Retroviral Elements and Their Hosts: Insertional Mutagenesis in the Mouse Germ Line

Irina A. Maksakova, Mark T. Romanish, Liane Gagnier, Catherine A. Dunn, Louie N. van de Lagemaat, Dixie L. Mager*

ABSTRACT

The inbred mouse is an invaluable model for human biology and disease. Nevertheless, when considering genetic mechanisms of variation and disease, it is important to appreciate the significant differences in the spectra of spontaneous mutations that distinguish these species. While insertions of transposable elements are responsible for only ~0.1% of de novo mutations in humans, the figure is 100-fold higher in the laboratory mouse. This striking difference is largely due to the ongoing activity of mouse endogenous retroviral elements. Here we briefly review mouse endogenous retroviruses (ERVs) and their influence on gene expression, analyze mechanisms of interaction between ERVs and the host cell, and summarize the variety of mutations caused by ERV insertions. The prevalence of mouse ERV activity indicates that the genome of the laboratory mouse is presently behind in the “arms race” against invasion.

Introduction

The activity of transposable elements (TEs) places a variable mutational load upon their host species [1–3]. In species such as Drosophila, TEs comprise approximately 10% of heterochromatic [4] and only 2%–3% of euchromatic DNA [4,5] but cause over 50% of de novo mutations [6]. In contrast, nearly half of the human genome is TE-derived but de novo disease-causing insertions are rare [7,8]. TE activity in the laboratory mouse falls in the middle of these two extremes [8,9], largely because of the activity of endogenous retroviruses (ERVs) and other elements with long terminal repeats (LTRs), which together make up 8%–10% of the genome [7,9–12] (Box 1). A striking difference between the mouse and the human repertoire of ERVs/LTR elements is that the mouse contains many “active” LTR retroelements and a few potentially infectious ERVs that are closely related to exogenous mouse retroviruses [9,13]. Unlike inbred mice, infectious human ERVs have not been described, no new insertions have been found, and there are no ERVs closely related to human exogenous retroviruses [10,11,14]. In addition to LTR elements, the major classes of retrotransposons in mammals are the non-autonomous short interspersed elements (SINEs) and the autonomous long interspersed elements (LINEs) [7,9,12]. The retrotransposition and genomic effects of these non-LTR retroelements have been extensively discussed in a number of recent reviews [2,15,16].

Since the mouse is widely used as a disease model, it is important to understand the mutagenic events affecting this species and how they differ from those in humans. This article examines mouse ERVs and other LTR retroelements, focusing on insertional mutagenesis of the germ line. For the purposes of this review, LTR retroelements, which amplify via intracellular retrotransposition, and true exogenous retroviruses, which amplify by extracellular infections and retrotransposition, will be considered together as “ERVs” as they have a common evolutionary origin [13]. We discuss mutational mechanisms of different families of ERVs, illustrating significant differences in their effects on genes. We also discuss host responses to curtail ERV activity and, in some cases, to adopt ERVs for normal cell functions. Finally, we present the view that inbred mice are in a transitory state in which ERVs are not at equilibrium with their host genome.

Prevalence of ERV-Induced Mouse Insertional Mutations

Present-day activities are markedly different between human and mouse ERVs. A recent review lists 48 TE insertion mutations in human [17], and estimates of the frequency of novel human TE insertions range from one retrotransposition per 8–33 births [2,18–20], all of them due to L1-mediated non-LTR retrotransposons (reviewed in [21,22]). Given that ~47,000 mutant alleles have been characterized (according to the Human Gene Mutation Database [http://www.hgmd.cf.ac.uk/]; August 2005), these numbers suggest that ~0.1% of human spontaneous mutations are due to TE insertions, but none are due to ERV

Citation: Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, et al. (2006) Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. PLoS Genet 2(1): e2.

