REGULATION OF THE ETN/MUSD FAMILY OF ACTIVE MOUSE LONG TERMINAL REPEAT RETROTRANSPOSONS

by

Irina Arielevna Maksakova

B.Sc. (Hon.), Novosibirsk State University, Russia, 2001

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY in The Faculty of Graduate Studies (Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA (Vancouver)

July 2008

© Irina Arielevna Maksakova, 2008
ABSTRACT

Long terminal repeat (LTR) retrotransposons account for approximately 10% of mouse and 8% of human genomes and may play a role in modifying gene expression. Many species harbor retrotransposon families encompassing both autonomous and non-autonomous members. Specifically, the mouse Early Transposon (ETn) family members lack all retroviral genes but are transcriptionally and retrotranspositionally active, causing over 20 known insertional germline mutations. ETns owe their retrotransposition potential to proteins encoded by structurally intact MusD retrotransposons with whom they share LTRs. ETn elements are transcribed at a much higher level than MusD retrotransposons in embryos and undifferentiated cells, suggesting their evasion of host restriction mechanisms. However, mechanisms responsible for the replicative success of non-autonomous retrotransposon subfamilies over their coding-competent relatives are poorly understood.

In the first stage of my research, I analyzed regulatory sequences in an ETn LTR responsible for its high promoter activity in the undifferentiated cell line P19. I found that three GC-boxes that may function as Sp1/Sp3 binding sites act synergistically and are indispensable for undifferentiated cell-specific promoter activity of the LTR. Sp1 binding partners may be responsible for the restricted ETn expression. Moreover, I have shown that unlike many retroviruses, ETn elements possess multiple transcription initiation sites and that they have amplified via intracellular retrotransposition in the P19 teratocarcinoma cell line.

In the next step of my research, I performed analysis of epigenetic mechanisms as a means of ERV suppression. Specifically, I showed that in embryonic stem cells, autonomous MusD retrotransposons are epigenetically suppressed to a greater degree than non-autonomous ETn retrotransposons, illustrated by a higher level of DNA methylation and a
lower level of active histone modifications. I hypothesize that MusD elements may be silenced by DNA methylation and repressive chromatin spreading into the LTR from the CpG-rich internal retroviral sequence absent in ETn elements.

I propose that internal structure largely devoid of high CG content enables ETn elements to evade host-imposed transcriptional repression, contributing to their high mutagenic activity in the mouse germline.
TABLE OF CONTENTS

Abstract.................................................................................................................................... ii
Table of contents .................................................................................................................... iv
List of tables............................................................................................................................ ix
List of figures...........................................................................................................................x
List of abbreviations ...............................................................................................................x
Acknowledgements ............................................................................................................... xv
Co-authorship statement..................................................................................................... xvi

CHAPTER 1. Introduction ........................................................................................................ 1

1.1. Genomic transposable elements ...................................................................................2

1.2. ERV structure and lifecycle..........................................................................................2

1.2.1. Proviral structure........................................................................................................3

1.2.2. Retroviral life cycle....................................................................................................4

1.3. Mouse ERV classification .............................................................................................5

1.3.1. Class I retroviruses....................................................................................................5

1.3.2. Class II retroviruses ..................................................................................................5

1.3.2.1. IAP retrotransposons...........................................................................................5

1.3.2.2. ETn/MusD retrotransposons...............................................................................7

1.3.3. Class III retroviruses..................................................................................................8

1.4. ERV sequences as part of the host genome .................................................................9

1.4.1. ERV expression patterns in the mouse .................................................................9

1.4.2. Regulatory role of ERVs in development.............................................................10

1.4.3. Mutagenic mechanisms of ERV insertions.........................................................11
1.4.4. Detrimental consequences of ERV expression

1.5. ERV activity and host restriction mechanisms in mammals

1.5.1. Comparison of human and mouse ERV activity and related restriction mechanisms

1.5.2. ERV activity in different mammalian genomes

1.5.3. ERV activity in mouse strains

1.5.4. Retrotranspositional success of non-autonomous mouse ERV families

1.6. Epigenetic mechanisms of ERV silencing

1.6.1. DNA methylation and DNA methyltransferases

1.6.2. Histone modifications and histone methyltransferases

1.7. Retroviral silencing in cell lines

1.7.1. Succession of epigenetic modifications during retroviral silencing

1.7.2. Role of Dnmts in ERV regulation in cell lines

1.7.3. Role of histone modifications in ERV regulation in cell lines

1.8. ERV regulation in development

1.8.1. Epigenetic reprogramming and ERV regulation in mouse preimplantation development

1.8.2. Epigenetic reprogramming and ERV regulation in mouse germ cell development

1.9. ERV regulation by small RNAs

1.10. Thesis objectives

1.11. References

CHAPTER 2. Transcriptional regulation of ETn elements, an active family of mouse LTR-retrotransposons
2.1. Introduction .................................................................................................................39

2.2. Materials and methods ...............................................................................................39

2.2.1. Cell culture, transient transfections and luciferase assays ....................................39

2.2.2. Plasmids and constructs .........................................................................................40

2.2.3. Preparation of RNA, Northern blotting and 5’-RACE ..........................................42

2.2.4. Preparation of genomic DNA and Southern blotting ............................................43

2.2.5. Nuclear extraction and electrophoretic mobility shift assays (EMSA) ....................44

2.3. Results and discussion ...............................................................................................45

2.3.1. Transcription of endogenous ETn and MusD in ES and EC cell lines .....................45

2.3.2. Promoter activity of ETnII and MusD LTRs in P19 and NIH3T3 cell lines ...............46

2.3.3. Regions in the ETnII LTR crucial for high promoter ability in P19 cells ..................49

2.3.4. The three ETnII LTR critical regions bind Sp1 and Sp3 transcription factors .........51

2.3.5. Mutational analysis of the ETnII LTR ...................................................................52

2.3.6. High promoter activity of the 3’-deleted ETnII LTR lacking the putative TATA-box ........................................................................................................54

2.3.7. Determination of ETnII transcription initiation sites .............................................56

2.3.8. Apparent amplification of ETnII elements in the permissive P19 cell line .............59

2.3.9. Role of Sp1 and Sp3 transcription factors in ETn expression ..................................62

2.3.6. References ..............................................................................................................64

CHAPTER 3. Activity of the mouse ETn retrotransposon subfamily is associated with escape from epigenetic silencing ..............................................................................67

3.1. Introduction .................................................................................................................68

3.2. Materials and methods ...............................................................................................69
3.2.1. Cell culture...............................................................................................................69
3.2.2. Plasmids and constructs ...........................................................................................69
3.2.3. In vitro patch-methylation ........................................................................................70
3.2.4. Transient transfections and luciferase assays ..........................................................71
3.2.5. Preparation of RNA and Northern blotting ..............................................................72
3.2.6. Preparation of genomic DNA and Southern blotting ...............................................72
3.2.7. Chromatin immunoprecipitation (ChIP) ..................................................................73
3.2.8. Quantitative real-time PCR ......................................................................................73

3.3. Results...........................................................................................................................74

3.3.1. Differences in LTR sequence do not explain higher transcript levels of ETnII elements ...................................................................................................................74
3.3.2. Neither ETnII nor MusD internal regions possess strong cis-acting regulatory signals ..............................................................................................................................................................77
3.3.3. Demethylation of MusD LTR and common region in the DNA methyltransferase knockout (KO) ES cell lines ..............................................................................................................................................................78
3.3.4. Elevated transcript levels of MusD and ETnII elements in the Dnmt KO ES cell line ..................................................................................................................................................................................81
3.3.5. More activating and less suppressive histone marks on ETnII compared with MusD elements ..................................................................................................................................................................................82
3.3.6. ETnII and MusD LTRs are silenced by CpG methylation. ...........................................84
3.3.7. ETnII internal region has a lower density of CpG dinucleotides compared with MusD ..................................................................................................................................................................................85

3.4. Discussion .....................................................................................................................87
LIST OF TABLES

Table 1.1  Overview of IAP and ETn mutagenic germline insertions............... 11
Table 2.1  List of primers................................................................. 41
Table 2.2  Description of the LTRs used in the study............................... 47
Table 3.1  List of primers................................................................. 70
LIST OF FIGURES

Figure 1.1. Proviral structure .................................................. 3
Figure 1.2. Molecular mechanisms of ERV-induced mutations ............... 12
Figure 1.3. Genome reprogramming and retrotransposon methylation in mouse preimplantation development ................................................. 24
Figure 1.4. Genome reprogramming and retrotransposon methylation in the mouse germline .......................................................... 25
Figure 2.1. ETn and MusD transcript abundance in undifferentiated and differentiated cell lines ............................................................. 46
Figure 2.2. Promoter activity of ETnII and MusD LTRs in P19 and NIH3T3 cells. 48
Figure 2.3. Promoter activity of 5’-deleted ETnII LTRs in P19 and NIH3T3 cells. .................. 49
Figure 2.4. Sequence of the ETnII LTR#6, position of deletion constructs and EMSA oligonucleotides ..................................................... 50
Figure 2.5. Binding of Sp1 and Sp3 transcription factors to the regions crucial for high LTR promoter activity ............................................. 51
Figure 2.6. Promoter activity of Sp-binding site mutant LTRs in P19 and NIH3T3 cells ................................................................. 53
Figure 2.7. Promoter activity of 3’-deleted ETnII LTRs in P19 and NIH3T3 cells .. 55
Figure 2.8. Mapping of transcription initiation sites in ETnII LTRs by 5’-RACE ................................................................. 57
Figure 2.9. Amplification of ETnII elements in the P19 cell line ................. 60
Figure 3.1. Alignment of the first 93 bp of the ETnII and MusD LTRs ............ 76
Figure 3.2. Structure of ETnII and MusD retrotransposon subfamilies .......... 77
| Figure 3.3. | Effect of ETn and MusD internal regions on LTR promoter activity | 78 |
| Figure 3.4. | DNA Methylation of ETnII and MusD promoter regions in *Dnmt KO* cell lines | 80 |
| Figure 3.5. | ETnII and MusD expression in *Dnmt KO* cell lines | 81 |
| Figure 3.6. | ChIP in ETnII and MusD common region | 83 |
| Figure 3.7. | Promoter activity of *in vitro* patch-methylated LTRs | 85 |
| Figure 3.8. | CpG composition of ETnII and MusD retrotransposons | 86 |
| Figure 3.9. | A model for amplification of deleted LTR-retrotransposon variants | 90 |
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>Capsid</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco's modified Eagle’s medium</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>Dnmt</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>dsRNA</td>
<td>Double-stranded RNA</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>enFeLV</td>
<td>Endogenous feline leukaemia virus</td>
</tr>
<tr>
<td>Env</td>
<td>Envelope</td>
</tr>
<tr>
<td>EOS</td>
<td>Vector with ETn, Oct4 and Sox2 regulatory elements</td>
</tr>
<tr>
<td>ERV</td>
<td>Endogenous retrovirus</td>
</tr>
<tr>
<td>ES</td>
<td>Embryonic stem</td>
</tr>
<tr>
<td>ETn</td>
<td>Early Transposon</td>
</tr>
<tr>
<td>FGO</td>
<td>Full grown oocyte</td>
</tr>
<tr>
<td>Gag</td>
<td>Group-specific antigen</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HMTase</td>
<td>Histone methyltransferase</td>
</tr>
<tr>
<td>Hprt1</td>
<td>Hypoxanthine phosphoribosyltransferase 1</td>
</tr>
<tr>
<td>H3</td>
<td>Histone 3</td>
</tr>
<tr>
<td>H3K9</td>
<td>Lysine 9 of histone 3</td>
</tr>
<tr>
<td>H3K9Ac</td>
<td>Acetylated lysine 9 of histone 3</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>Trimethylated lysine 9 of histone 3</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>H4</td>
<td>Histone 4</td>
</tr>
<tr>
<td>H4Ac</td>
<td>Acetylated histone 4</td>
</tr>
<tr>
<td>H4K20me3</td>
<td>Trimethylated lysine 20 of histone 4</td>
</tr>
<tr>
<td>IAP</td>
<td>Intracisternal A type particle</td>
</tr>
<tr>
<td>INT</td>
<td>Integrase</td>
</tr>
<tr>
<td>iPS</td>
<td>Induced pluripotent stem</td>
</tr>
<tr>
<td>KO</td>
<td>Knock-out</td>
</tr>
<tr>
<td>KoRV</td>
<td>Koala retrovirus</td>
</tr>
<tr>
<td>LINE</td>
<td>Long Interspersed Nuclear Element</td>
</tr>
<tr>
<td>LTR</td>
<td>Long Terminal Repeat</td>
</tr>
<tr>
<td>MA</td>
<td>Matrix</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblast</td>
</tr>
<tr>
<td>MLV</td>
<td>Moloney Murine Leukemia Virus</td>
</tr>
<tr>
<td>MMTV</td>
<td>Mouse Mammary Tumour Virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MT</td>
<td>Mouse Transcript retrotransposon</td>
</tr>
<tr>
<td>MusD</td>
<td>Mouse type D retrotransposon</td>
</tr>
<tr>
<td>NC</td>
<td>Nucleocapsid</td>
</tr>
<tr>
<td>NKG2A</td>
<td>NK-cell specific G2A receptor</td>
</tr>
<tr>
<td>NPC</td>
<td>Neural progenitor cell</td>
</tr>
<tr>
<td>P</td>
<td>Promoter</td>
</tr>
<tr>
<td>pA</td>
<td>Polyadenylation signal</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PBS</td>
<td>Primer binding site</td>
</tr>
<tr>
<td>PGC</td>
<td>Primordial germ cell</td>
</tr>
<tr>
<td>piRNA</td>
<td>PIWI-interacting RNA</td>
</tr>
<tr>
<td>Pol</td>
<td>Polymerase</td>
</tr>
<tr>
<td>PPT</td>
<td>Polypurine tract</td>
</tr>
<tr>
<td>Pro</td>
<td>Protease</td>
</tr>
<tr>
<td>PtERV</td>
<td>Primate Endogenous Retrovirus</td>
</tr>
<tr>
<td>R</td>
<td>Repeated sequence of the LTR</td>
</tr>
<tr>
<td>REPBASE</td>
<td>Repetitive DNA database</td>
</tr>
<tr>
<td>RMCE</td>
<td>Recombinase-mediated cassette exchange</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RT</td>
<td>Reverse transcriptase</td>
</tr>
<tr>
<td>SINE</td>
<td>Short Interspersed Nuclear Element</td>
</tr>
<tr>
<td>SU</td>
<td>Surface glycoprotein</td>
</tr>
<tr>
<td>TE</td>
<td>Transposable element</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TM</td>
<td>Transmembrane protein</td>
</tr>
<tr>
<td>TS</td>
<td>Trophoblast stem</td>
</tr>
<tr>
<td>U3</td>
<td>Unique sequence to the 3'-end of LTR</td>
</tr>
<tr>
<td>U5</td>
<td>Unique sequence to the 5'-end of LTR</td>
</tr>
<tr>
<td>5’-RACE</td>
<td>Rapid amplification of 5’-ends</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

My sincere and most heart-felt thanks go to my wonderful supervisor, Dixie Mager, whose seemingly endless support and encouragement motivated me to ask challenging questions and gave strength to search for answers. I thank all the lab members for creating a friendly and helpful work environment and Matt Lorincz for his instrumental input in shaping my interest in epigenetic regulation.

Last, but not least, I thank my mom and dad for believing in me and encouraging me every step of the way, and my dear husband Evgeny for his constant emotional support. I would not have been able to make it without you.
CO-AUTHORSHIP STATEMENT

Research described in chapters 2 and 3 was designed by I.A. Maksakova and D.L. Mager. I.A.M. performed the experiments, I.A.M and D.L.M. analysed results. Y. Zhang carried out multiple alignment presented in Figure 3.1. I.A.M. drafted the manuscripts described in chapters 2 and 3 and D.L.M. finalized the manuscripts.
CHAPTER 1

INTRODUCTION$^{1,2}$

---

$^1$ A version of this chapter has been published.

$^2$ A version of this chapter will be submitted for publication.
1.1. Genomic transposable elements

Transposable elements (TEs) have been found in all eukaryotic species and in many cases comprise a substantial fraction of their genomes. TEs can be subdivided into DNA transposons which move via a “cut-and-paste” mechanism and constitute a minor fraction of the mammalian genomes, and retrotransposons, which amplify via an RNA intermediate through the process of reverse transcription. Retrotransposons are subdivided into non-Long Terminal Repeat (LTR) retrotransposons and LTR retrotransposons which include Endogenous RetroVirus-like sequences (ERVs). LTR elements/ERVs make up 8-10% of the mouse and human genomes (49, 50). The most abundant retrotransposon classes in mammals are representatives of the non-LTR retrotransposons, the non-autonomous Short Interspersed Nuclear Elements (SINEs) and the autonomous Long Interspersed Nuclear Elements (LINEs) which collectively comprise 28% of the mouse and 35% of the human genomes (49, 50).

1.2. ERV structure and lifecycle

ERVs are derived from exogenous retroviruses that at some point in evolution integrated into the genome of a germline cell. They are transmitted vertically to offspring and amplify either by intracellular retrotransposition or by rounds of re-infection (1, 10, 45). Multiple ERV families, grouped into three classes, exist in the mouse genome (12, 50, 87). The majority of all characterized ERV germline insertional mutations, however, are caused by Class II Intracisternal A type particles (IAP) and Early Transposon (ETn)/MusD elements, the latter family being the focus of this thesis and thus described in greater detail below.
1.2.1. Proviral structure

The integrated form of a retrovirus, or provirus, is flanked by LTRs and has three open reading frames – *gag* (group-specific antigen)-*pro* (protease), *pol* (polymerase) and *env* (envelope) (Fig. 1.1A). *Gag* polyprotein is processed into mature structural proteins of the virus, matrix (MA), capsid (CA) and nucleocapsid (NC). *Pro* encodes aspartic protease which processes the proteins encoded by *gag*, *pro*, *pol* and in some cases also *env*. *Pol* encodes reverse transcriptase (RT) with DNA polymerase and RNaseH activities and integrase (INT) which mediates proviral integration into the genome. *Env* encodes the virion surface glycoprotein (SU) and transmembrane (TM) protein which form a complex that interacts with cellular receptor proteins (Fig. 1.1B). Most ERVs completely lack or have a non-functional *env* gene. The primer binding site (PBS) and the polypurine tract (PPT) are required for the process of reverse transcription (12).

![Figure 1.1. Proviral structure.](image)

**Figure 1.1. Proviral structure.**

(A) Structure of the integrated provirus and proteins encoded by viral genes. LTR, Long Terminal Repeat; P, promoter; pA, polyadenylation signal; PBS, primer binding site; PPT, polypurine tract. (B) Structure of the viral particle. TM, transmembrane protein; SU, surface glycoprotein; NC, nucleocapsid; CA, capsid; MA, matrix; RT, reverse transcriptase; INT, integrase.

