COBRA, and allelic expression analysis with respect to Cyb5r1 (Figure 4.1 and Table 4.1). In section 4.2, bisulfite sequencing (Table 4.2) and qRT-PCR (Figure 4.3) for Lair1 was completed by Liane Gagnier. In section 4.3, I completed all experiments except for qRT-PCR for CdGAP expression in B6 tissues which was completed by Liane Gagnier (Figure 4.11 and Figure 4.14).

This research was approved by the UBC Research Ethics Board. The certificate number is the following: A09-0372.
# Table of Contents

Abstract .......................................................................................................................... ii
Preface ............................................................................................................................ iii
Table of contents ........................................................................................................... iv
List of tables ................................................................................................................... vii
List of figures ................................................................................................................ viii
List of abbreviations .................................................................................................... x
Acknowledgements ...................................................................................................... xii
Dedication .................................................................................................................... xiii

1 Introduction ................................................................................................................. 1
  1.1 Overview of transposable elements ....................................................................... 1
  1.2 Overview of endogenous retroviruses .................................................................. 2
    1.2.1 ERV activity and structure in the human and mouse genome ....................... 2
    1.2.2 ERV classes in the mouse genome .................................................................. 3
    1.2.3 Impact of ERVs in the mouse genome ............................................................... 4
  1.3 Epigenetic repression of ERVs ......................................................................... 4
    1.3.1 Overview of DNA methylation ....................................................................... 4
      1.3.1.1 DNA methyltransferases ......................................................................... 5
      1.3.1.2 DNA methylation dynamics during development ......................................... 6
      1.3.1.3 DNA hydroxymethylation as a novel epigenetic mark ................................. 7
      1.3.1.4 DNA methylation and gene expression ....................................................... 7
    1.3.2 Overview of histone modifications ................................................................. 8
      1.3.2.1 Histone modifications and gene expression .................................................. 9
    1.3.3 ERV silencing by DNA methylation and histone modifications ....................... 10
    1.3.4 Relationship between DNA methylation and histone modifications ............... 12
  1.4 Spreading of epigenetic silencing from ERVs ......................................................... 13
    1.4.1 Previous evidence of spreading ...................................................................... 13
    1.4.2 Genomic defenses against spreading of epigenetic marks .............................. 14
  1.5 Thesis objectives .................................................................................................... 16

2 Materials and methods .............................................................................................. 20
  2.1 ES cell culture ...................................................................................................... 20
  2.2 DNMT1/SETDB1 siRNA knockdown experiments .............................................. 20
2.3 DNA/RNA extraction from cells/tissues ................................................................. 20
2.4 Reverse transcription reactions for cDNA ............................................................. 20
2.5 Allelic quantification of gene expression ............................................................... 20
2.6 Quantitative RT-PCR/PCR ................................................................................... 21
2.7 Bisulfite conversion/PCR/Sequencing .................................................................. 21
2.8 COBRA ................................................................................................................. 22
2.9 MeDIP ..................................................................................................................... 22
2.10 ChIP-Whole tissues ............................................................................................. 23
2.11 ChIP-ESCs .......................................................................................................... 25
2.12 Data sources ......................................................................................................... 26
2.13 Primers ................................................................................................................. 26

3 Results-The impact of ERVs on nearby genes is a locus-specific event .................. 36

3.1 B3galtl ....................................................................................................................... 36
   3.1.1 Function and expression of B3galtl ................................................................. 36
   3.1.2 Spreading of repressive marks from upstream LTR in ESCs ......................... 37
   3.1.3 Impact of repressive mark spreading on B3galtl expression in ESCs ............... 38
   3.1.4 Lack of repressive mark spreading from LTR in tissues .................................. 38
   3.1.5 B3galtl expression not significantly rescued upon DNMT1 and SETDB1 knockdown in ESCs ................................................................. 39

3.2 Overview of analyzed genes with nearby ERV insertions which are not affected by “spreading” ............................................................................................................. 40
   3.2.1 Hus1 .................................................................................................................. 40
   3.2.2 Pnpt1 ................................................................................................................. 41
   3.2.3 Mthfd2l ............................................................................................................. 42
   3.2.4 Atxn1l ................................................................................................................. 43
   3.2.5 Catsper3 ............................................................................................................. 43

4 Results-Cases of ERV insertions which exhibit differential methylation .................... 59

4.1 Cyb5r1 ..................................................................................................................... 59
4.2 Lair1 .......................................................................................................................... 60
4.3 CdGAP ....................................................................................................................... 61
   4.3.1 CdGAP function and expression ................................................................. 61
   4.3.2 CdGAP CpG island promoter is unmethylated in ESCs and tissues ............... 61
4.3.3 Differential epigenetic modifications of IAP LTRs in ESCs and tissues ........................................... 61
4.3.4 Intervening region is differentially methylated in ESCs and tissues ................................................. 62
4.3.5 CdGAP is differentially expressed in ESCs and tissues ........................................................................ 63
4.3.6 Presence of IAP insertion does not significantly impact CdGAP expression ........................................ 64
4.3.7 Relationship between CdGAP expression and 3’LTR methylation ...................................................... 64
4.3.8 CdGAP CpG island promoter and upstream region are associated with different chromatin environments in ESCs and tissues .................................................................................. 65
4.4 IAP 3’LTRs not generally subject to hypomethylation ............................................................................ 66
4.5 CpG island structure not significantly different between case genes .................................................... 67
5 Discussion/Conclusions ............................................................................................................................ 86
5.1 DNA methylation/repressive chromatin spreading from ERVs as a dynamic and locus-specific event in the B3galtil case .............................................................................................................. 86
5.2 Role of CpG islands and CpG island shores in maintaining an active chromatin state ......................... 87
5.3 The impact of high gene expression on the local chromatin environment ............................................... 88
5.4 A trade-off: ERV-mediated gene silencing vs. Gene-mediated ERV hypomethylation ........................ 88
5.5 Future directions ....................................................................................................................................... 89
5.6 Epigenetics and cancer .......................................................................................................................... 90
5.7 Significance of work ............................................................................................................................... 90
References ...................................................................................................................................................... 91
List of tables

Table 2.1  Age or source of mouse cells and tissues used for experiments ........................................... 27
Table 2.2  Enzymes used for COBRA ...................................................................................................... 28
Table 2.3  Antibodies used for MedIP .................................................................................................... 28
Table 2.4  Antibodies used for tissue ChIP and ESCs ChIP ................................................................. 28
Table 2.5  Sources used for data analysis and publicly available data.................................................... 29
Table 2.6  Primers used for various experiments..................................................................................... 30
Table 3.1  Average % methylation of Hus1 CpG island promoter, region upstream of promoter, and LTR in F1 hybrid cells and tissues ................................................................. 54
Table 3.2  Average % methylation of Pnpt1 CpG island promoter and 5’ and 3’ LTR of upstream IAP in cells and tissues............................................................................................................. 55
Table 3.3  Average % methylation of Mthfd2l CpG island promoter and 5’ and 3’ LTR of upstream ETn in cells and tissues ............................................................................................................ 56
Table 3.4  Average % methylation of Atxn1l CpG island promoter, region between CpG island and insertion, and 5’ and 3’ LTR of IAP in tissues ........................................................................... 57
Table 3.5  Average % methylation of region upstream of Catsper3 TSS and 5’ and 3’ LTR of IAP ....... 58
Table 4.1  Average % methylation of Cyb5r1 CpG island promoter, upstream region, and 5’ and 3’ LTR of ETn insertion................................................................................................................. 82
Table 4.2  Average % methylation of 5’ and 3’ LTR of ETn located in first intron of Lair1 ................. 83
Table 4.3  Average % methylation of 5’ and 3’ LTR of IAPs far away from genes......................... 84
Table 4.4  CpG island promoter structure of relevant case genes....................................................... 85
List of figures

Figure 1.1 Common affects of ERV insertions on host genes .......................................................... 18
Figure 1.2 Example of relationship between DNA methylation, transcription factor binding, and gene expression .................................................................................................................................. 18
Figure 1.3 Structure of nucleosome, histone core, and examples of post-translational modifications of amino tails ...................................................................................................................................... 19
Figure 1.4 Example of ERV-mediated gene silencing by spreading of DNA methylation ............ 19
Figure 3.1 Cases of genes with nearby ERV insertions and status of ERV in four different mouse strains discussed in chapter 3 ........................................................................................................................................... 45
Figure 3.2 Cases of genes with nearby ERV insertions and status of ERV in four different mouse strains discussed in chapter 4 ........................................................................................................................................... 46
Figure 3.3 DNA methylation spreading is observed from an upstream antisense solitary LTR into the B3galtl CpG island promoter in F1 hybrid ESCs and J1 ESCs ........................................................................................................ 47
Figure 3.4 Spreading of repressive chromatin marks is observed from an upstream antisense solitary LTR into the B3galtl CpG island promoter in J1 ESCs ........................................................................................................ 48
Figure 3.5 Presence of upstream LTR impacts B3galtl expression in F1 hybrid and J1 ESCs ........ 49
Figure 3.6 Spreading of DNA methylation from LTR into B3galtl CpG island promoter is not observed in tissues .................................................................................................................................................. 50
Figure 3.7 Knockdown of DNMT1 and SETDB1 in J1 ESCs does not significantly rescue B3galtl expression .................................................................................................................................................. 51
Figure 3.8 Residual DNA methylation may remain following DNMT1 and SETDB1 knockdown in J1 ESCs .................................................................................................................................................. 52
Figure 3.9 Upstream LTR does not negatively impact Hus1 expression as determined by allelic quantification .................................................................................................................................................. 53
Figure 4.1 Allelic quantification results show higher Cyb5r1 expression from LTR-containing allele in embryo .................................................................................................................................................. 53
Figure 4.2 RNA seq data generated by Lorincz lab and predicted transcripts in J1 and TT2 mouse ESCs .................................................................................................................................................. 68
Figure 4.3 Relative Lair1 expression measured by qRT-PCR in various B6 tissues ...................... 69
Figure 4.4 CpG island promoter of CdGAP is unmethylated despite presence of upstream insertion in ESCs .................................................................................................................................................. 70
Figure 4.5  CpG island promoter of CdGAP is unmethylated despite presence of upstream insertion in tissues .......................................................................................................................................................... 71

Figure 4.6  5’ and 3’ LTR of full-length IAP upstream of CdGAP CpG island promoter is methylated in ESCs ........................................................................................................................................................................ 72

Figure 4.7  5’ and 3’ LTR of full-length IAP upstream of CdGAP CpG island promoter is differentially methylated in tissues ........................................................................................................................................................................ 73

Figure 4.8  5’ and 3’ LTR of full-length IAP upstream of CdGAP is marked by different histone modifications in ESCs and tissues ........................................................................................................................................................................ 74

Figure 4.9  Intervening region is differentially methylated in ESCs with and without the upstream insertion ........................................................................................................................................................................ 75

Figure 4.10  Intervening region is unmethylated in tissues regardless of upstream insertion........ 76

Figure 4.11  CdGAP expression is reduced in B6 ESCs ........................................................................................................................................................................ 77

Figure 4.12  CpG island promoter of CdGAP is enriched for active and repressive marks in ESCs consistent with a bivalent promoter ........................................................................................................................................................................ 77

Figure 4.13  Full-length IAP insertion does not significantly impact CdGAP expression in tissues...... 78

Figure 4.14  Correlation observed between CdGAP expression and methylation of the upstream 3’LTR of the IAP ........................................................................................................................................................................ 79

Figure 4.15  IAP insertion affects nearby chromatin environment in ESCs ........................................ 80

Figure 4.16  CdGAP expression in tissues is associated with accumulation of active histone marks and loss of repressive marks ........................................................................................................................................................................ 81
List of abbreviations

TEs-transposable elements
SINEs-short interspersed nucleotide elements
SVA-SINE/VNTR/Alu
LINEs-long interspersed nucleotide elements
LTR-long terminal repeat
ERVs-endogenous retroviruses
IAP-intracisternal A particle
ETn-early transposon
DNMT-DNA methyltransferase
DNMT1-DNA methyltransferase 1
DNMT3a-DNA methyltransferase 3a
DNMT3b-DNA methyltransferase 3b
Bps-base pairs
ESCs-embryonic stem cells
ICF-immunodeficiency-centromeric instability-facial anomalies
DNMT1o-DNA methyltransferase 1-oocyte specific isoform
5-hmC-DNA hydroxymethylation
5-mC-DNA methylation
CTCF-CCCTC binding factor
MBPs-methyl CpG binding proteins
TDMR-tissue specific differentially methylated region
HATs-histone acetyltransferases
HDACs-histone deacetyltransferases
Sir-silent information regulator
HP1-heterochromatin protein 1
HKMTs-histone lysine methyltransferases
PRC1-polycomb repressive complex 1
PRC2-polycomb repressive complex 2
MLV-murine leukemia virus
LSH1-lymphoid specific helicase
NP95-nuclear protein of 95 kda
Kbs-kilobases
Chip-seq-chromatin immunoprecipitation-sequencing
RT-reverse transcription
SNPs-single nucleotide polymorphisms
qRT-PCR-quantitative reverse transcription polymerase chain reaction
COBRA-combined bisulfite restriction analysis
MedIP-methylated DNA immunoprecipitation
ChIP-chromatin immunoprecipitation
TSS-transcriptional start site
NPCs-neural precursor cells
MEFs-mouse embryonic fibroblasts
KD-knockdown
Acknowledgements

First, and foremost, I would like to acknowledge my fantastic supervisor, Dixie Mager, for taking me as a graduate student in her lab despite my limited research background. It has truly been a wonderful learning experience. Dixie has been continuously involved and encouraging from the first day I arrived in Vancouver whether it be asking how my classes are going, discussing data, or checking up on me while completing my thesis draft.

I would like to thank my top-notch committee members, Matt Lorincz and Carolyn Brown. Their advice and critiques have helped me to grow as a graduate student and has been vital in keeping my project on the right course.

I would also like to thank our wonderful technician Liane Gagnier who never fails to be helpful in all aspects of research from overcoming technical difficulties with an experiment or locating a reagent in the freezer.

I would like to acknowledge all of the past and present members of the Mager lab who have made the lab a fun place to work. I have learned so much from all of you. A special thanks to Rita who has been tremendously helpful with experiments, and whose excitement about science is contagious.

To my family, who have always encouraged my pursuit of education, even if it meant moving 3,000 miles away.

And last but not least, my best friend and husband, Jesse. I would not have been able to do this without your love and support.
Dedication

To two very special women, Nana & Aunt Mil, who both passed away during the course of my graduate studies. I am truly lucky to have had you both in my life. I miss and love you both.
1 Introduction

1.1 Overview of Transposable Elements (TEs)

Barbara McClintock’s initial discovery of “jumping genes” in the maize genome has lead to the identification of what are now referred to as transposable elements (TEs), present in nearly all organisms including bacteria, plants, and mammals (Goodier and Kazazian, 2008; Gogvadze and Buzdin, 2009). Modern day sequencing technologies have allowed the contribution of TEs to the genome to be estimated. It is well-known that these mobile elements make up a significant portion of host genetic material; for example, close to 50% of both the human and mouse genomes are composed of transposable elements or remnants of such sequences (Lander et al., 2001; Waterston et al., 2002).

The description of TEs as “jumping genes” is due to the ability of these pieces of DNA to move to different locations in the host genome. Mammalian TEs can be classified into two different groups based on the mechanism by which they transpose or move to different locations in the genome. DNA transposons belong to a group of transposable elements known as “class II elements” that move to different locations in the genome by a “cut” and “paste” mechanism. This process involves excision of the element and re-insertion into a different genomic location. Roughly 3% of the human genome is composed of DNA transposons; however none are known to be currently active. A more significant portion of the mammalian genome is composed of “class I elements” also known as retrotransposons which move in the genome by a “copy” and “paste” mechanism. This mechanism involves transcription of the element to generate RNA which is reverse transcribed into a complementary DNA copy that can insert in a new location in the host genome thereby increasing its copy number (Goodier and Kazazian, 2008; Gogvadze and Buzdin, 2009). Retrotransposons can be further divided into the following two categories: autonomous elements which encode the proteins necessary for transposition and nonautonomous elements which require other elements for their mobility. Short interspersed nuclear elements (SINEs) and SINE/VNTR/Alu (SVA) elements are examples of nonautonomous elements. Autonomous elements include long interspersed nuclear elements (LINEs) and long terminal repeat (LTR) retrotransposons including endogenous retroviruses (ERVs) (Goodier and Kazazian, 2008; Gogvadze and Buzdin, 2009). ERVs, specifically in the mouse genome, will be the focus of this thesis.
1.2 Overview of Endogenous Retroviruses (ERVs)

1.2.1 ERV activity and structure in the human and mouse genome

ERVs or LTR retrotransposons make up approximately 10% of the mouse and human genomes and are hypothesized to have originated as ancient remnants of retroviral infections in the germline during evolution (Maksakova et al., 2006; Gogvadze and Buzdin, 2009). A retrovirus becomes “endogenized” following infection of a germ cell or totipotent cell in early development which results in inheritance of the provirus insertion from one generation to the next (Jern and Coffin, 2008). Most ERVs are described as “non-infectious” meaning that transposition events are limited to within a single cell (Boeke and Stoye, 1997). Indeed, most families of ERVs and related sequences have amplified in the genome by rounds of intracellular retrotransposition, as opposed to re-infections (Boeke and Stoye, 1997; Maksakova et al., 2006; Jern and Coffin, 2008). While the genomic contribution of ERVs is quite similar between mice and humans, the activity of these mobile elements in the two species is remarkably different. Most ERVs in the human genome are old and incapable of transposition due to extensive deletions; thus it is not surprising that new ERV insertions causing mutations or disease have not been identified (Maksakova et al., 2006; Gogvadze and Buzdin, 2009). Strikingly, in the mouse genome, approximately 10% of all germ-line spontaneous mutations are due to ERV insertions. Therefore, some ERV families are extremely active in the mouse genome (Maksakova et al., 2006).

The 5’ and 3’ end of a coding-competent provirus consists of an LTR or long terminal repeat sequence which contains transcriptional regulatory sequences necessary for transcription, including the promoter, enhancers, and the polyadenylation signal (Kuff and Lueders, 1988; Jern and Coffin, 2008). Genes such as gag, pro, pol, and env are located between the LTRs and compose the “bulk” of the provirus which is typically 7-9 kb in length (Kuff and Lueders, 1988; Jern and Coffin, 2008). These genes encode proteins necessary for both viral assembly as well as infection (Jern and Coffin, 2008). Despite structural similarity to the proviral form of exogenous retroviruses, a significant proportion of ERVs have acquired mutations and/or deletions that have occurred over multiple generations while being maintained in the host genome which explains their general lack of infectious ability (Boeke and Stoye, 1997). In addition to full-length copies, a significant number of ERVs in the human and mouse genome exist as solo-LTRs. Solo-LTRs are the result of homologous recombination between two LTRs of an ERV copy and approximately 90% of all LTR or ERV related sequences in the human genome exist as solo LTRs (Medstrand, P. and Mager, D.L., 2003). It is hypothesized that this recombination mechanism may
be important in reducing the potential strain of these elements on the host genome (Boeke and Stoye, 1997).

1.2.2 ERV Classes in the Mouse Genome

Mouse ERVs can be divided into three different classes based on their similarity to exogenous retroviruses. IAP (intracisternal A particle) and (ETn) Early transposon/MusD elements belong to the class II ERVs, which account for approximately 3% of the mouse genome. While the large group of class II ERVs consists of a wide variety of other ERV families, these will not be the focus of this particular thesis. IAP and ETn/MusD elements are some of the most interesting families to study since these elements are responsible for the majority of de novo germ line mutations due to ERV insertions in the mouse genome (Maksakova et al., 2006; Stocking and Kozak, 2008).

The C57BL/6 (B6) mouse haploid genome consists of approximately 1000 full-length or partly deleted IAP copies with several thousand related solitary LTRs (Maksakova et al., 2006; Stocking and Kozak, 2008; Qin et al., 2010). In the 1960’s, IAPs were first identified as a result of electron microscopy of mouse tumor cells (Kuff et al., 1968). These elements are flanked by LTRs and contain coding sequences for gag and pol proteins in addition to env proteins (Stocking and Kozak, 2008). Interestingly, IAP RNA has been detected in oocytes and pre-implantation embryos in addition to normal mouse tissues, specifically thymus (Piko et al., 1984; Kuff and Fewell, 1985).

ETn copies are present in the haploid mouse genome at a lower frequency of less than 300 copies (Maksakova et al., 2006). ETn elements were originally identified in the 1980’s as a result of analyzing cDNA libraries from mouse embryonic carcinoma cells (Brulet et al., 1983). Initial analysis of the nucleotide sequence of these elements revealed an absence of a long, open reading frame in addition to a lack of sequence similarity with other retro-elements (Sonigo et al., 1987). Interestingly, following identification of MusD elements in the mouse genome, sequence comparisons with ETns revealed significant similarity with respect to their 5’ and 3’ LTRs in addition to a portion of the internal region of the elements. Furthermore, MusD elements encode the proteins necessary for transposition and have been shown to specifically mobilize ETn elements in retrotransposition assays (Mager and Freeman, 2000; Ribet et al., 2004). As indicated by their name, ETns are expressed early in embryonic development; specifically in embryos at days 3.5-7.5 (Sonigo et al., 1987; Rowe and Trono, 2011).