DOI: 10.1371/journal.pgen.0020002

Copyright: © 2006 Maksakova et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: Dnmt, DNA methyltransferase; ERV, endogenous retrovirus; ES, embryonic stem; ETn, early transposon; IAP, intracisternal A particle; LINE, long interspersed element; LTR, long terminal repeat; MLV, murine leukemia virus; RNAi, RNA interference; SINE, short interspersed element; siRNA, short interfering RNA; TE, transposable element

Irina A. Maksakova, Mark T. Romanish, Liane Gagnier, Catherine A. Dunn, Louie N. van de Lagemaat, and Dixie L. Mager are at the Terry Fox Laboratory, BC Cancer Agency, Vancouver, British Columbia, Canada, and Department of Medical Genetics, University of British Columbia, Canada. * To whom correspondence should be addressed. E-mail: dmager@bccrc.ca
Box 1. Classification of Mouse ERVs
A plethora of ERV families exist in the mouse and are grouped into three major classes (class I, II, and III) [9,13,80]. ERVs with infectious counterparts, namely MLV and mouse mammary tumor virus, have been studied for decades and are the subject of many excellent reviews [13,81,82]. Here we focus on the less well appreciated class II IAP and ETn elements, which account for the majority of characterized germ line insertional mutations, and briefly touch on representatives of other classes.

Class I retroviruses. The class I type C/gammaretroviruses, composing about 0.7% of the genome [9] and grouped based on similarity to MLV [13], were first isolated from lymphomas of AKR mice [83]. MLV entered the germ line of mice approximately 1.5 million years ago and its copy number ranges from 25 to 70 depending on the mouse strain. MLV proviruses are subdivided based on their host ranges determined by their env genes, but only a few encode replication-competent viruses [13,84]. Class I also includes several other families, but it is unclear if any have fully coding-competent members [13].

Class II retroviruses. The prototype of the much more numerous class II types B and D/betaretroviral group, composing about 3% of the genome [9], is mouse mammary tumor virus [13]. 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 [80,85].

One of the most extensively studied noninfectious families of ERVs in the mouse is the IAP family. Though there is variation in copy numbers between different mouse strains, about 700 full-length and 300 partially deleted elements are present in the haploid mouse genome. Type I elements encompass full-length members as well as four deleted classes [37]. Of these, the IA1 subclass, which has a 1.9-kb deletion in gag-pol, is the most abundant deleted form in the mouse genome and is also responsible for the majority of IAP insertional mutations [32,33]. Type II elements differ from type I elements by a characteristic 500-bp length difference and only comprise partially deleted members [37]. IAP elements were thought to lack an env gene until about 200 env-containing elements were discovered [86].

Most ERVs are highly transcribed during early zygotic divisions and in germ cells, resulting in an increased likelihood of new heritable proviral integrations (Box 3). Although genomic copy numbers of murine leukemia virus (MLV) are low (see Box 1), this family is the most active mouse ERV on a per provirus basis. New MLV provirus acquisitions are found in 2%–75% of the progeny in the highly susceptible SWRJ-RF/J hybrid mice [24–26]. AKR mice appear to gain one new ecotropic MLV provirus every 50–100 generations [13]. Five germ line mutations or strain variants due to insertions of MLVs have been well characterized in other lab strains (Table S1). The somatic effects of MLV, as well as mouse mammary tumor virus, in activating oncoproteins via insertional mutagenesis are well known [27]. Indeed, mapping common retroviral integration sites in mouse tumor systems has proven a powerful strategy to identify new genes involved in cancer [28,29].

Intracisternal A particle (IAP) and Early Transposon (ETn)/MusD elements are present in much higher copy numbers than MLVs (Box 1) and are responsible for the majority of ERV-induced de novo germ line mutations (Tables 1 and S1). In addition, IAP elements are frequent insertional mutagens in somatic cells, particularly in leukemia, plasmacytoma, and myeloma cell lines, and can activate oncogenes or cytokine genes [30]. Notably, very few new ETn insertions have been