LTRs are comprised of three subunits: R (repeated sequence), U5 (unique to the 5'-end) and U3 (unique to the 3'-end) and contain regulatory sequences necessary for production of retroviral transcripts, such as transcription initiation and polyadenylation sites. In addition
to the promoter bound by transcription factors that direct transcription efficiency and cell specificity of the virus, both U3 and U5 regions may contain enhancer regions capable of up-regulating transcription. (Fig. 1.1) (12).

1.2.2. Retroviral life cycle

Retroviruses comprise a large and diverse family of enveloped RNA viruses. The main characteristic of this family is its replicative strategy, which includes reverse transcription of the virion (+)RNA into linear double-stranded DNA followed by subsequent integration of the proviral DNA into the genome (135). Retroviruses enter the host cell via fusion of viral and cellular membranes resulting from attachment of the viral surface glycoprotein to a plasma membrane receptor. Reverse transcription takes place in the cytoplasm. The viral DNA is then translocated into the nucleus and integrated into chromosomal DNA with the help of integrase to form a provirus. Upon integration, the provirus obtains the status of a cellular gene and is transcribed by cellular RNA polymerase II and replicated by cellular enzymes along with chromosomal DNA. Nevertheless, the LTR regulatory sequences stay in control of proviral transcription which generates spliced mRNAs along with full-length unspliced viral genomic RNA. The primary transcript of retroviral DNA is capped at its 5’-end and polyadenylated at its 3’-end, closely resembling a cellular mRNA (12). ERVs represent remnants of ancient retroviral infections of the germline. Most ERVs are only capable of retrotransposing within the cell; some families, however, maintain full coding potential and infectivity (10, 12, 128).
1.3. Mouse ERV classification

1.3.1. Class I retroviruses

The Class I/Type C/gammaretroviruses, comprising about 0.7% of the mouse genome (50), are grouped based on their similarity to Moloney Murine Leukemia Virus (MLV) (12) first isolated from lymphomas of AKR mice (41). MLV entered the mouse germline approximately 1.5 million years ago, resulting in current copy numbers from 25 to 70 depending on the mouse strain. MLV proviruses are subdivided based on their host ranges determined by their envelope genes but only a few encode replication-competent viruses (12, 125). Another ERV family, GLN, was recently shown to encode functional retroviral particles (113). Class I also includes several other families but it is unclear if any have fully coding-competent members (12).

1.3.2. Class II retroviruses

The prototype of the much more numerous Class II/Type B&D/betaretroviral group, comprising about 3% of the mouse genome (50), is Mouse Mammary Tumour Virus (MMTV) (12). A wide variety of other mouse betaretroviral ERV families also exist, some of which retain coding capacity and appear to have entered the genome quite recently (5, 87).

1.3.2.1. IAP retrotransposons

One of the most extensively studied non-infectious families of ERVs in the mouse is the IAP family. They were discovered and characterized in the 1960s-1970s based on electron microscopy as large, 85-100 nm in diameter, particles assembled on the endoplasmic reticulum that bud into the cisternae in normal and malignant mouse mammary glands (142).
and early mouse embryos (24). While there is variation in copy numbers between different mouse strains, about ~700 full-length, ~1700 partially deleted elements and ~2500 solitary LTRs are present in the haploid mouse genome (148). IAP elements were thought to lack an envelope gene until about 200 env-containing elements were discovered (109).

Type I IAP elements encompass full-length members as well as four deleted classes (64). Of these, the IΔ1 subclass with a 1.9 kb deletion in gag-pol region is the most abundant deleted form in the mouse genome and is also responsible for the majority of IAP insertional mutations (52, 83, 108). A very recent report explains retrotranspositional advantage of this particular non-autonomous type by cooperation between autonomous and non-autonomous IAP proteins. Incredibly, a gag-pol fusion protein produced by IΔ1 elements recruits a functional IAP protein to the site of IΔ1 IAP transcription, expediting its packaging into a retroviral particle and enhancing IΔ1 reverse transcription and retrotransposition efficiency (116). Recently, a single fully active copy of an IAP element, capable of forming and releasing mature virions, was found in the mouse genome (112). This research had confirmed that IAP retrotransposons derived from an infectious retrovirus via passive loss of their envelope gene and gain of an endoplasmic reticulum targeting signal. These modifications assisted IAPs in acquisition of a true retrotransposon status by improving their intracellular amplification (112).

While an IAP LTR-promoted transgene is expressed exclusively in spermatogonia (33), endogenous IAP expression is also detected in oocytes and preimplantation embryos (24, 104, 127). Some adult mouse tissues, particularly thymus, also exhibit IAP transcription, which, in addition, is abundant in tumours and tumour cell lines (32, 64). IAP expression is also elevated upon differentiation of embryonic teratocarcinoma cells into endoderm-like
cells (47). However, it is difficult to estimate whether transcriptional induction observed in these cases reflects up-regulation of the majority of ERVs in the genome. Elevated transcript levels may be accounted for by only a small ERV subpopulation, if not just a few isolated loci particularly prone to de-repression due to permissive chromatin constitution.

1.3.2.2. ETn/MusD retrotransposons

The other major family of active mouse LTR elements and the focus of this thesis is the ETn family, first described in 1983 by Brulet et al. as a family of middle repetitive sequences transcribed during early mouse embryogenesis with expression peaking between 3.5 and 7.5 days (19, 20). ETns are flanked by LTRs and do not contain retroviral open reading frames. However, some remain transposition-competent, having induced over 20 novel insertional mutations (83). Two major subtypes of ETn elements, I and II, differ in the 3' portion of the LTR and a 5' internal segment. Although ETnI elements are more numerous, ETnII elements are currently more active (8, 83).

The lack of coding potential in ETn elements raised questions as to how they could retrotranspose, but it is now clear that a related coding-competent family of endogenous betaretroviruses, termed MusD (81), that share nearly identical LTRs and some internal sequence with ETns, provide the proteins necessary for ETn retrotransposition (110). Due to the close relationship of these subfamilies, I will be referring to this family as ETn/MusD throughout this thesis. No MusD element has an env-related sequence, suggesting amplification exclusively via retrotransposition. Indeed, it has recently been shown that, during evolution from an exogenous retrovirus, MusD elements lost the myristoylation and plasma membrane targeting signals, which benefited and assisted their exclusively
intracellular life cycle (111). As with IAP elements, copy numbers vary between strains (148), but the 2007 release of the C57BL/6 genome has ~300 ETnI/ETnII, ~100 MusD elements and ~1450 solitary LTRs (148).

ETn/MusD retrotransposons are highly transcribed in E3.5 to E7.5 mouse embryos (19, 20) and developing tissues of E7.5-E13.5 embryos in a tissue- and stage-specific manner (78). They are also expressed in embryonic stem (ES) cell lines (9), embryonic carcinoma (EC) (20) and several plasmocytoma cell lines (121), as well as primary acute myeloid leukemia (AML) cells (130). However, primary cells of other tumour types, such as hepatoma and lymphoma, lack ETn/MusD expression (130). ETn/MusD expression is low or undetectable in differentiated cell lines tested, including the embryonic 3T3 fibroblast line (20). Moreover, their expression level drops dramatically upon differentiation of the teratocarcinoma cell line (48).

1.3.3. Class III retroviruses

Mouse Class III elements, which may include some active ERVs, contain the MuERV-L family, present in up to 200 proviral copies of about 7.5 kb per haploid genome (11) and different subgroups of the highly repetitive non-autonomous ORR1 and MT MaLR elements (123). Class III elements comprise approximately 5.4% of the genome (50), significantly contribute to the early mouse transcriptome (103) and affect gene regulation. However, only one mutation due to insertion of a Class III MaLR element has been documented (83) and they are not as polymorphic as ETn or IAP retrotransposons (Y. Zhang, personal communication).
1.4. ERV sequences as part of the host genome

1.4.1. ERV expression patterns in the mouse

Acquisition of new heritable proviruses requires expression in the germline. Indeed, expression of ERV families, while repressed in differentiated tissues, seems to be tolerated to some extent in germ cells, preimplantation embryos and placenta (33, 78, 103, 127, 134) despite the fact that many host surveillance mechanisms target ERVs at these particular stages. Not surprisingly, retroviral elements that have successfully colonized the host genome possess regulatory sequences tailored to expression in cells either giving rise to a new organism, such as germ cells, or those in immediate contact with it, such as placenta. This strategy employed by ERVs is aimed at long-term survival in the host genome. This is quite contrary to the strategy of rapid short-term amplification of exogenous retroviruses, readily infecting and replicating in somatic tissues.

IAP and ETn elements are the most active retrotransposon families in the mouse. However, while IAP elements are expressed in many mouse tumours and cell lines (64), expression of ETns is more restricted, elevated only in undifferentiated embryonic carcinoma and embryonic stem cells and primary AML cells (130). Transcript levels of ETn and IAP elements are highest during early embryogenesis (19, 64, 78, 134). This time-specific restriction placed on ERV transcription, limited largely to a narrow window of early embryogenesis, is suggestive of extremely tight regulation imposed by complex mechanisms of the host genome in an effort to prevent somatic insertional mutagenesis.

It is possible that MLV and MMTV ERVs have remained at low copy numbers (12, 125) due to poor expression in the germline and/or inefficient intracellular retrotransposition,
since studies have shown a much higher retrotransposition rate for IAP and MusD ERVs compared with MLV (29, 43, 110).

1.4.2. Regulatory role of ERVs in development

The vast majority of ERV insertions are likely selectively neutral or detrimental for the host and subject to negative selection. However, ERVs are speculated to have played a role in speciation, contributing to protein-coding genes (14, 17, 58, 61), regulatory motifs implicated in embryogenesis (103, 120) and tissue-specific expression (137). ERVs, encompassing multiple transcription control elements, may be viewed as ready-made mobile regulatory cassettes which, when integrated, may confer a new expression profile to the adjacent gene (91, 137).

Given the inherent disposition of ERVs towards expression in the early embryo, it is tempting to speculate that mammalian genomes may have adapted the highly specific regulatory sequences within ERVs for coordinated regulation of developmental genes during early embryogenesis (103). As no active transcription takes place in the newly fertilized oocyte, the processes of fertilization, gamete nuclei reprogramming and zygotic genome activation are controlled by maternal transcripts contributed by the oocyte. Embryonic genome activation starts in the late 2-cell stage and continues on until late morula (68). Incredibly, 13% of all transcripts in the full-grown mouse oocyte (FGO) are contributed by ERVs, predominantly of the MT subfamily of MaLR retrotransposons (103) which comprise less than 5% of the genome (50). In addition to LTR-retrotransposon transcripts, many gene transcripts initiating in the LTR of Mouse Transcript (MT) retrotransposons are found in the FGO, while LTRs of other families are more prominent in promoting transcripts in the 2-cell
embryo. Most intriguingly, such chimeric transcripts, found in the FGO and preimplantation embryos, are not detectable in other developmental stages or adult tissues, and their disappearance by the 8-cell stage suggested they are maternal mRNAs promoted by MT LTRs (103). This interesting data supports the concept of LTR-retrotransposons possessing inherent promoter activity tailored to the oocyte and the early embryo.

1.4.3. Mutagenic mechanisms of ERV insertions

Most commonly, reported germline mutations due to ERV insertions occur in an intron, disrupting gene expression by causing premature polyadenylation, aberrant splicing or ectopic transcription driven by the ERV LTR (Fig 1.2) (83). In some cases, small amounts of normal gene transcripts and protein can still be detected. IAP and ETn elements show significant differences in their effects on genes (Table 1.1). For ETn insertions, the most commonly reported defect is premature polyadenylation within the ETn, coupled with aberrant splicing due to a few commonly used cryptic splice signals. IAP elements also disrupt splicing but have an additional effect on gene expression by driving ectopic transcription (83).

Table 1.1. Overview of IAP and ETn mutagenic germline insertions

<table>
<thead>
<tr>
<th>Family</th>
<th>IAP</th>
<th>ETn/MusD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic copy #</td>
<td>~1400</td>
<td>300-400</td>
</tr>
<tr>
<td># Mutagenic insertions*</td>
<td>32</td>
<td>23</td>
</tr>
<tr>
<td>Strain bias</td>
<td>C3H (&gt;50%)</td>
<td>A/J (~30%)</td>
</tr>
<tr>
<td>Active subtype</td>
<td>IAP IΔ1</td>
<td>ETnII-β</td>
</tr>
<tr>
<td>Promoter effects</td>
<td>Yes (antisense)</td>
<td>No</td>
</tr>
<tr>
<td>Polyadenylation and splicing effects</td>
<td>Common</td>
<td>Common</td>
</tr>
<tr>
<td>Somatic insertions</td>
<td>Common</td>
<td>Rare</td>
</tr>
<tr>
<td>Intronic orientation relative to affected gene</td>
<td>Mixed</td>
<td>Sense (15/16 cases)</td>
</tr>
</tbody>
</table>

*Characterized new germ-line insertions, including strain variants
Many of the mutant alleles caused by IAP LTR-driven gene expression show variable expressivity among genetically identical mice and have been termed metastable epialleles (107). The variable expressivity is due to stochastic establishment of a specific methylation state of the 5’ LTR. If the LTR is mostly methylated, its promoter is inactive and little or no effect on the gene is observed. However, if the LTR is unmethylated, its promoter drives ectopic gene expression resulting in the mutant phenotype. Such cases have been extensively studied by Whitelaw and co-workers who have proposed the intriguing theory that phenotypic variation in mammals could in part be due to incomplete and variable silencing of retrotransposons in somatic cells (141).

Figure 1.2. Molecular mechanisms of ERV-induced mutations.
An integrated ERV, gene exons and resulting transcripts are depicted. (A) Premature polyadenylation (B) Ectopic transcription induced by an ERV promoter (C) Ectopic transcription induced by an ERV enhancer

1.4.4. Detrimental consequences of ERV expression

As mentioned above, most immediate effects of novel retroviral integrations, including germline mutations (83) and cancerous transformations (22, 46) are neutral or harmful for the host. Many reports indicate that aberrant retrotransposon expression in
somatic cells accompanies cancerous transformations (46, 70, 90, 95) and autoimmune
diseases (3, 27) in both human and mouse. Given the typical germ cell-specific expression
pattern, new ERV integrations can occur in somatic cells only if their transcription has been
activated by, for example, demethylation, which takes place on a global scale in transformed
cells (40, 126, 146). This would mean that insertions, some of which can target proto-
oncogenes (46), may compromise genome integrity and play a role in subsequent
transformations. Demethylation and transcriptional activation of ERVs may result not only
from cancer-induced genome-wide demethylation, but also from aging (7, 105). Reports of
IAP-induced aberrant transcripts in tissues of aging mice indicate that other processes besides
cancer may result in ERV deregulation.

1.5. ERV activity and host restriction mechanisms in mammals

Waves of ERV amplification are normally countered by host defences (37) and
negative selection that quench ERV activity until a new variant or a “master” element,
capable of evading this suppression, appears to further instigate amplification (16, 61). ERV
transcription is restricted in most differentiated tissues of animals and plants due to silencing
directed by DNA methylation, histone modifications, and RNA interference (86, 122). The
high rate of ERV germline and somatic insertional mutations in the laboratory mouse
indicates that at least some inbred strains are currently in an active phase of genomic ERV
expansion. In contrast, ERV-like elements in humans, while present in comparable overall
numbers, have long ago ceased activity (49).
1.5.1. Comparison of human and mouse ERV activity and related restriction mechanisms

While nearly half of the human genome is TE-derived, de novo disease-causing integrations, estimated at ~0.1% of all spontaneous mutations (18, 57, 58, 75), are rare and all of them are LINE1-mediated (28, 31). A small number of de novo LINE insertions have also been reported in the mouse (31, 59), but the ongoing LTR retrotransposon and ERV activity dominates and accounts for 10-12% of all mouse mutations (83). In contrast, no infectious human ERVs or new insertions have been found and there are no ERVs closely related to human exogenous retroviruses (6, 56, 82).

Inactivity of ERVs in human may reflect the presence of human-specific restriction mechanisms protecting our genomes from deleterious ERV retrotranspositions. Alternatively, active ERV variants may have been eliminated from the human population by random chance due to multiple evolutionary bottlenecks. PtERV1, an ancient ERV abundant in chimpanzee and gorilla, is absent from the human genome likely due to protection conferred by an intrinsic immune protein TRIM5α. According to recent research, all modern humans descended from ancestors who resisted PtERV1 infection conferred by a human variant of TRIM5α, which can prevent infection by exogenous PtERV1 (53). This version of the protein, however, makes humans more susceptible to other retroviruses, such as HIV (53). Fixation of a specific TRIM5α variant may serve as evidence for selection in the human population.

Another possible reason for reduced ERV activity in human in comparison to mouse may lie in the APOBEC family of cytidine deaminases involved in RNA editing and DNA mutagenesis. Surprisingly, while the mouse genome has only one gene of this family, the
human genome has seven, all functioning in diverse tissue types, and some implicated in anti-retroviral activity aimed at different stages of the retroviral lifecycle (26). It is conceivable that evolutionary expansion of the APOBEC gene family may have contributed to reduction in active retroviral content of the human genome.

1.5.2. ERV activity in different mammalian genomes

ERVs of other mammalian species display variable activity, although most harbour at least one retrotransposition-competent ERV family. The most recent germline acquisition of an ERV is demonstrated by the koala retrovirus (KoRV), present at variable copy numbers in some koala populations and completely devoid from others (133). Moreover, KoRV transcript levels in blood are associated with high incidence of leukemia, lymphoma and chlamydiosis (132). Similar findings were described for mouse MLV ERVs, with high-leukaemic mouse strains displaying high MLV transcript levels in the embryo and during lifetime (12). Hamsters, close mouse relatives, possess ERVs similar in sequence to MLV and IAP, both of which appear to be transcriptionally active and form viral particles, though none of them were found to be infectious (12). MMTV, new integrations of which may cause mammary tumours in mice, can also be transmitted both vertically and horizontally, via infectious particles in mother’s milk (128). Similar to mouse MLV, endogenous feline leukemia viruses (enFeLVs) are polymorphic between breeds (131) and can form infectious retroviruses by recombination with exogenous forms (115). The multitude of active ERVs in mammalian genomes implies that lack of retrotransposition-competent forms in human is an exception rather than the rule and suggests that multiple evolutionary bottlenecks faced by human ancestors have shaped the ERV load we have today.
1.5.3. ERV activity in mouse strains

Different mouse strains have different predisposition to ERV transcription and movement. In AKR mice, a high-leukaemic mouse strain, MLV transcripts are detected at high levels in the embryo and throughout the life of the animal, likely originating from one ancestral provirus in this strain (12). In low-leukemic mouse strains, such as BALB/c, C3H/He and C57BL/6, MLV proviruses are transcribed at significantly lower or undetectable levels (114). However, IAP retrotransposons are most active in C3H/He strain (83), suggesting that strain-specific differences in host restriction mechanisms may affect different retrotransposon or retrovirus families to a different degree.