In addition to their high mutagenic potential, multiple studies have shown variability of IAP and ETn RNA levels between different mouse strains providing evidence of these elements as polymorphic (Kuff and Fewell, 1985; Sonigo et al., 1987). Confirmation of these elements as highly polymorphic has come
from a recent genome-wide study of ERV insertions in four different mouse strains using publicly available sequence data. Interestingly, these polymorphic ERV insertions could play an important role in phenotypic variability and disease susceptibility between mouse strains (Zhang et al., 2008).

### 1.2.3 Impact of ERVs in the mouse genome

ERVs, and TEs in general, can affect the host genome through a variety of mechanisms. A TE can insert into the coding region of a gene; thus resulting in insertional mutagenesis. Insertion of a TE can result in the element acting as an alternative promoter or enhancer of a host gene (van de Lagemaat et al., 2003; Goodier and Kazazian, 2008; Jern and Coffin, 2008; Cohen et al., 2009). Interestingly, a majority of ERV-induced germ line mutations in the mouse genome are a result of an insertion into a host gene intron which can lead to aberrant splicing, premature polyadenylation, and ectopic gene expression (Figure 1.1) (Maksakova et al., 2006). A well-studied example in the mouse genome of the relationship between ERV insertions and host gene expression occurs at the agouti locus which features a full-length IAP insertion 100 kilobases upstream of the gene promoter. A promoter located within the 5’LTR of the IAP has the ability to drive ectopic transcription of the agouti gene which results in various abnormal phenotypes such as yellow fur, obesity, and increased likelihood of developing tumors (Whitelaw and Martin, 2001; Maksakova et al., 2006). jkJk

### 1.3 Epigenetic repression of ERVs

As TEs can negatively impact the host genome in a variety of ways, it has been suggested that epigenetic mechanisms have evolved primarily to silence these mobile elements at both the transcriptional and post-transcriptional level (Slotkin and Martienssen, 2007; Maksakova et al., 2008). Epigenetic mechanisms can be described as modifications to DNA that do not alter the actual DNA sequence (Dambacher et al., 2010). Restraining TE activity at the transcriptional level is achieved by epigenetic marks such as DNA methylation and repressive histone modifications (Slotkin and Martienssen, 2007; Maksakova et al., 2008). RNA interference and host protein products also act in restraining ERV activity; however these mechanisms will not be the focus of this thesis (Maksakova et al., 2008).

#### 1.3.1 Overview of DNA methylation

In addition to TE silencing, DNA methylation also plays a role in gene regulation, imprinting, as well as X chromosome inactivation (Miranda and Jones, 2007). DNA methylation involves the addition
of a methyl group to the 5-position of a cytosine base which usually occurs in the context of a “CG” dinucleotide in mammals (Jones and Takai, 2001). Genome-wide methylation studies have estimated that a significant portion of all CG dinucleotides are marked by DNA methylation in the mouse and human genome (Edwards et al., 2010). In comparison to other nucleotides, CpG sites are under-represented in the genome due to spontaneous mutation of methylcytosine to thymine. Regions in the genome known as “CpG islands”, often found in gene promoters, are an interesting exception, and are normally quite resistant to DNA methylation regardless of the expression status of the corresponding gene. While the definition is dynamic, CpG islands are characterized as stretches of DNA greater than 200 base pairs (bps) in length with high GC content in addition to a significant number of CpG dinucleotides. An estimated 70% of human gene promoters overlap with a CpG island and these genes are often associated with house-keeping function (Jones and Takai, 2001; Miranda and Jones, 2007; Illingworth and Bird, 2009).

### 1.3.1.1 DNA methyltransferases (DNMTs)

The addition of methyl groups to DNA is achieved by the activity of enzymes known as DNA methyltransferases (DNMTs) (Law and Jacobsen, 2010). DNMT1 (DNA methyltransferase 1), DNMT3a (DNA methyltransferase 3a), and DNMT3b (DNA methyltransferase 3b) are the main enzymes involved in the establishment and maintenance of DNA methylation patterns in mammals. *De novo* methylation is carried out by DNMT3a and DNMT3b which act during early embryonic development to establish methylation patterns. High expression of these two enzymes is limited to undifferentiated embryonic stem cells (ESCs), which further points to their role in establishing DNA methylation patterns in early development. DNMT1 is the major maintenance DNA methyltransferase in mammals which specifically recognizes hemi-methylated DNA and is important in maintaining methylation patterns following DNA replication (Okano et al., 1999; Miranda and Jones, 2007).

Early studies involving generation of DNMT knockout mice have provided evidence for the importance of these enzymes in DNA methylation, TE silencing, and overall development. For example, both DNMT1 and DNMT3a knockout mice exhibit early embryonic lethality whereas DNMT3b knockout mice die shortly after birth (Okano et al., 1999; Kaneda et al., 2004). Upon analysis of the embryos, significant developmental abnormalities were observed (Okano et al., 1999). Furthermore, DNMT1 knockout embryos exhibit extensive hypomethylation of IAP elements and a 50-100 fold elevation in IAP transcript levels in addition to hypomethylation of all other sequences examined (Walsh et al., 1998; Walsh and Bestor, 1999; Slotkin and Martienssen, 2007; Maksakova et al., 2008). While DNMT3a or
DNMT3b null mice retain relatively normal *de novo* methylation activity, the double knockout is unable to methylate introduced proviral DNA. Interestingly, mutations of DNMT3b in humans have been associated with the development of Immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome which is characterized by chromosome instability and developmental problems (Okano *et al*., 1999).

**1.3.1.2 DNA methylation dynamics during development**

DNA methylation patterns are dynamic throughout early embryonic development; however, this modification becomes a relatively stable mark in somatic cells and tissues that is maintained from one cell to the next (Sarraf and Stancheva, 2004). Following fertilization and zygote formation, DNA methylation is erased (Monk *et al*., 1987; Kafri *et al*., 1992). While the paternal genome is characterized by rapid erasure of DNA methylation patterns, the maternal genome exhibits passive reduction in methylation concurrent with cellular division (Monk *et al*., 1987; Rougier *et al*., 1998; Oswald *et al*., 2000; Santos and Dean, 2004). Despite global de-methylation of the genome and detection of IAP expression observed in pre-implantation embryos, these retroelements remain relatively highly methylated similar to what is observed for imprinted genes (Piko *et al*., 1984; Lane *et al*., 2003; Svoboda *et al*., 2004). Interestingly, high levels of DNMT1 oocyte specific isoform (DNMT1o) have been found in early embryos. Furthermore, inactivation of this isoform results in slight de-methylation of IAP LTR sequences. Therefore, DNMT1o may play an important role in maintaining ERV methylation in the midst of global re-setting of DNA methylation patterns (Mertineit *et al*., 1998; Gaudet *et al*., 2004).

PGC7/Stella is a recently identified protein also found in early embryos. Embryos deficient for this factor exhibit a reduction in IAP methylation levels which provides evidence for this protein as important in protecting ERVs from DNA methylation reprogramming (Sato *et al*., 2002; Nakamura *et al*., 2007). At the blastocyst stage following implantation, global methylation patterns are reset across the genome except at CpG islands which remain methylation-free as described above (Kafri *et al*., 1992; Santos *et al*., 2002). Interestingly, similar waves of DNA methylation erasure and *de novo* methylation also occur in primordial germ cells. Methylation levels specifically at IAPs in embryonic day 13.5 primordial germ cells have been estimated to be approximately 30%, while hypermethylation of these elements occurs around the time of birth (Lane *et al*., 2003; Kato *et al*., 2007).
1.3.1.3 DNA hydroxymethylation as a novel epigenetic mark

The TET family of proteins in humans, which are responsible for DNA hydroxymethylation, was first identified as a result of a search for homologs of JBP1 and JBP2 which are found in trypanosomes and act by modifying the 5-methyl group of thymine (Tahiliani et al., 2009). Both TET1 and TET2 are highly expressed in mouse ESCs and act by addition of a hydroxyl group to 5-methylcytosine. DNA hydroxymethylation (5-hmC) has been estimated to occur at 5% of all CpG sites in mouse ESCs and has also been found in some adult mouse tissues such as brain (Ficz et al., 2011). Recent studies involving the generation of Tet1 null mouse ESCs have provided insight into the role of this gene in development. Despite reduced 5-hmC levels and slight changes in gene expression patterns, Tet1 negative ESCs maintain pluripotency. Furthermore, Tet1 knockout mice are viable and fertile indicating deficiency of this enzyme does not significantly impact postnatal development. However, whether or not these mice develop problems later in life has yet to be investigated (Dawlaty et al., 2011). Genome-wide studies of 5-hmC levels in mouse ESCs have shown enrichment of this mark in euchromatic genomic regions such as CpG island promoters. RNA-seq data has provided evidence for a relationship between 5-hmC at gene promoters and high levels of transcription compared to promoters modified by DNA methylation (5-mC) or promoters without any modification. Furthermore, 5-hmC at gene exons positivity correlates with gene transcription. Interestingly, Dnmt1/3a/3b triple knockout ESCs show significant reduction in 5hmC levels which suggest that the presence of 5-mC is required for 5-hmC (Ficz et al., 2011). While slight 5hmc enrichment has been found for satellite repeats, no published studies have fully characterized this mark with respect to ERVs in the mouse genome (Ficz et al., 2011).

1.3.1.4 DNA methylation and gene expression

DNA methylation specifically at CpG island promoter regions is often associated with gene silencing. Evidence has been provided for DNA methylation acting in blocking access of transcription factors to promoter/enhancer regions (Siegfried and Simon, 2010). Previous studies have shown that at the maternal Igf2 locus, binding of CCCTC-binding factor (CTCF) is observed between an upstream enhancer and the gene promoter. CTCF-binding prevents interaction of the enhancer with the promoter; thus blocking gene expression (Figure 1.2). However, the paternal Igf2 locus exhibits Igf2 expression. Interestingly, the region between the enhancer and promoter is methylated at the paternal locus which prevents CTCF binding. Therefore, the enhancer is able to interact with the gene promoter (Figure 1.2). This study and others suggests binding of transcription factors may act to block access of DNMTs (Jones and Takai, 2001; Miranda and Jones, 2007).
In addition to preventing transcription factor binding, DNA methylation can also recruit methyl-CpG-binding proteins (MBPs) that specifically recognize and bind to methylated DNA. MBPs have the ability to recruit other chromatin remodeling proteins that can ultimately lead to the formation of heterochromatin (Jones and Takai, 2001; Miranda and Jones, 2007). Furthermore, DNA methylation has also been shown to affect nucleosome occupancy at promoter regions. Therefore, methylation can modulate the chromatin environment in a way that prevents access of transcriptional machinery such as RNA polymerase II to DNA (Siegfried and Simon, 2010).

Previous studies have also provided evidence for a role of DNA methylation in tissue-specific gene expression. A study of mouse promoter regions found tissue-specific differentially methylated regions (T-DMRS) that exhibited different methylation patterns in different tissues. Interestingly, some of these regions corresponded to genes encoding tissue-specific transcription factors (Yagi et al., 2008).

While methylation of CpG island promoter regions including proximal regions downstream of the TSS have been shown to block transcription, DNA methylation in gene body regions has been shown to be positively correlated with gene expression (Appanah et al., 2007; Siegfried and Simon, 2010). It has been proposed that DNA methylation within genes may act to block spurious transcription initiation (Siegfried and Simon, 2010).

1.3.2 Overview of histone modifications

Eukaryotic chromatin is composed of units known as nucleosomes which consist of a histone protein core around which 146 bps of DNA is wrapped. The histone core is an octamer complex which consists of two of each of the core histone proteins. The core histones include H2A, H2B, H3, and H4. Each histone harbors a protruding “amino” tail where various post-translational modifications can take place including methylation, acetylation, phosphorylation, and ubiquitination (Figure 1.3). Depending on the type of modification and the position at which it takes place on the amino tail, the mark can either be active or repressive (Strahl and Allis, 2000; Grewal and Moazed, 2003; Suganuma and Workman, 2008; Wang et al., 2008). Furthermore, proteins can specifically recognize and bind to modified histones which can ultimately modulate the chromatin environment (Dambacher et al., 2010). The most widely studied and best characterized histone modifications include acetylation and methylation of lysine residues corresponding to core histones H3 and H4 (Figure 1.3) (Grewal and Moazed, 2003; Wang et al., 2008).
1.3.2.1 Histone modifications and gene expression

Histone acetylation is carried out by enzymes known as histone acetyltransferases (HATs) (Strahl and Allis, 2000). Interestingly, HATs act by removing the positive charge associated with the lysine residue with which they interact. This charge neutralization has been suggested to weaken the interaction between the histones and DNA; perhaps resulting in an open chromatin configuration (Bannister and Kouzarides, 2011). Many studies have provided evidence for a relationship between increased levels of histone acetylation and transcriptional activity (Grewal and Moazed, 2003). A study involving mapping of histone acetylation in human T-cells found a concentration of this mark at promoter regions; specifically CpG islands (Roh et al., 2005). Removal of acetylation is carried out by histone deacetylases (HDACs) (Strahl and Allis, 2000). Decreased acetylation of histone tails has been associated with transcriptional repression and compaction of chromatin structure (Grewal and Moazed, 2003; Bannister and Kouzarides, 2011). Interestingly, in vitro studies in yeast have shown that Sir (silent information regulator) proteins are able to bind to de-acetylated histone tails and are involved in the formation of heterochromatin (Grewal and Moazed, 2003).

In addition to acetylation, histone tails can also be modified by methylation with the most commonly studied modification occurring at lysine residues (Dambacher et al., 2010). In contrast to acetylation, methylation does not change the positive charge of the histone protein (Bannister and Kouzarides, 2011). Methylation can result in transcriptional activation or repression depending on the position on the lysine tail where methylation takes place in addition to how many methyl groups are present; one, two, or three resulting in mono, di, or tri-methylation. It has been well-established that H3K4me3 is concentrated at the promoter regions of transcriptionally active genes (Dambacher et al., 2010). Furthermore, H3K4me3 is specifically localized to CpG island promoters that lack DNA methylation (Thomson et al., 2010). Interestingly, it has been shown that the PHD finger protein specifically recognizes and binds to H3K4me3; subsequently recruiting HATs. Furthermore, CHD1 has also been shown to specifically bind to H3K4me3. This enzyme functions in re-positioning of nucleosomes (Bannister and Kouzarides, 2011). Interestingly, some gene promoters in ESCs are marked by both H3K4me3 in addition to H3K27me3, which is a known repressive mark. The presence of these opposing marks at promoter regions results in the establishment of a bivalent domain. These bivalent domains are often associated with genes that function in developmental processes. While bivalent genes are not expressed in ESCs, they are “poised” for rapid activation upon differentiation (Dambacher et al., 2010). Other studied histone marks include H3K4me1 and H3K36me3 which have been shown to be specifically associated with enhancers and regions of active transcription, respectively. H3K9me3,
which will be discussed more in detail in a later section, is a well-characterized repressive histone mark that is recognized by HP1 (heterochromatin protein 1) which functions in heterochromatin formation. Histone lysine methylation is carried out by HKMTs (histone lysine methyltransferases), which are highly specific enzymes (Bannister and Kouzarides, 2011).

In addition to post-translational modifications of histone tails, core histones can be replaced by histone variants which can play a role in gene expression and chromatin dynamics. One of the most widely studied histone variants is H2A.Z which is conserved in all eukaryotes. Interestingly, H2A.Z has been associated with both transcriptional activation and repression. Previous studies have shown that H2A.Z is correlated with increased nucleosome mobility which may allow access of certain DNA binding proteins. While studies in yeast and plants have shown a negative correlation between H2A.Z enrichment and transcription, H2A.Z has been found at gene promoter regions in humans and is positively correlated with gene expression. Interestingly, studies in yeast have shown H2A.Z enrichment at euchromatin-heterochromatin boundaries (Zlatanova and Thakar, 2008).

### 1.3.3 ERV silencing by DNA methylation and histone modifications

Early studies in mice have shown IAP LTRs as heavily methylated in various cells and tissues. Furthermore, Dnmt1 deficient mouse embryos exhibit significant accumulation of IAP transcripts likely a result of the absence of methylation at LTR sequences (Walsh et al., 1998). Following deletion of Dnmt1 in mouse ESCs, a significant upregulation of IAP transcript and protein levels was not observed. However, following differentiation of these cells, a drastic increase in IAP expression was observed similar to what was observed in embryos. Therefore, these results suggest different mechanisms of ERV silencing in ESCs and differentiated cell types (Hutnick et al., 2010).

In addition to DNA methylation, histone modifications also play an important role in TE silencing. A recent genome-wide study in the mouse genome showed enrichment of repressive histone marks H3K9me3 and H4K20me3 at IAP and ETn/MusD elements in ESCs. However, a significant reduction in enrichment of these two repressive marks was observed in differentiated cell types such as embryonic fibroblasts and neural precursors (Mikkelsen et al., 2007; Rowe and Trono, 2011). SETDB1 is the histone methyltransferase responsible for depositing the H3K9me3 mark, while Suv420h1 and Suv420h2 are the histone methyltransferases responsible for depositing the H4K20me3 mark. Upon deletion of SETDB1 in mouse ESCs, class II ERVs such as IAP and ETn/MusD elements in addition to class I ERVs are significantly upregulated (Matsui et al., 2010; Karimi et al., 2011; Rowe and Trono, 2011). In triple Dnmt knockout mouse ESCs lacking DNA methylation, H3K9me3 and H4K20me3 enrichment was
maintained at ERV sequences. Furthermore, ERV expression was not significantly upregulated in these cells (Matsui et al., 2010). This is in concordance with the study mentioned above which suggested an ERV silencing mechanism independent of DNA methylation in ESCs (Hutnick et al., 2010). Interestingly, evidence has been provided for Polycomb group proteins as playing a role in ERV silencing in mouse ESCs. In mouse ESCs deficient for both Polycomb group repressive complexes, PRC1 (Polycomb repressive complex 1) and PRC2 (Polycomb repressive complex 2), an increase in expression of LTR-retroelements was observed. Interestingly, enzymatic activity of PRC2 is responsible for depositing the H3K27me3 mark. Therefore, H3K27me3 may contribute to ERV silencing in ESCs (Leeb et al., 2010; Rowe and Trono, 2011). In addition to histone methylation, histone deacetylation has been shown to play a role in ERV silencing. Treatment of mouse ESCs with HDAC inhibitors has been shown to induce murine leukemia virus (MLV) expression and upregulation of IAP expression has been associated with enrichment of H4-acetylation (Lorincz et al., 2000; Rowe et al., 2010).

In addition to DNA methylation and repressive histone modifications, evidence has been provided for a role of certain proteins in ERV silencing. The TRIM family of proteins consists of a protein referred to as KAP1 or Trim28 which is recruited to specific sequences via KRAB-zinc finger proteins (Rowe et al., 2010; Rowe and Trono, 2011). In KAP1 negative mouse ESCs, H3K9me3 enrichment was significantly reduced and various ERV families including IAPs and MusD exhibited de-repression. DNA methylation, however, was maintained at these sequences (Rowe et al., 2010; Rowe and Trono, 2011). Similar to H3K9me3 enrichment, KAP1 binding was uninterrupted at ERVs in triple Dnmt knockout ESCs (Matsui et al., 2010; Rowe and Trono, 2011). In mouse embryos deficient for the chromatin remodeling factor known as lymphoid specific helicase (LSH1), retro-elements exhibit a loss of DNA methylation, an increase in active histone modifications, and de-repression. Furthermore, ESCs deficient for this protein are unable to de novo methylate introduced proviral DNA (Dennis et al., 2001; Yan et al., 2003; Zhu et al., 2006). Interestingly, NP95 (nuclear protein of 95 kDa) has been shown to bind to methylated DNA and form a complex with Dnmt1. Furthermore, Np95 null embryos exhibit hypomethylation and de-repression of various retrotransposons including IAP elements (Sharif et al., 2007). Therefore, these results suggest a potential role for both LSH1 and NP95 in ERV silencing in early development (Rowe and Trono, 2011). HP1 (heterochromatin binding protein-1) has been shown to interact and specifically bind to H3K9-methylated tails (Bannister et al., 2001). However, upon knockout of several isoforms of HP1 in mouse ESCs, significant upregulation of ERV expression was not observed (Maksakova et al., 2011).
1.3.4 Relationship between DNA methylation & histone modifications

In contrast to DNA methylation patterns which are characterized as relatively stable in somatic cells and tissues, histone modifications can be more dynamic and reversible epigenetic marks (Sarraf and Stancheva, 2004; Cedar and Bergman, 2009). These two epigenetic marks share an interesting and complex relationship with respect to epigenetic silencing. While some silenced genomic regions are associated with the presence of both of these modifications, other silenced regions are associated with only DNA methylation or solely repressive histone modifications. Histone modifications may play an important role in the establishment of DNA methylation patterns in early development. It has been suggested that RNA polymerase II, which associates with CpG islands, may recruit H3K4-methyltransferases. Interestingly, interaction of a non-catalytic Dnmt homolog known as Dnmt-3L with core histone H3 has been shown to be inhibited at CpG islands due to the presence of H3K4-methylation. Therefore, CpG islands may avoid global re-setting of DNA methylation patterns due to the presence of H3K4-methylation (Cedar and Bergman, 2009). Another intriguing relationship has been observed with respect to H2A.Z and DNA methylation in Arabidopsis. Loss of methylation results in a gain of H2A.Z enrichment whereas mutations of the complex that deposits H2A.Z results in genome-wide hypermethylation; thus supporting these marks as antagonistic (Zilberman et al., 2008).