activity. As in human, a small number of de novo germ line L1 insertions have been reported in mice (reviewed in [22,23]). However, the ongoing activity of ERVs accounts for the majority of new insertional mutations in the mouse. To provide a current estimate of the fraction of spontaneous mutations due to ERV insertion, we tabulated all documented cases and found 63 (Table S1). The Mouse Genome Informatics database (http://www.informatics.jax.org/) lists 1,489 spontaneous mutant alleles (as of August 2005). After removing unannotated cases and numerous non-independent entries and revertants derived from the nonagouti allele (Box 2), 519 spontaneous alleles with an annotated molecular mechanism remained. This list included 55 of our 63 ERV insertion mutations. Taken at face value, these figures suggest that 10%–12% of all mutations are due to ERV insertions, a fraction very similar to previously reported estimates of 10%–15% based on lower numbers [22,23]. Reversion of ERV-induced mutations has also been observed at a few loci due to LTR–LTR recombination (Box 2). It should be noted that the 10%–15% figure is likely an underestimate because of ascertainment bias. For example, point mutations in coding regions will be more readily detected than ERV insertions in introns or outside gene borders. Regardless of the precise figure, ERV activity in inbred mice is dramatically higher than in modern humans.
3.9–4.5 x10

Germ line reversions of the are caused by aberrant splicing due to an MLV integration into the mutation [94] and the mouse and human genomes [9,11]. In the case of the proviral forms and make up the bulk of retroviral material in the LTRs are typically present in much higher copy numbers than full mechanisms have been efficient throughout evolution. Indeed, solitary LTRs has been noted in early studies on MLV [93], and this parental germ line, leaving behind a solitary LTR. Such generation of recombination between the 5' and the 3' LTRs of the provirus in the reversion involves deletion of internal ERV sequences via homologous among mice carrying ERV insertional mutations. The mechanism of to affected gene

Intronic orientation relative Mixed Sense (15/16 cases)

Somatic insertions Common Rare

Active subtype IAP I

Strain bias C3H (50%) A/J (∼30%)

Behavioral effects Yes (antisense) No

Polyadenylation and splicing effects Common Common

Somatic insertions Common Rare

Intronic orientation relative to affected gene Table 1. Overview of IAP and ETn Mutagenic Germ Line Insertions

| Characteristic                              | Family |
|--------------------------------------------|--------|
|                                            | IAP    | ETn/MusD |
| Genomic copy number                        | ~1,000 | 300–400 |
| Number of mutagenic insertions*            | 32     | 23      |
| Strain bias                                | C3H (50%) A/J (∼30%) |
| Active subtype                             | IAP IΔ1 ETnII-β |
| Promoter effects                           | Yes (antisense) No |
| Polyadenylation and splicing effects        | Common Common |
| Somatic insertions                         | Common Rare |
| Intronic orientation relative to affected gene | Mixed Sense (15/16 cases) |

*Characterized new germ line insertions, including strain variants. For complete details, see Table S1.

Box 2. Reversion of ERV-Induced Mutations

In a few instances, reversion to a wild-type phenotype occurs among mice carrying ERV insertional mutations. The mechanism of reversion involves deletion of internal ERV sequences via homologous recombination between the 5' and the 3' LTRs of the provirus in the parental germ line, leaving behind a solitary LTR. Such generation of solitary LTRs has been noted in early studies on MLV [93], and this mechanism has been efficient throughout evolution. Indeed, solitary LTRs are typically present in much higher copy numbers than full proviral forms and make up the bulk of retroviral material in the mouse and human genomes [9,11]. In the case of the hairless mutation [94] and the dilute coat color mutation [95], both of which are caused by aberrant splicing due to an MLV integration into the intron, reversion to wild-type occurs via generation of a solitary LTR. Germ line reversions of the dilute mutation occur at a frequency of 3.9–4.5 x10^-6 events per gamete [96], and one somatic revertant, chimeric for about 50% of reverted cells, was encountered, with the frequency of somatic reversion estimated at 9 x 10^-7 per animal analyzed [96]. Occasionally, reversions can result in diverse phenotypes because of expression of different forms of transcripts. Such is the case with the nonagouti a allele, which encompasses an insertion of a 5.5-kb VL30 element containing 5.5 kb of additional internal sequence flanked by 526-bp direct repeats. By means of homologous recombination, a can revert to two dominant agouti alleles, black-and-tan (alt), containing only the VL30 element with a single internal 526-bp repeat, and the white-bellied agouti (Awe), which only has a single VL30 LTR [97]. There is no published evidence of other ERV-induced mutations reverting to wild-type, perhaps because of the difficulties associated with breeding the mutant animals, the time span required to detect reversions, and, for the somatic revertants, restricted tissues where the reversion could be detected.