The currently active ETn/MusD and IAP elements amplify via intracellular retrotransposition, thereby avoiding the “front line” defence mechanisms involving proteins such as Fv-1 and Fv-4, in place to inhibit early stages of exogenous infections (37). Other factors involved in antiviral defence, namely, TRIM (98) and APOBEC (26) family proteins, were found to act against endogenous retroviruses IAP and MusD (35). In addition, an allele of an apparent epigenetic modifier, Mdac, is implicated in DNA methylation and transcriptional suppression of MusD elements in select mouse strains (54). The same allele may play a role in protection of the mouse genome from other ERVs and may in part explain strain-specific variability in susceptibility to retroviral infection and/or retrotransposition. It is possible that difference in ERV activity between mouse strains is determined by combination and functionality of different alleles of multiple host resistance genes.
1.5.4. Retrotranspositional success of non-autonomous mouse ERV families

Inbred mice may represent a relatively transitory state in which host silencing mechanisms have not yet adapted to retrotransposition of new ERV variants. To illustrate this point, the IΔ1 partly deleted subtype is currently the most active IAP element but comprises a minor fraction of the total number of existing IAPs. This situation suggests that full-length IAPs amplified to high copy numbers during mouse evolution but have recently been essentially silenced. The IΔ1 subtype must have arisen recently and has been freed from suppression and allowed to retrotranspose – mainly in the C3H/HeJ strain. As discussed in Chapter 1.3.2.1, successful retrotransposition of the IΔ1 subtype is conferred by ability of the IΔ1 fusion protein to recruit functional IAP proteins to the site of IΔ1 transcription (116). A similar scenario is occurring with ETn/MusD elements where a minor population of ETnII-β elements is causing the bulk of current retrotranspositional activity (83), though the mechanism of such selective expansion is unclear. This relatively permissive phase of ERV expansion ongoing in inbred mice provides a rare opportunity to study how a mammalian host genome responds to new waves of invasion by mobile elements.

One of the aims of this thesis is to identify mechanisms employed by the host genome to suppress active transcription of the ETn/MusD family of ERVs and the means this family uses to evade the host defence system, successfully retrotranspose and colonize its host species.
1.6. Epigenetic mechanisms of ERV silencing

1.6.1. DNA methylation and DNA methyltransferases

As had been proposed previously (145), the primary role of DNA methylation may in fact be host defence against transposable elements, and, specifically, endogenous retroviruses. It has been speculated that roles for DNA methylation in gene regulation and X-inactivation are secondary adaptations, as some species successfully regulate their gene expression and compensate for X-chromosome imbalance in the absence of DNA methylation (80). Later reviews also favour the idea of gene silencing having originated as a means of silencing parasitic sequences (86). In addition to DNA methylation acting at the level of transcriptional repression, Matzke at al. propose that a second line of defence involving RNA degradation is acting at the post-transcriptional level. In fact, the two can be interrelated: double-stranded RNA (dsRNA) molecules derived from parasitic sequences may induce both transcript degradation and DNA methylation via RNA interference (RNAi) pathways. RNA-directed DNA methylation and heterochromatinization have been studied extensively in plants and yeast (reviewed in (85)), though the existence of these mechanisms in mammals remains controversial.

Among the numerous mechanisms of epigenetic repression, DNA methylation is the most well-studied. In the mammalian genome, a methyl group is predominantly targeted to the cytosine in the context of a CpG dinucleotide. Four enzymatically functional DNA methyltransferases (Dnmts), Dnmt1 (74), Dnmt3a, Dnmt3b (99) and Dnmt2 (144) have been identified in mammals (reviewed in (69)). These enzymes contain highly conserved DNA methyltransferase motifs but have different functional properties with respect to the substrate. Dnmt1 is the major maintenance Dnmt that functions primarily during DNA replication (73).
Dnmt3a and Dnmt3b are viewed as *de novo* Dnmts and are able to methylate both unmethylated and hemi-methylated DNA. Dnmt3l is a Dnmt-like protein with no enzymatic activity which, however, is necessary for Dnmt3a and Dnmt3b to realize their full methyltransferase potential (39). Dnmt2, possessing extremely low level of Dnmt activity, (44, 77), seems so play a role in RNA methylation (38, 106), though its exact functional niche in mammals remains to be determined.

### 1.6.2. Histone modifications and histone methyltransferases

In eukaryotic cells, the basic unit of chromatin is a nucleosome, which consists of 146 bp of DNA wrapped around two tetramers of core histone proteins H3, H4, H2A and H2B. Histone tails, especially those of histones H3 and H4, are subject to covalent posttranscriptional modifications, particularly at their lysine and arginine residues. Among others, the most common and well-studied ones are methylation and acetylation, frequently associated with transcriptional control or localization to specific genomic neighbourhoods (93). Moreover, most of these modifications are cross-regulated. Some histone modifications are mutually exclusive while others promote deposition of different histone marks on other amino acid residues of the same histone tail (67).

In animals, methylation of histone H3 tail at lysine 9 (H3K9me) has been implicated in silencing of repetitive elements (84, 92); however, the role of specific histone methyltransferases (HMTases) in ERV regulation remains controversial. At least five H3K9-specific histone methyltransferases have been identified in mammals: Suv39h1, Suv39h2, Eset/SETDB1, GLP/EuHMTase1 and G9a/EuHMTase2 (63). To add to the sophistication of epigenetic regulation, DNA methylation has also been shown to dictate histone methylation
and deacetylation, potentially promoting reduction of active and increase of repressive chromatin marks (51, 79, 100, 119).

1.7. Retroviral silencing in cell lines

1.7.1. Succession of epigenetic modifications during retroviral silencing

The majority of our basic knowledge on suppression of ERVs originates from work investigating retroviral and lentiviral vector silencing. ES cells provide a good model for studying silencing of endogenous and introduced proviruses due to their high level of de novo methyltransferase activity (72). Introduced proviruses are rapidly silenced in ES cells, suggesting that active epigenetic mechanisms of host defence are in place; in particular, chromatin modifications and their cross-talk with DNA methylation (reviewed in (34)).

Transcriptional silencing of most retroviruses occurs within 2-3 days after infection (60, 99) regardless of the presence of de novo DNA methyltransferases (99, 102, 143), but DNA methylation is detected only 8-10 days post infection (30, 60, 99), suggesting it is a secondary event in retroviral silencing. Other work had confirmed that that H3K9 methylation and even more so, DNA methylation, lie downstream of deacetylation and transcriptional silencing, and that deacetylated histones may be a requirement for histone and DNA methylation to take place (96).

It is now evident that DNA methylation is closely interrelated with H3K9 methylation in regulation of gene expression and heterochromatin maintenance (124). As discussed in the following sections, DNA methylation plays a pivotal role in ERV silencing (76, 138). Recently, significant association of ERVs with H3K9me2 and/or H3K9me3 marks was
shown in genome-wide studies in murine cells (84, 92), emphasizing the tight relationship of DNA methylation and histone modifications in regulation of ERVs.

1.7.2. Role of Dnmts in ERV regulation in cell lines

DNA methylation and Dnmt family members appear to be critically important, albeit to varying extents, for the maintenance of genome integrity and regulation of repetitive elements including ERVs. In ES cells deficient for Dnmt1, besides genome-wide demethylation, the endogenous loci of MLV (72, 99), IAP (99) and MaLR (76) ERVs are substantially demethylated, while the same repetitive sequences maintain heavy methylation in the wild type. However, the Dnmt1−/− ES cells efficiently methylate the newly integrated MLV proviruses (72, 99) and successfully re-methylate endogenous MaLR elements after transient exposure to the demethylating agent 5-azaC (76), emphasizing importance of de novo Dnmts in suppression of the incoming proviruses.

ES cells deficient for both de novo Dnmts, Dnmt3a and Dnmt3b, display only slight demethylation of IAP and MLV (99) elements but exhibit a much more pronounced effect on the previously unmethylated sequences: the introduced proviruses remain almost completely unmethylated (99), and endogenous MaLR elements are not re-methylated to normal levels following 5-azaC-induced demethylation (76). Importantly, deficiency in only one of de novo methyltransferases, either Dnmt3a or Dnmt3b, is not sufficient to reduce methylation of endogenous (99) or exogenous (99) proviruses, suggesting that in ES cells the two Dnmts may have redundant functions, and that DNA methylation is crucially important for ERV suppression.
1.7.3. Role of histone modifications in ERV regulation in cell lines

Several genome-wide, as well as sequence-specific, analyses of histone modifications associated with ERV sequences have been performed in recent years. In genome-wide screen for sequences binding to modified histones, H3K9me3 and H4K20me3, both associated with silencing, showed enrichment on IAP elements in ES cells (84, 92). Intriguingly, of all the ERV families in the mouse genome, even the more abundant ones, the majority of the H3K9me3 mark in ES cells was associated with the most transcriptionally and retrotranspositionally active elements, IAP and ETn/MusD retrotransposon families (83). Controversial as it is for the mammals, such selective targeting of active elements may result from RNAi-dependent silencing via targeting of repressive chromatin modifications, similar to the observations made in yeast, plants and Drosophila (88). Antisense ERV RNAs are produced in considerable amounts in ES cells (103, 127) and may potentially form double-stranded RNAs triggering RNAi. Small RNA-dependent silencing is discussed in greater detail in chapter 1.9.

Interesting results were derived from analysis of chromatin marks associated with IAP LTRs in different cell types and during differentiation. IAP transcription is substantially down-regulated in mouse embryonic fibroblast (MEF) and trophoblast stem (TS) cell lines (84) compared with ES cells. Surprisingly, this correlates with extinguished repressive chromatin modifications on IAP elements in MEF, TS (84) and neural progenitor cells (NPC) (92), suggesting that in fully differentiated cell lines, suppression is achieved by either stable DNA methylation or lack of transcription factors necessary for ERV expression. In this case, repressive chromatin may not be essential for the maintenance of ERV suppression.
1.8. ERV regulation in development

1.8.1. Epigenetic reprogramming and ERV regulation in mouse preimplantation development

Mammalian development is tightly associated with changes in global and local DNA and chromatin modifications. After fertilization, the paternal pronucleus undergoes rapid active, and the maternal pronucleus, slow passive demethylation. This is followed by passive demethylation of both genomes to the lowest methylation level reached in the morula (94, 117). The initiation of de novo methylation by Dnmt3a and Dnmt3b, which are up-regulated at this time (25, 62, 139), coincides with the onset of differentiation and results in hypermethylation of the inner cell mass (Fig. 1.3) which gives rise to all tissues of the adult and is also used to make ES cells (94, 117).

The kinetics of IAP methylation throughout early development, thoroughly examined by Lane et al. (66), showed resistance of this ERV family to demethylation throughout all stages of development. High IAP methylation level of approximately 95-98% in mature sperm and oocytes remained the same in the zygote and decreased to 62% by the blastocyst stage (66). This value, however, is still considerably higher than the genome average (94, 117). Surprisingly, the LINE1 non-LTR retrotransposon family of repetitive elements, though highly methylated in sperm, undergoes massive demethylation after fertilization (Fig. 1.3). It is therefore tempting to speculate that IAP elements and other ERVs may carry sequence elements or histone modifications similar to those of imprinted genes, preventing their demethylation in the zygote. High levels of DNA methylation on IAP elements and possibly other ERVs during the period of genome-wide demethylation may be maintained by a protein...
termed PGC7/Stella, since Stella−/− oocytes give rise to zygotes with over 2-fold reduced IAP methylation levels (97).

![Figure 1.3. Genome reprogramming and retrotransposon methylation in mouse preimplantation development.](image)

Relative levels of DNA methylation on IAP and LINE1 retrotransposons as well as single-copy genes in the maternal and paternal pronuclei are shown. LINE1 methylation differs in the maternal and paternal pronuclei, and both are shown with respective signs.

1.8.2. Epigenetic reprogramming and ERV regulation in mouse germ cell development

Both male and female primordial germ cells (PGCs) arise in the posterior primitive streak of the embryo at E7.5 and start migrating to the genital ridge, entering it around E11.5 (reviewed in (94)). Upon their arrival at the genital ridge, PGCs are subject to extensive epigenetic reprogramming resulting in erasure of parental imprints manifested by demethylation of maternally and paternally imprinted loci (Fig. 1.4) (118). In the mouse, male and female gonads are morphologically indistinguishable at E12.5. Male germ cells undergo mitotic arrest at E13.5 in G1 phase. They resume mitosis just after birth, and the first spermatogenic stages enter meiosis at least a week after birth. Female germ cells undergo one more round of DNA replication and enter meiosis. They are arrested after birth in prophase of meiosis I which resumes only before ovulation (89). In male germ cells, prospermatogonia, paternal imprints are established between E14.5 and birth (Fig. 1.4), while
in female germ cells, maternal imprints are established after birth during oocyte growth (118).

Similar to early embryogenesis, methylation of IAP, and to a lesser extent, LINE1 retrotransposons, is maintained during global demethylation occurring in germ cells (42, 71). Even at E12.5 prospermatogonia, IAP methylation level remains at 45-50% (55), staying substantially higher than LINE1 methylation at 13% (66). Subsequently, re-establishment of methylation, virtually complete by the time of birth, takes place on paternally methylated genes, IAP and LINE1 elements (Fig. 1.4) (55).

![Diagram showing CpG methylation levels and events](image)

**Figure 1.4. Genome reprogramming and retrotransposon methylation in the mouse germline.**

Relative levels of DNA methylation on IAP and LINE1 retrotransposons as well as single-copy genes are shown. Paternally imprinted genes are those differentially methylated in developing prospermatogonia during male embryogenesis. Maternally imprinted genes are those methylated during oocyte growth in the female germline after birth.

Intriguingly, transgenic mice with an IAP LTR driving expression of a reporter gene indicated that IAP promoter activity is restricted exclusively to undifferentiated spermatogonia in the testis of late-stage embryos (E16.0) or adults, and not detected in 2-cell embryos, blastocysts, E13.0 embryos or female germ cells (33). This restricted expression pattern suggests that either some of the IAP LTRs known to be transcriptionally active in other tissues, such as thymus or early embryo, possess regulatory elements not present in the
LTR used for the transgene, or that a limited number of elements evade host suppression due to position effects (33).

1.9. ERV regulation by small RNAs

The suppression mechanisms reviewed above fall into the category of transcriptional silencing imposed by DNA methylation or histone remodelling. Another mechanism of retrotransposon suppression is referred to as gene silencing directed by double-stranded RNA (dsRNA) molecules. Its end result is post-transcriptional or translational silencing occurring via degradation of target mRNA or abrogation of protein synthesis, or transcriptional silencing in the form of RNA-directed methylation and/or assembly of repressive chromatin structures on the corresponding DNA sequence. Retrotransposon silencing by RNA-directed DNA methylation or repressive chromatin establishment is a phenomenon described in plants, fungi, Drosophila, nematode and human cell lines (21, 85, 88, 122, 147). Though very tempting, the proof of existence of RNA-mediated DNA methylation in mammals remains elusive so far.

Recent advances in the field, however, had suggested that dsRNA may be involved in germline retrotransposon silencing in mammals by means of RNA interference (RNAi). RNAi is a mechanism of post-transcriptional repression conserved in eukaryotes, whereby dsRNA causes degradation of homologous mRNAs (2, 36). Argonaute proteins associate with small RNAs and guide post-transcriptional gene silencing by degradation or translational repression of corresponding mRNAs. Argonaute proteins in the mouse are subdivided into Argonaute and Piwi clades, the latter containing Miwi, Miwi2 and Mili, expressed predominantly in the male germline. A significant proportion of small RNAs
associated with Piwi proteins (piRNAs) in the male mouse germline are specific in sequence to retrotransposons and are produced from genomic clusters of transposable element fragments (4), similar to the pathway reported for Drosophila (15). Intriguingly, retrotransposon-related small RNAs corresponding in size to siRNAs (140) or piRNAs (129) have also been reported in oocyte. Moreover, Mili and Miwi2 mutant mice exhibit aberrantly up-regulated expression of IAP and LINE1 elements and lose DNA methylation on LINE1 elements in sperm (4, 23), suggestive of a crucial role for Piwi proteins in retrotransposon suppression in the germline. Recent research demonstrates that during the critical period of de novo methylation in male germline, LINE1 and IAP methylation is reduced by half in Mili and Miwi2 mutant germ cells, correlating with decrease in the amount of respective piRNAs (65). These data implicate Miwi2 and especially Mili proteins in de novo methylation of repeats in fetal testes. Moreover, similar defect in DNA methylation of retrotransposons is observed in Dnmt3l mutants (13), suggesting that Mili may guide Dnmt3l to repeats for de novo methylation. While the role of piRNAs in DNA methylation is suggested by Mili and Miwi2 mutants, more evidence is needed to unravel the exact mechanism of this phenomenon.

1.10. Thesis objectives

It is still unclear how and why some species successfully silence ERVs while others allow for their expression or are incapable of suppressing it. Unlike in human, ERVs are in an expansion phase in the mouse genome with specific IAP and ETn variants currently playing the dominant role (83). These elements have accumulated to hundreds or thousands of copies but only a few have a high probability of retrotransposing, apparently due to
specifics of the sequence of these currently active variants. Although ERV insertions are not a source of new mutations in human, understanding their regulation in mice may help shed light on the multiple host mechanisms acquired throughout evolution to suppress ERVs and the means ERVs employ to counteract this suppression.

In Chapter 2 of this thesis, I perform detailed analysis of the ETn/MusD LTR promoter, examining the regions essential for the undifferentiated cell-specific expression of this ERV family. I also attempt to identify transcription factors playing a role in ETn/MusD regulation and define sites of transcriptional initiation. This work enhances our understanding of the structure of the LTR promoter and paves the way for further research dedicated to regulation of ETn/MusD transcription and host-imposed transcriptional suppression.

In Chapter 3 of the thesis, I analyse the difference in ETn/MusD epigenetic regulation between expressing and non-expressing cell lines. I also discover and elaborate on the epigenetic differences between the two subfamilies, ETnII and MusD retrotransposons, of which ETnII is the most active retrotranspositional mutagen. The results of this work allow me to propose a new model of retrotransposon escape from host suppression and provide insight into host mechanisms involved in ERV restriction.
1.11. References


methylase independent and marked by a repressive histone code. EMBO J 19:5884-94.