In contrast to DNA methylation which is maintained following DNA replication, chromatin structure is likely disrupted at this time in order for replication machinery to access DNA. Therefore, DNA methylation may serve to re-establish histone marks following replication (Cedar and Bergman, 2009). DNA methylation and repressive histone modifications share an important role in terms of silencing genes associated with pluripotency upon differentiation. Upon binding of certain transcription factors to gene promoters, a repressive complex containing G9a is recruited. G9a is a histone lysine methyltransferase responsible for H3K9-methylation. Interestingly, G9 has also been shown to recruit to Dnmt3a/3b and function in de novo methylation (Cedar and Bergman, 2009; Leung et al., 2011). Gene repression associated with the presence of repressive histone modifications is not always associated with DNA methylation. Silencing mediated by Polycomb group proteins in ESCs is associated with CpG island promoters of developmental genes. While, EZH2 of the PRC2 complex catalyzes H3K27-trimethylation at these regions, these regions are generally protected from de novo methylation (Cedar and Bergman, 2009).
1.4 Spreading of epigenetic silencing from TEs

1.4.1 Previous evidence of spreading

In addition to effects discussed in section 1.2.3., another way in which TEs or specifically ERVs can potentially affect host gene expression is through a phenomenon known as spreading of epigenetic silencing. Upon insertion of an ERV, this sequence may be targeted by silencing mechanisms such as DNA methylation and repressive histone modifications that could accumulate and subsequently spread into nearby sequences. Therefore, a gene in close proximity to an ERV could be targeted by this spreading mechanism; potentially resulting in gene silencing (Figure 1.4) (Feschotte, 2008).

Initial evidence for TEs acting in the induction of silencing marks came from studies using plasmid constructs to determine if certain fragments could result in methylation of a reporter gene. An early study provided evidence for a region 400 bps upstream of the mouse *Aprt* house-keeping gene as acting as a center of *de novo* methylation. Interestingly, it was determined in a later study that this methylation center consisted of two B1 repetitive elements, a type of SINE. These repetitive elements were heavily methylated and were shown to induce *de novo* methylation of flanking regions in reporter-con structs (Mummaneni *et al.*, 1993; Yates *et al.*, 1999). A similar result was obtained for B1 and B2 repetitive sequences located upstream of the rat *AFP* gene which were shown to direct methylation of a reporter gene (Hasse and Schulz, 1994).

Evidence for the relationship between spreading of DNA methylation and gene expression has come from studies in plants. The *FWA* gene in *Arabidopsis thaliana* functions in flowering. Knockout of genes involved in establishing DNA methylation patterns resulted in reduced methylation of the *FWA* gene, ectopic *FWA* expression, and a late-flowering phenotype. Therefore, DNA methylation likely plays a role in the normal regulation of this gene. Interestingly, plants deficient for DNA methylation transfected with a transgene containing SINE-like repeats around the *FWA* TSS exhibited a normal phenotype; therefore providing evidence for repeat-mediated silencing of the *FWA* gene in *Arabidopsis* (Kinoshita *et al.*, 2007). More recently, evidence for DNA methylation spreading from a DNA transposon was observed in the melon genome. In female flowers, DNA methylation spreading was observed from a DNA transposon into the *CmWIP1* promoter located approximately 1.5 kilobases (kb) upstream resulting in gene inactivation, expression of a stamen inhibitor, and subsequent development of female flowers. In male flowers, which do not contain the DNA transposon upstream of *CmWIP1*, gene expression remains intact which ultimately results in the development of male flowers. Therefore, DNA transposon-induced epigenetic spreading likely plays an important role in sexual determination in the
melon plant (Martin et al., 2009). Furthermore, a genome-wide study in *Arabidopsis thaliana* showed that evolutionary older TEs marked by DNA methylation were located farther from genes in comparison to unmethylated and/or younger TEs. This result provides evidence for negative selection against methylated insertions near genes, possibly due to the harmful impact on host genes through spreading of methylation (Hollister and Gaut, 2009).

A recent study has provided direct evidence for spreading of repressive chromatin marks with respect to ERVs, specifically IAPs and ETns, in the mouse genome. Using H3K9me3 chromatin immunoprecipitation-sequencing (Chip-seq) data collected in mouse ESCs, enrichment of these repressive marks was able to be measured at regions directly flanking insertionally polymorphic mouse ERV insertions in order to determine if an ERV insertion resulted in local enrichment of H3K9me3. Interestingly, strong H3K9me3 enrichment was seen flanking polymorphic IAP copies. However, this enrichment was not present at the empty site in mouse ESCs lacking the insertion. Therefore, these results indicate that H3K9me3 enrichment is a result of the presence of the IAP. While a similar trend was observed for ETn copies, the result was not as significant. Interestingly, DNA methylation was observed at both full and empty sites indicating that this mark is independent of the presence of the ERV insertion in ESCs (Rebollo et al., in press).

### 1.4.2 Genomic defense against spreading of epigenetic marks

Due to the general resistance of CpG islands to acquiring DNA methylation, it has been suggested that these regions likely contain important sequence elements or transcription factor binding sites that protect against repressive modifications (Fan et al., 2007). Initial evidence for transcription factors as important in maintaining the methylation-free status of CpG islands came from studies in transgenic mice. Upon deletion or mutation of Sp1 transcription factor binding sites in the *Aprt* housekeeping gene CpG island promoter, significant levels of methylation were observed likely originating and spreading from upstream repetitive elements (Macleod et al., 1994; Yates et al., 1999). Furthermore, the presence of Sp1 and Sp3 binding sites in the long allele of the *RIL* gene in humans was consistent with the protection of the CpG island promoter of this gene from aberrant hypermethylation in cancer (Boumber et al., 2008). In addition to transcription factor binding at promoter regions, binding of certain factors outside of promoters may also serve to protect against encroaching repressive marks. A study completed in human brain tissue showed evidence for over-representation of binding sites for zinc finger proteins in 400 bp flanking regions of CpG island promoter genes defined as “methylation-resistant.” In addition to enrichment of transcription factor binding sites, these flanking regions also
exhibited significant enrichment for active histone marks such as H3K4me3 (Fan et al., 2007). Another zinc finger protein known as VEZF1 which was originally identified as binding to the chicken beta-globin locus has also been suggested to be important in preventing DNA methylation at CpG islands. This protein was not only found to be bound to the mouse Aprt CpG island promoter, but was also able to protect against and mediate removal of de novo methylation (Dickson et al., 2010).

Boundary elements which act to separate euchromatin and heterochromatin domains may also act in blocking the accumulation and spreading of repressive marks. In normal human cells, a chromosome boundary characterized by CTCF zinc finger protein binding, is located 2 kb upstream of the p16 tumour suppressor gene promoter. Upon disruption of CTCF binding, accumulation of the repressive mark H4K20me3 and loss of the active mark H2A.Z is observed at the gene promoter consistent with reduction in gene expression similar to what is observed in cancer cell lines. Therefore, absence of CTCF binding and spreading of heterochromatin may represent a mechanism by which p16 is inactivated in some forms of cancer (Witcher and Emerson, 2009). Interestingly, a recent genome-wide study in the human genome showed that gene promoters resistant to DNA methylation in cancer exhibited an increased frequency of retro-elements nearby compared to promoters prone to methylation. It was hypothesized that methylation-resistant genes may harbor more transcription factor binding sites or boundary elements that act to prevent methylation whereas methylation-prone genes do not have these “protecting” factors and are therefore more susceptible to potential silencing which results in stronger negative selection against nearby insertions (Estecio et al., 2010).

Evidence in yeast has suggested a role of H2A.Z in restricting the spread of heterochromatin from the silent HMR locus. In H2A.Z negative cells, spreading of heterochromatin was observed from the silent locus into adjacent genes; significantly reducing expression (Meneghini et al., 2003). Although CpG island promoters normally unmethylated regardless of expression of the corresponding gene, transcriptional activity may also play a role in protection against repressive mark accumulation. The dynamics of spreading were studied in yeast using a fragment known to induce heterochromatin. Upon insertion of the fragment near a highly expressed gene, a drastic reduction in the extent of repressive histone mark spreading was observed. However, a gene expressed to a lower extent also exhibited barrier activity (Wheeler et al., 2009). It has been speculated that a drop in transcriptional activity could potentially initiate spreading from a heterochromatin domain. Decreased transcription could result in decreased frequency of transcription factor binding which would normally protect against encroaching repressive mark. It has been hypothesized that spreading of repressive marks involves an interaction between DNA methylation and chromatin remodeling factors which alter the chromatin of nearby
sequences, making them more accessible to silencing marks (Turker, 2002). While this mechanism has yet to be elucidated, it is likely dependent on several factors.

1.5 **Thesis Objectives**

While TE-mediated spreading of repressive marks and subsequent gene silencing has been somewhat characterized in plants, this phenomenon has not yet been studied with respect to ERVs in mammalian genomes. Indeed, although it is often cited in reviews as a potential deleterious consequence of TE insertion, direct evidence for TE-induced local heterochromatin affecting gene expression in mammals is surprisingly scarce. I, therefore, sought to investigate the hypothesis that an ERV insertion in the mouse genome could result in the induction and subsequent spreading of silencing marks that could affect the expression of a nearby gene. To do this, I took advantage of our laboratory’s dataset of insertionally polymorphic mouse ERVs and chose those ERVs located less than 2 kb from the TSS of a gene (Zhang et al., 2008). By using appropriate F1 hybrid mice or ESCs, I could study alleles with and without a particular insertion but in the same mouse.

In chapter 3 of my thesis, I investigate the phenomenon of “spreading” with respect to various genes with nearby ERV insertions. I provide evidence for the impact of ERVs on nearby genes as a locus specific event with spreading of repressive marks being an uncommon occurrence. I show that CpG island promoters are generally unmethylated despite nearby repressive modifications. An exception I observed is the B3galtt gene where spreading of DNA methylation and repressive chromatin is observed from the upstream LTR into the CpG island promoter of the gene in ESCs.

In chapter 4 of my thesis, I discuss the observation of differential LTR methylation. For two genes, Lair1 and Cdgap, I address the correlation between gene expression and methylation of the nearby insertion. Furthermore, I investigate the hypothesis that a highly expressed gene may be able to impact the epigenetic state of a nearby ERV insertion through the accumulation and possible spreading of active chromatin marks.

In chapter 5 of my thesis, I discuss these results to further define the complex relationship between ERVs and nearby genes. I address how locus specific factors may mediate the extent of spreading or protection from silencing. Furthermore, I describe the significance of these results with respect to understanding how TEs can impact gene expression and vice versa. I also discuss how these findings may contribute to the field of epigenetics and cancer development.
Figure 1.1 Common affects of ERV insertions on host genes. (A) Insertion of an ERV in a gene intron can lead to aberrant splicing or premature polyadenylation. (B) Insertion of an ERV in a gene intron can also lead to ectopic gene expression, for instance, driven by an antisense promoter in the 5’LTR of the insertion.

Figure 1.2 Example of relationship between DNA methylation, transcription factor binding, and gene expression. (A) At maternal Igf2 locus, CTCF binding (moon) at unmethylated CpG sites (white circles) blocks action (dashed line) of enhancer (oval) which results in lack of gene transcription (“X” over TSS). (B) At paternal Igf2 locus, DNA methylation of CpG sites (black circles) blocks CTCF binding which allows interaction (dashed line) of enhancer (oval) with Igf2 gene, resulting in gene transcription.
Figure 1.3 Structure of nucleosome, histone core, and examples of post-translational modifications of histone amino tails. A nucleosome consists of a histone octamer complex (cylinder) around which DNA (dashed line) is wrapped. The histone octamer complex consists of two of each of the following core histones: H2A, H2B, H3 and H4. These core histones have protruding amino (NH2) tails (black line) that can be modified. The most commonly studied modifications are acetylation (Ac-triangle) and methylation (Me-circle) of lysine (K) residues. Adapted with permission from “Histone onco-modifications” (Fullgrabe et al., 2011).

Figure 1.4 Example of ERV-mediated gene silencing by spreading of DNA methylation. (A) ERV insertion is targeted by silencing mechanisms such as DNA methylation (black circles). (B) DNA methylation (black circles) at ERV insertion spreads into adjacent sequence and ultimately into nearby gene; shutting off its transcription as shown by “X” over gene TSS. Rectangle represents gene promoter region and square represents an exon.
2 Materials and methods

2.1 ES cell culture

C2 (C57BL/6) ESC pellets were provided by Nicole Hofs of the BC Cancer Research Center for Genetic Modeling. J1 (129S4/SvJae) and TT2 (C57BL/6xCBA) ES cell pellets were provided by Irina Maksakova of Matt Lorincz’s lab at the University of British Columbia.

2.2 DNMT1/SETDB1 siRNA knockdown experiments

DNMT1 and SETDB1 siRNA-mediated knockdowns were completed by Irina Maksakova as previously described (Maksakova et al., 2011).

2.3 DNA/RNA extraction from cells/tissues

DNA and RNA were extracted from cell pellets and whole tissues according to the protocol provided with the All Prep DNA/RNA extraction kit (Qiagen). DNase treatment of RNA was completed according to the protocol of the TURBO DNA-free kit (Applied Biosystems). Quantification and purification of DNA and RNA was measured using a Nanodrop spectrophotometer (Thermo Scientific). For strains of mice and age of mice when tissues retrieved, see table 2.1.

2.4 Reverse transcription reactions for cDNA

Reverse transcription (RT) reactions were performed according to the “Superscript III First-Strand Synthesis System” protocol (Invitrogen). Modifications to the protocol include the following: the cDNA synthesis step was completed for 60 minutes at 50°C, and the reaction was terminated by heating samples at 70°C for 15 minutes. For each sample, two RT reactions were completed. A reaction without reverse transcriptase was also completed for each sample in order to serve as a negative control.

2.5 Allelic quantification of gene expression

In order to distinguish between alleles of the F1 hybrid, single nucleotide polymorphisms (SNPs) were used. SNPs were found using the mouse phenome database provided by Jackson Laboratories (http://phenome.jax.org/). SNPs useful for determining allelic expression were those that would be in the mRNA product. PCR primers were designed flanking SNP(s) in order to include the polymorphism in the amplified region. PCR amplification was performed in the F1 hybrid tissue or cell type of interest.
using two cDNA samples and a genomic DNA sample. Gel electrophoresis was used in order to analyze PCR products and purification was completed according to the MinElute Gel Extraction kit protocol (Qiagen). Sequencing was completed by McGill University and Genome Quebec Innovation Centre Sequencing Platform (http://gqinnovationcenter.com/index.aspx). Sequencing data was then analyzed using the “Peak Picker” software program as previously described (Ge et al., 2005). A peak height ratio was determined with respect to the informative SNP(s) in the expressed mRNA. This value was then normalized to reference peak heights in the sequence. The genomic DNA sequence was used as a reference of an approximate 50:50 allelic ratio.

2.6 Quantitative RT-PCR

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was completed using the “7500 Real Time PCR System” (Applied Biosystems). Duplicate 10 μl reactions were completed for each sample using “Fast SYBR Green Master Mix” (Invitrogen). The PCR conditions were as follows: 95°C for 20 seconds, 45 cycles of 95°C for 3 seconds and 60°C for 30 seconds, 95°C for 15 seconds, 60°C for 1 minute, 95°C for 15 seconds, and 60°C for 15 seconds.

2.7 Bisulfite conversion/PCR/sequencing

Bisulfite conversion of approximately 500ng of genomic DNA was completed according to the protocol provided with the EZ DNA methylation kit (Zymo Research). Modifications to the protocol are as follows: after addition of the M-dilution buffer and water to the DNA, the sample was incubated at 50°C for 12 minutes. Following addition of the CT conversion agent, the DNA sample was incubated for 4 hours using previously described conditions (Reiss et al. 2007). Following the conversion reaction, DNA was eluted in 15 μl of M-elution buffer. Approximately 1-2 μl of converted DNA was used in subsequent PCR reactions with AmpliTaq Gold DNA Polymerase (Applied Biosystems). As described previously, one round of PCR followed by a second round with a nested primer was completed in order to obtain enough product for cloning and sequencing with respect to each region of interest (Reiss and Mager, 2007). Both rounds of PCR were completed with the following temperature conditions: 95°C for 10 minutes, 35 cycles of 95°C for 30 seconds; 50°C for 30 seconds; and 72°C for 30 seconds. An extension period of 7 minutes at 72°C was used for the first round of PCR, while the second round of PCR consisted of a 15 minute extension period in order to complete A-tailing of the PCR product for subsequent ligation and cloning reactions. Both rounds of PCR were completed in duplicate for each sample in order to avoid amplification bias of
methylated or unmethylated DNA. Products were analyzed following the second round of PCR by gel electrophoresis. Purification of PCR products was completed according the “Purelink Quick Gel Extraction kit” protocol (Invitrogen). Following confirmation of purified product through gel electrophoresis, samples were cloned using the “TOPO TA vector” kit (Invitrogen). Plasmid DNA extraction from colonies and sequencing was completed by McGill University and Genome Quebec Innovation Centre Sequencing Platform (http://gqinnovationcenter.com/index.aspx).

Analysis of generated sequences was completed by using a free online tool known as “Quma: Quantification tool for Methylation Analysis” (http://quma.cdb.riken.jp/). Sequences included in the results had to be unique with respect to at least one of the following criteria: methylation pattern, conversion rate, or number of mismatches. Therefore, each individual clone could be associated with a different cell.

In order to distinguish between alleles in F1 hybrid cells and tissues, SNPs were used between the two mouse strains found via the mouse phenome database from Jackson Laboratories (http://phenome.jax.org/). “C” versus “T” SNPs were avoided due to the conversion of unmethylated cytosines to thymines following bisulfite conversion. PCR primers were designed flanking SNP(s) in order to include the polymorphism in the amplified region. All sequences were analyzed for the SNP(s) polymorphic between the two mouse strains which allowed clones to be categorized according to allele. In some cases, a polymorphic ERV insertion instead of a SNP(s) was used in order to distinguish between alleles.

2.8 COBRA/Combined bisulfite restriction analysis

COBRA was performed as previously described (Reiss and Mager, 2007). Enzymes used and restriction sites are listed in table 2.2.

2.9 MeDIP/Methylated DNA immunoprecipitation

Approximately 1.5 μg of genomic DNA was prepared for sonication by adding NaPO₄ (10 mM) to bring the total sample volume to 300 μl. Sonication was completed on ice water using the Diagenode Bioruptor for 50 seconds (25 seconds on, 25 seconds off). Approximately 5 μl of sonicated material was used in order to check fragment size by gel electrophoresis.

Following sonication, samples were spiked with 2 μl of methylated and nonmethylated control DNA (1 μl of each in 299 μl H₂O; Invitrogen). Samples were denatured (8-10 minutes at 94°C) and immediately cooled on ice. Samples were then subdivided into equal aliquots including an input sample.
(stored at -20°C) and IP samples. The following components were added to each IP sample followed by an overnight incubation at 4°C with rotation: 50 μl IP buffer 10x (100 mM NaPO₄, pH 7.0, 1.4 M NaCl, 0.5% Triton X-100), 4 μl of antibody, 5 μl PMSF 100x, and 300 μl H₂O. Following overnight incubation, 80 μl of blocked Protein A beads (50% slurry-Millipore) was added to each IP sample followed by 2 hours of incubation at 4°C with rotation. Beads were then isolated by centrifugation (2 minutes at 2500 rpm) and washed twice with cold IP buffer (1x), once with cold IP buffer (300 mM NaCl), and once with TE buffer (2x). Washes with IP buffer were completed for 4 minutes with rotation at 4°C while the TE buffer wash was completed for 2 minutes. Following each wash, beads were isolated via centrifugation as described above and the supernatant was discarded. Beads were resuspended in 250 μl of digestion buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS) and 7 μl of proteinase K (10 mg/ml) followed by incubation for 3-4 hours at 55°C with vortexing.

DNA was extracted from IP and input samples using phenol-chloroform followed by ethanol precipitation (with 0.3 M sodium acetate and glycogen). Following precipitation, samples were washed once with 70% ethanol and resuspended in 30 μl of Tris-HCl pH 7.5 (10 mM). Samples were diluted 1:10 with sterile water and 2 ul of each sample was used in duplicate in quantitative PCR analyses. Enrichment was calculated as percentage of input using the following formula: Primer efficiency(C_{input} - C_{IP}) x 100 with primer efficiency being determined through a standard curve with dilutions of genomic DNA. In order to confirm specific enrichment of methylated DNA, qPCR was completed with primers specific for methylated and unmethylated DNA supplied by Invitrogen. See table 2.3 for antibodies used.

2.10 ChIP (Chromatin immunoprecipitation)-Whole tissues

The following protocol was performed as previously described (Wederell et al., 2008). After thawing of the frozen tissue sample on ice, a sterile razor blade was used in order to dissect the tissue using cold PBS solution. Tissue pieces were added to a Dounce homogenizer containing 1% formaldehyde in PBS and homogenized 20x. Following homogenization, the sample was incubated at room temperature with rocking/rotation for 10 minutes in order to complete cross-linking. The cross-linking reaction was stopped by addition of glycine (0.125 M) followed by a 5 minute room temperature incubation with rocking/rotation. Cell pellets were isolated by centrifugation (3000 rpm for 3 minutes at 4°C), washed twice with PBS solution, and pelleted again. Resuspension of cell pellets was completed using cold lysis buffer (10mM Tris-HCl ph 8.0, 10mM NaCl, 3mM MgCl₂, 0.50% NP40, 1x PIC) 4-5x the volume of the pellet. Cells were incubated on ice for 15 minutes and dounce homogenized 20x during the incubation period. Isolation of cell pellets was then completed by centrifugation (13200 rpm for 5
minutes at 4°C). Cell pellets were resuspended in cold nuclear lysis buffer (1% SDS, 5mM EDTA, 50 mM Tris-HCl ph 8.1, 1x PIC) 3-4x the volume of the pellet followed by incubation for 15 minutes on ice. Sample volume was brought to 500 µl with sterile water before proceeding to the sonication step. Samples were sonicated on ice water for a total of ten minutes (20 cycles of 30 seconds on, 40 seconds off) using the “Diagenode Bioruptor Sonicator”. Phenol-chloroform extraction of DNA was completed for a 15 ul aliquot of the digested sample followed by gel electrophoresis in order to check the nuclei size.