Genetic Background and ERV Subtype Influence on Insertion Probability

Not all mouse strains are equally susceptible to ERV insertions. Most of the IAP insertions have occurred in C3H/HeJ (Tables 1 and S1), and nearly all of these cases are of the IAP subtype IΔ1 [32,33]. It seems likely that one or a small number of IΔ1 IAP elements are active in this strain, possibly because of a favorable genomic context or escape from host suppression mechanisms (see below). The IΔ1 subtype, however, requires complementation in trans from coding-competent IAP elements [34], so the latter must also be expressed. A specific strain bias for ETn insertions is not as obvious, although six mutations have occurred in A/J mice and two in each of two other infrequently used strains (SELH/Bc and MRL/MpJ) (Table S1). As with IAP elements, this suggests that some strains harbor more “active” elements and/or allow more ETn or MusD expression. Where sufficient sequence is available, it has been found that most ETn insertions are of a particular structural subtype, ETnII-β, and are nearly identical, suggesting very few currently active elements ([35]; unpublished data). A very definite strain bias has been observed for germ line movement of MLVs, with AKR and SWR/J-RF/J mice being highly active strains (reviewed in [13]) (see Box 3). One of the explanations for such selective activity in particular strains may be the presence of a single highly transcribed master element in a favorable genomic context, as appears to be the case for AKR mice. The other explanation is the difference in host suppression factors, such as methylation levels and the presence of virus-suppressing loci. One long-studied locus involved in suppression of MLV and a number of other ERVs is Fv1 (see below).

As mentioned above, most de novo insertions of IAP and ETn elements are those of defective sequences lacking full coding potential. This fact is curious, given that, in assay systems, coding-competent IAP and MusD elements retrotranspose much more efficiently than proteins and retrotransposing RNA are encoded by the same template (cis preference) [34,36]. Expression patterns of defective and full-length IAP elements vary widely in different cells [30,32,37], but the IΔ1 deleted subtype is preferentially expressed in acute myeloid leukemia cell lines derived from C3H/HeJ mice, despite being present at lower genomic numbers than full-length forms [32]. Non-coding ETnII elements are transcribed at a much higher level than their coding-competent MusD relatives [35], probably explaining their higher likelihood to retrotranspose. Indeed, among the 23 characterized mutagenic ETn/MusD insertions, only two have been reported as MusD (Table S1). However, it is unclear why transcripts from defective elements would predominate in vivo. Possibly they are less likely to be recognized as retroviral elements and to be repressed by host cell silencing machinery.

Mutagenic Mechanisms of ERV Insertions

Most commonly, germ line 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 (Tables 1 and S1). In some cases, small amounts of normal gene transcripts and protein can still be detected. While the number of characterized MLV-induced mutations is too small to perceive general trends, IAP and ETn elements show significant differences in their effects on genes (Figure 1). For ETn insertions, the most commonly reported defect is
promoters active during early embryogenesis and in the germ line, driving speciation, as has been discussed by others [2,3,71,102]. Associated with genes differentially expressed in the two species chimpanzee but not human found that several such elements were human and primate placenta due to an alternative promoter provided key enzyme in estrogen biosynthesis, which has high expression in ERV that provides an enhancer [111], and the gene, which has acquired male-specific expression due to a MuRRS reported [71,74,107–110], such as the mouse CYP19 Slp (sex-limited protein) gene, encoding a CYP19 gene transcripts have not been well characterized.