CHAPTER 2
TRANSCRIPTIONAL REGULATION OF ETN ELEMENTS, AN
ACTIVE FAMILY OF MOUSE LTR-RETROTRANSPOSONS

1 A version of this chapter has been published in:
2.1. Introduction

LTR sequences have a key role in controlling the expression of retroviral elements and tissue tropism of retroviruses (19, 25, 33) employing a wide spectrum of cellular transcription factors (TFs). Despite the fact that ETn elements are active mouse mutagens, very little is known about the factors controlling their expression or that of their coding-competent MusD relatives. In this study, we have performed functional analysis of ETnII LTRs to gain insight into transcriptional regulation of this intriguing family of retrotransposons.

2.2. Materials and methods

2.2.1. Cell culture, transient transfections and luciferase assays

Mouse embryonic teratocarcinoma cell line P19 and mouse embryonic fibroblast cell line NIH3T3 (ATCC) were maintained in Dulbecco's modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum, penicillin and streptomycin. Cells were seeded 24 h prior to transfection into six-well plates at a density of $3 \times 10^5$ cells/well. Cells grown in monolayers were transfected with 1 μg of plasmid DNA and 100 ng of the Renilla luciferase vector pRL-TK using Lipofectamine-2000 (Invitrogen) according to manufacturer’s instructions. Cells were harvested into 1× passive lysis buffer (Promega) 24 h after transfection, and firefly and luciferase activities were measured using the dual-luciferase reporter assay system (Promega) according to manufacturer’s instructions. The data were standardized to the internal Renilla luciferase control and expressed with regard to the promoter efficiency of the promoterless pGL3-Basic (pGL3B) vector. The results are a mean and standard deviations of three separate experiments performed in duplicates.
2.2.2. Plasmids and constructs

ETnII and MusD LTRs were cloned into the KpnI-BglII cloning site of the pGL3B luciferase reporter vector (Promega). The 318 bp-long LTRs were amplified from C57BL/6 mouse genomic DNA by nested PCR, except for LTR #3 amplified from a BAC clone. All oligonucleotides used in plasmid construction are listed in Table 2.1. The following forward primers located in the flanking sequence were used for the 1st round PCR: LTR of elements #1 (AC074208) – primer wiz_2921as; #7 (AC079540) – 540_1s; #6 (AC090879) – 879_1s; #13 (AC079497) – 497_1s; # 10 (AL589871) – 871_1s; # 8 (AC087263) – 263_1s; # 9 (AF132039) – 039_1s. The reverse primer MusD/ETnII_361as was common for all 1st round PCR reactions. Nested PCR primers were identical for all LTRs with forward primer: IM_LTR_1s and reverse primer: IM_LTR_2as. Progressive 5’ deletion constructs of the ETnII LTR#6 were generated by PCR amplification with a common reverse primer IM_LTR_2as and forward primers IM_ETnII_18s, IM_ETnII_29s, IM_ETnII_43s, IM_ETnII_61s, IM_ETnII_61s-T, IM_ETnII_76s, IM_ETnII_140s and IM_ETnII_162s. 3’ deletion constructs of the ETnII LTR#6 were generated by PCR amplification with a common forward primer IM_LTR_1s and reverse primers IM_ETnII_84-as, IM_ETnII_164-as, IM_ETnII_197-as, IM_ETnII_234-as and IM_ETnII_277-as. The resulting LTR fragments were cloned into the KpnI-BglII site of the pGL3B reporter vector and sequenced.

To generate mutation in the first Sp1 binding site, in vitro mutagenesis was performed by amplification of ETnII LTR#6 with primers IM_LTR_mSp1-s and IM_LTR_2as. For the other two Sp1 binding site mutation constructs, the 3’ part of the ETnII LTR#6 was first amplified with a common primer IM_LTR_1s and the following mutating oligonucleotides: IM_mutSp1/32-as to generate mutation in the 2nd Sp1 binding site and IM_mutSp1/61-as for
mutation in the third Sp1 binding site. Products of these first reactions were used as forward primers for amplification of ETnlII LTR#6 with a common reverse primer IM_LTR_2as. All PCR reactions were carried out in standard conditions using Pfu DNA polymerase (Invitrogen). All insertions cloned into the pGL3B vector were sequenced.

Table 2.1. List of primers.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Details</th>
<th>Sequence 5’ to 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’actin</td>
<td>Antisense</td>
<td>5’-GCGCAAGTTAGGTTTTGTCAAAG</td>
</tr>
<tr>
<td>039_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-TCCCTTCCCAATCTCTCAGCA</td>
</tr>
<tr>
<td>263_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-GCCACAACGTAACAAACCT</td>
</tr>
<tr>
<td>497_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-GATCCAGCAACACCTCTCTTG</td>
</tr>
<tr>
<td>540_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-TGTAAGGACCTGAGTCTGGC</td>
</tr>
<tr>
<td>871_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-TTGTTGCAGTCTTATTGCG</td>
</tr>
<tr>
<td>879_1s</td>
<td>+EcoRI site</td>
<td>5’-GGCAATTC-GAAGGAGCTGGAAGCATCTG</td>
</tr>
<tr>
<td>ETnlII_665as</td>
<td>5’-RACE</td>
<td>5’-GCTATAAGGCCCAGAGAAATTTC</td>
</tr>
<tr>
<td>ETnlIgr1_3636as</td>
<td>Antisense</td>
<td>5’-CATT(T/C)(G/A)TTAGT(C/T)AGGGGGTATTAAGTGAC</td>
</tr>
<tr>
<td>GL-2</td>
<td>5’-RACE</td>
<td>5’-CTTTATGTTTTTGGCGTCTTCCA</td>
</tr>
<tr>
<td>IM_3as</td>
<td>5’-RACE</td>
<td>5’-ATC(T/C)CTCTGCCATTCTTCAAG</td>
</tr>
<tr>
<td>IM_ETnlII_84-as</td>
<td>+BglII Site</td>
<td>5’-GACAGATCT-CCAGCATGGGGCTGCAAG</td>
</tr>
<tr>
<td>IM_ETnlII_164-as</td>
<td>+BglII Site</td>
<td>5’-CGAGATCT-AAGGAAGGGAGGGAGGACC</td>
</tr>
<tr>
<td>IM_ETnlII_197-as</td>
<td>+BglII Site</td>
<td>5’-GGAGATCT-GGCTCAAATTTTATTGTGCG</td>
</tr>
<tr>
<td>IM_ETnlII_234-as</td>
<td>+BglII Site</td>
<td>5’-CGAGATCT-GAAGAAGAACATGTAGCCAAG</td>
</tr>
<tr>
<td>IM_ETnlII_277-as</td>
<td>+BglII Site</td>
<td>5’-CGAGATCT-GAAGAAGAACATGTAGCCAAG</td>
</tr>
<tr>
<td>IM_ETnlII_18s</td>
<td>+KpnI site</td>
<td>5’-GGTACC-CCAGCTGAAACCTTGCTGCT</td>
</tr>
<tr>
<td>IM_ETnlII_29s</td>
<td>+KpnI site</td>
<td>5’-GGTACC-CTGTTGCAGGGGTTGAG</td>
</tr>
<tr>
<td>IM_ETnlII_43s</td>
<td>+KpnI site</td>
<td>5’-GGTACC-GTGGAGCTTCTGCTATTGCG</td>
</tr>
<tr>
<td>IM_ETnlII_61s</td>
<td>+KpnI site</td>
<td>5’-GC GGTACC-TCTGTTCTGCAAGCCCCACTG</td>
</tr>
<tr>
<td>IM_ETnlII_61s-T</td>
<td>+KpnI site</td>
<td>5’-GC GGTACC-TCTGTTCTGCAAGCCCCACTG</td>
</tr>
<tr>
<td>IM_ETnlII_76s</td>
<td>+KpnI site</td>
<td>5’-GC GGTACC-CACTGCTGGAACCTGCGGAG</td>
</tr>
<tr>
<td>IM_ETnlII_140s</td>
<td>+KpnI site</td>
<td>5’-CG GGTACC-AAATCGGGTCCCCTCCCCC</td>
</tr>
<tr>
<td>IM_ETnlII_162s</td>
<td>+KpnI site</td>
<td>5’-CG GGTACC-CTTCTACACTGTGTCGCAAC</td>
</tr>
<tr>
<td>IM_ETnlII/MusD_329s</td>
<td></td>
<td>5’-GAACTGGGACCTGGAAGAATG</td>
</tr>
<tr>
<td>IM_ETnlII/MusD_624as</td>
<td></td>
<td>5’-TCTTCTTCCACTAAGGCAAGC</td>
</tr>
<tr>
<td>IM_LTR_1s</td>
<td>+KpnI site</td>
<td>5’-CGGGTGATCC-TGTAGTCCCTCCCGC(T/A)</td>
</tr>
<tr>
<td>IM_LTR_2as</td>
<td>+BglII site</td>
<td>5’-GACAGATCT-TGTTGCAGCCGCAAGCAG</td>
</tr>
<tr>
<td>IM_LTR_mSp1-s</td>
<td>+KpnI site</td>
<td>5’-GACGGTGATCC-TGTAGTCCCTCCATACCCGCA</td>
</tr>
<tr>
<td>IM_Luc_208as</td>
<td>5’-RACE</td>
<td>5’-CGATATGTGCAATCTGTAAAGGC</td>
</tr>
<tr>
<td>IM_mutSp1/1-s</td>
<td>Mutated nt.</td>
<td>5’-TCAGGGGATGGGGAGAGACTACA</td>
</tr>
<tr>
<td>IM_mutSp1/1-as</td>
<td>Mutated nt.</td>
<td>5’-TCAGGGGATGGGGAGAGACTACA</td>
</tr>
<tr>
<td>IM_mutSp1/32-s</td>
<td>Mutated nt.</td>
<td>5’-CTTGCTCAGGATGGAGCTTCTTG</td>
</tr>
<tr>
<td>IM_mutSp1/32-as</td>
<td>Mutated nt.</td>
<td>5’-GCAGGAAGCTCCATACCTGAGCAAG</td>
</tr>
<tr>
<td>IM_mutSp1/61-s</td>
<td>Mutated nt.</td>
<td>5’-TCGTTCTGCACACTACACCTGAGCAAG</td>
</tr>
<tr>
<td>IM_mutSp1/61-as</td>
<td>Mutated nt.</td>
<td>5’-TCCAGCAGTGAGTGGCAGAAGCA</td>
</tr>
<tr>
<td>Primer name</td>
<td>Details</td>
<td>Sequence 5’ to 3’</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>3’actin</td>
<td>Antisense</td>
<td>5’-GCGCAAGTTAGGTTTTGTCAAAG</td>
</tr>
<tr>
<td>IM_Sp1/1-s</td>
<td>Binding site</td>
<td>5’-TGTAGTCTCC TC CC CCCAGCCTGA</td>
</tr>
<tr>
<td>IM_Sp1/1-as</td>
<td>Binding site</td>
<td>5’-TCAGGCTGGGGAGGGAGACTACA</td>
</tr>
<tr>
<td>IM_Sp1/32-s</td>
<td>Binding site</td>
<td>5’-CTTGCTCAGGGGTGAGGCTTCCCTGC</td>
</tr>
<tr>
<td>IM_Sp1/32-as</td>
<td>Binding site</td>
<td>5’-GCAGGAAGCTCCACCCCCTGAGCAAG</td>
</tr>
<tr>
<td>IM_Sp1/61-s</td>
<td>Binding site</td>
<td>5’-TCGTTTCTGCCACGCCCACGTTGGA</td>
</tr>
<tr>
<td>IM_Sp1/61-as</td>
<td>Binding site</td>
<td>5’-TCCAGCAGTGGCGTGCGAGAAAGC</td>
</tr>
<tr>
<td>MusDEtnII_361as</td>
<td>+EcoRI site</td>
<td>5’-GCGAATTC-ATCTCTCTGCCATTTT(C/T)AGG</td>
</tr>
<tr>
<td>MusD_823s</td>
<td></td>
<td>5’-CACTCAGGCGAGAAATTTACC</td>
</tr>
<tr>
<td>MusD_1192as</td>
<td></td>
<td>5’-TCATTAGTCAGCCTCTGAGGTT</td>
</tr>
<tr>
<td>STD1</td>
<td></td>
<td>5’-TGAAGAAACGGCAAAGGATTTCCTTGGA</td>
</tr>
<tr>
<td>STD2</td>
<td></td>
<td>5’-ATTACCCAGCTCCACTGCTGGA</td>
</tr>
<tr>
<td>wiz_2921as</td>
<td>+EcoRI site</td>
<td>5’-GCGAATTCT-GTGAGATGGGTGAGTCCCAG</td>
</tr>
</tbody>
</table>

### 2.2.3. Preparation of RNA, Northern blotting and 5’-RACE

P19 and NIH3T3 cells grown to 90% confluency were lysed in 1ml TRIzol™ (Invitrogen) per 10 cm², total RNA was extracted according to the TRIzol™ protocol. Two samples of ES cell line R1 RNA were generously provided by Lynn Mar and Dr. Laura Sly.

For Northern blotting, 10 μg of RNA for each lane was denatured, electrophoresed in 1.2% agarose 3.7% formaldehyde gel in 1x MOPS buffer, transferred overnight to a Zeta-probe nylon membrane (Bio-Rad) and UV-cross-linked. Probes specific for ETnI and MusD elements were synthesized by PCR using following primers: ETnI: forward – STD1, reverse – STD2; MusD: forward – MusD_823s, reverse – MusD_1192as. Amplified DNA fragments were α-32P-labeled using a Random Primers DNA Labeling System (Invitrogen). α-32P-end-labeled antisense oligonucleotide probes were used as probes for ETnII-β elements (ETnIIgr1_3636as) (3) and β-actin (3’actin). Membranes were prehybridized in ExpressHyb (BD Biosciences) for 2-4 hours at 68°C for fragment probes and 54°C for oligonucleotide
probes, hybridized overnight at the same temperatures in fresh ExpressHyb, washed according to manufacturer’s instructions and exposed to film.

5'-RACE was carried out using the FirstChoice RLM-RACE kit (Ambion) with 10 µg of total RNA from untreated P19 cell line or P19 cells transfected with expression constructs according to manufacturer’s instructions. Two rounds of nested PCR were carried out with forward primers specific for the 5'-RNA adapter and reverse primers specific for an ETnII element. Endogenously expressed ETnII elements were cloned using ETnI/II_665as as the reverse primer for the first round PCR and IM_3as as the reverse primer for the second round PCR. Both primers were located downstream of 5’ LTR, prohibiting amplification of solitary LTRs. Transfected full-length and 3’ deleted constructs were cloned using primers specific for the luciferase gene: IM_Luc_208as as the reverse primer for the first round PCR and GL-2 as the reverse primer for the second round PCR. Specific amplification products were cloned into the pGEM-T vector (Promega). Several independent clones derived from each PCR product were sequenced.

2.2.4. Preparation of genomic DNA and Southern blotting

P19 and NIH3T3 cells grown to 90% confluency were lysed in 750µl DNAzol™ (Invitrogen) per 10 cm², extraction of genomic DNA was carried out according to the DNAzol™ protocol. Genomic DNA from ES cell lines R1, EK.ECC, AB-1 and 671 was a gift from Lynn Mar; genomic DNA from mouse strains 129S1/Sv-+Tyr-cKitlSl-J/+ (stock #000090), 129X1/SvJ (stock #000691) and C3H/HeJ (stock # 000659) was obtained from the Jackson Laboratory (Bar Harbor, Maine, USA). 1µg gDNA for each lane was digested overnight, ran overnight in a 0.8% 1xTAE agarose gel and transferred overnight in 0.4 M
NaOH to a Zeta-probe nylon membrane (Bio-Rad). A probe specific for both ETnII and MusD elements was amplified by PCR from the plasmid containing an ETnII clone using following primers: forward – IM_ETnII/MusD_329s, reverse – IM_ETnII/MusD_624as and \( \alpha^{32}P \)-labeled using a Random Primers DNA Labeling System (Invitrogen). Membrane was prehybridized in ExpressHyb (BD Biosciences) at 60°C for 3-4 hours, hybridized overnight at the same temperature in fresh ExpressHyb, washed according to manufacturer’s instructions and exposed to film.

### 2.2.5. Nuclear extraction and electrophoretic mobility shift assays (EMSA)

Nuclear extracts were prepared from P19 and NIH3T3 cells as follows: cells were washed in PBS, pelleted at 5000 rpm for 5 min at 4°C and resuspended in lysis buffer (10 mM Tris-HCl (pH 8.0), 60 mM KCl, 1 mM EDTA, 1 mM dithiothreitol (DTT), protease inhibitors, 0.1% NP-40). After incubation on ice for 5 min, the lysates were spun at 2500 rpm at 4°C for 4 min. The pelleted nuclei were washed in lysis buffer without NP-40 and pelleted at 2,500 rpm for 4 min at 4°C. The nuclear pellet was resuspended in nuclear extraction buffer (20 mM Tris-HCl pH 8.0), 420 mM NaCl, 1.5 mM MgCl\(_2\), 0.2 mM EDTA, and 25% glycerol). After incubation on ice for 20 min, the nuclei were briefly vortexed and spun at 13,000 rpm for 6 min at 4°C. Supernatant was removed and used as a nuclear extract. Protein concentrations were determined using the Bio-Rad protein assay (Bio-Rad).

Electrophoretic mobility shift assays (EMSA) were performed with \( \gamma^{32}P \)-labeled double stranded oligonucleotides identical to the three fragments of the ETnII LTR#6 containing putative Sp1/Sp3 binding sites (I, II and III). 5 \( \mu \)g nuclear extracts were preincubated with 20 \( \mu \)l reaction mixture containing 0.5 \( \mu \)g of poly(dl-dC) in 10 mM Hepes,
4 mM DTT, 0.2 mM EDTA, 0.1 mM NaCl, 0.1 mg/ml BSA, 4% glycerol. Samples were incubated for 20 min on ice with either 200-fold excess of unlabeled competitor oligonucleotide or an antibody specific for Sp1 or Sp3 (sc-59 X and sc-644 X, Santa Cruz Biotechnology) for supershift assays. Double stranded oligonucleotides were labeled using $\gamma^{32}$ATP and T4 polynucleotide kinase according to manufacturer’s protocols (Invitrogen) and purified on MicroSpin™ G-25 columns (Amersham Pharmacia Biotech). $\gamma^{32}$P-labeled oligonucleotide probe was added to the reaction mixture, and the incubation was continued on ice for another 30 min. Reaction products were fractionated on a 5% polyacrylamide gel and ran at 4°C in 0.5x TBE buffer at 8-15 mA for 3 h. Gels were transferred onto a Whatman paper, dried and autoradiographed.