Following sonication, samples were centrifuged in order to pellet cellular debris (13200 rpm for 10 minutes at 4°C). The supernatant was isolated and approximately 30 ug of chromatin was aliquoted for each immunoprecipitation reaction in addition to 20 µl of Protein A beads (50% slurry), 1 µl of 10x PIC, and ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl, 16.7 mM Tris-Cl ph 8.1) 3-4x the volume of chromatin added. Following an incubation/rotation period of 1 hour at 4°C, the samples were centrifuged (4000 rpm for 2 minutes at 4°C) and the supernatant was added to new siliconized tubes. The antibody of interest (3 ug) was then added to the respective sample and incubated overnight at 4°C with rotation. The next day, 20 µl of Protein A beads (50% slurry-Millipore) was added each IP sample and incubated with rotation for 4 hours at 4°C followed by centrifugation (4000 rpm for 2 minutes at 4°C). The supernatant was discarded for all samples except for the IgG control which was kept as the input DNA and stored at -20°C.

The beads corresponding to the various IP samples were washed at 4°C with 500 µl of the following solutions for 5 minutes each with rotation followed by centrifugation (4000 rpm for 2 minutes 4°C): low salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-Cl ph 8.1, and 150 mM NaCl), high salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-Cl ph 8.1, 500 mM NaCl), lithium chloride buffer (0.25 M LiCl, 1% NP-40, 1% deoxylate, 1 mM EDTA, 10 mM Tris-Cl ph 8.1), and 2x TE buffer (10 mM EDTA, 10 mM Tris-Cl ph 8.0). In order to release DNA from the beads, 125 µl of fresh elution buffer (1% SDS, 0.1 M NaHCO₃) was added to each sample followed by incubation at room temperature for 15 minutes with rotation and vortexing. In order to pellet the beads, samples were centrifuged (6000 rpm for 90 seconds) and the supernatant was transferred to new tubes. The elution step was repeated with another 125 µl of elution buffer followed by incubation/rotation for 30 minutes. A reverse cross-linking step and protease K treatment was completed overnight at 65°C for all samples including input (0.2 M NaCl, 40 mM Tris-HCl ph 6.5, 10 mM EDTA, 2 µl 10 mg/ml Prot K, 1.25 ul 20mg/ml RNAse A).
DNA was purified from all samples using phenol-chloroform extraction (2x) followed by ethanol precipitation (with glycogen and 3 M sodium acetate) at -80°C. Samples were washed with 95% ethanol and resuspended in 35 μl of sterile water. Samples were diluted 1:10 with water and 2 μl of each sample was used in duplicate for quantitative PCR (qPCR). Enrichment was calculated as relative to input using the following formula: Primer efficiency = \( \frac{C_{\text{IP}} - C_{\text{input}}} {C_{\text{input}}} \) with primer efficiency being determined by a standard curve with dilutions of input DNA. See table 2.4 for antibodies used.

### 2.11 ChIP-ESCs

This protocol was completed as previously described (Rebollo et al., in press). After 10 million ESCs were thawed and harvested on ice, 250 μl of ice-cold dousing buffer (10 mM Tris-Cl ph 7.5, 4 mM MgCl₂, 1 mM CaCl₂, 1x PIC) was added and cells were homogenized by pipetting up and down. Cells were then transferred to a 1.5 ml tube and further homogenized 20x using a 1 ml syringe. Enzymatic digestion of DNA was completed by adding micrococcal nuclease (150 U/mL) to homogenized cells followed by incubation at 37°C for 7 minutes. EDTA (10 mM) was then added to the sample in order to stop digestion followed by incubation on ice for 5 minutes.

Cells were lysed by adding 1 ml of hypotonic lysis buffer (0.2 mM EDTA ph 8.0, 0.1 mM benzamidine, 0.1 mM PMSF, 1.5 mM DTT, 1x PIC) to the sample followed by incubation on ice for 1 hour with vortexing at 10 minute intervals. After incubation, cells debris was pelleted by centrifugation (3000g for 5 minutes at 4°C) and supernatant was transferred to a 1.5 ml tube. Phenol-chloroform extraction of DNA was completed for a 50 μl aliquot of the digested sample followed by gel electrophoresis in order to check the nuclei size. A preclearing step was completed by adding 100 μl of blocked Protein A beads (50% slurry) to the digested chromatin sample followed by rotation at 4°C for 2 hours and centrifugation (4000 rpm for 2 minutes at 4°C) and transfer of supernatant to a new eppendorf tube. The supernatant was then split into equal volumes for each antibody plus the input sample (stored at -20°C). IP buffer (10 mM Tris-HCl ph 8.0, 1% Triton X-100, 0.1% Deoxycholate, 0.1% SDS, 90 mM NaCl, 2 mM EDTA, 1x PIC) was added to each IP sample to bring the total volume to 325 μl per sample. Antibody incubation was completed by adding 3 μg of antibody of interest to each sample and rotating at 4°C for 1 hour followed by addition of 20 μl of Protein A (50% slurry-Millipore) to each sample and overnight incubation at 4°C with rotation. After overnight incubation, beads were pelleted by centrifugation (4000 rpm for 2 minutes at 4°C) and supernatant was discarded.

Beads corresponding to each IP sample were washed with 400 μl of ChIP wash buffer (20 mM Tris-HCl ph 8.0, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, and 1x PIC), vortexed briefly, and
incubated at 4°C for 3 minutes with rotation. Samples were centrifuged as described above followed by an addition wash with ChIP wash buffer then a final wash with final ChIP wash buffer (20 mM Tris-HCl pH 8.0, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, and 1x PIC). Both washes were completed as described above; however, prior to the final wash step, IP samples with beads were transferred to new 1.5 ml tubes. 100 μl of elution buffer (100 mM NaHCO₃, 1% SDS) was then added to each IP sample (with just beads) and buffer was also added to the input sample for a final volume of 200 μl. RNase (0.5 μl of 10 μg/μl) was added to all samples followed by incubation at 68°C for 2 hours with brief periods of vortexing. DNA from IP samples was isolated via centrifugation (4000 rpm for 2minutes) and transfer of supernatant to a new tube. A second elution of DNA from the IP samples was completed by adding 100 μl of elution buffer to the beads followed by incubation at 68°C for 5 minutes and centrifugation as described above. Supernatants of the same sample from the two elutions were then pooled together.

DNA for all samples included input was purified using the QIAquick PCR purification kit (Qiagen) with the following modifications: following the PE wash, samples were centrifuged for 2 minutes, columns were air dried for 1 minute prior to addition of elution buffer, elution buffer was pre-heated to a temperature of 42°C and 60 μl was added to the column followed by incubation at room temperature for 1 minute before centrifugation. Samples were diluted 1:5 with water and 2 μl of each sample was used in duplicate for quantitative pcr (qPCR) analyses. Enrichment was calculated as relative to input using the following formula: Primer efficiency \((C_{input} - C_{IP})\) with primer efficiency being determined by a standard curve with dilutions of input DNA. See table 2.4 for antibodies used.

### 2.12 Data Sources, etc.

For a list of sources used for data analysis and publicly available data, see table 2.5

### 2.13 Primers

For a list of primers used in all experiments, see table 2.6.
Table 2.1 Age or source of mouse cells and tissues used for experiments

<table>
<thead>
<tr>
<th>Strain used</th>
<th>Cells or tissues</th>
<th>Age/Source</th>
<th>Experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6x129 (F1 hybrid)</td>
<td>ES cells</td>
<td>Obtained from Danny Chui of BCCRC</td>
<td>Bisulfite sequencing, COBRA</td>
</tr>
<tr>
<td>(C57BL/6x129Sv/Ev Tac)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V26.2 (C57BL/6)</td>
<td>ES Cells</td>
<td>Obtained from Danny Chui of BCCRC</td>
<td>Bisulfite sequencing, qRT-PCR</td>
</tr>
<tr>
<td>C2 (C57BL/6)</td>
<td>ES cells</td>
<td>Obtained from Nicole Hofs of BCCRC</td>
<td>ChIP</td>
</tr>
<tr>
<td>J1 (129S4/SvJae)</td>
<td>ES cells</td>
<td>Obtained from Irina Maksakova of Lorincz Lab at UBC</td>
<td>Bisulfite sequencing, ChIP, Knockdown studies, qRT-PCR, MedIP</td>
</tr>
<tr>
<td>TT2 ESCs (C57BL/6xCBA)</td>
<td>ES cells</td>
<td>Obtained from Irina Maksakova of Lorincz Lab at UBC</td>
<td>Bisulfite Sequencing, ChIP, Knockdown studies, qRT-PCR</td>
</tr>
<tr>
<td>B6xAJ (F1 hybrid)</td>
<td>Embryos</td>
<td>Day 7.5</td>
<td>Bisulfite sequencing, COBRA, Allelic quantification</td>
</tr>
<tr>
<td>(C57BL/6xAJ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6x129 (F1 hybrid)</td>
<td>Adult tissues</td>
<td>7 weeks</td>
<td>Bisulfite sequencing, COBRA, Allelic quantification</td>
</tr>
<tr>
<td>(C57BL/6x129Sv1/SvImJ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6xAJ (F1 hybrid)</td>
<td>Adult tissues</td>
<td>8 weeks</td>
<td>Bisulfite sequencing, COBRA, Allelic quantification</td>
</tr>
<tr>
<td>(C57BL/6xAJ)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B6 (C57BL/6)</td>
<td>Adult tissues</td>
<td>7-8 weeks</td>
<td>Bisulfite sequencing, qRT-PCR, ChIP</td>
</tr>
<tr>
<td>AJ</td>
<td>Adult tissues</td>
<td>7-8 weeks</td>
<td>Bisulfite sequencing, qRT-PCR, ChIP</td>
</tr>
</tbody>
</table>
### Table 2.2 Enzymes used for COBRA

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Restriction Site (-)</th>
<th>Methylation Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taq1</td>
<td>T-CG-A</td>
<td>Cut=methylated</td>
</tr>
<tr>
<td>Rsal</td>
<td>GT-AC</td>
<td>Cut=methylated</td>
</tr>
<tr>
<td>Hinfl</td>
<td>G-ANT-C</td>
<td>Cut=methylated</td>
</tr>
<tr>
<td>BstBII</td>
<td>TT-CGAA</td>
<td>Cut=unmethylated</td>
</tr>
<tr>
<td>Acl1</td>
<td>AA-CGTT</td>
<td>Cut=unmethylated</td>
</tr>
<tr>
<td>XmnI</td>
<td>GAANN-NNTTC</td>
<td>Cut=methylated</td>
</tr>
<tr>
<td>MboI</td>
<td>-GATC</td>
<td>Cut=methylated</td>
</tr>
</tbody>
</table>

### Table 2.3 Antibodies used for meDIP

<table>
<thead>
<tr>
<th>Antibody Name</th>
<th>Host</th>
<th>Amount (µg)</th>
<th>Amount (µl)</th>
<th>Catalog Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG</td>
<td>Rabbit</td>
<td>4 µg</td>
<td>4 µl</td>
<td>Millipore-12-1370</td>
</tr>
<tr>
<td>5-mC (5-methyl-cytosine)</td>
<td>Mouse</td>
<td>4 µg</td>
<td>4 µl</td>
<td>CalbioChem-NA81</td>
</tr>
</tbody>
</table>

### Table 2.4 Antibodies used for ChIP in tissues and ESCs

<table>
<thead>
<tr>
<th>Antibody Name</th>
<th>Host</th>
<th>Amount (µg)</th>
<th>Amount (µl)</th>
<th>Catalog Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG</td>
<td>Rabbit</td>
<td>3 µg</td>
<td>3 µl</td>
<td>Millipore-12-1370</td>
</tr>
<tr>
<td>H3Ac</td>
<td>Rabbit</td>
<td>3 µg</td>
<td>3 µl</td>
<td>Millipore-06-599</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Rabbit</td>
<td>3 µg</td>
<td>3 µl</td>
<td>Millipore-17-614</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Mouse</td>
<td>3 µg</td>
<td>15 µl</td>
<td>Abcam-6002</td>
</tr>
</tbody>
</table>
Table 2.5 Sources used for data analysis and publicly available data

<table>
<thead>
<tr>
<th>Source</th>
<th>Web Site/Relevant paper</th>
<th>Use(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse Phenome Database</td>
<td><a href="http://phenome.jax.org/">http://phenome.jax.org/</a></td>
<td>SNP database</td>
</tr>
<tr>
<td>McGill University and Genome Quebec Innovation Centre</td>
<td><a href="http://gginnovationcenter.com/services/nanuq.aspx?l=e&amp;s=seq">http://gginnovationcenter.com/services/nanuq.aspx?l=e&amp;s=seq</a></td>
<td>Nanuq sequencing service-bisulfite sequencing</td>
</tr>
<tr>
<td>Quma: Quantification Tool for Methylation Analysis</td>
<td><a href="http://quma.cdb.riken.jp/">http://quma.cdb.riken.jp/</a></td>
<td>Analysis of bisulfite sequencing data</td>
</tr>
<tr>
<td>Peak Picker Software</td>
<td><a href="http://genome.cshlp.org/content/15/11/1584.full">http://genome.cshlp.org/content/15/11/1584.full</a></td>
<td>Allelic quantification of expression</td>
</tr>
<tr>
<td>UCSC Genome Browser</td>
<td><a href="http://genome.ucsc.edu/">http://genome.ucsc.edu/</a></td>
<td>CpG Island definition/structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publicly available tracks: Broad H3 ChIP-seq data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Custom loaded tracks (Lorincz Lab): Mouse ESCs RNA-seq</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOPHAT/cufflinks program (predicted transcripts)</td>
</tr>
<tr>
<td>Gene (Region)</td>
<td>Primer 1 (5’-3’)</td>
<td>Primer 2 (5’-3’)</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>B3galtl (+LTR)</td>
<td>TTTTTTTTTAAGGAGAATTITTTT</td>
<td>TTATTTTTTGATTGGTGTAGTTTAT</td>
</tr>
<tr>
<td>B3galtl (-LTR)</td>
<td>TGTTTTTTATAAAAATTGAAGATAAGTTTTTAT</td>
<td>GATAGTAAGGATTTATTTGTTAGGG</td>
</tr>
<tr>
<td>B3galtl (Exon)</td>
<td>ACCGTAACCTGGGAACATCT</td>
<td>Not applicable</td>
</tr>
<tr>
<td>B3galtl (LTR)</td>
<td>AAGTAGAACAATCGGTGACCAAC</td>
<td>Not applicable</td>
</tr>
<tr>
<td>B3galtl (CGI)</td>
<td>GTTTGTCAGAGGTCAAGAAAG</td>
<td>Not applicable</td>
</tr>
<tr>
<td>B3galtl (Exon)</td>
<td>AAGCAGAGATGGCCCTCTA</td>
<td>Not applicable</td>
</tr>
<tr>
<td>DNMT1 (Exon)</td>
<td>CCTAGTCGTCGCTAGCGAGAA</td>
<td>Not applicable</td>
</tr>
<tr>
<td>SETDB1 (Exon)</td>
<td>CTTCTGGCTCTGAGCCGGTATG</td>
<td>Not applicable</td>
</tr>
<tr>
<td>IAP</td>
<td>GCTCCTGAAGATGTGAAGCAATAAAG</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Hus1 (CGI)</td>
<td>TGTTTTTTTTAGTATTTTGGTTGG</td>
<td>GTTTAGAATTITTTGGGTGTTTGT</td>
</tr>
<tr>
<td>Gene (Region)</td>
<td>Primer 1 (5'-3')</td>
<td>Primer 2 (5'-3')</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><em>Hus1</em> (Int. Reg.)</td>
<td>TTTTTTAAGGATTATATTTGTA</td>
<td>GGTATTGGGATTATATTTTTT</td>
</tr>
<tr>
<td><em>Hus1</em> (LTR)</td>
<td>GTGGTTGTAAGTTAATTATTAATTTT</td>
<td>GGGGAGTTTTTGGTTTAGTATTTAAT</td>
</tr>
<tr>
<td><em>Hus1</em> (Exon)</td>
<td>AGGAAATGAAAGCTGCCAGA</td>
<td>ATGGACACACACATCCCAC</td>
</tr>
<tr>
<td><em>Pnpt1</em> (CGI)</td>
<td>AGGTTTTTTTTTAATTTATTTTTT</td>
<td>TTTCACCTATATCCCCAACATCAAAT</td>
</tr>
<tr>
<td><em>Pnpt1</em> (3'LTR)</td>
<td>GAGAGTTGTAAGTTAAGTAGATTTGTAATG</td>
<td>ATAAAAAAATCATATTTTCCTATAA</td>
</tr>
<tr>
<td><em>Pnpt1</em> (5'LTR)</td>
<td>TAGGTGGTAGATATTAGTTTGGTTTAAAAG</td>
<td>TTTAAAATTTGATTGATATTATATAT ATTTGG</td>
</tr>
<tr>
<td><em>Mthfd2l</em> (CGI)</td>
<td>GAGATTAGGGATTAGAGTGGGTTT</td>
<td>AAACAAAAACTAAAAAAACAAAAA</td>
</tr>
<tr>
<td><em>Mthfd2l</em> (5'LTR)</td>
<td>ATGGGGTGATTTGATTGATTTTTTTTAAATA</td>
<td>TTTATTGGTAGTTAATTAGTGAATTAG</td>
</tr>
<tr>
<td><em>Mthfd2l</em> (3'LTR)</td>
<td>TTTATTTGGAATTTGATTGATT</td>
<td>ATATTGGGATTATTGTATAGTATTAT</td>
</tr>
<tr>
<td><em>Atxn1l</em> (CGI)</td>
<td>GAATTGGTTGAGTTTGGTTTAA</td>
<td>TTTGGTTAATAGTTGTTTGTATT</td>
</tr>
</tbody>
</table>
Table 2.6 (continued) Primer sequences used in various experiments

<table>
<thead>
<tr>
<th>Gene (Region)</th>
<th>Primer 1 (5’-3’)</th>
<th>Primer 2 (5’-3’)</th>
<th>Primer 3 (5’-3’)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Atxn1l</em> (Int. Reg.)</td>
<td>GGGGTGGTGTGTTATGATTTTATA</td>
<td>TTGTGTAGGATTATTAGAAAAAG</td>
<td>CTAAACACAAAAACACTATT AACCAA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Atxn1l</em> (5’LTR)</td>
<td>GAATTGTTTGTTGTTTTTGGTTTTAA</td>
<td>TTTGGTTAATAGTGGTTTTTGTGTTTA</td>
<td>TCCCCAATAAAAATCTAAATTC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Atxn1l</em> (3’LTR)</td>
<td>TGGGTGGTTTATTTTTGTTTTTATTAAA</td>
<td>AAACTACATTCCCCAAATCACAACACT</td>
<td>TAAAAAACAAAACATCACAATTTTCCT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Catsper3</em> (5’LTR)</td>
<td>ATATAGTAGATGATTATTTTTGTT</td>
<td>GGGGTGGGATAATGTTATTTTATT</td>
<td>ACTATCTTAATTTCTTTTTAAAAAATTC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Catsper3</em> (3’LTR)</td>
<td>GGGAGATATGTTATTATTATTATTATTATTATTATTATGTTTTT</td>
<td>CACTCTAACCCTACTCAACAACAA</td>
<td>AATCATTACCTTCCCCAAATTACA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Catsper3</em> (Int. Reg.)</td>
<td>GGGTGGTGTGTTATTTTTTATGTGTTTTGTTT</td>
<td>TACAAGATTACAAACCAATAAACATTACACCCATTACACTTAAAATCAACCTAAATTCTTTT</td>
<td>CTTTTAAATTTAAATTCAATTCCT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Catsper3</em> (Prom.)</td>
<td>ATATAGTAGATGATTATTTTTGTT</td>
<td>ATATGTTTTTGTTGATAATATTGTTTTT</td>
<td>ACTAAATCTTAAAACACCACCTAAATTCTCT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Cyb5r1</em> (+ 3’LTR)</td>
<td>GGTAGAATTATTTAGTGATGATAAATTTTTT</td>
<td>TTTTTTAGTTGAAAATTGTGTTT</td>
<td>CTAAATCCCACAACCTAAATCCCAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Cyb5r1</em> (- 3’LTR)</td>
<td>TTTGAGGTGGTTGGAAGAGGTATT</td>
<td>TTTTTGGTTTTTTTTGAAAGAGAAGA</td>
<td>CTAAATCCCACAACCTAAATCCCAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td><em>Cyb5r1</em> (5’LTR)</td>
<td>TGATTGATTGGATGAATATTGTTTATAG</td>
<td>AAGGATTAGGAAATGAGGTTTTT</td>
<td>AAACATCTTTCTACAACCAACTAAACAC</td>
<td>Bisulfite Sequencing</td>
</tr>
</tbody>
</table>
Table 2.6 (continued) Primer sequences used in various experiments