A striking difference between the effects of IAP and ETn elements is their tendency to drive ectopic gene expression (Figure 1). For IAP elements, nine cases of LTR-driven gene expression have been reported. Interestingly, eight of these nine cases are driven from an antisense promoter located in the 5′ IAP LTR (Figure 2B; Tables 1 and S1). Many of the mutant alleles caused by IAP LTR-driven gene expression show variable expressivity among genetically identical mice and have therefore been termed metastable epialleles [38]. The variable expressivity is due to stochastic establishment of the 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 premature polyadenylation within the ETn, coupled with aberrant splicing due to a few commonly used cryptic splice signals (Figures 1 and 2A). IAP insertions within introns also typically cause aberrant splicing but use a wider variety of cryptic splice signals. In addition, compared to ETns, fewer cases of premature polyadenylation within IAP elements are well documented (Figure 1), but, in many cases, all aberrant gene transcripts have not been well characterized.

A growing number of studies have shown that LTR elements are particularly well suited to donation of enhancers or promoters and, if fixed, can assume roles in gene regulation. Many examples of mammalian genes regulated by ERVs/LTRs or other TEs have been reported [71,74,107–110], such as the mouse Slp (sex-limited protein) gene, which has acquired male-specific expression due to a MuRRS ERV that provides an enhancer [111], and the CYP19 gene, encoding a key enzyme in estrogen biosynthesis, which has high expression in human and primate placenta due to an alternative promoter provided by an ancient LTR [74]. A recent study documenting ERVs present in chimpanzee but not human found that several such elements were associated with genes differentially expressed in the two species [112], raising the possibility that ERVs and other TEs may be critical in driving speciation, as has been discussed by others [2,3,71,102].

Most ERV LTRs seem to be extremely powerful, ready-made promoters active during early embryogenesis and in the germ line, and recent work has fueled speculation that LTR retroelements could play a role in expression of genes essential for early development. Knowles and co-workers reported that about 13% of cDNAs from full-grown mouse oocytes contained retroviral sequences, the majority of which were derived from the poorly understood class III MT MaLR family [52], but by the two-cell stage, transcripts of another class III family, MuERV-L, started to predominate [52,113]. Notably, MT, MuERV-L, and some other LTRs were found to act as alternative promoters for subsets of host genes in full-grown oocytes and cleavage-stage embryos, apparently controlling synchronous, developmentally regulated expression of these genes [52]. An independent study reported that blocking MuERV-L expression inhibits embryonic development to the four-cell stage [113]. Such findings have prompted the proposal that differential gene expression driven by LTRs may trigger sequential reprogramming and genome remodeling during embryonic development [52]. This idea is intriguing, but since LTR retrotranspon families and insertion patterns are generally not conserved across divergent species, it is difficult to envisage a scenario in which such elements evolved to play a critical role in common developmental processes. It is easier to imagine LTRs/TEs being involved in species-specific processes. Nevertheless, it is clear that McClintock’s original theory of TEs as “controlling elements” [114] and Britten and Davidson’s postulation of repetitive elements as regulatory units [115], views that have been little appreciated for decades, are now gaining increasing attention.
mutant phenotype. Such cases have been extensively studied by Whitelaw and coworkers, 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 [39].

It is unclear why no instances of ETn-promoted ectopic gene expression have been observed, but the lack of such cases could be explained by inactivity of the ETn LTR promoter in somatic cells due to heavy methylation or lack of necessary transcription factors. Expression studies (Box 3) indicate that at least some IAP elements are transcribed in various cell types, a property that would increase the probability of such elements providing promoter function. It is possible that the presence of the cryptic antisense promoter in the IAP LTR also increases the likelihood that an IAP element 5' of a gene will provide promoter function (Figure 2B).