2.3. Results and discussion

2.3.1. Transcription of endogenous ETn and MusD in ES and EC cell lines

We performed RT-PCR (data not shown) and Northern blotting analysis to estimate the amount of endogenous ETn and MusD transcripts in the embryonic teratocarcinoma (EC) cell line P19, two samples of embryonic stem (ES) cell line R1 and the differentiated embryonic fibroblast NIH3T3 cell line (Fig. 2.1A). For ETnII, an oligonucleotide probe specific for the most abundant subtype, ETnII-β, was used. This probe spans a small deletion found only in ETnII-β elements (Fig. 2.1B) (3). ETn and MusD RNA was abundant in P19 and R1 cells, with both groups of elements expressed at a higher level in P19 cells. The NIH3T3 cell line, as expected, did not express ETn or MusD elements. Single major bands depicted by Northern blotting clearly denote a preference for transcription of a group of elements of a specific size, suggesting other subtypes are either transcriptionally inert or
transcribed at far lower levels. Similar results were previously demonstrated for ETnI and ETnII elements combined (44). Thus, our analysis has confirmed previously reported high levels of ETn and MusD expression in undifferentiated EC and ES cell lines (3, 7) and their absence in the differentiated NIH3T3 cell line (7).

![Figure 2.1](image)

**Figure 2.1. ETn and MusD transcript abundance in undifferentiated and differentiated cell lines.**

(A) Northern blot. 10 µg RNA isolated from EC P19 (lane1), ES R1 (lanes 2 and 3) and fibroblast NIH3T3 cells (lane 4) was hybridized with ETnI, ETnII-β, MusD or β-actin probes. ETnII-β and MusD probes produced single bands, and the ETnI probe produced a single major band of 5.6 kb and two minor bands of 7.2 kb and 6.2 kb. The blot was initially hybridized with an ETnII-β oligonucleotide probe, then stripped and rehybridised in turn with other indicated probes. (B) Structures of ETn elements and location of probes used in Northern blotting.

2.3.2. Promoter activity of ETnII and MusD LTRs in P19 and NIH3T3 cell lines

Since ETnII elements are most frequently found at the sites of recent insertions and are more highly transcribed than MusD elements (2, 3), we analyzed the promoter efficiencies of ETnII and MusD LTRs in differentiated versus undifferentiated cell lines. The 5’ LTRs of four ETnII and four MusD elements belonging to different subfamilies were chosen for the study. These elements possess various degrees of nucleotide identity between
their 3’ and 5’ LTRs, with more recently inserted elements displaying 100% identity (Table 2.2).

Table 2.2. Description of the LTRs used in the study.

<table>
<thead>
<tr>
<th>#</th>
<th>% identity of 5’ and 3’ LTR 2)</th>
<th>accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ETnII</strong></td>
<td>1 100 % (317/317)</td>
<td>AC074208</td>
</tr>
<tr>
<td></td>
<td>7 100 % (317/317)</td>
<td>AC079540</td>
</tr>
<tr>
<td></td>
<td>6 100 % (318/318)</td>
<td>AC090879</td>
</tr>
<tr>
<td></td>
<td>13 94 % (301/318)</td>
<td>AC079497 (13 bp deletion in 5'LTR)</td>
</tr>
<tr>
<td><strong>MusD</strong></td>
<td>3 100 % (319/319)</td>
<td>AC084696</td>
</tr>
<tr>
<td></td>
<td>10 99 % (316/319)</td>
<td>AL589871</td>
</tr>
<tr>
<td></td>
<td>8 96 % (308/319)</td>
<td>AC087263</td>
</tr>
<tr>
<td></td>
<td>9 97 % (310/319)</td>
<td>AF132039</td>
</tr>
</tbody>
</table>

1) element number used in this article  
2) nucleotide identity given in brackets

The LTRs were cloned into the pGL3B luciferase reporter vector, transfected into P19 and NIH3T3 cells and analyzed for their promoter efficiency by measuring luciferase gene expression (Fig. 2.2). While the sample size is small, we observed that the LTRs of younger elements had higher promoter activities, suggesting degeneration of promoter-competent LTRs with time due to accumulation of mutations either by random genetic drift or by negative selection aimed at the elimination of deleterious sequences.

CpG methylation is an important factor in suppression of LTR-driven transcription, as shown for intracisternal A particle elements and other classes of mouse ERVs (5, 12, 27, 45). As discussed in Chapter 1.10., a wave of global mammalian CpG methylation occurs during cellular differentiation around the time of implantation (46), correlating with the reduction in ETn transcription. However, despite the lack of CpG methylation of the expression plasmids, transiently transfected LTR promoter constructs revealed fifteen times lower promoter
activity in NIH3T3 compared to P19 cells (Fig. 2.2), suggesting an important role for cell type-specific transcription factors.

![Figure 2.2. Promoter activity of ETnII and MusD LTRs in P19 and NIH3T3 cells.](chart)

Four ETnII and four MusD LTRs (see Table 1.1 for accession numbers) cloned into a pGL3B luciferase reporter vector were assayed for promoter efficiency by transient transfections into P19 and NIH3T3 cells.

We propose that, in addition to possible suppressing effects of methylation, ETn transcription in NIH3T3 cells may be restricted by the lack of certain activating transcription factors present in the P19 cell line or by the presence of suppressor transcription factors in the NIH3T3 cell line. This is further supported by the recently published evidence that CpG demethylation increases the level of transcription initiating from human endogenous retrovirus-K LTRs only in the context of the specific transcription factor pool present in germ cell tumors (16). However, since the absolute difference in LTR promoter activities between P19 and NIH3T3 cells is difficult to determine, our subsequent experiments focus on the relative promoter activity within each cell line of partially deleted or mutated LTRs.
2.3.3. Regions in the ETnII LTR crucial for high promoter ability in P19 cells

To identify regulatory regions in the ETn LTRs responsible for their exceptionally high promoter activity in the P19 cells, we performed functional analysis of the ETnII LTR#6 (accession number AC090879) which displayed the highest promoter activity among LTRs tested (see Fig. 2.2). The promoter activities of the 5’ LTR deletion constructs were analyzed in P19 and NIH3T3 cells. We identified three regions in the ETnII LTR essential for high LTR promoter efficiency in P19 cells. All were positioned within the first 76 bp of the LTR, 190 to 120 bp upstream of the previously mapped transcription start site (13) (Fig. 2.3A, 2.3B).

Figure 2.3. Promoter activity of 5’-deleted ETnII LTRs in P19 and NIH3T3 cells.

(A) Graphical representation of the promoter efficiency of 5’-deleted LTR#6 constructs. Numbers indicate the span of the construct, with base pairs corresponding to those in the LTR#6. The putative TATA-box and transcription start site identified previously (13) are shown. Ovals represent possible Sp1 binding sites. Sp1 binding site mutated in the 61(T)-317 construct is depicted in black. (B) Promoter activity of the LTR deletion constructs in P19 and NIH3T3 cells as percentage of promoter activity of the full-length LTR#6.
Regions overlapping nucleotides 1-18, 29-43 and 61-76 bp are clearly responsible for P19-specific expression, since these deletions lead to a severe reduction of activity in P19 (down to 5% of the wild-type level) but not NIH3T3 cells. All of the three critical regions contain predicted recognition sequences for Sp1 and Sp3 transcription factors (Fig. 2.4). Notably, a single-nucleotide C→T mutation positioned 72 bp from the beginning of the LTR, abrogating the third Sp1 binding site, is found in ETnII LTR #13 and MusD LTRs #8 and #9 and may account for their lower promoter activities (see Fig. 2.2). A 5’-deletion construct incorporating this mutation (61(T)-317) demonstrates significantly reduced promoter activity in P19 cells in comparison to the wild-type deletion construct (61-317). We hypothesize that the second and third regions described above are necessary for high P19-specific and, possibly, embryo-specific expression of ETn elements.

Figure 2.4. Sequence of the ETnII LTR#6, position of deletion constructs and EMSA oligonucleotides.
The transcription start site, as identified in (13), is indicated as +1. The predicted TATA-like sequence and polyadenylation signals (13) are boxed and U3, R and U5 regions indicated. Oligonucleotides used as probes in EMSA (I, II and III) are underlined. Sp1/Sp3 recognition sequences are highlighted in light gray, core regions specified in reverse text.
2.3.4. The three ETnII LTR critical regions bind Sp1 and Sp3 transcription factors

To determine whether the three regions critical for high ETnII LTR promoter activity bind Sp transcription factors, we performed electrophoretic mobility shift assays (EMSA) with double-stranded DNA oligonucleotides corresponding to each of the three regions (Fig. 2.4). Supershift results demonstrate efficient binding of Sp1 and Sp3 transcription factors from both P19 and NIH3T3 nuclear extracts to all of the three candidate regions, resulting in extremely similar band patterns. However, incubation of the first segment with NIH3T3 nuclear extract produced an extra band, marked with an asterisk, indicating the presence of additional complexes (Fig. 2.5), and suggestive of repression. This hypothesis, however, is unlikely based on deletion analysis results described above and mutational analysis described in the next section.

**Figure 2.5. Binding of Sp1 and Sp3 transcription factors to the regions crucial for high LTR promoter activity.**
γ-32P-labeled oligonucleotides identical in sequence to the regions essential for high LTR promoter ability in P19 cells (I, II and III) were incubated with nuclear extracts from P19 and NIH3T3 cells. Competition assays were performed with 200-fold excess of wild-type unlabeled oligonucleotide. Antibodies against Sp1 and Sp3 supershifted specific bands.
Binding of Sp1 and Sp3 to these regions in both P19 and NIH3T3 cells suggests that Sp1-dependent transcription may be mediated by cofactors that bind to Sp proteins. This hypothesis is supported by numerous recent reports of heterodimerization of various transcription factors with Sp1, resulting in transcriptional activation through Sp1 binding sites (14, 30, 36).

2.3.5. Mutational analysis of the ETnII LTR

To assess the importance of the Sp1/Sp3 binding sites for the P19 cell-specific promoter activity of ETnII LTRs, we introduced mutations abolishing Sp1/Sp3 binding sites into the ETnII LTR#6 and assayed the promoter efficiency of these constructs. Prior to transfection experiments, we employed EMSA to demonstrate that mutations identical to the ones introduced into the LTR completely abolished binding of proteins to these regions (data not shown). Mutation of either the first or third Sp binding sites did not lead to a significant reduction of promoter activity, whereas mutation of the second Sp binding site led to a 2-fold decrease in promoter efficiency. Notably, combined mutation of sites 2 and 3 was almost as dramatic as the triple mutation and reduced promoter activity in P19 cells to the level of the wild-type LTR promoter activity observed in NIH3T3 cells (Fig. 2.6A, 2.6B, 2.6C). Promoter activity of the triple mutant was only 9% of the wild-type level in P19 cells, while the same combination of mutations reduced promoter activity in NIH3T3 cells to a lesser degree. NIH3T3 cells are clearly less responsive to the abolition of Sp1/Sp3 recognition sequences compared to P19 cells, confirming the crucial role of these regions in transcriptional activation of ETnII LTRs in the P19 cell line and, presumably, in ES cells and embryonic tissues.
Figure 2.6. Promoter activity of Sp-binding site mutant LTRs in P19 and NIH3T3 cells.

(A) Graphical representation of the promoter efficiency of LTR#6 mutated at Sp1/Sp3 recognition sites. The putative TATA-box and transcription start site identified previously (13) are shown. Ovals represent possible Sp1 binding sites. Crossed ovals represent mutated Sp1 binding sites. (B) The core of the wild-type Sp1 recognition sequences is boxed and the mutated nucleotides are represented in lower case. (C) Promoter activity of the mutated LTR constructs in P19 and NIH3T3 cells as percentage of promoter activity of the wild-type LTR#6.

The results of these deletion and mutation analyses are suggestive of activation of LTR promoter ability in P19 cells versus its suppression in NIH3T3 cells, since we registered no increase in the promoter efficiency in NIH3T3 cells upon the deletion regions which could
contain binding sites for the putative transcriptional repressor (see Fig. 2.3). Despite high responsiveness of the LTR reporter constructs to the mutation of Sp1/Sp3 binding sites, cotransfection of P19 and NIH3T3 cell lines with expression plasmids encoding Sp1 and Sp3 proteins (kind gifts from Paul Gardner, Tom Shenk and Guntram Suske) together and separately in varying proportions did not significantly alter the promoter efficiency of ETnII or MusD LTRs (data not shown). This result suggests that ubiquitously expressed Sp1 and Sp3 are not limiting factors for LTR transcription efficiency and indicates possible involvement of other tissue-specific co-regulating factors. This is confirmed by our unpublished data that Sp1 and Sp3 expression determined by RT-PCR is similar in both cell lines. It is also possible that other transcription factors which employ the same or overlapping binding sites as the Sp family proteins are specifically expressed in undifferentiated cells and are the ones responsible for transcriptional activation of ETn and MusD elements.

2.3.6. High promoter activity of the 3’-deleted ETnII LTR lacking the putative TATA-box

To determine whether LTR sequences 3’ to the previously reported transcription start site (44) play a role in transcriptional activation, functional analysis of the 3’ deleted constructs of the ETnII LTR#6 was carried out. The promoter activity of the LTR 3’ deletion constructs was analyzed in P19 and NIH3T3 cells. Deletions of the LTR from the 3’ end caused a mild drop in the promoter efficiency of the LTR in both P19 and NIH3T3 cell lines, suggesting a lack of cell-specific regulatory sequences in the region (Fig. 2.7). Notably, the construct spanning positions 1-164 bp of the LTR had surprisingly high promoter activity in the P19 cell line, despite the deletion of its 3’ sequence including a previously predicted non-
canonical TATA-box (13). This may be due to an alternative transcription start site 5' to the weak TATA-box, since GC-rich sequences present in the 5' end of the LTR may serve as a TATA-less promoter. Previous studies identified a single site of transcription initiation at the border of the U3 and R regions in the ETn LTR in primary AML (44) and carcinoma cells (13). However, the family of elements investigated in those studies were ETnI elements, which differ significantly from the ETnII elements studied here in the 3' end of their LTR and part of the downstream sequence (3, 38). Although the presumed U3 region is nearly identical in both families, highly divergent U5 and downstream sequences may play a role in variable transcription initiation between ETnI and ETnII elements. Our data is suggestive of an alternative transcription start site in ETnII LTRs that functions in the P19 cell line.

Figure 2.7. Promoter activity of 3’-deleted ETnII LTRs in P19 and NIH3T3 cells. (A) Graphical representation of the promoter efficiency of 3’-deleted LTR#6 constructs. The putative TATA-box and transcription start site identified previously (1) are shown. Ovals represent possible Sp1 binding sites. (B) Promoter activity of the LTR deletion constructs in P19 and NIH3T3 cells as percentage of promoter activity of the full-length LTR#6.
2.3.7. Determination of ETnII transcription initiation sites

To determine transcription start sites, we performed 5’-rapid amplification of cDNA ends (5’-RACE) of full-length and 3’-deleted ETnII LTR constructs transfected into P19 cells (Fig. 2.8A) and endogenous ETnII elements in P19 cells (Fig. 2.8B). The 3’-deletion construct spanning 1-164 bp of the LTR appeared to possess multiple transcription start sites upstream of the previously predicted non-consensus TATA-like sequence, explaining its high promoter activity in the P19 cells. The transcription initiation sites of the 1-164 bp construct largely overlap with those of the transfected full-length LTR#6 and fall primarily into the region spanning base pairs 149-161 of the LTR (Fig. 2.8B). It is therefore possible that a band detected in EMSA experiment by Tanaka et al. (44) with an oligonucleotide spanning base pairs 147-169 of the LTR contains polymerase complexes involved in initiation of transcription.

We have also observed multiple transcription initiation sites within the LTRs of endogenously expressed ETnII elements (Fig. 2.8B). This is a very common trait of promoters containing a weak TATA-box and a GC-rich upstream sequence; in fact, as many as 85% of all human genes have been reported to possess highly variable transcription start sites (42). The transcription start site most frequently used by endogenous ETnII elements in P19 cells does not exactly correspond to the sites employed by the transfected full-length LTR#6 reporter construct, though there is an overlap. This may be due to the fact that ETnII elements with LTRs other than LTR#6 account for the majority of transcripts in P19 cells, as seen from the sequence of 5’-RACE clones. The sequence differences between the LTRs may result in variable efficacies of transcription initiation sites. It is also possible that the
ETnII sequence downstream of the LTR, present in the endogenous ETnII elements but not in the LTR reporter constructs, may affect transcription initiation.

**A.**

<table>
<thead>
<tr>
<th>ETnII#6 AC090879 LTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGTAGTCTCCCCCTCCCCCAGCCTGAAACCTGCTTCACCTGCTCAT 60</td>
</tr>
<tr>
<td>TCGTTCTGCCCCGCACTGGCTGAACTCCGGAGCCACACCTGCTCATGG 120</td>
</tr>
<tr>
<td>ACCAGAGATTATTCGGCGGGAATCGGGTCCCCTCCCCCTTCCTTCATAACTGGTGTCACA 180</td>
</tr>
<tr>
<td>ACAATAAAATTTGAGCCTTGATCAGAGTAACTGTCTTGGCTACATTCTTTCTTCCGCCCC 240</td>
</tr>
<tr>
<td>GTCTAGATTCCTCTCTTACAGCTCGAGCGGCCTTCTCAGTCGAACCGTTCACGTTGCGAG 300</td>
</tr>
<tr>
<td>CTGCTGGCGGCCGCAACA 318</td>
</tr>
</tbody>
</table>

∇ - transfected 3’deleted LTR construct (1-164 bp)
¶ - transfected full-length LTR construct

**B.**

<table>
<thead>
<tr>
<th>Adcy1 ETnII insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGTAGTCTCCCCCTCCCCCAGCCTGAAACCTGCTTCACCTGCTCAT 60</td>
</tr>
<tr>
<td>TCGTTCTGCCCCGCACTGGCTGAACTCCGGAGCCACACCTGCTCATGG 120</td>
</tr>
<tr>
<td>ACCAGAGATTATTCGGCGGGAATCGGGTCCCCTCCCCCTTCCTTCATAACTGGTGTCACA 180</td>
</tr>
<tr>
<td>ACAATAAAATTTGAGCCTTGATCAGAGTAACTGTCTTGGCTACATTCTTTCTTCCGCCCC 240</td>
</tr>
<tr>
<td>GTCTAGATTCCTCTCTTACAGCTCGAGCGGCCTTCTCAGTCGAACCGTTCACGTTGCGAG 300</td>
</tr>
<tr>
<td>CTGCTGGCGGCCGCAACA 318</td>
</tr>
</tbody>
</table>

↓ - Endogenous ETnII in P19 cells

**Figure 2.8. Mapping of transcription initiation sites in ETnII LTRs by 5’-RACE.**

(A) Transcription initiation sites of transfected full-length and 3’-deleted ETnII LTR#6 in P19 cells. Symbols represent 5’ termini of individual RACE clones and numbers indicate that certain clones were found multiple times. The 1-164 3’-deleted construct is indicated with an arrow. The previously noted TATA-like element (13) is boxed and a second such motif is underlined. (B) Transcription initiation sites of endogenous ETnII elements in P19 cells. Arrows represent 5’ termini of RACE clones and numbers indicate that certain clones were found multiple times. This LTR sequence is that of the new ETnII insertion into the Adcy1 locus (17). The previously noted TATA-like element (13) is boxed and a second such motif is underlined.
The majority of the transcription initiation sites of endogenous ETnII elements in P19 cells are clustered in the region TCACAACAAT (176-185 bp), with all three C-A dinucleotides in this sequence employed for initiation of transcription (Fig. 2.8B). This sequence resembles a loose initiator element consensus YYA+1N(T/A)YY, where Y is a pyrimidine (for review, see (18, 39)). In the absence of a strong TATA-box, initiator sequences are known to effectively drive transcription in the presence of upstream activator Sp1 binding sites (for review, see (39)). Thus, ETnII elements appear to rely on Sp1/Sp3 recognition sequences in the absence of a strict TATA-box for transcription from multiple initiation sites. However, we have also noted another potential TATA-box at nucleotide positions 124-133 of the LTR (underlined in Fig. 2.8) that may play a role in initiation of some transcripts. Although the 5’-RACE procedure used is designed to detect only transcripts with a 5’ cap (22), it is possible that some of the transcription start sites downstream of the previously identified site (13) may be due to failure of PCR or reverse transcription during the RACE procedure. However, we are confident that the majority of the transcription start sites identified, especially those found more than once and those mapping 5’ to the previously defined beginning of the R region, are legitimate.