<table>
<thead>
<tr>
<th>Gene (Region)</th>
<th>Primer 1 (5’-3’)</th>
<th>Primer 2 (5’-3’)</th>
<th>Primer 3 (5’-3’)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>CdGAP (CGI)</td>
<td>GGGGGTATATTATTTTTGTTT</td>
<td>TTTAGTTTTTTTTTTGTTTTT</td>
<td>ACAAAAAACAATCCATAAACAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>TTTTTTTGTTTTTTTGTATT</td>
<td>AAGATGGGGGTTTTGGAATTGTGATT</td>
<td>AAATACCTACCTGTTCAAACC TATATATA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>TATATTGGTTAAAAATAGTTAATA</td>
<td>GGGAGGAGAAGTTTTGTTATTATTATTAA</td>
<td>ATTACAAAAACTAAACCTA CAATAAA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>AAAAGAATAAAAAAGGTTTTTAAATATTTAAGGA</td>
<td>ACATAAATTCCAATAACTACAAAACC</td>
<td>AAAAAAACATCAATAATCAATAA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td></td>
<td>CAGGTCACATCCAAACGCACT</td>
<td>Not applicable</td>
<td>CAGCCTCTGTGGGGCTAACT</td>
<td>ChIP</td>
</tr>
<tr>
<td></td>
<td>TGAGTGCAGTCGAAAGGAATAGAA</td>
<td>Not applicable</td>
<td>CTCTGGACTAGCAGAGCCACTT</td>
<td>ChIP</td>
</tr>
<tr>
<td></td>
<td>GGCATGTTTTTCTGTTTAGT</td>
<td>Not applicable</td>
<td>TGAGTTGTCAAGGCACACTAAGC</td>
<td>ChIP</td>
</tr>
<tr>
<td></td>
<td>CGTTCTGTAATAAAACAGGAGTG</td>
<td>Not applicable</td>
<td>TAGAAAGGTTTCGTTTTTCTC</td>
<td>ChIP</td>
</tr>
<tr>
<td></td>
<td>CCAACTTAGAACACAGGATGCA</td>
<td>Not applicable</td>
<td>AGAGGCAGACAGGTACTGC</td>
<td>ChIP</td>
</tr>
<tr>
<td>Chr. 11 IAP (3’LTR)</td>
<td>TTTGTTTTTTTTTTTTTTTTTTT</td>
<td>ATTACTCTAAATAAAAAACCTATAAACCT</td>
<td>Bisulfite Sequencing</td>
<td></td>
</tr>
<tr>
<td>Gene (Region)</td>
<td>Primer 1 (5'–3')</td>
<td>Primer 2 (5'–3')</td>
<td>Primer 3 (5'–3')</td>
<td>Method</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Chr. 11 IAP (5’LTR)</td>
<td>GTAGTTGAATTTTTTTTTAATAGTTTGTGTA</td>
<td>CAAAACCTCCAAATTAAAAACTCAA</td>
<td>ATTTCCCAAAAAATAACAAACTC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 6 IAP (3’LTR)</td>
<td>TTTAATTTGATTTATTTGTATTTTTTAAA</td>
<td>TAAGTAAAGTTTGAATGGA GGGTTATTT</td>
<td>AAAAACATATCATCTTCTCAT AAAAATCAATA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 6 IAP (5’LTR)</td>
<td>ATGTGAAAAAGAATAAAGGTTTTTTAA</td>
<td>CAAAAACAAAAAACAACCAA CTTTCT</td>
<td>CTTATCAAAACATTTCATTATA CTACAAAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 10 IAP (3’LTR)</td>
<td>GTGTATTAGTAGTAAGAT AAATAAGATT</td>
<td>ATGAATAGATAAATAAGATTTT AGAATTTAAA</td>
<td>CCTTTCTAAAAAAACATAT CATCTTTCTAA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 10 IAP (5’LTR)</td>
<td>AGTTTTGGTTTGGGAATGAAGTATT</td>
<td>AAAAAAACAATACCCTTTTTTCTTAC</td>
<td>TTTCTCTAAATAAAAAAA CAATCCC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 1 IAP (3’LTR)</td>
<td>TTATTTTGTGGGTAATACTTTTGTGTTT</td>
<td>AATTGAATTTTTTTTGTATTGTTG</td>
<td>TAAAATTTAACCCTCTATCT CCCTCTCATTA</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 1 IAP (5’LTR)</td>
<td>GTAGTTGAATTTTTTTTTAATAG TTTGTGTTA</td>
<td>ATATTTTTATATCCCCCATAT ACCTTAAT</td>
<td>AACTAAAACCTTTAAAAACTCTTTAAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 7 IAP (3’LTR)</td>
<td>TGGTTTATTTTGTATTTATTAAAAG AAAAAGGGG</td>
<td>ATCAATCTTTAATAACTTTAATT CAAAACCA</td>
<td>CAAAAATATATATATCTAAA TCTAATTCCT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 7 IAP (5’LTR)</td>
<td>AATAGAAAAATTTAAGAATGGTTTTTTGT</td>
<td>TGATAAAAATTGGTTTGAAT AAAAATTATA</td>
<td>AATTCTAATTTCTAAAAATAAA TATCTCCT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Chr. 17 IAP (3’LTR)</td>
<td>GTGGTTTATTTTTATTTAATAA GGATAGGG</td>
<td>CTCTACTATAAAACCTCCATT AAAAAACTA</td>
<td>ATAAAAATATATACCTTTCCC TCCCTAC</td>
<td>Bisulfite Sequencing</td>
</tr>
</tbody>
</table>
Table 2.6 (continued) Primer sequences used in various experiments

<table>
<thead>
<tr>
<th>Gene (Region)</th>
<th>Primer 1 (5’-3’)</th>
<th>Primer 2 (5’-3’)</th>
<th>Primer 3 (5’-3’)</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr. 17 IAP (5’ LTR)</td>
<td>TATTGGGTAAAGGGAATTAATAAGTTTTT</td>
<td>GTTTGATTAGATTTTTTTTTTTTTATATGTTT</td>
<td>CCCCAAACAAAAAATTCTAAA TCTATACATAT</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Lair1 (5’ LTR)</td>
<td>AGGGAAGTTAGGTATATTTTTTTTTTTTTA</td>
<td>Not used</td>
<td>ACAATATCTCTCTCAAAAATCTTTAC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Lair1 (3’ LTR)</td>
<td>TTTTTTTATGGGATTAGATTATT</td>
<td>Not used</td>
<td>ACTATAAAACAAATTATACCTACCC</td>
<td>Bisulfite Sequencing</td>
</tr>
<tr>
<td>Lair1 (Exon)</td>
<td>TTCCTGAGGACAGATGGACA</td>
<td>Not applicable</td>
<td>CTGTGAGGGAGTGATGATGT</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>GAPDH (Exon)</td>
<td>AGAGAGGGAGGAGGGAATG</td>
<td>Not applicable</td>
<td>AACAGGGAGGAGCAGAGCAGAC</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>ACTB (Exon)</td>
<td>TCACCAAGGGTGAGAAAGTC</td>
<td>N/A</td>
<td>GGCCTTTTTAGGACTCTG</td>
<td>qRT-PCR</td>
</tr>
<tr>
<td>ALDH (Exon)</td>
<td>ATGGCCAGGGCGAGGAGT</td>
<td>N/A</td>
<td>ACTGCAGACGACCCTGTGAA</td>
<td>qRT-PCR</td>
</tr>
</tbody>
</table>
3 Results-The impact of ERVs on nearby genes is a locus-specific event

To investigate the impact of ERVs on nearby genes, nine genes were selected which had an ERV insertion located less than 2 kb away from the gene TSS (Figures 3.1 and 3.2). While most genes had a nearby polymorphic insertion, being absent in at least one of the four mouse strains, two genes, Atxn1l and Lair1, had a nearby insertion which was common among all four mouse strains (Figures 3.1 and 3.2). Interestingly, most genes, excluding Catsper3 and Lair1, have a CpG island promoter (Figures 3.1 and 3.2). Therefore, I was interested to see if CpG island promoters would be subject to ERV-mediated silencing since they are usually unmethylated regardless of gene expression. In this chapter, I will discuss results pertaining to six different genes listed in Figure 3.1. I will first discuss results pertaining to the B3galtl gene and provide evidence for ERV-mediated repressive mark spreading into the CpG island promoter of this gene. I will then provide an overview of genes analyzed whose CpG island promoters remain unmethylated regardless of a nearby methylated insertion. The remaining three genes, listed in Figure 3.2, will be discussed in chapter 4.

3.1 B3galtl

3.1.1 Function and expression patterns of B3galtl

B3galtl (β1, 3-galactosyltransferase-like gene) is a type of glycosyltransferase which functions in the modifications of proteins (Heinonen et al., 2006; Lesnik Oberstein et al., 2006). This type of post-translational modification which involves the addition of sugar moieties is known to be vital to cell signaling pathways. B3galtl is a highly conserved gene with orthologs in various eukaryotes ranging from insects to humans (Lesnik Oberstein et al., 2006). The mouse and human coding regions exhibit approximately 84% amino acid sequence identity in addition to similar expression patterns of the gene (Lesnik Oberstein et al., 2006). Previously published northern blot results measuring B3galtl mRNA levels in various adult Naval Medical Research Institute (NMRI) mouse tissues showed high expression corresponding to heart, kidney, liver, skin and small intestine whereas lowest expression was observed in spleen, testis, stomach, and muscle (Lesnik Oberstein et al., 2006). Interestingly, point mutations in the gene resulting in a truncated protein have been implicated in Peter’s plus syndrome in humans; a disorder characterized by various developmental defects (Heinonen et al., 2006).
3.1.2 Spreading of repressive marks from upstream LTR in ESCs

In the 129 allele of C57BL/6x129S6/SvEv Tac (B6x129) F1 hybrid mice, an antisense solitary LTR is located 368 bps upstream of the transcriptional start site (TSS) of B3galtl which has a promoter characterized as a CpG island (Figure 3.1). This LTR is not present in B6 mice. In order to determine if DNA methylation was spreading from the LTR into the CpG island promoter of the gene, bisulfite sequencing of the LTR or the empty site, the region upstream of the promoter, and the CpG island promoter region was completed by Liane Gagnier in B6x129 F1 hybrid mice. The presence of the LTR itself was used to distinguish between alleles with and without the insertion. Bisulfite sequencing results corresponding to the 129 allele in ESCs showed DNA methylation of the LTR, the intervening region between the LTR and the TSS, as well as the CpG island promoter region in a majority of cells (Figure 3.3). In cells that exhibited a lack of methylation at the CpG island despite the upstream insertion, the intervening region was also unmethylated (Figure 3.3). This region may therefore be important in determining whether or not methylation spreads into the CpG island promoter region.

Bisulfite sequencing results in ESCs corresponding to the C57BL/6 (B6) allele which does not contain the upstream insertion showed an overall absence of methylation at the B3galtl CpG island promoter region (Figure 3.3). Therefore, these results indicate a strong correlation between methylation of the B3galtl promoter in ESCs and the presence of the upstream insertion in the 129 allele.

DNA methylation spreading from the LTR upstream of B3galtl was further investigated by Rita Rebollo who completed bisulfite sequencing for the same regions mentioned above in J1 ESCs (129 origin) which are homozygous for the insertion. This data showed DNA methylation spreading from the LTR into the gene’s CpG island promoter in approximately 50% of cells (Figure 3.3). Similar to results observed in the 129 allele in ESCs, J1 ESCs which exhibited a lack at DNA methylation at the B3galtl promoter usually exhibited a lack of methylation at the intervening region between the CpG island and the insertion (Figure 3.3). This observation further supports the idea that this region is important in determining whether or not spreading into the CpG island occurs. Also similar to results in F1 hybrid ESCs, TT2 ESCs (B6xCBA F1) which are negative for the insertion exhibit an overall lack of DNA methylation at the B3galtl CpG island promoter region (Figure 3.3).

In order to correlate the observed spreading of DNA methylation with the spread of repressive chromatin, ChIP was completed in J1 and TT2 ESCs by Rita Rebollo. ChIP results in J1 ESCs showed relative enrichment of the active mark H3K4me3 and repressive marks H3K27me3, H3K9me3, and H4K20me3 at the region proximal to the LTR and at the B3galtl CpG island promoter (Figure 3.4). Interestingly, it has been previously shown that both H3K9me3 and H4K20me3 specifically mark repeats
such as ERVs in ESCs (Mikkelsen et al., 2007; Matsui et al., 2010). TT2 ESCs, however, exhibited a lack of enrichment of H3K9me3 and H4K20me3 at the empty site and the gene’s CpG island promoter region (Figure 3.4). These results indicated that enrichment and spreading of these two repressive marks into the B3galTL CpG island promoter was specifically associated with the presence of the insertion similar to DNA methylation. In addition to being observed in J1 ESCs, H3K27me3 enrichment is also found at the B3galTL promoter and upstream region in TT2 ESCs which do not contain the upstream LTR (Figure 3.4). While H3K27me3 is known as a repressive modification, it does not seem to be specifically associated with the presence of the LTR at the B3galTL locus.

### 3.1.3 Impact of repressive mark spreading on B3galTL expression in ESCs

To investigate whether the presence of the upstream LTR had an impact on B3galTL transcription in F1 hybrid ESCs, allelic quantification of gene expression was completed by Liane Gagnier. The results showed reduced gene expression from the LTR-containing allele which correlates with the presence of DNA methylation at the B3galTL CpG island promoter observed by bisulfite sequencing (Figure 3.5). A similar reduction in B3galTL expression in J1 ESCs, containing the insertion, in comparison to TT2 ESCs, without the insertion, was shown by qRT-PCR (Figure 3.5). These results provided evidence for a relationship between the presence of the LTR insertion, repressive marks at the B3galTL CpG island promoter, and reduced B3galTL gene expression.

### 3.1.4 Lack of repressive mark spreading from LTR in tissues

To determine whether spreading of repressive marks from the LTR into the CpG island promoter of B3galTL could be observed in tissues, bisulfite sequencing was completed by Liane Gagnier using DNA from brain and kidney from F1 hybrid mice. Surprisingly, despite methylation of the LTR in the 129 allele, the CpG island promoter of B3galTL remained unmethylated in both tissue types similar to the results obtained for the same region in the B6 allele without the insertion (Figure 3.6). Interestingly, bisulfite sequencing results corresponding to the intervening region between the LTR and B3galTL promoter in ESCs showed variable methylation in both tissues (Figure 3.6). This result indicates that DNA methylation may be spreading from the LTR into the transition region, however, it stops before reaching the gene’s CpG island promoter.

Liane Gagnier completed allelic quantification of B3galTL expression in brain in order to investigate the impact of the LTR on gene expression. The results showed approximately equal
expression from both alleles, with and without the insertion (Data not shown). According to these results, the presence of the LTR is not significantly impacting \textit{B3galtl} allelic expression, which correlates with the lack of DNA methylation spreading observed from the LTR into the gene’s CpG island promoter.

\textbf{3.1.5 \textit{B3galtl} expression is not significantly rescued upon DNMT1 and SETDB1 knockdown in ESCs}

Due to the observation that spreading of repressive modifications from the LTR into the \textit{B3galtl} CpG island promoter was limited to ESCs, I sought to investigate possible epigenetic differences between ESCs and tissues. A previous study examined relative enrichment of H3K9me3 at various ERV families in ESCs and differentiated cell types such as NPCs (neural precursor cells) and MEFs (mouse embryonic fibroblasts). Significant enrichment of H3K9me3 at IAPs was observed in ESCs whereas enrichment was lost in differentiated cell types (Mikkelsen \textit{et al.}, 2007; Matsui \textit{et al.}, 2010). Furthermore, it has been previously shown that H3K9me3 is required for ERV silencing in ESCs (Matsui \textit{et al.}, 2010). Therefore, I developed the hypothesis that H3K9me3 is required for DNA methylation spreading, and the lack of spreading into the \textit{B3galtl} CpG island promoter observed in tissues is due to the loss of H3K9me3 enrichment at the upstream LTR insertion. In order to investigate if H3K9me3 was required for DNA methylation spreading, a transient siRNA knockdown (KD) of SETDB1 and DNMT1 was completed by Irina Maksakova in J1 ESCs which are homozygous for the LTR insertion. SETDB1 is the enzyme responsible for depositing the H3K9me3 mark and DNMT1 is the major maintenance methyltransferase in mammals (Li \textit{et al.}, 1992; Matsui \textit{et al.}, 2010). We hypothesized that upon knockdown of these two enzymes, \textit{B3galtl} expression would be significantly increased due to the absence of repressive modifications spreading from the LTR into the gene’s CpG island promoter.

To investigate this hypothesis, I first completed qRT-PCR to measure \textit{DNMT1} and \textit{SETDB1} expression in J1 ESCs transfected with scrambled, DNMT1, SETDB1, or DNMT1 and SETDB1 siRNA. The results showed \textit{DNMT1} expression as significantly reduced in cells transfected with \textit{DNMT1} or \textit{DNMT1} and \textit{SETDB1} siRNA (Figure 3.7). Also, \textit{SETDB1} expression was significantly reduced in cells transfected with \textit{SETDB1} or \textit{DNMT1} and \textit{SETDB1} siRNA (Figure 3.7). These results provided evidence for the knockdowns as both specific and efficient at day 1. Surprisingly, qRT-PCR results of \textit{B3galtl} expression at day 4 in J1 ESCs transfected with both \textit{DNMT1} and \textit{SETDB1} siRNA showed no significant increase compared to gene expression in J1 ESCs transfected with scramble siRNA (Figure 3.7).

Results of an earlier study had shown that upon efficient knockdown of both \textit{DNMT1} and \textit{SETDB1} in ESCs, expression of the IAP-ez subfamily was significantly upregulated at day 4 compared to
cells transfected with only SETDB1 siRNA (Karimi et al., 2011). Therefore, to further evaluate the efficiency of the provided knockdown samples, I measured overall IAP expression at day 4 by qRT-PCR. The results showed that while IAP expression was increased in J1 ESCs transfected with DNMT1 and SETDB1 siRNA compared to SETDB1 siRNA alone, the increase was not as significant as observed in the earlier study (Figure 3.8). I further evaluated knockdown efficiency by completing meDIP to examine 5-mC enrichment at the region proximal to the LTR and the B3galTl CpG island promoter on day 4 of the DNMT1 and SETDB1 knockdown in J1 ESCs. The results showed a significant drop in 5-mC enrichment at both regions in the DNMT1 and SETDB1 knockdown sample compared to J1 ESCs transfected with scramble siRNA (Figure 3.8). Despite the significant drop in methylation observed at the LTR and B3galTl promoter, 5mC enrichment is still higher than the IgG background control at both regions (Figure 3.8). Therefore, both of these results may indicate the presence of residual H3K9me3 or DNA methylation at the upstream LTR which is able to spread into the CpG island promoter of B3galTl and subsequently affect expression. However, additional experiments could be performed to investigate this scenario further.

3.2 Overview of analyzed genes with nearby ERV insertions which are not affected by “spreading”

Due to the relatively large amount of raw data generated for genes with a nearby polymorphic ERV insertion, this section will consist of a brief overview of results for genes where I did not observe any significant evidence for ERV-mediated repressive mark spreading (Figure 3.1). While the raw bisulfite sequencing results will not be presented for these cases, the methylation data will be presented in a table in order to keep the results section of this thesis at a manageable length. Methylation data will be represented as “average % methylation” which is equal to % methylation for each clone summed and divided by the total number of clones. Therefore, a region that is 100% methylated exhibits methylation of all CpG sites in all sequenced clones.

3.2.1 Hus1

*Hus1* encodes a cell cycle check point protein that together with Rad1 and Rad9 forms what is known as the “9-1-1” complex. This complex is a vital component of the DNA damage response pathway (Levitt et al., 2005). A previous study that measured *Hus1* expression in various mouse tissues showed ubiquitous expression patterns. However, tissues such as thymus, ovary, muscle, brain, and the embryo were sites of particularly high *Hus1* expression (Weiss et al., 1999). In B6 mice, a solitary sense
LTR is located 1515 bps upstream of the TSS of Hus1 which has a CpG island promoter (Figure 3.1). To determine if methylation was spreading from the LTR into the CpG island promoter, I completed bisulfite sequencing corresponding to the LTR, the intervening region between the LTR and the gene TSS, and the CpG island promoter of the gene.

The average % methylation of the Hus1 CpG island is less than 2% in F1 hybrid ESCs, embryo, thymus, and brain corresponding to both alleles, regardless of the presence of the LTR upstream (Table 3.1). The upstream LTR exhibits an average % methylation of close to 100% in both F1 hybrid ESCs and brain which indicates that either DNA methylation is restricted to the LTR or perhaps it is spreading from the LTR but stops before reaching the CpG island (Table 3.1). In F1 hybrid ESCs and brain, the intervening region upstream of the CpG island promoter, exhibits an average methylation level of 70-90% regardless of the presence of the insertion (Table 3.1). In F1 hybrid embryo tissue, the average methylation (86%) of this region corresponding to the LTR-containing allele is higher than the same region in the allele negative for the insertion (50%) (Table 3.1). However, due to the lack of a large number of sequenced clones, it is difficult to determine if this methylation difference is significant. Overall, the results indicate that the Hus1 CpG island promoter is unmethylated and the region upstream of the promoter is methylated regardless of the presence of the insertion.