Host Silencing Mechanisms

Transcriptional gene silencing. To guard against harmful genomic consequences of ERVs and other TEs, an arsenal of cellular defense strategies has evolved to counteract their amplification (Figure 3; Table 2). Transcriptional gene silencing is a principle mechanism for controlling TEs in a broad range of species including mammals, flowering plants, and those fungi whose genomes contain m^5C [40]. The best-documented mechanism, DNA methylation of promoters, can directly impede access of transcription factors or lead to an inactive form of chromatin at target loci [41]. Indeed, a majority of genomic CpG dinucleotides and 5-methyl cytosines reside within ERVs and other retroelements in mammals [42]. Several lines of evidence confirm that genomic hypomethylation and TE activation are interrelated. DNA methyltransferase (Dnmt) mutant mice with mutant Dnmt1 or Dnmt3 do not maintain and initiate methylation at existing or new proviral loci, respectively [43,44]. In fact, both MLVs and IAPs become substantially demethylated [43], and IAP transcripts are expressed up to 100-fold higher in Dnmt1/C0 mice relative to wild-type [45]. While Dnmt1 is necessary after DNA replication, Dnmt3a and Dnmt3b are essential in the germ line and during development to establish the

DOI: 10.1371/journal.pgen.0020002.g001

Figure 1. Mutagenic Mechanisms of IAP and ETn Insertions
IAP and ETn insertions were classified by their mechanism of gene disruption. Well documented instances of aberrant transcription initiation (5'-terminus) and polyadenylation (3'-terminus) were counted, as well as aberrant splicing and exon skipping (internal disruption). Insertions that cause gene disruption by multiple mechanisms (Table S1) were counted once in each relevant class.

DOI: 10.1371/journal.pgen.0020002.g002

Figure 2. Common Effects of ETn and IAP Insertions on Gene Expression
(A) ETn effects on gene transcript processing. The most common patterns of aberrant transcript processing caused by ETns in gene introns are shown. The natural LTR polyadenylation (polyA) site and a second cryptic polyadenylation site in the internal region, along with four cryptic splice acceptors (SA) and a donor site (SD), are involved in most cases. The number of such cases is an underestimate, since several reports lack sufficient detail of aberrant transcripts. In some cases, several aberrant forms have been found. Boxes denote gene exons, thin lines denote introns, and thick lines denote spliced mRNAs, with direction of transcription from left to right. For clarity, cryptic splice acceptor sites in the 3' LTR are not shown since no documented splicing events involving these sites were found. Intronic mutagenic ETns and the affected gene are most often found in the same orientation (15 of 16 cases).

(B) IAP promoter effects on gene transcription. Ectopic gene expression driven by an antisense promoter in the 5' LTR of an IAP has been reported in eight cases. In some cases, the IAP is located a significant distance upstream of the gene.
above the gene locus. Various potential aberrant or chimeric transcripts further rounds of retrotransposition or retroviral infection, is shown the gene. A full-length retroviral transcript, which could be packaged for normal gene transcript driven by the native promoter (P) is shown below donor and acceptor sites. Spliced RNA is depicted with dashed lines. A through introduction of polyadenylation sites, promoters, and splice details and other examples. An ERV or LTR element within an intron is shown to illustrate common gene-disruptive effects of such sequences and RNA processing block, uncoating block, Fv4; and APOBEC3. See text and Table 2 for more details and other examples. An ERV or LTR element within an intron is shown to blocked by pre-infection MLV. Antiviral protein Viral RNA expression MLV. Induction of spraying T lymphocyte response Post-infection Friend virus. Poorly understood function affecting viral capsid Early post-entry HIV, SIV, MLV. Role in antibody response against infection Post-infection Polytropic MLV. TEs (class I and II LTRs, LINEs, SINEs, and centromeric repeats), and their transcripts are abundant. This effect appears specific to repetitive sequences. A further level of silencing is mediated by histone methylation. Intriguingly, different families of repeats were found to have characteristic depressive histone methylation patterns [50]. Furthermore, histone methyltransferase knockout ES cells exhibited a loss of methylation repertoire [44]. This activity is largely restricted to dispersed and tandem repeats [46]. Dnmt3a and Dnmt3b knockout embryonic stem (ES) cells are unable to establish methylation at new MLV integrations. Knockout ES cells and embryos exhibit a general decrease in methylation at centromere repeats, MLVs, IAPs, and L1s [44]. In addition to the research on promoter methylation, there is a study showing that intragenic methylation reduces the elongation efficiency of RNA polymerase II [47], which suggests that the methylated state of TEs within introns might affect gene expression.