Most well-studied exogenous retroviruses have strong TATA-boxes and a single major transcriptional initiation site which defines the beginning of the R region, though multiple transcription start sites were found to be functional in exogenous and endogenous retroviruses (10, 35, 40). In addition, multiple transcription initiation sites in the ETnII LTR were found in the R1 ES cell line (data not shown), suggesting this may be a common feature of ETnII elements in different cell lines. It is possible that families of retroviral elements that
reside in the genome gradually assume more flexibility in transcriptional control, compared to their exogenous counterparts.

It is worth noting that, despite the multitude of copies of ETnII elements in the genome, the transcribed elements are highly similar to one another, as based on the sequences derived from the 5’-RACE clones. Most of the transcripts in P19 cells were identical to the newly retrotransposed ETnII-β element which caused a mutation in the Adcy1 gene in an ICR mouse (17) (accession number AF108230), suggesting they belong to a small subfamily of the most active and evolutionary young retroviral elements. Since the P19 cell line is derived from the C3H/HeJ mouse background, and ETn elements are highly polymorphic among mouse strains (2), it is difficult to positively identify the ETn element responsible for the majority of the transcripts based on the published C57BL/6 genome. The 5’-RACE sequences are most similar to an ETn element on chromosome 13 at the position Ch13:110665922-110671471 in the May 2005 draft of the mouse C57BL/6 genome. This is the same element identified previously as being identical to ETnII sequences highly transcribed in the early embryo from different mouse strains, LM and SELH (accession number AC079540) (3). However, using genomic PCR analysis, we have found that this particular element is polymorphic in different mouse strains (2) and is not present in P19 cells or in the parental strain C3H/HeJ (data not shown).

2.3.8. Apparent amplification of ETnII elements in the permissive P19 cell line

To determine whether the high transcription rate of ETnII and MusD elements resulted in copy number elevation in permissive cell lines, we performed genomic Southern blotting to compare ETnII and MusD copy numbers in the DNA of P19, R1 and NIH3T3
cells and their respective parental mouse strains. DNA from the NIH/Swiss parental mouse strain of the NIH3T3 cell line was not analyzed due to unavailability. The approximate expected size of the fragments, given the heterogeneous nature of ETns, comprised 1400 bp for the ETnII elements and 2100-2350 bp for the MusD elements (Fig. 2.9A). Southern blotting results (Fig. 2.9B) indicate a significant increase in ETnII copy number in the P19 cell line, likely due to high level of transcription (see Fig. 2.1) and retrotransposition.

**Figure 2.9. Amplification of ETnII elements in the P19 cell line.**

(A) Digestion scheme. 1 μg genomic DNA was digested with restriction endonucleases PstI and XbaI; XbaI cuts inside the LTR and PstI – in the body of the element, resulting in different fragment lengths for ETnII (approximately 1.4 kb) and MusD (approximately 2.1-2.35 kb) elements. The probe used is specific for both ETnII and MusD elements and is located in the common region just 3’ of the 5’ LTR. (B) Genomic southern blotting and hybridization. Genomic DNA from the parental mouse strain of the EC P19 cell line – C3H/HeJ (lane1), EC P19 cell line (lane 2), parental mouse strains of the ES cell lines, 129X1/SvJ (lane 3) and 129S1/Sv-1^{Tyr-c KIfn^{StJ}+/+} (lane 4), ES cell line R1 (lane 5), ES cell line EK,ECC (lane 6), ES cell line AB-1 (lane 7), ES cell line 671 (lane 8) and fibroblast cell line NIH3T3 (lane 9) was digested with PstI and XbaI. The blot was hybridized to the probe specific for both ETnII and MusD elements.
Though the majority of transcripts in the P19 cell line are very similar (see above section), it is unclear whether the new ETnII copies are the products of a single, highly transcribed source element, or are derived from one or multiple proviruses which originated from this element. No increase in ETnII proviruses is evident in any of the ES cell lines at this resolution. These data suggest, however, that commonly used ES cell lines may accumulate new ETn copies if grown for many generations. Methods capable of detecting new single ETn integrations are necessary to assess the frequency of such events in ES cells.

A possible reason for amplification of ETnII elements in P19 versus ES cells is an increased probability of retrotransposition due to a longer culturing period, given that P19 was established in 1982 (23) whereas ES cell lines derived from the 129 mouse strain were established in the early 1990s (8, 24, 26) and are generally maintained in culture for very short periods of time. Accumulation of ETnII copies in P19 cells could also be explained by expression of one or more specific coding-competent MusD elements in this cell line. Only a small number of MusD elements are fully intact (1) and able to facilitate ETn retrotransposition (34). The finding that ETnII elements have amplified in preference to their MusD relatives is likely due to the fact that ETnII transcripts are generally much more abundant than MusD transcripts (3), making the former more likely to retrotranspose.

Finally, it is possible that the observed ETnII amplification is due to inherent genomic instability of the teratocarcinoma P19 cell line, which may have led to DNA amplifications via segmental duplications or other rearrangements. In this case, it is expected that MusD elements would also be amplified, since both ETnII and MusD elements are distributed randomly in the genome (unpublished observations). Given that the MusD copy number has not noticeably increased in these cells, thus serving as an internal control, it is unlikely that
DNA amplifications unrelated to retrotransposition explain the increase in ETnII element number. However, the remote possibility that only a small DNA region containing ETnII element(s) and no MusD elements has selectively amplified in the P19 genome cannot be ruled out.

### 2.3.9. Role of Sp1 and Sp3 transcription factors in ETn expression

This study has identified three regions containing Sp1/Sp3 binding sites as indispensable for high promoter activity of the ETnII LTR in the ETn-expressing P19 cell line. Mutation of all three of these sites reduces LTR promoter activity 10-fold, to the level observed in the non-expressing NIH3T3 cell line.

While Sp1 is a ubiquitously expressed transcription factor, its expression may vary more than 100-fold between different tissues and developmental stages (37). It binds GC- and GT-rich boxes and is involved in the transcriptional regulation of many housekeeping, tissue-specific and developmental genes (for review, see (41)) as well as some viral genes (11, 32, 36). Sp1 binding sites have been implicated not only in protection of CpG islands from de novo methylation, but also in providing signals for direct DNA demodification (6, 9, 21). Another member of the Sp family of TFs, Sp3, that has been shown to bind the same recognition sequences, either compensates for Sp1 or competes with it. Since some of the Sp3 isoforms may function as transcriptional repressors, the ratio of Sp1 to Sp3 is frequently responsible for the degree of gene activation (for review, see (41)).

The Sp1 transcription factor is a candidate for being a major regulator of the onset of zygotic genome activation, since it is required for the hsp70 promoter activation in two-cell mouse embryos (4). Despite being ubiquitously expressed, Sp1 seems to be involved in
transcriptional activation of some embryonic-specific genes, such as Oct3/4 (9, 28, 31), telomerase (29) and FGF-4 (15, 20), as well as ST8Sia-I (43), which is expressed in P19 cells but not in 3T3 cells. This versatility of Sp1 apparently stems from the multitude of its binding partners, since Sp1 can be transactivated by a number of other transcription factors (14, 30, 36). Given that Sp1 and Sp3 levels are comparable between ETn-expressing P19 and ETn non-expressing NIH3T3 cell lines (data not shown) and transient cotransfection experiments with Sp1 and Sp3 expression vectors did not significantly alter LTR promoter efficiency in either of the cell lines (data not shown), we suggest that the high transcription rate of ETn elements and high LTR promoter activity in the P19 cell line is dependent on the availability of Sp1/Sp3-binding transcription factors. Presumably, high ETn transcriptional activity in other types of undifferentiated cells, including ES cells and embryonic tissues, relies on the same mechanism.
2.3.6. References

CHAPTER 3

ACTIVITY OF THE MOUSE ETN RETROTRANSPOSON SUBFAMILY IS ASSOCIATED WITH ESCAPE FROM EPIGENETIC SILENCING

---

1 A version of this chapter will be submitted for publication.

Maksakova, I. A., and Mager, D. L. Activity of the mouse ETn retrotransposon subfamily is associated with escape from epigenetic silencing.
3.1. Introduction

Long terminal repeat (LTR)-retrotransposon and endogenous retroviral (ERV) families typically include autonomous members as well as a heterogeneous collection of non-autonomous or defective elements with deletions or rearrangements of the coding sequence (2, 20, 23, 33, 42, 50). While most non-autonomous variants do not retrotranspose, some form distinct subfamilies as the result of having amplified to significant numbers in the genome. For example, the non-autonomous BARE-2 elements in barley have higher copy numbers and transcript levels compared with autonomous BARE-1 elements (48). For other yet unconfirmed, but highly plausible pairs of non-autonomous/autonomous LTR-retrotransposons, such as Dasheng/RIRE2 (19), Spip/RIRE3, and Squiq/RIRE8 (49) of rice, and Retand-1/Retand-2 of Silene latifolia (22), the non-autonomous members also have similar or higher copy numbers than their probable autonomous counterparts. Families of LTR elements in mammals show similar characteristics. In human, certain partially deleted elements of the HERV-H family have amplified to much greater numbers than full-length forms (18, 52). In mouse, a partially deleted subclass of the retroviral-like Intracisternal A-type Particle (IAP) family is transcribed at a high level (3, 14) and is most often found at the site of new insertions (30), suggesting a higher current rate of retrotranspositional activity than that of the fully coding-competent IAPs. The reasons for genomic amplification of specific defective subfamilies, often in preference to the coding-competent elements on which they depend for retrotransposition, are largely unknown.

As a model system to study this phenomenon, we have chosen the ETn/MusD retrotransposon family in the mouse. ETns lack all retroviral genes and instead carry a non-retroviral and non-coding sequence of unknown origin (see Figure 3.2A) but are
transcriptionally active and owe their retrotranspositional potential to the proteins encoded by their structurally intact cousins, the MusD elements (39). The level of ETnII transcripts in embryos and undifferentiated cells, such as embryonic stem (ES) and embryonic carcinoma (EC) cells, is much higher than the level of MusD transcripts, despite the lower copy number of ETnII elements (4). This chapter addresses whether ETn and MusD elements harbour different epigenetic marks affecting their expression.

3.2. Materials and methods

3.2.1. Cell culture

Mouse embryonic teratocarcinoma cell line P19 and mouse embryonic fibroblast cell line NIH3T3 (ATCC) were maintained in Dulbecco's modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum for P19 or 10% calf serum for NIH3T3, penicillin and streptomycin.

3.2.2. Plasmids and constructs

For luciferase assays, ETnII#7 (AC079540, element position: 130750-136290) and MusD#3 (AC084696, element position: 139824-132348) internal sequences were cloned into the BamHI-SalI restriction site opposite to the ETnII#7 LTR-containing KpnI-BglII cloning site of the pGL3-Basic (pGL3B) luciferase reporter vector (Promega). LTR cloning was described previously (29). Internal regions were cloned from C57/BL6 genomic DNA in standard PCR conditions using the primers below. ETnII, 1st round: 540_1s and ETnIIgr1_3610as; 2nd round: IM_ETn_637s-BamHI and IM_ETn_3593as-SalI for the forward-oriented insert, IM_INT_637s-SalI and IM_ETn_3593as-BamHI for the reverse-
oriented insert. MusD, 1st round: 696_1s and MusD_4094as; 2nd round: IM_MusD_636s-SalI and IM_MusD_3992as-BamHI for the reverse-oriented insert, IM_MusD_636s-SalI and IM_MusD_3992as-SalI, with subsequent selection of correctly oriented clones. Inserts were confirmed by sequencing. See primer sequences in the Table 3.1.

**Table 3.1. List of primers.**

<table>
<thead>
<tr>
<th>Cloning:</th>
<th>Probe synthesis:</th>
</tr>
</thead>
<tbody>
<tr>
<td>540_1s</td>
<td>ETnII/MusD_465-s</td>
</tr>
<tr>
<td>696_1s</td>
<td>GCTGAAGGACCTGAGTCTGAGGAGTGCATGGATCAGGCGG</td>
</tr>
<tr>
<td>ETnIlg1_3610as</td>
<td>ETnII/MusD_652-as</td>
</tr>
<tr>
<td>ETn_637s-BamHI</td>
<td>ACCGCTGATCCATCATGACTGAC</td>
</tr>
<tr>
<td>ETn_3593as-BamHI</td>
<td>mGAPDH_139s</td>
</tr>
<tr>
<td>ETn_3593as-SalI</td>
<td>ACGACCCCTTCATGTGAC</td>
</tr>
<tr>
<td>INT_637s-SalI</td>
<td>mGAPDH_331as</td>
</tr>
<tr>
<td>INT_2045s-SalI</td>
<td>CTCAGCAATATCATAAAATAAGGTTC</td>
</tr>
<tr>
<td>MusD_636s-SalI</td>
<td>ChIP</td>
</tr>
<tr>
<td>MusD_3992as-BamHI</td>
<td>ETnII_662-as</td>
</tr>
<tr>
<td>MusD_3992as-SalI</td>
<td>ACTGGGGCAATCGGCTATTAC</td>
</tr>
<tr>
<td>MusD_4094as</td>
<td>ETn-MusD_514-s</td>
</tr>
<tr>
<td></td>
<td>GCAACCGGCTGATCCATGAC</td>
</tr>
</tbody>
</table>

**3.2.3. In vitro patch-methylation**

Constructs with ETnII and MusD LTRs cloned into the KpnI-BglII restriction site of pGL3B luciferase reporter vector were described previously (29). Two ETnII LTRs, LTR#7 and LTR#13 (AC079497, element position: 103995-96963) and two MusD LTRs, LTR#3 and LTR#9 (AF132039, element position: 79690-87160), were used. ETnII and MusD LTR-containing pGL3B vectors were methylated in vitro with SssI methyltransferase according to manufacturer’s recommendations; mock-methylation was performed similarly with the exclusion of SssI. Complete CpG methylation, or its absence, was confirmed by digestion.
with a methylation-sensitive enzyme ApaLI which has a restriction site in all LTRs. Methylated and mock-methylated LTRs were released from the respective plasmids by the double KpnI/BglII digestion and gel-isolated using a MinElute™ gel extraction kit (Qiagen). In parallel, an unmethylated pGL3B vector was double-digested with KpnI and BglII, purified with a QIAquick PCR-purification kit (Qiagen), treated with calf intestinal phosphatase (NEB) and again purified with a PCR-purification kit (Qiagen). Each methylated and mock-methylated LTR was ligated into the unmethylated vector backbone in 1:1 molar ratio with high concentration T4 DNA ligase (NEB) for 2 days at 16°C. As a promoterless control, pGL3B digested with a single KpnI enzyme, purified, then re-ligated, was used.

3.2.4. Transient transfections and luciferase assays

With large DNA fragments introduced into plasmid DNA, we saw evidence of reduced plasmid DNA uptake via lipofection. Such effects have been described previously and shown to account for more than 10-fold reduction in luciferase expression, inversely correlating with insert size (53). We cloned a 3 kb and a 3.7 kb lambda phage “stuffer” DNA fragments into the SalI restriction site of the promoterless pGL3B vector to serve as a promoterless size-adjusted control, and normalized pGL3B-LTR#7 luciferase activity to non-modified pGL3B, since the LTR is only 317 bp long, while the luciferase activity of the 3.15kb ETnII internal region-containing constructs was normalized to the 3 kb lambda phage-containing pGL3B, and that of the 3.68 kb MusD internal region-containing constructs, to the 3.7 kb lambda phage-containing pGL3B.

For luciferase assays, P19 cells were seeded into 24-well plates at a density of $3 \times 10^4$ cells per well 24 h prior to transfection. Cells grown in monolayer were transfected with 500
ng of test plasmid DNA and 50 ng of the Renilla luciferase vector pRL-TK using 0.5 μl of Lipofectamine-2000 (Invitrogen) per well for the analysis of the internal ETnII and MusD regions, or 1 μl of Lipofectamine-2000 per well for the analysis of the methylated LTR activity. Cells were washed 24 h after transfection in phosphate-buffered saline and harvested in 100 μl 1× passive lysis buffer (Promega). Firefly and Renilla luciferase activities were measured using the dual-luciferase reporter assay system (Promega).

The data were first standardized to the internal Renilla luciferase control and expressed relative to the activity of the promoterless cut and re-ligated pGL3B vector for the methylated LTR analysis, or the pGL3B vector with a λ-phage “stuffer” DNA of the corresponding size for the ETnII and MusD internal region analysis. The results are a mean and standard deviations of three separate experiments performed in duplicates.

3.2.5. Preparation of RNA and Northern blotting

DNA methyltransferase knockout ES cell lines were generously provided by Dr. En Li. Total RNA was isolated using RNeasy mini kit (Qiagen) according to the manufacturer’s protocol. Northern blotting and hybridizations were performed as described previously (29). A probe common for both ETnII and MusD was synthesized using primers ETnII/MusD_465-s and ETnII/MusD_652-as. A GAPDH probe was synthesized using primers mGAPDH-139s and mGAPDH_331as.