To determine if the presence of the upstream solitary LTR was negatively impacting Hus1 expression, I completed allelic quantification to quantify expression from each allele, with and without the insertion. Allelic quantification results in F1 hybrid ESCs, embryo, and thymus showed that the presence of the LTR did not significantly reduce Hus1 expression from the B6 allele (Figure 3.9). This result is not surprising since spreading of DNA methylation from the LTR into the Hus1 CpG island promoter was not observed. Interestingly, Hus1 expression is actually slightly higher from the B6 allele compared to the 129 or AJ allele without the insertion (Figure 3.9). Therefore, it is possible that the upstream LTR could be acting as an enhancer of Hus1 expression, but this possibility was not investigated further.

3.2.2 Pnpt1

Pnpt1 encodes an enzyme known as “pnpase” which is widely conserved and found in bacteria, plants, and mammals (Chen et al., 2006). Previous studies have provided evidence for a role of this enzyme in the import of RNA into the mitochondria as well as mRNA processing and stability (Walter et al., 2002; Wang et al., 2010). While Pnpt1 expression is likely important in all cells and tissue types, a
previous study specifically showed high gene expression in post-implantation embryos (Leszczyniecka et al., 2003).

**Pnpt1** has a promoter characterized as a CpG island and in B6 and 129 mice, a full-length sense IAP insertion is located 973 bps away from the TSS of the gene (Figure 3.1). To determine if DNA methylation was spreading from the IAP into the Pnpt1 CpG island promoter, I completed bisulfite sequencing using DNA from cells and tissues of F1 hybrid (B6xAJ) mice. Average % methylation of the CpG island promoter is less than 2% in F1 hybrid (B6xAJ) thymus tissue with respect to both alleles with and without the insertion (Table 3.2). A similar trend was observed in B6 ESCs which are homozygous for the upstream insertion (Table 3.2). To investigate if the full-length insertion upstream of Pnpt1 was marked by repressive modifications, I completed bisulfite sequencing of the 5’ and 3’ LTR of the IAP. Both LTRs exhibited an average % methylation of close to 100% in F1 hybrid thymus tissue corresponding to the B6 allele (Table 3.2). The 3’LTR of the IAP positioned closest to the gene exhibited similar methylation levels in F1 hybrid embryo and in B6 ESCs (Table 3.2). Therefore, these results indicate that the insertion upstream of Pnpt1 is targeted by DNA methylation and spreading of this mark is either not occurring or spreading of this mark stops before reaching the gene promoter which is unmethylated.

### 3.2.3 Mthfd2l

In B6 and AJ mice, a full-length antisense ETn is located 754 bps upstream of the TSS of Mthfd2l which has a CpG island promoter (Figure 3.1). This gene encodes an enzyme which functions in metabolic reactions in the cytoplasm and mitochondria and is widely expressed in various adult rat tissues such as brain, heart, kidney, lung, and testis (Bolusani et al., 2011). To investigate whether the insertion was marked by repressive modifications and if these modifications were spreading, COBRA and bisulfite sequencing were completed by Liane Gagnier to examine the methylation status of the ETn and the nearby CpG island promoter of Mthfd2l.

COBRA results show that, in a variety of cells and tissues such as B6 ESCs and embryo in addition to F1 hybrid brain, the CpG island promoter region of the gene is unmethylated regardless of the presence of the nearby insertion (Table 3.3). In contrast, the average % methylation of the 5’ and 3’ LTR of the ETn is approximately 80% in F1 hybrid brain tissue (Table 3.3). A similar methylation pattern was observed in B6 embryo at the 5’LTR which is the LTR positioned closest to the gene (Table 3.3). Interestingly, the insertion is more variably methylated in B6 ESCs as represented by the average methylation level of around 50% for both LTRs (Table 3.3). This observation is supported by previous
work which has provided evidence for ETns as being variably methylated in different cells and individuals (Reiss et al., 2010). Overall, these results indicate that the CpG island promoter of Mthfd2l remains unmethylated regardless of a methylated ETn insertion upstream.

### 3.2.4 Atxn1l

Atxn1l was initially identified as a paralog of Atxn1 which is involved in the development of SCA1 which is a neurodegenerative disease. Interestingly, overexpression of Atxn1l was found to reduce neurotoxicity in mice with a phenotype that mimics SCA1 pathology (Bowman et al., 2007). Therefore, Atxn1l is a gene which may be especially important in the brain.

Located 1316 bps upstream of the TSS of Atxn1l is a full-length antisense IAP. The promoter of Atxn1l is characterized as a CpG island (Figure 3.1). In contrast to other genes mentioned above, the IAP insertion is not known to be polymorphic since it is present in all four mouse strains analyzed previously (Zhang et al., 2008). COBRA and bisulfite sequencing were completed by Liane Gagnier to determine if DNA methylation was spreading from the insertion and subsequently into the Atxn1l CpG island promoter. COBRA results show a lack of methylation at the CpG island promoter in tissues such as liver and brain, while the average methylation level of the 5’LTR, positioned closest to the gene, is close to 100% (Table 3.4). Interestingly, the intervening region between the insertion and the gene TSS exhibits an average methylation level of 64% in brain tissue; indicating that perhaps methylation is spreading from the insertion (Table 3.4). However, this region cannot be compared to the same region without the upstream insertion since the insertion is not polymorphic in the four analyzed mouse strains. These results indicate that despite methylation of the insertion and possible spreading of this mark into the transition region, the CpG island promoter of Atxn1l remains unmethylated.

### 3.2.5 Catsper3

Catsper3 belongs to the sperm cation channel-like protein family which is a group of proteins specifically expressed in testis. Previous studies have provided evidence for calcium channels as vital for proper sperm function. Interestingly, a previous study in mice has shown the time course of Catsper3 expression which begins in the testis at post natal day 15 and gradually increases with age. Furthermore, this study established a correlation between high Catsper3 expression and high sperm motility in human sperm samples (Li et al., 2007).

In B6 mice, a full-length sense IAP insertion is located 832 bps upstream of the Catsper3 TSS (Figure 3.1). In contrast to above-mentioned genes, this gene has a non-CpG island promoter and its
expression is tissue specific, being exclusively expressed in the testis. To investigate if DNA methylation was spreading from the IAP into the gene promoter, I completed bisulfite sequencing in F1 hybrid testis and kidney, where \textit{Catsper3} is expressed and not expressed, respectively. In both tissues for the LTR-containing allele (B6), the 5’ and 3’LTR of the IAP exhibit an average % methylation of close to 100% (Table 3.5). A similar average % methylation is also observed for the intervening region between the insertion and the TSS in kidney and testis (Table 3.5). Furthermore, the region upstream of the TSS corresponding to the AJ allele, negative for the insertion, exhibits an average methylation level of close to 100% in F1 hybrid testis (Table 3.5). Therefore, the region upstream of the \textit{Catsper3} TSS is methylated regardless of the presence of the insertion and regardless of whether or not the gene is expressed.

To determine if the presence of the IAP was negatively impacting testis-specific \textit{Catsper3} expression, Liane Gagnier completed qRT-PCR to measure gene expression in B6 and AJ testis, homozygous and negative for the insertion, respectively. The results show approximately equal \textit{Catsper3} expression in B6 versus AJ testis, indicating that the presence of the IAP is not significantly impacting tissue-specific expression (Data not shown).
Figure 3.1 Cases of genes with nearby ERV insertions and status of ERV in four different mouse strains which are discussed in chapter 3. All genes, excluding Catpser3, have a CpG island promoter and an ERV insertion located less than 2 kb from the gene TSS. All genes are polymorphic with the exception of the case featuring the Atxn1l gene where a common IAP insert is located upstream. Large arrow represents solitary LTR while rectangle represents full-length insertion. Box represents gene promoter region and small arrow indicates location of TSS.
Figure 3.2 Cases of genes with nearby ERV insertions and status of ERV in four different mouse strains which are discussed in chapter 4. All genes, excluding Lair1, have a CpG island promoter and an ERV insertion located less than 2 kb from the gene TSS. All genes are polymorphic with the exception of the case featuring the Lair1 gene where a common ETn insertion is located in the first intron. Rectangle represents full-length insertion. Box represents gene promoter region and small arrow indicates location of TSS.
Figure 3.3 DNA methylation spreading is observed from an upstream antisense solitary LTR into the *B3galtl* CpG island promoter in LTR-containing ESCs. (A) In the 129 allele of F1 hybrid ESCs and in J1 ESCs, DNA methylation is observed at the LTR, the intervening region between the LTR and the promoter, and at the CpG island promoter of the gene in some clones. (B) The B6 allele of F1 hybrid ESCs and TT2 ESCs, which do not contain the upstream insertion, showed relatively low levels of DNA methylation at the CpG island promoter of *B3galtl*. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites. Bisulfite sequencing was completed by Liane Gagnier and Rita Rebollo.
Figure 3.4 Spreading of repressive chromatin marks is observed from an upstream antisense solitary LTR into the B3galtl CpG island promoter in J1 ESCs. (A) In J1 ESCs, homozygous for the insertion, enrichment of active mark H3K4me3 in addition to repressive marks H3K27me3, H3K9me3, and H4K20me3 is observed at the region flanking the LTR in addition to the CpG island. (B) In TT2 ESCs, which are negative for the insertion, enrichment of H3K9me3 and H4K20me3 is absent from the empty site in addition to the CpG island promoter region. Numbers and arrows indicate locations of qPCR primers. Error bars represent standard deviation of two biological replicates. ChIP qPCR was completed by Rita Rebollo.
Figure 3.5 Presence of upstream LTR impacts B3galtsl expression in F1 hybrid and J1 ESCs. (A) Allelic results in F1 hybrid ESCs show reduced B3galtsl from LTR-containing allele in black. (B) qRT-PCR results showed significantly reduced B3galtsl transcripts in J1 ESCs homozygous for the upstream insertion as compared to TT2 ESCs, negative for the upstream LTR insertion. Relative expression shown relative to J1 and normalized to ACTB (beta-actin) and ALDH (aldehyde dehydrogenase). Error bars represent standard deviation between two cDNA samples. Allelic quantification was completed by Liane Gagnier and qRT-PCR was completed by Rita Rebollo.
Figure 3.6 Spreading of DNA methylation from LTR into B3galTl CpG island promoter is not observed in tissues (A) Despite DNA methylation of the upstream LTR in the 129 allele in brain and kidney, this repressive mark does not spread into the gene promoter. (B) Similar to results observed in ESCs with respect the B6 allele, the CpG island promoter of B3galTl remains unmethylated in the absence of the upstream insertion. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites. Bisulfite sequencing was completed by Liane Gagnier.
Figure 3.7 Knockdown of DNMT1 and SETDB1 in J1 ESCs does not significantly rescue B3galTL expression.  
(A) In day 1 knockdown (KD) samples, DNMT1 expression is significantly reduced in J1 ESCs transfected with DNMT1 or DNMT1 and SETDB1 siRNA.  
(B) At day 1, SETDB1 expression is significantly reduced in J1 ESCs transfected with SETDB1 or DNMT1 and SETDB1 siRNA.  
(C) In day 4 knockdown (KD) samples, B3galTL expression is not significantly increased upon knockdown of DNMT1 & SETDB1 compared to cells transfected with scramble siRNA.  
Values are shown relative to scramble and normalized to ACTB and GAPDH.  
Error bars represent standard deviation between two cDNA samples.  
Knockdown experiments were completed by Irina Maksakova and I completed qRT-PCR.
Figure 3.8 Residual DNA methylation may remain following DNMT1 and SETDB1 knockdown (KD) in J1 ESCs. (A) At day 4, IAP expression is slightly higher in J1 ESCs transfected with DNMT1 and SETDB1 siRNA compared to cells transfected with SETDB1 alone. However, this result is not as dramatic as expected. Expression is shown relative to scramble and is normalized to GAPDH. Error bars presented as standard deviation of two cDNA samples. (B) At day 4, 5mC enrichment is significantly reduced at the LTR and B3galt1 promoter in J1 ESCs transfected with DNMT1 and SETDB1 siRNA compared to cells transfected with scramblesiRNA. Enrichment is shown as % of input DNA (Note: medIP has been completed with only one biological sample). Knockdown experiments were completed by Irina Maksakova and I completed qRT-PCR and MedIP-qPCR.
Figure 3.9 Upstream LTR does not negatively impact Hus1 expression as determined by allelic quantification. Hus1 expression is slightly higher in B6 allele (LTR-containing) in ESCs, thymus, and embryo; possibly indicating LTR as acting as enhancer of gene expression. I completed allelic quantification.
Table 3.1 Average % methylation of *Hus1* CpG island promoter, region upstream of promoter, and LTR in F1 hybrid cells and tissues. I completed bisulfite sequencing and COBRA to determine average % methylation.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present</th>
<th>LTR %Me†</th>
<th>Int. Region %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCs</td>
<td>B6 allele (B6x129 F1)</td>
<td>+</td>
<td>A-88% (11)*</td>
<td>A-90% (13)</td>
<td>A-2% (11)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D#</td>
<td>B-N/D</td>
<td>B-0% [1]**</td>
</tr>
<tr>
<td>ESCs</td>
<td>129 allele (B6x129 F1)</td>
<td>-</td>
<td>N/A#</td>
<td>A-78% (7)</td>
<td>A-1% (14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Embryo</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>N/D</td>
<td>A-86% (6)</td>
<td>A-1% (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Embryo</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A</td>
<td>A-50% (6)</td>
<td>A-1% (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Thymus</td>
<td>B6 allele (B6x129 F1)</td>
<td>+</td>
<td>N/D</td>
<td>N/D</td>
<td>A-1% (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Thymus</td>
<td>129 allele (B6x129 F1)</td>
<td>-</td>
<td>N/A</td>
<td>N/D</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Brain</td>
<td>B6 allele (B6x129 F1)</td>
<td>+</td>
<td>A-97% (13)</td>
<td>A-79% (17)</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-N/D</td>
<td></td>
</tr>
<tr>
<td>Brain</td>
<td>129 allele (B6x129 F1)</td>
<td>-</td>
<td>N/A</td>
<td>A-86% (7)</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td></td>
</tr>
</tbody>
</table>

+ indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

> %Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

** Numbers in brackets indicate number of CpG sites interrogated by COBRA

= N/D indicates bisulfite sequencing was not completed

# N/A indicates bisulfite sequencing was not completed because region was not present in allele
**Table 3.2** Average % methylation of *Pnpt1* CpG island promoter and upstream IAP in F1 hybrid cells and tissues. I completed bisulfite sequencing to determine average % methylation with the exception of B6 heart data, collected by Liane Gagnier.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present &amp;</th>
<th>5’LTR %Me&gt;</th>
<th>3’ LTR %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCs</td>
<td>B6</td>
<td>+</td>
<td>N/D</td>
<td>94% (11)*</td>
<td>2% (7)</td>
</tr>
<tr>
<td>Embryo</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>N/D</td>
<td>87% (16)</td>
<td>N/D</td>
</tr>
<tr>
<td>Thymus</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>97% (12)</td>
<td>95% (12)</td>
<td>1% (10)</td>
</tr>
<tr>
<td>Thymus</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A#</td>
<td>N/A</td>
<td>1% (11)</td>
</tr>
<tr>
<td>Heart</td>
<td>B6</td>
<td>+</td>
<td>N/D</td>
<td>96% (15)</td>
<td>N/D</td>
</tr>
</tbody>
</table>

& + indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

> %Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

= N/D indicates bisulfite sequencing was not completed

# N/A indicates bisulfite sequencing was not completed because region was not present in allele
Table 3.3 Average % methylation of *Mthfd2l* CpG island promoter and 5’ and 3’ LTR of upstream ETn in cells and tissues. Bisulfite sequencing and COBRA completed by Liane Gagnier in order to determine average % methylation.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present &amp;</th>
<th>3’LTR %Me*</th>
<th>5’ LTR %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCs</td>
<td>B6</td>
<td>+</td>
<td>A-41% (23)*</td>
<td>A-46% (16)</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-N/D</td>
<td>B-0% [1]**</td>
</tr>
<tr>
<td>Embryo</td>
<td>B6</td>
<td>+</td>
<td>N/D</td>
<td>A-85% (13)</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-0%</td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Brain</td>
<td>B6</td>
<td>+</td>
<td>A-78% (26)</td>
<td>A-80% (24)</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td>B-0%</td>
<td>B-0% [1]</td>
</tr>
<tr>
<td>Brain</td>
<td>AJ</td>
<td>+</td>
<td>N/D</td>
<td>A-78% (6)</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

& indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

* %Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

** Numbers in brackets indicate number of CpG sites interrogated by COBRA

= N/D indicates bisulfite sequencing was not completed
Table 3.4 Average % methylation of *Atxn1* CpG island promoter, region between CpG island and insertion, and 5’ and 3’ LTR of IAP in tissues. Bisulfite sequencing and COBRA completed by Liane Gagnier in order to determine average % methylation.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present</th>
<th>3’LTR %Me</th>
<th>5’LTR %Me</th>
<th>Int. Region %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver</td>
<td>B6</td>
<td>+</td>
<td>A-N/D†</td>
<td>A-96% (19)*</td>
<td>N/D</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-60-80% [2]**</td>
<td>B-60-80% [1]</td>
<td></td>
<td>B-0% [2]</td>
</tr>
<tr>
<td>Brain</td>
<td>B6</td>
<td>+</td>
<td>N/D</td>
<td>A-96% (16)</td>
<td>A-64% (26)</td>
<td>A-N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B-N/D</td>
<td></td>
<td>B-0% [2]</td>
</tr>
</tbody>
</table>

&
+ indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

%Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

** Numbers in brackets indicate number of CpG sites interrogated by COBRA

= N/D indicates bisulfite sequencing was not completed

# N/A indicates bisulfite sequencing was not completed because region was not present in allele
Table 3.5 Average % methylation of 5’ and 3’ LTR of IAP and region upstream of Catsper3 TSS in expressing and non-expressing tissues. I completed bisulfite sequencing in order to determine average % methylation.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present*</th>
<th>5’LTR %Me*</th>
<th>3’ LTR %Me</th>
<th>Int. Region %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kidney</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>98% (23)*</td>
<td>98% (12)</td>
<td>96% (17)</td>
<td>N/A#</td>
</tr>
<tr>
<td>Kidney</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/D=</td>
</tr>
<tr>
<td>Testis</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>96% (24)</td>
<td>98% (8)</td>
<td>97% (15)</td>
<td>N/A</td>
</tr>
<tr>
<td>Testis</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>96% (14)</td>
</tr>
</tbody>
</table>

*: + indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

*: %Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

*: Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

*: N/D indicates bisulfite sequencing was not completed

*: N/A indicates bisulfite sequencing was not completed because region was not present in allele
4 Results- Cases of ERV insertions which exhibit differential methylation

In the previous chapter, I provided evidence for ERV-mediated gene silencing with respect to the gene known as B3gal1. However, in most other cases, the CpG island promoter of the corresponding gene remained unmethylated despite a nearby methylated insertion. In chapter 4, I will present results pertaining to three different genes (shown in Figure 3.2) where the nearby insertion is differentially methylated, with the 3’LTR, closest to the gene, exhibiting hypomethylation. First, I will discuss results pertaining to the Cyb5r1 gene where the upstream ETn is likely acting as a promoter/enhancer of the gene. Next, I will discuss two cases featuring the CdGAP and Lair1 genes, where an interesting relationship is observed with respect to 3’LTR hypomethylation and expression of the nearby gene.

4.1 Cyb5r1

Cyb5r1 encodes the enzyme cytochrome b5 reductase 1 which is a type of cytochrome b5 reductase. Cytochrome b5 reductases play an important role in metabolism in addition to drug transformation. Interestingly, these enzymes also play an important role in red blood cells by being involved in the conversion of methemoglobin to hemoglobin. Previous studies have provided evidence for a link between mutations in Cyb5r3, a cytochrome b5 reductase, and the development of methemoglobinemia. This disease is characterized by a reduction in hemoglobin levels and ultimately a decreased capacity of the blood as far as carrying oxygen; ultimately leading to cyanosis, mental retardation, and even death in some severe cases (http://www.ncbi.nlm.nih.gov/omim; (Wang et al., 2000). Cyb5r3 is characterized as a housekeeping gene which provides evidence for Cyb5r1 as also being ubiquitously expressed (Wang et al., 2000).

In B6 mice, a full-length sense ETn is located 398 bps upstream of the Cyb5r1 TSS, which has a promoter characterized as a CpG island (Figure 3.2). To determine if DNA methylation was spreading from the full-length ETn into the Cyb5r1 CpG island promoter, bisulfite sequencing in addition to COBRA was completed by Liane Gagnier. Regardless of the presence of the upstream ETn, the CpG island promoter of Cyb5r1 and the upstream region exhibited a relatively low average % methylation in B6 ESCs, and F1 hybrid tissues such as embryo, spleen, and testis (Table 4.1). Interestingly, average % methylation results in F1 hybrid tissues show differential methylation of the ETn with the 3’LTR positioned closest to the gene exhibiting significantly lower methylation levels (Table 4.1). The 5’LTR located furthest from the gene was associated with an overall higher average % methylation in various
tissues. A similar trend of differential methylation with respect to the ETn was also observed in B6 ESCs; however the result is not as drastic which may be due to a lack of a large number of bisulfite sequenced clones (Table 3.6).