3.2.6. Preparation of genomic DNA and Southern blotting

DNMT KO ES cell line genomic DNA was extracted using DNAzol™ (Invitrogen) according to the protocol. DNA was dissolved in 200 μl 8mM NaOH per 1ml DNAzol™ and
analyzed for purity based on the ratio of the absorbance at 260/280 nm wavelength. Southern blotting and hybridizations were performed as described previously (29), with the probe specific for both ETnII and MusD elements synthesized using primers ETnII/MusD_465-s and ETnII/MusD_652-as.

3.2.7. Chromatin immunoprecipitation (ChIP)

ChIPs were performed with a ChIP assay kit (Upstate Biotechnology) according to the manufacturer’s instructions. The following antibodies were used for immunoprecipitations: polyclonal anti-acetyl-histone H3 (Lys9) (#07-352), polyclonal anti-acetyl-histone H4 (Lys5, -8, -12, and -16) (#06-866) and polyclonal anti-trimethyl-histone H3 (Lys9) (#07-442) from Upstate Biotechnology. The DNA was purified using QIAGen PCR purification kit (Qiagen) and eluted in 50 µl of elution buffer.

3.2.8. Quantitative real-time PCR

A primer in the common region, ETn-MusD_514-s, together with either an ETnII-specific primer, ETnII_662-as, or a MusD-specific primer, MusD_657-as, was used in quantitative real-time PCR to amplify a 149 or a 143-bp common region of the population of ETnII or MusD elements in the genome, respectively. As a positive control, the housekeeping gene hypoxanthine phosphoribosyltransferase 1 (Hprt1) was amplified; as a negative control, a CpG-rich region of mouse NKG2A was amplified (see primer sequences in (41)). Forty-five rounds of amplification with SYBR Green PCR Master Mix (Applied Biosystems) were performed. The default 7500 System SDS software version 1.2.10 (7500 RealTime PCR System; Applied Biosystems) cycle was used with the exception that
amplification was performed at 65°C in a total volume of 25 µl. Dissociation curve analysis was performed after the end of the PCR to confirm the presence of a single and specific product.

3.3. Results

3.3.1. Differences in LTR sequence do not explain higher transcript levels of ETnII elements

As discussed above, there is an approximately 30-fold higher level of ETnII compared with MusD transcripts in mouse embryos and embryonic stem (ES) cells (4, 29) (see Figure 3.2A for ETnII and MusD structure). Sequencing of cDNAs indicated that several, though closely related, ETnII elements are expressed (Chapter 2.3.7 and (4, 29)), suggesting that the higher level of ETn transcripts is not due to a single, highly transcribed element. We previously studied ETnII and MusD LTR promoter activities using transient transfection assays in embryonic carcinoma (EC) cells and found that the average promoter activity of the four ETnII LTRs tested was 3-fold higher than the average activity of the four MusD LTRs tested, suggesting that slight LTR differences may contribute to the difference in transcript levels (Chapter 2.3.2 and (29)). However, because only a few, randomly chosen LTRs of each type were tested, the results may not reflect the average promoter activity of the whole ETnII or MusD populations. In our previous study, we identified three Sp1/Sp3 binding sites in the first 76 bp of the LTR and demonstrated that this region of the LTR is most important in conferring high promoter activity in EC cells (Chapter 2.3.5 and (29)). Indicative of the synergistic effect of the sites, mutation of all three of them resulted in a drop of the LTR promoter activity to 9% of the original level.
To determine if differences in this region may contribute to higher ETnII transcription, we aligned 5′ LTRs from the 69 MusD and 32 ETnII elements identified in the assembled C57BL/6 genome based on the beginning of the 5′ LTR and the region specific for either ETnII or MusD elements (Figure 3.1).

Figure 3.1. (Next page) Alignment of the first 93 bp of the ETnII and MusD LTRs. The alignment of the first 93 bp of 32 ETnII and 69 MusD LTRs is shown. ETnII and MusD element sequences were recovered from the mm8 version of the C57BL/6 genome based on the 5′ end of the LTR and a region different between ETnII and MusD. The core regions of three Sp1/Sp3 binding sites important for the LTR promoter activity are boxed, and the beginning of the deleted construct 76-317 with abolished promoter activity (Chapter 2.3.3) is shown with an arrow. LTRs with sequence identical to the consensus in the 93 bp region are indicated with triangles on the right-hand side: blue for ETnII and green for MusD. Colour code: adenine (a)-red; cytosine (c)-green; guanine (g)-yellow; thymine (t)-blue.
As summarized in Figure 3.2B, we detected a number of differences in the core regions of Sp1/Sp3 binding sites important for the high LTR promoter activity. However, the overall number of MusD elements with a consensus sequence in the first examined 93 bp of the LTR is higher than that of ETnII elements. Therefore, it is unlikely that LTR sequence differences are the primary factor that results in higher ETnII transcript levels.

![Figure 3.2](image)

**Figure 3.2.** (A) Structure of ETnII and MusD retrotransposon subfamilies. (B) Conservation of Sp1/Sp3 sites critical for LTR activity. The presence of the intact core of each of the three Sp1/Sp3 binding sites in the ETnII and MusD elements shown to be important for the LTR activity (Chapter 2.3.5 and (29)) is indicated for all of the ETnII and MusD 5' LTRs found in the C57/BL6 genome. The number of elements displaying the consensus sequence in the first 93 bp of the LTR is shown.

### 3.3.2. Neither ETnII nor MusD internal regions possess strong *cis*-acting regulatory signals

To analyze the internal ETnII and MusD sequences for possible enhancer or suppressor elements, we tested the ability of these regions to affect LTR promoter activity using transient transfection assays. We examined the activity of the ETnII LTR#7 driving the
expression of the Firefly luciferase reporter gene in the pGL3B vector. Reporter constructs included either a 3,148 bp ETnII or a 3,575 bp MusD internal sequence not directly linked to the promoter (Fig. 3.3A). The MusD chosen for this study is a “young” coding-competent element (39) and the ETnII element chosen belongs to the most abundant and transcriptionally active ETnII-β subtype (4).

As shown in Figure 3.3B, neither ETnII non-retroviral nor MusD retroviral internal sequences possess strong enhancer or suppressor properties when assessed in both orientations; this data was confirmed by deletion analysis of the ETnII internal region (data not shown). Therefore, the distinct internal regions of ETnII and MusD elements do not appear to harbor cis-acting motifs that significantly affect LTR activity.

**Figure 3.3.** Effect of ETn and MusD internal regions on LTR promoter activity.

(A) Cloning of ETnII and MusD internal regions. IntETn and intMusD – regions cloned into the reporter constructs in forward (Fw) and reverse (Rv) orientations. (B) Relative luciferase activity of the ETnII and MusD internal region-containing LTR-promoted constructs.

3.3.3. Demethylation of MusD LTR and common region in the DNA methyltransferase knockout (KO) ES cell lines

Because large sections of the internal regions of ETnII and MusD elements are completely unrelated (4), it is possible that these elements may be recognized differently by
factors responsible for epigenetic modifications. To examine this possibility, we performed genomic Southern blotting using the strategy described below. Several methylation-sensitive enzymes that cut DNA only if it is unmethylated were chosen (Fig. 3.4A). The recognition sites of those enzymes in the LTR and the 5′ internal common region are present in the majority of ETnII and MusD elements, due to the relatively young age of this family, allowing us to assay a large population of these ERVs in the genome. The second, methylation non-sensitive enzyme that cuts at a different distance in ETnII and MusD elements, permits the analysis of the relative frequency of unmethylated sites in both ETnII and MusD elements in one lane using a single probe that recognizes both subfamilies. A methylation non-sensitive enzyme XbaI, cutting in the LTR, was used as a control for overall ETnII and MusD copy numbers.

To examine the methylation of ETnII and MusD in permissive cells, we analyzed the methylation status of LTRs and common regions of ETnII and MusD elements in wild-type and DNA methyltransferase (Dnmt) deficient ES cell lines (24, 36). As shown in Figure 3.4B, the 3′-most sites in the LTR and the common region of MusD elements are heavily methylated in the wild-type, Dnmt3a−/− and Dnmt3b−/− cell lines. However, the Dnmt3a−/−3b−/− ES line and especially the Dnmt1−/− line, displayed a near-complete demethylation of MusD elements. ETnII elements, already in a highly unmethylated state, were not noticeably affected (Fig. 3.4B).

Demethylation in Dnmt1−/− ES cells and embryos has been shown previously for other retrotransposon families, such as IAP, MaLR and LINE (16, 25, 51). The unique setup of this Southern blotting system allows for the analysis of methylation levels of the whole population of ETnII and MusD elements in the genome at several CpG dinucleotides within
sites recognized by methylation-sensitive restriction enzymes. Because the examined regions are nearly identical in ETnII and MusD elements, we suggest that the marked difference in levels of CpG methylation between these two subfamilies results from the influence of the distinctly different adjacent regions – the ETnII non-retroviral and the MusD retroviral sequences.

**A.**

![DNA Methylation of ETnII and MusD promoter regions in Dnmt KO cell lines.](image)

(A) Schematic representation of the digestion for genomic Southern blotting. Expected fragment sizes are indicated; sensitivity of the enzymes used to CpG methylation is given.

(B) Genomic Southern blotting of the digests. Lane number corresponds to the enzyme number in Fig. 3.4A; DNA from *Dnmt* w/t (J1), *Dnmt3a*−/− (3a−/−), *Dnmt3b*−/− (3b−/−), *Dnmt3a*−/−*b*−/− (3a−/−3b−/−), and *Dnmt1*−/− (1−/−), is used. Bands indicating expected sizes of unmethylated MusD elements are marked by asterisks.
3.3.4. Elevated transcript levels of MusD and ETnII elements in the Dnmt KO ES cell line

To determine if the extensive demethylation of MusD retrotransposons observed in the Dnmt1 KO ES cell line (Fig. 3.4) causes elevation in their transcript levels, we analyzed MusD and ETnII element expression by Northern blotting using RNA isolated from the J1 wild-type, the double Dnmt3a−/−3b−/−, and the Dnmt1−/− ES cells. Similarly to other retrotransposons, such as IAP (10, 51), MALR (25) or LINE1 (47), MusD expression is dramatically elevated in the Dnmt1−/− ES cells (Fig. 3.5), supporting the hypothesis that CpG methylation plays a major role in MusD suppression.

![Figure 3.5. ETnII and MusD expression in Dnmt KO cell lines.](image)

Northern blotting of Dnmt w/t (J1), Dnmt3a−/−3b−/− (3a−/−3b−/−), and Dnmt1−/−(1−/−) RNA. Hybridization was performed with a probe located in the common region which recognizes both ETnII and MusD. Ethidium bromide-stained gel is shown as a loading control.

We also detected a substantial increase in ETnII transcription in the Dnmt1−/− ES cells, despite little detectable effect of Dnmt1 KO on ETnII DNA methylation seen in Figure 3.4B. This finding suggests either induction of ETnII transcription from a very limited number of previously silenced ETnII copies, their demethylation undetected due to Southern blot limitations, or the ability of Dnmt1 to perform a repressor role independent of its DNA methyltransferase activity. The latter can be achieved by Dnmt1 recruiting histone
deacetylases and methyltransferases or transcriptional repressors to the LTR, exemplified by LSH-mediated transcriptional silencing dependent on binding the catalytically inactive DNMT1 (35).

3.3.5. More activating and less suppressive histone marks on ETnII compared with MusD elements.

Since dense CpG methylation has been shown to induce histone hypoacetylation (12, 28, 44), we analyzed the state of histone acetylation and methylation in the common region of ETnII and MusD retrotransposons. ES cell line R1 which expresses ETn/MusD elements and fibroblast cell line NIH3T3 which lacks ETn/MusD expression (see Fig. 2.1) were subjected to chromatin immunoprecipitation (ChIP) analysis with antibodies against histone 3 (H3) acetylated at Lys9 (K9) and histone 4 (H4) acetylated at multiple residues, both modifications hallmarks of active chromatin; and an antibody against H3 trimethylated at Lys9 – a hallmark of silent chromatin (reviewed in (5)). Precipitation reactions with no antibody and the input fraction before immunoprecipitation were used as negative and positive controls, respectively. After immunoprecipitation and reversal of the cross-linking, enrichment for ETnII and MusD in each sample was monitored by quantitative real-time PCR. Due to the technical restrictions on amplicon length, we were limited to analyzing the ~150 bp at the 3′ end of the ~300 bp ETnII/MusD common region using a common forward primer and an ETnII- or a MusD-specific reverse primer just outside the common region (Fig. 3.6A). A promoter region of the house-keeping Hprt gene was amplified as an additional positive control, and a GC-rich region of the NK-cell specific gene NKG2A, deacetylated in non-expressing cells (41), was amplified as an additional negative control.
In line with our expectations, ETnII elements, hypomethylated in the common region in ES cells (see Fig. 3.4B), displayed higher levels of histone acetylation and lower levels of H3K9 trimethylation, while MusD elements, methylated to a higher extent, carried less active and more suppressive histone marks in the permissive R1 ES cell line (Fig. 3.6B).

![Position of ETnII and MusD ChIP primers]

**Figure 3.6. ChIP in ETnII and MusD common region.** Amplification of NKG2A, ETnII, MusD and Hprt promoter region fragments from DNA material immunoprecipitated with the following antibodies: acetylated H3K9; acetylated H4; trimethylated H3K9. ETn/MusD-expressing R1 and non-expressing NIH3T3 cell lines were used. The results are presented as a relative amount of acetylated or methylated histones associated with the target samples with respect to the input material. Quantitative PCRs on DNA from three separate immunoprecipitations were performed in duplicates; a representative experiment is shown. Error bars are standard deviations between PCR replicates.

In the three different immunoprecipitations, the association of ETnII with H3K9-acetylated histones was 2.9 to 6-fold higher compared with MusD, the difference ranging between 2.2 to 3.2-fold for H4-acetylated histones. Conversely, MusD had a 1.5 to 1.7-fold
higher association with H3K9me3 histones in the three immunoprecipitations. Interestingly, no such difference was observed in the non-expressing NIH3T3 cell line, suggesting targeted epigenetic regulation in the cells capable of ETn/MusD expression. Similar reduction in both active and repression chromatin marks in differentiated compared with ES cells was reported for IAP retrotransposons (31, 34).

Since hyperacetylated nucleosomes are associated with a higher rate of transcription (38), it is possible that the lower MusD compared with ETnII expression is the result of lower MusD acetylation level imposed by high CpG methylation. In a genome-wide chromatin screen, ETn and IAP sequences were found to be enriched for H3K9me3 and H4K20me3 in ES cells (34). Since REPBASE (http://www.girinst.org) used by the authors to identify the classes of repeats does not distinguish between ETn and MusD, designating both sub-families as ETn, it is probable that most of the elements picked up by the screen were, in fact, the more numerous MusD retrotransposons. These results show that, despite nearly identical sequence of the LTRs and the common region, ETnII and MusD elements are treated differently by genomic epigenetic mechanisms.

3.3.6. ETnII and MusD LTRs are silenced by CpG methylation.

To determine if ETnII and MusD LTR promoter activity is indeed suppressed by CpG methylation, we performed transient transfection assays with in situ methylated LTR reporter constructs. Only the LTR sequence was methylated according to the patch-methylation protocol (Fig. 3.7A) to avoid gene silencing due to the reporter gene methylation. The methyltransferase SssI, which adds a methyl group to every cytosine in a CpG context, was
used. ETnII and MusD LTRs that have exhibited the highest and the lowest promoter activities in the previous studies (29) were chosen for patch-methylation in vitro.

Figure 3.7. Promoter activity of in vitro patch-methylated LTRs.

(A) In vitro patch-methylation scheme. Mock-methylated constructs were produced by omitting SsoI methyltransferase in the first step. (B) Relative luciferase activity of methylated and mock-methylated LTR-promoted constructs. Data is a mean and standard deviation of three experiments performed in duplicates.

Two ETnII LTRs, LTR#7 (AC079540) and LTR#13 (AC079497), and two MusD LTRs, LTR#3 (AC084696) and LTR#9 (AF132039) were subjected to in vitro methylation and promoter assays. The results in Figure 3.8B show that in vitro methylation causes a significant (3.1-5.6-fold) reduction in the promoter activity of all ETnII and MusD LTRs in P19 cells, suggesting that cytosine methylation of these LTRs in vivo will suppress their promoter activity.

### 3.3.7. ETnII internal region has a lower density of CpG dinucleotides compared with MusD

To determine if ETnII and MusD elements could be differently recognized by the repressive epigenetic machinery due to dissimilarity of their sequence composition, we
analyzed two representative elements from the genome using MethPrimer (http://www.urogene.org) (Fig. 3.8).

**Figure 3.8. CpG composition of ETnII and MusD retrotransposons.**
The Y axis represents overall GC content. CpG dinucleotides, depicted by vertical bars, were mapped in one representative ETnII element #7 (AC079540) and a representative MusD element #3 (AC084696). CpG islands, predicted by a default setting of the MethPrimer program, are shown as grey areas.

The ETnII non-retroviral region has 0.9 CpGs per 100 base pairs, compared with 2.9 CpGs per 100 base pairs of MusD retroviral region. Moreover, the same MusD sequence has an overall GC content of approximately 55%, whereas an ETnII non-retroviral region has a GC content of around 35%. This difference suggests that the abundance of CpG dinucleotides in the MusD retroviral region make it more susceptible to DNA methylation, which spreads into the LTR promoter, suppressing transcription. This idea is supported by the fact that, in MusD elements, the 5’-most methylation-sensitive enzyme restriction sites, such as ApaLI, have lower levels of methylation than the internal sites closer to the CpG-rich retroviral region, suggesting methylation spreading. The long stretch of CpG-rich DNA characteristic of MusD elements may be more readily recognized by the cellular epigenetic
machinery. Consequently, methylation or repressive chromatin may spread from the retroviral region into the LTR promoter of MusD.

3.4. Discussion

Eukaryotic genomes have evolved multiple lines of defense against active retrotransposons and retroviruses, including restriction at the level of virus entry and uncoating, reverse transcription and trafficking of the viral RNA, transcription of the integrated provirus and processing of its transcript (see (15) for review). DNA methylation has long been regarded as a major means of transcriptional suppression. It is abundant on repetitive sequences, including tandem repeats, such as pericentromeric major and minor satellites, and interspersed repeats, which constitute DNA transposons, LTR-retrotransposons or ERVs, Long interspersed nuclear elements (LINEs), and Short interspersed nuclear elements (SINEs). In fact, it has been suggested that most of the genome’s methylated CpGs lie within transposable elements (54). Suppressed in most normal tissues, various groups of transposable elements are known to be re-activated in the absence of DNA methyltransferases (6, 25, 36, 51), supporting the notion that DNA methylation is indeed a main defense mechanism against transposable elements in general and endogenous retroviruses in particular (54).

Suppression at the level of transcription, directed by DNA methylation, histone modifications, and RNA interference, has resulted in ERV silencing in most differentiated tissues of animals and plants. Though some ERVs and LTR retrotransposons are expressed highly in sperm (11), oocytes (37) and early embryos (26, 37, 40, 46), their transcript levels
in those cells and tissues are presumably much lower than if all the hundreds or thousands of copies of multiple families were free of all transcriptional repression.

We propose that non-autonomous ETnII elements owe their high expression and retrotransposition activity to the loss of the CpG-rich retroviral region present in MusD. DNA methylation or repressive chromatin structure initiating from the retroviral region may induce silencing of MusD elements, resulting in the observed lower transcription level of MusD compared with ETnII elements in early embryos and permissive cell lines (4). A similar situation was reported for the \textit{rho}-globin gene in chicken embryonic erythroid cells, when during normal developmental silencing, \textit{de novo} methylation originates in the CpG-dense transcribed region and then spreads into the promoter, silencing transcription and resulting in complete DNA methylation in the adult erythroid cells (45). Not surprisingly, a higher methylation level of MusD is accompanied by a higher level of suppressive and lower level of activating histone marks. Indeed, it is entirely possible that CpG methylation of MusD sequences is not the primary suppressive mark, since many viral vectors are silenced via chromatin remodeling before the onset of DNA methylation (reviewed in (13)).

Transcriptional suppression of MusD retrotransposons may occur via several mechanisms. Firstly, the elongation efficiency of the transcript may be directly influenced by CpG methylation downstream of the promoter in the MusD retroviral region (27). In addition, histone hypoacetylation associated with dense CpG methylation (12, 28, 44) may cause reduction in MusD transcription rate (38). It has also been shown that downstream methylation, even in the context of an unmethylated LTR, can virtually abolish transcription (17). Alternatively, the downstream DNA methylation and hyperacetylated chromatin may spread into the promoter region, inhibiting transcription factor binding (32) or Polymerase II
loading (44). It also remains to be determined whether piRNAs, implicated in suppression of transposable elements in *Drosophila* (7) and mice (1, 8) may play a role in inducing DNA methylation on MusD retrotransposon sequences or reducing transcript levels via RNA degradation.

A common feature of the evolution of retrotransposon families involves the emergence of certain non-autonomous retrotransposon variants that gain advantage over their coding-competent counterparts by different mechanisms. These mechanisms may include, but are not limited to, acquisition of stronger enhancer and promoter elements, reduced binding of a repressive histone variant macroH2A, as reported for a deleted variant of MLV ERVs (9), production of a fusion protein facilitating retrotransposition as recently found of IΔ1 IAP elements (43) or evasion of epigenetic suppression, as proposed here. The MusD/ETn family illustrates an intriguing way for a new retrotransposon variant to gain an advantage by increasing its likelihood of transcription and, consequentially, retrotransposition. This route involves the evolution of deleted or mutated elements, which retain structural features necessary for retrotransposition, assisted *in trans* by proteins from coding-competent members of the same family, but which have lost at least some retroviral sequence to enable escape from host silencing mechanisms (Fig. 3.9).

In the arms race between parasitic transposable elements and their host (30), the newly emerged ETnII element, devoid of most internal retroviral sequences that are *foci* for repressive epigenetic marks, has apparently found a gap in the defence systems of the mouse genome. The exploitation of this gap by ETnII elements is currently ongoing, resulting in multiple novel germ-line mutations (30). This element will likely continue to amplify in the genome until the disappearance of the last source of MusD proteins required for
retrotransposition; or until the epigenetic and/or genetic machinery of the host evolves mechanisms to suppress it.

**Figure 3.9. A model for amplification of deleted LTR-retrotransposon variants.**

(1) The transcription of a coding-competent LTR-retrotransposon is partially suppressed by epigenetic machinery. (2) The viral mRNA is translated to produce matrix, capsid and nucleocapsid, reverse transcriptase and integrase proteins. (3) Virion RNA is packaged with the viral proteins into a retroviral particle where reverse transcription takes place. (4) The fully-functional ERV integrates into a new location in the genome. (5) Mutations and deletions are frequently acquired during reverse transcription; most render the newly integrated retroelement non-functional. (6) Some deletions or mutations offer a newly integrated provirus an advantage with respect to evasion of epigenetic (or other) host restriction mechanisms. Though non-coding, these proviruses are transcribed at a much higher level compared with their coding-competent predecessor. (7) The deleted/mutated virion RNA is packaged with the viral proteins produced by the full-length retrotransposons and reverse transcribed. (8) Due to the higher transcription rate of the deleted/mutated variant, a greater number of virion RNAs is reverse transcribed and integrated into the new genomic locations, resulting in amplification of the deleted/mutated variant.
3.5. References


CHAPTER 4

DISCUSSION AND CONCLUDING REMARKS
4.1. ETnII evasion of host-imposed silencing

While no current activity and no new mutations due to ERV insertions have been found in human, understanding how ERVs are regulated in the mouse genome may help identify host mechanisms of ERV suppression that have developed during evolution. Eukaryotic genomes have acquired multiple defence mechanisms against active retrotransposons and retroviruses (12). The major means of transcriptional suppression, DNA methylation and histone modifications, have lead to successful silencing of ERVs in most differentiated tissues of animals and plants, though the means by which some ERV variants escape this suppression are poorly understood and require further examination.

In this thesis, my aim was to shed light on mechanisms allowing select ERV subfamilies to circumvent host cell control and successfully amplify in the genome. The ETn/MusD family of retrotransposons is a perfect system for this study, since the coding-competent MusD retrotransposons are suppressed to a significant degree in permissive undifferentiated cells, whereas the evolutionary younger non-autonomous ETn retrotransposons, utilising retroviral proteins produced by MusD elements for their retrotransposition, avoid transcriptional suppression.

In Chapter 2 of this thesis, I analysed whether subtle differences between ETnII and MusD LTRs may account for the dramatic difference in expression levels of these subfamilies. Though promoter activities of select ETnII LTRs were slightly higher than those of MusD LTRs, I concluded that LTR sequence differences, while unquestionably contributing to some of the expression difference, are unlikely to fully account for the 30-fold expression difference between ETnII and MusD elements. Since only four LTRs belonging to each subfamily were analysed, the data may not be representative of the broad
ETnII and MusD populations in the mouse genome. However, Chapter 3 addresses this question. Additional inspection of ETnII and MusD LTRs showed that overall number of MusD LTRs retaining regions critical for LTR promoter activity is greater, thus confirming that the major reason for the higher ETnII compared with MusD expression lies outside the LTR, in the internal sequence of these elements.

Continuing to investigate the problem, I hypothesized that internal regions, substantially different between the two subfamilies, may provide enhancer or suppressor functions, modulating ETnII or MusD promoter activity in cis. However, as described in Chapter 3, no significant cis-acting elements were found in the internal region sequence, suggesting that they may affect ETnII and MusD expression in a completely different fashion, namely, by means of recruiting different epigenetic modifications.

Indeed, as data presented in Chapter 3 demonstrates, regions highly similar in sequence between ETnII and MusD elements are methylated to a higher degree and are associated with less activating and more repressive histone modifications in MusD retrotransposons. According to my hypothesis, this may be the result of DNA methylation and/or repressive chromatin spreading from the internal CpG-rich region of MusD elements, while the younger ETnII subfamily, devoid of CpG-rich sequence, escapes this suppression. This hypothesis is supported by the fact that CpG dinucleotides are more highly methylated in the region closer to the MusD retroviral gag sequence compared with the 5'-end of the LTR.

The data presented in this thesis sheds light on a novel way of evasion of epigenetic silencing by a subpopulation of ERVs, namely, ETns, via their loss of the CpG-rich internal region targeting them for suppression. The optimal way to demonstrate selective targeting of
MusD elements by epigenetic mechanisms would have been to compare the effect of the internal ETnII and MusD sequences on the LTR promoter in stably transfected ES or EC cells. Such attempts were made. However, transfection of EC cells with a circular reporter plasmid containing either an ETnII or a MusD internal sequence adjacent to an LTR resulted in tandem integrations of the constructs, complicating the analysis of the results. Another strategy was to perform site-directed construct integration by means of recombinase-mediated cassette exchange (RMCE) (9). Unfortunately, an ES cell line with a well-characterised RMCE locus was not available, and attempts to demonstrate preferential \textit{de novo} methylation of the MusD retroviral region versus the ETnII internal region in RMCE-adapted mouse erythroleukemia (MEL) cells failed, likely due to the extremely low \textit{de novo} methylation ability of differentiated cell lines (22, 31). However, preliminary results showed evidence of differential chromatin modification in the promoter region of integrated ETnII and MusD constructs (data not shown). Subsequent experiments should give a clearer indication of whether a retroviral MusD sequence may recruit repressive chromatin modifications to the promoter and whether an ETnII internal sequence may escape such chromatin-imposed silencing. In addition, a MEL cell line may be generated to express \textit{de novo} methyltransferases to assess whether MusD-containing integrated constructs may be preferentially methylated in comparison to ETnII-containing constructs.

Further research could focus on demonstration of selective \textit{de novo} methylation of MusD retroviral regions in bulk cultures of retrovirally transduced ES cells. While retroviral transduction results in random integration of reporter constructs, multiple integrations should yield an average of promoter activity of LTRs adjacent to either an ETnII or a MusD internal sequence regardless of the integration site. This approach could demonstrate selective \textit{de}
novo methylation of MusD elements and provide confirmation that evasion of suppression by ETnII elements is associated with their avoidance of epigenetic silencing. Another issue that may need further confirmation is spreading of DNA methylation and silenced chromatin from internal MusD and ETnII sequences into an LTR promoter. This could be addressed separately from the experiment described above by monitoring DNA methylation and histone modification state of the common region or the LTR in vectors containing pre-methylated ETnII or MusD internal regions adjacent to LTRs virally introduced into ES cells. The obvious problem with this approach is inability to control the site of integration which may affect the behaviour of the provirus with respect to its methylation levels and thus influence its activity. Another concern is that the promoter may appear to be resistant to methylation due to active transcription and abundance of bound transcription factors.

An approach that may be undertaken to directly address the issue of impact of ETnII or MusD internal sequences on their promoter activity involves episomal replicating plasmids. Introducing ETnII or MusD-containing constructs into such a vector would reduce the ambiguity of random insertions observed with lenti- or retroviral transductions, subject to repressive effects of the genomic context of their integration site. However, the most frequently used Epstein-Barr virus-based episomal vectors do not replicate in mouse cell lines (45), and matrix attachment region-containing episomes require the S/MAR region to be transcribed (11) thus selecting for unmethylated episomes. Since human ES cells are difficult to sustain in undifferentiated state necessary for their de novo methylation as well as ETn promoter activity, I would have to resort to differentiated human cells supplemented with de novo methyltransferases, such as Dnmt3a, to mimic the processes occurring in cells with Dnmt capacity, as has been done previously (14).
Recent research suggests involvement of ERV-derived piRNAs in establishment of DNA methylation patterns in the male germline which determine ERV silencing (18). It would be interesting to examine small RNA sequences isolated from male fetal germ cells from the critical stage immediately preceding *de novo* methylation in the germline and map them to ETn or MusD retrotransposons, thus uncovering whether prevalence of MusD-specific piRNAs may play a role in their selective silencing later in development.

**4.2. Undifferentiated cell-specific ETn LTR promoter activity**

As I have shown in Chapter 2 of this thesis, ETn and MusD LTRs have promoter activity tailored specifically for expression in undifferentiated cells. This is mediated, at least in part, by GC-boxes corresponding to Sp1/Sp3 binding sites, and while the proteins themselves are ubiquitously expressed, they may function through binding of undifferentiated cell-specific factors that transactivate the promoter. Indeed, Sp1/Sp3 binding sites have been implicated in zygotic genome activation (2) and transcriptional activation of some embryonic-specific genes, such as Oct3/4 (10, 29, 33), telomerase (30) and FGF-4 (19, 23). While Sp1 binding is not affected by DNA methylation (5, 13) according to some studies, others report on inhibition of Sp1/Sp3 binding by DNA methylation (3, 7, 20). A tight relationship between DNA methylation and Sp1 binding appears to be responsible for tissue-specific expression of different genes and, possibly, ETn/MusD ERVs. Alternatively, GC-boxes may function as binding sites for pluripotent-specific activating transcription factors other than Sp1/Sp3. Further research may focus on identifying transcription factors responsible for undifferentiated cell-specific promoter activity of ETn/MusD LTRs.
In a recent collaboration between our lab and the lab of Dr. James Ellis, this undifferentiated cell-specific activity was put to use by employing an ETn LTR as a stem-cell specific promoter for reporter gene expression in self-inactivating retro- and lentiviral vectors. Several modifications, including deletion of the LTR polyA signal and addition of Oct4 or Sox2 core enhancer elements produced optimized EOS (ETn, Oct4, Sox2) vectors with ES cell-specific expression (A. Hotta, J. Ellis & others, unpublished results). Due to the ETn LTR promoter, these vectors are highly expressed in mouse and human ES cells but not fibroblasts, this expression extinguished upon ES cell differentiation. Mouse (40, 44) and human (39, 46) somatic cells can be reprogrammed into induced pluripotent stem (iPS) cells by retroviral introduction of several pluripotency-inducing genes. Derivation efficiency of iPS cells can be enhanced by selection, using a stably integrated marker or an antibiotic resistance gene under control of a pluripotent cell-specific EOS promoter (A. Hotta, J. Ellis & others, unpublished results). EOS-driven pluripotent marker gene expression can be of value in studies dedicated to pluripotency, cell differentiation and de-differentiation. Technology of iPS production may assist in modelling human diseases and production of patient-specific cells and tissues for transplantation.

4.3. Regulatory effects of ERVs on genes

Many ERV families, including ETn/MusD (21), MuERV-L (17, 36, 38) and IAP retrotransposons (34, 38) are preferentially expressed during early mouse embryogenesis and have regulatory motifs specifically tailored to the expression in germline (8) or preimplantation embryo (32). It is tempting to speculate that a low level of ERV expression in these cells may be not only tolerated but even encouraged. It has been shown that reverse
transcriptase, abundant in the zygote due to LINE1 and LTR retrotransposon activity, is essential for normal early embryonic development, its depletion resulting in developmental arrest at 2- and 4-cell stages (1). In addition, new ERV retrotransposition events in germline and preimplantation embryo may give rise to an individual with new germline mutations (24) which, provided they are not detrimental, offer material for natural selection and are advantageous for the species.

In recent years, ERVs are taking a centre stage in study of gene regulation. In mouse ES cells, nearly 80% of the unique sequences precipitated with H3K9me3 and H4K20me3, otherwise predominantly associated with ERVs, were also found within 2 kb of LTRs (27), suggesting spreading of repressive chromatin from retroviral sequence into proximal regions and lending support to the concept of ERV-induced gene silencing. Since thousands of mouse and human ERVs/LTRs are located within gene borders (25, 42, 47), the disruptive consequence of their presence may affect gene expression, contributing to differences between individuals, mouse strains and species. Indeed, ETn and IAP elements are highly polymorphic between different mouse strains, possibly contributing to some of the strain-specific phenotypic characteristics (47). An additional layer of regulation of such polymorphic insertions is guided by epigenetics, when ectopic gene expression, driven by an LTR, may or may not manifest itself depending on the LTR methylation state (26).

High levels of ERV polymorphisms between species and mouse strains may be used in fingerprinting assays to establish phylogenetic connections between species or trace genealogy of the mouse strains. While there are very few ERV polymorphisms in human, studies are underway to determine if they may predispose individuals towards certain diseases (28). The more abundant polymorphic mouse ERVs, some of which may be
responsible for phenotypic differences between mouse strains (47), require more detailed investigation. Future studies of the impact of ERVs on the mouse genome may include examination of gene expression profiles of different mouse strains with respect to presence or absence of a polymorphic ERV. The next step would be to determine correlation between epigenetic status of ERVs and expression of adjacent genes, as well as ERV effects on splicing, polyadenylation and transcription initiation.

4.4. Conclusions

In recent years, multiple pairs of plant and mammalian autonomous and non-autonomous LTR retrotransposons were identified. While some cases require further functional confirmation (15, 16, 41, 43), trans-complementation experiments have unequivocally shown dependence of some non-autonomous elements on proteins encoded by their autonomous relatives (6, 35). The mechanisms responsible for a frequently observed dominance of non-autonomous forms with respect to transcription rate, retrotransposition or genomic copy number, are still unclear. Recent research has hinted that multiple and diverse mechanisms may be in place. Binding of a repressive histone variant macroH2A to non-autonomous MLV ERVs is reduced compared with binding to autonomous copies (4). In another system, a fusion protein spanning the deletion in non-autonomous IΔ1 IAP elements facilitates retrotransposition by recruiting the wildtype protein to the site of IΔ1 RNA synthesis (37). A third possible reason, namely, transcriptional dominance of the defective family due to depletion of the CpG-rich retroviral sequence, is presented here.

The research described in this thesis demonstrates that loss of CpG-rich retroviral sequence may be one of the routes retrotransposons take to circumvent epigenetic silencing
imposed by host surveillance mechanisms. These findings may suggest a solution to alleviating silencing of retro- and lentiviral vectors via reduction of their CpG content.

**4.5. Contributions to research and development**

In course of my research dedicated to transcriptional regulation of mouse ETn/MusD retrotransposons, I have shown that:

1. GC-boxes corresponding to Sp1/Sp3 binding sites are crucial for pluripotent-cell specific activity of ETnII LTR and act synergistically (Chapter 2).

2. ETnII retrotransposons are capable of amplifying in copy number in EC cells, suggesting that retrotranspositions may also take place in ES cells, thus affecting their genotype (Chapter 2).

3. ETn retrotransposons, unlike many retroviruses, possess multiple transcription start sites (Chapter 2).

4. Methylation of the LTR reduces its promoter activity (Chapter 3) and may be responsible for ETn and MusD silencing in differentiated cells and MusD suppression in undifferentiated cells (Chapter 3).

5. MusD retrotransposons are transcriptionally suppressed by DNA methylation and repressive chromatin initiating from their retroviral CpG-rich sequence, while ETn elements lack this sequence, thus avoiding suppression by epigenetic mechanisms (Chapter 3).

6. Normalizing transiently transfected plasmids to the promoterless vector of the corresponding size is essential to account for the drop in transfection efficiency observed with large plasmids (Chapter 3.2.4, Materials and methods).
Overall, my research has contributed to understanding of mechanisms guiding transcriptional regulation and suppression of ERVs in different cell types and mechanisms employed by ERVs to circumvent this suppression. Ultimately, this brings us one step closer to understanding mechanisms of host control and surveillance that may have protected the human population from invasion by ERVs during recent evolutionary history.
4.6. References


APPENDIX

PUBLICATIONS AND MANUSCRIPTS SUBMITTED FOR PUBLICATION

RESEARCH ARTICLES


REVIEW ARTICLES