To determine if the presence of the upstream ETn insertion was affecting \textit{Cyb5r1} expression, allelic quantification of expression was completed by Liane Gagnier. Allelic quantification results in F1 hybrid (B6xAJ) spleen and testis showed approximately equal expression from each allele (Figure 4.1). However, results in F1 hybrid (B6xAJ) embryo showed higher transcript levels from the allele with the ETn (Figure 4.1). Therefore, these results indicate that the ETn may be acting as a promoter or enhancer specifically in the embryo. Indeed, RNA-seq data generated by Matt Lorincz’s lab show sequence reads initiating from the 3’LTR into the gene in J1 and TT2 ESCs which are homozygous and heterozygous for the insertion, respectively (Figure 4.2). Furthermore, transcripts predicted in TT2 ESCs based on RNA-seq data show fusion transcripts between the 3’LTR and the \textit{Cyb5r1} gene (Figure 4.2). Therefore, this data provides further support for the ETn as acting as an embryo-specific promoter of the \textit{Cyb5r1} gene.

4.2 \textit{Lair1}

\textit{Lair1} is a gene which encodes a receptor in mice and humans which has been found on a variety of cells in the immune system and functions in the inhibition of certain immune system responses (Lebbink \textit{et al.}, 2004). In B6 mice, a full-length antisense ETn insertion is located 565 bps from the TSS of \textit{Lair1} (Figure 3.2). In contrast to the \textit{Cyb5r1} case, this gene does not have a CpG island promoter and the ETn insertion is actually located within the first intron of the gene (Figure 3.2).

To investigate the methylation status of the ETn in the first intron of \textit{Lair1}, Liane Gagnier completed bisulfite sequencing of the insertion in a panel of B6 cells and tissues which are homozygous for the insertion. Similar to the \textit{Cyb5r1} case, the ETn exhibited differential methylation in spleen tissue with the 3’LTR which is positioned closest to the TSS of \textit{Lair1} showing relatively low average % methylation (Table 4.2). The 5’LTR, however, exhibited over 90% average methylation (Table 4.2). Interestingly, 3’LTR hypomethylation was limited to spleen tissue, since ESCs, heart, and muscle exhibited over 50% average methylation of the 3’LTR (Table 4.2).

Liane Gagnier next completed qRT-PCR to determine if \textit{Lair1} expression could correlate with the observed differences in 3’LTR methylation. Interestingly, \textit{Lair1} expression is significantly higher in spleen compared to ESCs, heart, and muscle (Figure 4.3). Therefore, these results provide evidence for an interesting relationship between \textit{Lair1} expression and hypomethylation of the nearby 3’LTR.
4.3 *CdGAP*

4.3.1 *CdGAP* function & expression

*CdGAP*, also known as Cdc42-GTPase activating protein, belongs to the RhoGAP family of proteins which are important regulators of RhoGTPases. RhoGTPases are involved in a variety of important functions such as cell growth, cell proliferation, and cell migration. RhoGAPs serve as “off” switches by converting the active GTP bound form of GTPases to the inactive GDP bound form. Northern blot results of *CdGAP* expression in tissues of Balb/c mice showed that heart and lung are sites of high expression while tissues such as kidney, liver, and spleen exhibited lower levels of expression. Interestingly, truncating mutations of *CdGAP* have been found to be associated with Adams-Oliver syndrome in humans characterized by limb defects, the absence of hair and skin from the scalp, and a variety of other developmental abnormalities (Lamarche-Vane and Hall, 1998; Southgate *et al.*, 2011). These results point to the importance of *CdGAP* expression in normal development.

4.3.2 *CdGAP* CpG island promoter is unmethylated in ESCs and tissues

In B6 mice, a full-length sense IAP is located 1337 bps upstream of the TSS of *CdGAP* which has a promoter characterized as a CpG island (Figure 3.2). In order to investigate if DNA methylation was spreading from the full-length IAP insertion into the CpG island promoter of *CdGAP*, bisulfite sequencing was first completed at the CpG island promoter. The results in F1 hybrid (B6x129) ESCs showed that the *CdGAP* CpG island promoter is unmethylated in alleles with and without the upstream insertion (Figure 4.4). Similar results were observed at the CpG island promoter in B6 ESCs, which are homozygous for the upstream insertion (Figure 4.4). A lack of DNA methylation at the *CdGAP* CpG island promoter was also observed in both alleles in F1 hybrid (B6x129) tissues such as thymus and brain (Figure 4.5). Also, bisulfite sequencing results in B6 lung showed a similar trend at the CpG island promoter of *CdGAP* (Figure 4.5). Therefore, these results indicate that the *CdGAP* CpG island promoter remains unmethylated despite the presence of the upstream insertion.

4.3.3 Differential epigenetic modifications of IAP LTRs in ESCs and tissues

In order to investigate if the full-length IAP upstream of *CdGAP* was targeted by DNA methylation, bisulfite sequencing was completed at the 5’LTR of the IAP located approximately 7 kb
away from the CdGAP CpG island promoter and the 3’LTR of the IAP located approximately 1 kb from the CpG island promoter. The results corresponding to the B6 allele of F1 hybrid (B6x129) ESCs showed DNA methylation of both IAP LTRs (Figure 4.6). Bisulfite sequencing results in B6 ESCs, which are homozygous for the insertion, showed the 3’LTR as methylated similar to what was observed in F1 hybrid ESCs (Figure 4.6). Therefore, in ESCs, the IAP upstream of CdGAP is marked by DNA methylation while the CpG island promoter of the gene remains unmethylated.

In contrast to the results observed in ESCs, a different methylation pattern of the IAP upstream of CdGAP was observed in tissues. Bisulfite sequencing of DNA from F1 hybrid tissues such as brain and thymus corresponding to the B6 allele showed differential methylation of the IAP (Figure 4.7). The 5’LTR, located furthest from the gene, was methylated while the 3’LTR, located closest to the gene, showed low levels of methylation in a majority of sequenced clones (Figure 4.7). Furthermore, bisulfite sequencing results in B6 tissues such as lung, homozygous for the upstream insertion, also showed a low level of DNA methylation at the 3’LTR (Figure 4.7).

Due to the observation of differential DNA methylation at the 5’ and 3’ LTR of the IAP, I next investigated if the LTRs were also differentially marked by histone modifications by completing ChIP. ChIP results in B6 lung tissue showed enrichment of active marks such as H3-acetylation and H3K4me3 at the 3’LTR of the IAP closest to CdGAP, whereas enrichment of these active marks was reduced at the 5’LTR (Figure 4.8). Enrichment of the repressive mark H3K27me3 was not observed at either the 5’ or 3’LTR (Figure 4.8). Therefore, the 5’LTR of the IAP may be associated with other repressive chromatin marks or it may be silenced solely by DNA methylation in tissues. ChIP results in B6 ESCs, showed a lack of significant enrichment of H3 acetylation and H3K4me3 at both LTRs which correlates with the presence of DNA methylation as shown previously by bisulfite sequencing (Figure 4.8). Interestingly, the 3’LTR of the IAP showed enrichment for H3K27me3, while enrichment was not observed at the 5’LTR (4.8). While both LTRs are marked by DNA methylation in ESCs, they may be associated with different repressive chromatin marks. Overall, these results provide evidence for a relationship between a lack of DNA methylation at the 3’LTR of the upstream IAP and the presence of active histone marks in tissues.

4.3.4 Intervening region is differentially methylated in ESCs vs. tissues

Since methylation of the both the 5’ and 3’ LTR of the upstream IAP was observed in ESCs, I completed bisulfite sequencing of the intervening region between the IAP and the CdGAP CpG island promoter to investigate if DNA methylation was spreading from the IAP into this region. Bisulfite
sequencing data corresponding to the intervening region in B6 ESCs showed an interesting methylation pattern. A majority of the six CpG sites in the intervening region which are positioned closest to the IAP were methylated while the seven CpG sites in the intervening region closest to the CdGAP CpG island promoter showed relatively low levels of DNA methylation in most sequenced clones (Figure 4.9). Therefore, this result provided evidence that DNA methylation was spreading from the IAP insertion into the intervening region in ESCs. A boundary, however, may be present in this region which is acting to prevent DNA methylation from spreading further and subsequently into the CdGAP CpG island promoter. Bisulfite sequencing results corresponding to the same region in J1 ESCs, which are negative for the upstream insertion, showed lower levels of methylation (Figure 4.9). Therefore, the observed methylation of the intervening region in B6 ESCs may be due to the presence of the nearby insertion.

I next hypothesized that the intervening region in B6 tissues such as lung would be relatively unmethylated since previous bisulfite sequencing results showed an absence of DNA methylation of the upstream 3’LTR of the IAP. As expected, bisulfite sequencing results in B6 lung showed a lack of DNA methylation at the intervening region, similar to results obtained for the same region in AJ lung which is negative for the upstream insertion (Figure 4.10). Therefore, while DNA methylation is potentially spreading from the IAP towards the CdGAP CpG island promoter in ESCs, spreading of DNA methylation is not observed in tissues, which is not surprising due to the absence of DNA methylation observed at the 3’LTR of the IAP.

4.3.5 CdGAP is differentially expressed in ESCs and tissues

I next hypothesized that differences in CdGAP expression could possibly influence the different methylation patterns observed in ESCs and tissues with respect to the intervening region and the 3’LTR of the upstream IAP. To investigate expression differences, I completed qRT-PCR using cDNA from B6 ESCs and B6 lung, homozygous for the IAP insertion upstream of CdGAP. Interestingly, the results showed gene expression as being significantly lower in ESCs compared to lung tissue (Figure 4.11). Lack of CdGAP expression in ESCs was further confirmed by RNA seq data generated by Matt Lorincz’s lab. In both J1 and TT2 ESCs, which are negative and homozygous for the insertion, respectively, a significant number of reads were not observed. This result is consistent with a previous study that provided ChIP-seq evidence for the CdGAP CpG island promoter as bivalent in ESCs (Figure 4.12). Bivalent promoters are characterized by enrichment for active mark H3K4me3 and repressive mark H3K27me3. Genes with bivalent promoters are described as not being expressed in ESCs, but are “poised” for expression upon differentiation (Mikkelsen et al., 2007).
CdGAP expression in lung is consistent with a previous study completed in Balb/c mice which showed this tissue as a site of high CdGAP expression (Lamarche-Vane and Hall, 1998). Therefore, these expression results provide evidence for a relationship between CdGAP expression and the methylation status of the upstream 3’LTR of the IAP. An attractive hypothesis could be that as a result of low CdGAP expression, DNA methylation starts to spread from the IAP; however the CpG island promoter and perhaps an upstream boundary are resistant to the encroaching repressive mark. When CdGAP is highly expressed, the CpG island promoter and nearby nucleosomes become enriched for active marks, which ultimately affects the epigenetic state of the nearby 3’LTR of the upstream IAP. Since the 5’LTR of the IAP is located several kb away from the CpG island promoter, it remains marked by repressive modifications.

### 4.3.6 Presence of IAP insertion does not significantly impact CdGAP expression

To investigate if the presence of the IAP was impacting CdGAP expression, qRT-PCR was completed to compare gene expression in B6 tissues, homozygous for the insertion, to AJ tissues, negative for the insertion. The results showed no significant difference in CdGAP expression in B6 versus AJ kidney (Figure 4.13). A similar result was observed in B6 and AJ lung tissue (Figure 4.13). Therefore, these results provide evidence that the hypomethylated 3’LTR in tissues is likely not affecting gene expression by providing a promoter or enhancer.

### 4.3.7 Relationship between CdGAP expression and upstream 3’LTR methylation

To further establish a relationship between CdGAP expression and DNA methylation of the upstream 3’LTR, CdGAP expression was measured in a variety of tissues via qRT-PCR by Liane Gagnier. I then completed bisulfite sequencing of the upstream 3’LTR in the same tissues. The results showed an interesting correlation between expression and methylation. B6 heart tissue which exhibited relatively high CdGAP expression similar to what was observed in lung was associated with low DNA methylation at the upstream 3’LTR in some clones (Figure 4.14). In contrast, in B6 kidney tissue where CdGAP expression was significantly lower, the upstream 3’LTR exhibited higher DNA methylation levels similar to what was observed in B6 ESCs (Figure 4.14). Interestingly, in B6 thymus where CdGAP expression is low, bisulfite sequencing results of the 3’LTR showed an absence of DNA methylation in some sequenced clones (Figure 4.14). As bisulfite sequencing and qRT-PCR corresponding to each tissue were
not completed in the same mouse, it is difficult to establish a direct correlation between gene expression and LTR methylation. However, overall results point to an interesting trend involving increased CdGAP expression and hypomethylation of the nearby 3’LTR.

4.3.8 CdGAP CpG island promoter and upstream region are associated with different chromatin environments in ESCs and tissues

To further establish a role for CdGAP expression potentially affecting the epigenetic state of the upstream 3’LTR, I completed ChIP in order to examine the chromatin environment in ESCs and lung where CdGAP is not expressed or expressed, respectively. ChIP results corresponding to the CdGAP CpG island promoter show enrichment for active mark H3K4me3 and repressive mark H3K27me3 in ESCs. Enrichment of these marks at the CpG island promoter was observed in both B6 and J1 ESCs which are homozygous and negative for the IAP insertion, respectively (Figure 4.12). Therefore, these results indicate the CpG island promoter as bivalent in ESCs regardless of the presence of the insertion. These results are consistent with previous studies which have shown the CdGAP CpG island promoter as bivalent in ESCs (Mikkelsen et al., 2007). Furthermore, these results are consistent with the lack of CdGAP expression observed in ESCs via qRT-PCR.

Interestingly, ChIP results in B6 ESCs corresponding to the transition region show a reduction in H3K4me3 enrichment in addition to increased enrichment of H3K27me3 (Figure 4.15). This enrichment trend is also observed at the upstream 3’LTR with increased H3K27me3 enrichment while H3K4me3 enrichment is reduced close to background levels (Figure 4.15). While the intervening region was characterized by decreased H3K4me3 in B6 ESCs, H3K4me3 and H3-acetylation enrichment as well as slight H3K27me3 enrichment were maintained in this region in J1 ESCs which are negative for the insertion (Figure 4.15). Therefore, the 3’LTR in B6 ESCs is not only associated with DNA methylation, but enrichment for H3K27me3. These repressive modifications may be spreading into the intervening region, but the CdGAP promoter and the immediate upstream region appear resistant to the encroaching repressive marks.

A different chromatin environment with respect to the CpG island promoter of CdGAP and the upstream region was observed in tissues such as lung. In B6 lung which contains the insertion upstream of CdGAP, enrichment of H3K4me3 and H3-acetylation is observed at the CpG island promoter whereas H3K27me3 enrichment is not observed (Figure 4.16). Similar results are observed in AJ lung which is negative for the insertion upstream of CdGAP (Figure 4.16). These results are consistent with qRT-PCR results which showed increased CdGAP expression in lung compared to ESCs. I next investigated
enrichment of these marks at the upstream intervening region in order to determine if the absence of DNA methylation previously observed could be correlated with the presence of active histone marks. Despite lower enrichment of active marks at this region in comparison to the CpG island promoter, observed enrichment was higher than background levels (Figure 4.16). This enrichment was observed in both B6 and AJ lung, homozygous and negative for the insertion respectively (Figure 4.16). Enrichment of H3K4me3 and H3-acetylation was also observed in B6 lung at the 3’LTR which is consistent with the observed hypomethylation of this region as previously shown by bisulfite sequencing (Figure 4.16). In contrast to what was observed in ESCs, the repressive mark H3K27me3 is not enriched at the CdGAP CpG island promoter or the upstream region in both B6 and AJ lung (Figure 4.16). The presence of active marks and absence of repressive marks at the CpG island promoter, intervening region, and 3’LTR is consistent with CdGAP expression in lung in addition to hypomethylation shown at these regions. Therefore, these results show that in ESCs consistent with a lack of CdGAP expression, the CpG island promoter is bivalent. Outside of the CpG island promoter, active marks are reduced, while H3K27me3 enrichment remains. In lung tissue characterized by CdGAP expression, H3K27me3 enrichment is lost, and active marks are not only found at the promoter but also the upstream region which contains the 3’LTR of the IAP insertion. The presence of active marks such as H3K4me3 throughout this region correlates with the bisulfite sequencing data since a negative correlation is known to exist between H3K4me3 and DNA methylation (Miranda and Jones, 2007). Therefore, CdGAP expression is associated with the accumulation of active marks which may ultimately result in 3’LTR hypomethylation.

4.4 IAP 3’LTRs not generally subject to hypomethylation

In order to determine if 3’LTRAas of IAPs are generally subject to hypomethylation in the genome, Ying Zhang compiled a list of full-length IAPs located 10-20 kbs away from genes. I then completed bisulfite sequencing for 6 cases to measure the methylation status of the 5’ and 3’ LTRs. The results show high levels of methylation of both the 3’ and 5’ LTRs of IAP far away from genes indicating that 3’LTRs of IAPs do not seem to be more variably methylated than 5’LTRs (Table 4.3). Therefore, the 3’LTR hypomethylation that we are observing is likely related to the presence of a nearby gene.
4.5 CpG island structure not significantly different between case genes

While we observed CpG island promoters as generally unmethylated, we noted interesting differences in the methylation state of the nearby insertions. While most insertions were methylated, we observed a few cases of 3’LTR hypomethylation. We sought to determine if differences in CpG island structure of the promoter could potentially explain differences in methylation of the upstream insertion. For example, we could hypothesize that a longer CpG island consisting of a high number of CpG sites may correspond with higher enrichment of H3K4me3 and a more active chromatin environment which could subsequently affect the methylation state of the nearby insertion. Overall, however, we do not observe any significant differences in CpG island length, number of CpG sites, GC content, or observed/expected CpG ratio between the different cases (Table 4.4). However, if we compare between two similar cases, Pnpt1 and CdGAP, which both feature a full-length IAP insertion roughly 1 kb away from their TSSs, we observe an interesting difference. The CpG island promoter of CdGAP, where differential methylation of the IAP was observed, is almost 400 bps longer than the Pnpt1 CpG island promoter which features a methylated insertion upstream (Table 4.4). While this may provide evidence for the CdGAP CpG island as a “stronger” promoter, it is likely that other factors are involved in the observed methylation differences.
Figure 4.1 Allelic quantification results show higher Cyb5r1 expression from LTR-containing allele in embryo. Cyb5r1 expression is not significantly different between alleles in adult tissues such as testis and spleen. Allelic quantification completed by Liane Gagnier.

Figure 4.2 RNA seq data generated by Lorincz lab and predicted transcripts in J1 and TT2 mouse ESCs. RNA seq reads are show initiating in the ETn and going into the Cyb5r1 gene in both J1 and TT2 ESCs. Transcripts predicted based on RNA seq data in TT2 ESCS show transcripts coming from 3’LTR of ETn and splicing into Cyb5r1 gene. Predicted transcripts provided by Mohammad M. Karimi.
Figure 4.3 Relative *Lair1* expression measured by qRT-PCR in various B6 tissues. High *Lair1* expression is observed in spleen tissue whereas lower levels of expression are seen in ESCs, heart, and muscle. Expression is shown relative to B6 ESC values and normalized to *GAPDH*. Error bars represent standard deviation of two biological replicates. Note: only one biological replicate used for muscle. qRT-PCR performed by Liane Gagnier.
Figure 4.4 CpG island promoter of CdGAP is unmethylated despite presence of upstream insertion in ESCs. (A) In the B6 allele of F1 hybrid (B6x129) ESCs, the CpG island promoter of CdGAP is relatively unmethylated despite the presence of the upstream insertion. A similar result is also observed for the CpG island promoter in B6 ESCs; homozygous for the insertion. (B) Results corresponding to the 129 allele of F1 hybrid (B6x129) ESCs, which is negative for the insertion, also show a lack of DNA methylation at the CdGAP CpG island promoter. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
Figure 4.5 *CdGAP* CpG island promoter is unmethylated despite presence of upstream insertion in tissues. (A) In the B6 allele of F1 hybrid (B6x129) thymus and brain, bisulfite sequencing results show that the CpG island of *CdGAP* is relatively unmethylated despite the presence of the upstream insertion. A similar result is also observed for the CpG island promoter in B6 lung; homozygous for the upstream insertion. (B) Results corresponding to the 129 allele of F1 hybrid (B6x129) thymus, which is negative for the insertion, also show an absence of DNA methylation at the *CdGAP* CpG island promoter. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
Figure 4.6 Full-length IAP upstream of *CdGAP* CpG island promoter is methylated in ESCs. (A) In F1 hybrid (B6x129) ESCs corresponding to the B6 allele, both the 3'LTR and the 5'LTR of the IAP are methylated. (B) In B6 ESCs which are homozygous for the upstream insertion, the 3'LTR closest to the CpG island promoter of *CdGAP* is methylated, consistent with bisulfite sequencing results observed in F1 hybrid (B6x129) ESCs. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
Figure 4.7 Full-length IAP upstream of \textit{CdGAP} CpG island promoter is differentially methylated in tissues. In the B6 allele of F1 hybrid (B6x129) thymus (A) and brain (B), the 5’LTR located furthest from the CpG island promoter of \textit{CdGAP} is methylated whereas lower levels of methylation are observed at the 3’LTR of the IAP located closest to the CpG island promoter. In B6 lung (C), which is homozygous for the insertion, a similar methylation pattern is shown with the 3’LTR of the IAP. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
Figure 4.8 Full-length IAP upstream of CdGAP is marked by different histone modifications in ESCs and tissues. (A) ChIP results in B6 lung show enrichment of active marks H3-acetylation and H3K4me3 overlapping the 3’LTR, while enrichment of this active marks is reduced at the 5’LTR. (B) Both the 5’LTR and 3’LTR upstream of CdGAP in B6 ESCs are not significantly enriched for active marks H3-acetylation and H3K4me3 which correlates with bisulfite sequencing data. The 3’LTR exhibits enrichment for H3K27me3, while the 5’LTR lacks significant enrichment for this repressive mark. Enrichment is shown as relative to input DNA. (Note: H3K27me3 enrichment is shown as negative to represent it as a repressive mark.) Arrows indicate approximate location of qPCR primers.
Figure 4.9 Intervening region is differentially methylated in ESCs with and without the upstream insertion. (A) In B6 ESCs which are homozygous for the upstream insertion, the 6 CpG sites in the intervening region which are positioned closest to the upstream 3’LTR are mostly methylated. The 7 CpG sites closest to the CpG island promoter of \textit{CdGAP}, however, are relatively unmethylated. (B) In J1 ESCs which are negative for the upstream insertion, the intervening region is unmethylated in some sequenced clones. While some clones exhibit methylation of the 6 CpG sites furthest from the CpG island promoter, the overall methylation of this region is lower than observed for the same region in B6 ESCs. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
Figure 4.10 Intervening region is unmethylated in tissues regardless of upstream insertion. (A) In B6 lung tissues which is homozygous for the upstream insertion, the intervening is unmethylated in a majority of sequenced clones. (B) In AJ lung tissue which is negative for the upstream insertion, the intervening region is relatively unmethylated in most sequenced clones similar to what was observed in J1 ESCs. Bisulfite sequencing data is represented as white or black-colored circles representing the individual CpG site as unmethylated or methylated, respectively. The length of the line between circles represents the relative distance between CpG sites.
**Figure 4.11** *CdGAP* expression is significantly reduced in B6 ESCs. qRT-PCR results showed significantly lower *CdGAP* expression in B6 ESCs compared to B6 lung. Expression data is shown as relative to *beta-actin (ACTB)*. qRT-PCR was completed by Liane Gagnier.

**Figure 4.12** The CpG island promoter of *CdGAP* is enriched for active and repressive marks in ESCs consistent with a bivalent promoter. Previously published histone chip-seq results in mouse hybrid (B6x129) ESCs show enrichment of active mark H3K4me3 and repressive mark H3K27me3 overlapping the CpG island promoter (green bar) of *CdGAP*. Results in B6x129 brain tissue show loss of H3K27me3 and enrichment of H3K4me3 at the CpG island promoter consistent with *CdGAP* expression in tissues(Image from UCSC Genome Browser-Mikkelsen, *et al.* 2007).
Figure 4.13 Full-length IAP insertion does not significantly impact $CdGAP$ expression in tissues. No significant difference in $CdGAP$ expression is observed between B6 and AJ kidney tissue. A similar result is seen for B6 lung versus AJ lung tissue. Expression is shown normalized to $ALDH$. Error bars represent standard deviation of the mean of two cDNA samples.
Figure 4.14 Correlation observed between CdGAP expression and methylation of the upstream 3'LTR of the IAP. In B6 ESCs, characterized by lack of CdGAP expression, the upstream 3'LTR is mostly methylated. In B6 thymus, which exhibits low CdGAP expression, some cells exhibit a lack of methylation at the 3'LTR. This trend of some cells exhibiting a lack of methylation at the 3'LTR is also observed in B6 kidney and B6 heart where CdGAP is expressed at a low level or variably expressed, respectively. In B6 lung, characterized by consistently high CdGAP expression, the upstream 3'LTR is mostly unmethylated in all clones. Bisulfite sequencing and qRT-PCR not completed in some mice.
Figure 4.15 IAP insertion affects nearby chromatin environment in ESCs. (A) & (B) In B6 ESCs, the CpG island promoter of CdGAP is associated with enrichment for H3K4me3, H3K27me3, and H3-acetylation consistent with a bivalent domain. The region upstream of the CpG island and the 3\'LTR of the IAP are associated with decreased enrichment of H3K4me3 and increased enrichment for H3K27me3. (C) & (D) In J1 ESCs, the CpG island promoter is associated with H3K4me3, H3K27me3, and H3-acetylation consistent with a bivalent domain. The region upstream of the promoter is associated with active mark enrichment in contrast to what was observed for B6 ESCs. Enrichment is shown as relative to input DNA. B6 ESCs data represented by 3 biological replicates whereas J1 ESCs data represented by 2 biological replicates. H3K27me3 enrichment shown as negative to represent it as a negative mark. Error bars shown as standard error of the mean.
Figure 4.16 

CdGAP expression in tissues is associated with accumulation of active histone marks and loss of repressive marks. (top) In B6 lung, the CdGAP CpG island promoter in addition to the upstream region including the 3’LTR of the IAP are characterized by loss of H3K27me3 enrichment and maintained enrichment for active marks such as H3K4me3 and H3Ac. (bottom) In AJ lung, the CpG island promoter and upstream region are characterized by loss of H3K27me3 and enrichment for active marks similarly to B6 lung. However, active mark enrichment decreases with increasing distance from the TSS in contrast to B6 lung. H3K27me3 enrichment shown as negative to represent it as a negative mark. B6 lung data represented by 3 biological replicates whereas AJ lung data represented by 2 biological replicates Error bars shown as standard error of the mean.
Table 4.1 Average % methylation of *Cyb5r1* CpG island promoter, intervening region, and upstream ETn. COBRA and bisulfite sequencing completed by Liane Gagnier.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present &amp;</th>
<th>5’LTR %Me*</th>
<th>3’ LTR %Me</th>
<th>Int. Region %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCs</td>
<td>B6</td>
<td>+</td>
<td>A-56% (7)* B-N/D=</td>
<td>A-45% (5)</td>
<td>A-0% (5)</td>
<td>A-4% (5)</td>
</tr>
<tr>
<td>Embryo</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A#</td>
<td>N/A</td>
<td>A-14% (20)</td>
<td>A-1% (20)</td>
</tr>
<tr>
<td>Testis</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A</td>
<td>N/A</td>
<td>N/D</td>
<td>A-N/D B-0% [1]</td>
</tr>
<tr>
<td>Spleen</td>
<td>AJ allele (B6xAJ F1)</td>
<td>-</td>
<td>N/A</td>
<td>N/A</td>
<td>N/D</td>
<td>A-N/D B-0% [1]</td>
</tr>
</tbody>
</table>

& + indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

> %Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

** Numbers in brackets indicate number of CpG sites interrogated by COBRA

= N/D indicates bisulfite sequencing was not completed

# N/A indicates bisulfite sequencing was not completed because region was not present in allele
Table 4.2 Average % methylation of ETn upstream of *Lair1*. COBRA and bisulfite sequencing completed by Liane Gagnier.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele/Strain</th>
<th>LTR Present &amp;</th>
<th>5’LTR %Me*</th>
<th>3’ LTR %Me</th>
<th>Promoter %Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCs</td>
<td>B6</td>
<td>+</td>
<td>A-N/D</td>
<td>A-96% (13)*</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-60-80% [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart</td>
<td>B6</td>
<td>+</td>
<td>A-N/D</td>
<td>A-54% (8)</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-60-80% [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle</td>
<td>B6</td>
<td>+</td>
<td>N/D</td>
<td>A-96% (7)</td>
<td>N/D</td>
</tr>
<tr>
<td>Spleen</td>
<td>B6</td>
<td>+</td>
<td>A-93% (12)</td>
<td>A-12% (18)</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-60-80% [3]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thymus</td>
<td>B6</td>
<td>+</td>
<td>A-N/D</td>
<td>N/D</td>
<td>N/D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>B-60-80% [3]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

& + indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele

%Me corresponds to average % methylation determined by bisulfite sequencing (A) or COBRA (B)

* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me

** Numbers in brackets indicate number of CpG sites interrogated by COBRA

= N/D indicates bisulfite sequencing was not completed
Table 4.3 Average % methylation of IAPs far away from genes determined by bisulfite sequencing.

<table>
<thead>
<tr>
<th>Cells/Tissue Analyzed</th>
<th>Allele (Strain)</th>
<th>LTR Present*</th>
<th>Chromosome Location</th>
<th>5’LTR -%Me&gt;</th>
<th>3’LTR -%Me</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thymus</td>
<td>B6 (C57BL/6)</td>
<td>+</td>
<td>11</td>
<td>98% (10)*</td>
<td>97% (12)</td>
</tr>
<tr>
<td>Thymus</td>
<td>B6 (C57BL/6)</td>
<td>+</td>
<td>7</td>
<td>99% (11)</td>
<td>98% (10)</td>
</tr>
<tr>
<td>Brain</td>
<td>B6 allele (B6xAJ F1)</td>
<td>+</td>
<td>17</td>
<td>93% (15)</td>
<td>83% (12)</td>
</tr>
<tr>
<td>Brain</td>
<td>B6 allele (B6xAJ F1)</td>
<td>-</td>
<td>6</td>
<td>97% (14)</td>
<td>100% (11)</td>
</tr>
<tr>
<td>ESCs</td>
<td>B6 (C57BL/6)</td>
<td>+</td>
<td>10</td>
<td>N/D*</td>
<td>85% (14)</td>
</tr>
<tr>
<td>ESCs</td>
<td>B6 (C57BL/6)</td>
<td>+</td>
<td>1</td>
<td>N/D</td>
<td>94% (6)</td>
</tr>
</tbody>
</table>

* indicates LTR is present in corresponding allele; - indicates LTR is absent in corresponding allele.  
> %Me corresponds to average % methylation determined by bisulfite sequencing.  
* Numbers in parentheses indicate number of bisulfite sequence clones used to determine %Me.  
N/D indicates bisulfite sequencing was not completed.
Table 4.4 CpG island structure of case genes as defined by the UCSC genome browser

<table>
<thead>
<tr>
<th>Gene</th>
<th>CGI Length</th>
<th>CGI CpG Sites</th>
<th>CGI GC Content</th>
<th>CGI Obs/Exp CpG Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3galtl</td>
<td>758 bps</td>
<td>100</td>
<td>73%</td>
<td>1</td>
</tr>
<tr>
<td>Hus1</td>
<td>409 bps</td>
<td>41</td>
<td>69%</td>
<td>0.84</td>
</tr>
<tr>
<td>Pnpt1</td>
<td>280 bps</td>
<td>27</td>
<td>66%</td>
<td>0.88</td>
</tr>
<tr>
<td>Mthfd2l</td>
<td>325 bps</td>
<td>45</td>
<td>72%</td>
<td>1.06</td>
</tr>
<tr>
<td>Atxn1l</td>
<td>596 bps</td>
<td>48</td>
<td>65%</td>
<td>0.77</td>
</tr>
<tr>
<td>Cyb5r1</td>
<td>223 bps</td>
<td>21</td>
<td>70%</td>
<td>0.76</td>
</tr>
<tr>
<td>CdGAP</td>
<td>653 bps</td>
<td>63</td>
<td>62%</td>
<td>0.99</td>
</tr>
</tbody>
</table>
5 Discussion and conclusions

It is well-known that transposable elements are targeted by silencing mechanisms such as DNA methylation and repressive histone modifications in order to prevent these elements from wreaking havoc in the genome by disrupting or altering gene expression (Slotkin and Martienssen, 2007). Despite epigenetic modifications serving to restrain TE activity, it has been suggested that these silencing marks could potentially “spread” from the TE into adjacent sequences, thus altering the nearby chromatin environment. Recently, direct evidence for this phenomenon was provided in the mouse genome where IAP ERV insertions were shown to result in the induction and spread of the repressive chromatin mark H3K9me3 in ESCs for at least one kb and up to five kb in some cases (Rebollo et al., in press).

While TE-mediated gene silencing via “spreading” has been shown in plants, a similar effect has yet to be fully studied in mammalian genomes (Kinoshita et al., 2007; Martin et al., 2009). Therefore, I sought to investigate if an ERV insertion could induce spreading of repressive marks into a nearby gene, thus affecting gene expression. I initially chose to study polymorphic mouse ERVs located less than 2 kb away from the TSS of a gene. By utilizing polymorphic insertions, I was able to compare between genomic regions with the insertion versus the same region in another allele or another mouse strain negative for the insertion.

5.1 DNA methylation/repressive chromatin spreading from ERVs as a dynamic and locus-specific event in the B3galtl case

B3galtl is a particularly interesting gene as it is the only case where we actually observed spreading of repressive histone marks and DNA methylation initiating from the LTR into the CpG island promoter of the gene which correlated with reduced gene transcription. The observation of “spreading” and reduced gene expression in only one case can lead to the conclusion that spreading of repressive marks from ERVs into nearby genes is likely an uncommon event. This is not surprising since evidence has been provided in plants for negative selection against methylated TE insertions near genes (Hollister and Gaut, 2009). We observe in LTR-containing ESCs that not all sequenced clones exhibit B3galtl CpG island promoter methylation. Furthermore, while B3galtl expression is reduced in ESCs with the insertion, it is not completely abolished which allows us to conclude that the affect of spreading at the B3galtl locus may be subtle. Indeed, this gene is likely not essential in development since mouse strains containing the insertion are in fact viable. This could explain why the methylated insertion has been maintained in close proximity to the gene. It is also possible that this insertion is under weak negative
selection since, through analysis of genomic distribution patterns, our lab has shown that at least some young, polymorphic IAPs are likely still under selection (Zhang et al., 2008). In addition to likely being a locus-specific phenomenon, “spreading” is also perhaps a tissues-specific event since the B3gal tl CpG island promoter remains unmethylated despite an upstream methylated LTR in adult tissues. While the questionable results of the DNMT1/SETDB1 siRNA knockdown experiments cannot allow us to conclude whether or not H3K9me3 is required for “spreading”, we can hypothesize that other factors are likely involved. For example, the region adjacent to the LTR insertion appears important in determining whether or not methylation spreads into the CpG island promoter. When this “intervening” region is methylated in ESCs, the CpG island is almost invariably methylated. This region may therefore harbor transcription factor binding sites or some sort of boundary that is more efficient in protecting the CpG island promoter of B3gal tl in tissues. A related explanation is that B3gal tl expression differences between ESCs and tissues could account for the limited occurrence of spreading. For example, it has been previously suggested that reduction in transcription levels could lead to reduced frequency of transcription factor binding which could promote the spreading of heterochromatin (Turker, 2002). Therefore, possible lower B3gal tl expression in ESCs could result in the CpG island promoter being more susceptible to encroaching repressive marks. However, qRT-PCR, completed by Liane Gagnier, did not detect any significant differences in B3gal tl expression in ESCs versus tissues. The overall lack of observed “spreading”, however, is likely due to the selection of cases with ERVs and genes in close proximity. While ERV-induced spreading of repressive marks into nearby genes is rare, this phenomenon may be a frequent occurrence in genomic regions devoid of “spreading” barriers.

5.2 Role of CpG islands and CpG island shores in maintaining an active chromatin state

With respect to the various case genes, the most commonly observed result is the presence of a methylated ERV insertion in close proximity to a gene harboring a CpG island promoter which remains unmethylated. Furthermore, in the CdGAP case, even when the proximal region upstream of the promoter is methylated in ESCs perhaps due to “spreading” from the insertion, the CpG island promoter retains its unmethylated state. Therefore, we can conclude that in addition to resisting genome-wide de novo methylation, CpG islands are also efficient in maintaining an open chromatin state despite encroaching repressive marks. In addition to CpG islands antagonizing repressive marks, CpG island shores of methylation-resistant genes in humans have also been shown to play an important role in maintaining an open chromatin environment via transcription factor binding and H3K4me3 enrichment.
(Fan et al., 2007). Indeed, we observe in two cases, not only hypomethylation of the CpG island shore/intervening region, but also hypomethylation of the 3’LTR of the upstream insertion; located closest to the gene. Therefore, similar to the idea of repressive chromatin silencing, the active state of a CpG island may be able to propagate into the nearby region.

5.3 The impact of high gene expression on the local chromatin environment

In two cases, CdGAP and Lair1, we show that high gene expression is correlated with hypomethylation of the 3’LTR of the nearby insertion. Since we did not observe any evidence for the insertions as acting as enhancers/promoters of gene expression, we can conclude that gene activity may be influencing the epigenetic state of the ERV. While CpG island promoter regions are generally associated with the presence of H3K4me3 regardless of expression of the corresponding gene, several studies have shown a correlation between gene transcriptional activity and H3K4me3 levels (Barski et al., 2007). This is also similar to what is observed for bivalent genes that become activated upon differentiation and are characterized by loss of repressive mark H3K27me3 and increased levels of H3K4me3 (Ku et al., 2008). While we are unable to truly compare enrichment of active marks in ESCs and tissues due to different ChIP techniques, we can conclude that CdGAP expression is associated with loss of H3K27me3 and enrichment of active marks not only at the promoter, but also in the upstream region which contains the LTR. While a mechanism of active mark spreading has yet to be characterized, it has been shown that RNA pol II binding, which may recruit H3K4-methyltransferases, correlates with gene transcription levels (Cedar and Bergman, 2009). Therefore, higher gene expression may lead to more RNA pol II binding which may result in the accumulation of H3K4me3.

5.4 A trade-off: ERV-mediated gene silencing vs. Gene-mediated ERV hypomethylation

A previous study has provided evidence for an evolutionary trade-off in plants that involves restraining TE activity by epigenetic mechanisms and the potential for this silencing to negatively affect nearby genes; resulting in selection against methylated TE insertions near genes (Hollister and Gaut, 2009). This is in concordance with our results since our case list of polymorphic ERV insertions near genes is indeed short, suggesting strong negative selection against ERVs located near genes. Indeed, Ying Zhang in our lab has shown through bioinformatics analysis that both IAPs and ETns are markedly underrepresented within 2-3 kb of gene TSSs, compared to random expectations (Zhang, Y., personal
While we do observe cases of methylated ERV insertions near genes, we also observe 3’LTR hypomethylation in some cases that does not have an obvious affect on activity of the nearby gene. Therefore, we can begin to think about “what is the worst scenario”? Is it worse to have a methylated ERV insertion near a gene which could result in spreading of epigenetic silencing? Or, is it of higher risk to have hypomethylation of an ERV that could result in an active TE which could further propagate in the genome? It is particularly interesting to note that we only observe 3’LTR hypomethylation and 5’LTR hypermethylation, which could mean that the insertion is not truly capable of retrotransposition since proviral transcription begins in the 5’LTR. It is possible that ERVs oriented such that the 5’LTR is close to the TSS would be more strongly selected against. While heavy methylation of the 5’ LTR in the cases I have examined here would suppress transcription of the ERV element itself, we do observe transcription effects mediated by the undermethylated 3’LTR near the Cyb5r1 gene that results in fusion transcripts between the LTR and the gene in ESCs, the functional consequences of which have not been explored.

5.5 Future directions

I think it would be of interest to further investigate the requirement of H3K9me3 in methylation spreading in ESCs. While I did observe an increase in IAP expression at day 4 upon knockdown of DNMT1 and SETDB1, it was not as significant as has been shown by previous studies. Therefore, it may be worthwhile to repeat the experiment with new knockdown samples. Furthermore, it would be interesting to examine DNA methylation at individual CpG sites for the B3galtl case by bisulfite sequencing in KD cells to further examine the relationship between DNA methylation and H3K9me3 in “spreading”. Bisulfite sequencing results indicated that the intervening region between the CpG island promoter of B3galtl and the LTR may be important in determining whether or not methylation spreads into the promoter. Therefore, it would be of interest to characterize this region further perhaps by looking for CTCF binding sites which could be important in blocking the spread of repressive chromatin.

While we have observed several cases where the gene CpG island promoter is unmethylated and the nearby insertion is methylated, we have not yet characterized the chromatin environment with respect to histone modifications. It would be interesting to more fully characterize the CpG island shore/intervening region to see if DNA methylation and/or histone modifications are spreading and if any important transcription factor binding sites or boundary elements are present in these regions.

With respect to the CdGAP case, it would be interesting to induce gene expression in ESCs to establish a more direct link between expression and hypomethylation of the 3’LTR. Furthermore, it
would be interesting to study the \textit{CdGAP} case in differentiated cells deficient for H3K4me3 to see if loss of this active histone mark would lead to a reduction in \textit{CdGAP} gene expression.

\section*{5.6 Epigenetics and cancer}

It is widely accepted that cancer development is associated with hypomethylation of transposable elements and hypermethylation of tumour suppressor genes which oftentimes have CpG island promoters (Esteller, 2008). Therefore, many researchers have become interested in what makes genes, normally resistant to methylation, susceptible to cancer-mediated silencing. A recent study in the human genome showed that genes prone to methylation in cancer exhibited a lower frequency of transposable elements near their TSSs compared to genes viewed as methylation resistant. It was hypothesized that methylation prone genes have less transcription factor binding sites or other elements that would potentially protect against heterochromatin spreading. Thus, negative selection acts against TEs insertions near these genes (Estecio \textit{et al.}, 2010). Therefore, understanding what makes genes different in their susceptibility to spreading or protection from repressive marks may contribute to identifying genes that are susceptible to becoming silenced in cancer.

\section*{5.7 Significance of work}

This work is significant in that it contributes to understanding the impact of transposable elements on gene expression. I sought to investigate if an ERV could exert an epigenetic influence on a nearby gene; subsequently negatively impacting its expression. While we observed one example of heterochromatic forces initiating from an LTR affecting a nearby gene, a majority of the genes examined are able to maintain a euchromatic state despite a silenced ERV nearby. Therefore, the extent to which two distinct epigenetic domains may interact and influence one another is likely an ongoing battle that may shift in favor of the ERV or gene depending on locus-specific factors such as the ERV-TSS distance, the presence of a CpG island, and gene expression.
References


91


Walter, M., Kilian, J. and Kudla, J. (2002). PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. The EMBO journal 21, 6905-6914.