It is well established that genomic methylation can serve to recruit chromatin-remodeling proteins [41]. The SWI/SNF family members are components of the trithorax group protein complex and are responsible for maintaining transcriptional activity. A SWI/SNF mammalian catalytic subunit, Brm (SWI/SNF-related, matrix associated, actin-dependent regulator of chromatin), is involved in increased transcription of retroviral RNA, but this is alleviated in cells lacking this protein [48]. Moreover, Brm-deficient cells treated with histone deacetylase inhibitors are unable to silence transcription of retroviral genes. These results suggest that Brm-type SWI/SNF is essential for TE expression and that histone deacetylation is crucial for silencing. Paradoxically, Lsh (lymphoid-specific helicase), also a SWI/SNF family member, preferentially associates with repeats and contributes to their silencing [49]. Lsh<sup>−/−</sup> mice are hyperacetylated at histones overlaying TEs (class I and II LTRs, LINEs, SINEs, and centromeric repeats), and their transcripts are abundant. This defect appears specific to repetitive sequences. A further level of silencing is mediated by histone methylation. Intriguingly, different families of repeats were found to have characteristic reproductive histone methylation patterns [50]. Furthermore, histone methyltransferase knockout ES cells exhibited a loss of methylation repertoire [44].

| Species | Gene | Function | Time of Block | Virus/Repeat Restricted | References |
|---------|------|----------|---------------|-------------------------|------------|
| h, ch, a, m | APOBEC3 (g/f) | Cytidine deaminase, RNA editing | During RT | HIV, SIV, MLV, IAP, MusD | [66] |
| Mammals | Dnmt1 | Maintenance methylation (adult) | Viral RNA expression | IAP, MLV | [42,44] |
| Mammals | Dnmt1o | Maintenance methylation (embryonic) | Viral RNA expression | IAP | [116] |
| Mammals | Dnmt3a | De novo methylation (germ/ES cells), maternal/paternal imprinting | Viral RNA expression | MLV, L1, IAP | [43,117] |
| Mammals | Dnmt3b | De novo methylation (ES cells/early embryos) | Viral RNA expression | MLV, L1, IAP, centromeric repeats | [43] |
| Mammals | Dnmt3li | De novo methylation (spermatogonial precursors) | Viral RNA expression | IAP, LINE1 | [118] |
| m | Edg2 (Rec1) | G-protein-coupled receptor | Pre-infection | Not available | [61] |
| m | Fv1 | Endogenous gag-related, binds assembled virions inducing dominant negative effect | Post-infection, pre-integration | MLV | [65] |
| m | Mst1r (Fv2) | C-met-related tyrosine kinase, limits target cell proliferation | Post-infection | Friend virus | [119] |
| m | Fv4 (Akvr) | Endogenous env-related, binds cell receptors | Pre-infection | Ecotropic MLV | [120] |
| m | Fv5 | Limitation of target cell proliferation | Post-infection | Friend virus | [121] |
| m | Lsh | ATP-dependent chromatin-remodeling protein | Viral RNA expression | LTR ERVs, LINE, SINE, satellite | [49] |
| m | Nxf1 | Nuclear export factor, role in correctly processing pre-mRNAs possessing retroviral sequence | Post-integration | IAP, MLV, LINE? | [67] |
| Mammals | Ref1/Lv1/Trim5 | Poorly understood function affecting viral capsid | Early post-entry | HIV, SIV, MLV | [68] |
| m | Rf1/2/3 | Induction of cytotoxic T lymphocyte response | Post-infection | Friend virus | [122,123] |
| m | Rmdc | Endogenous env-related, binds cell receptors | Pre-infection | Polytropic MLV | [124] |
| m | Rmv12/3 | Role in antibody response against infection | Post-infection | MLV | [125] |
| m | Scl20a2 (Rom1) | Cell membrane transporter proteins | Pre-infection | Exogenous amphotropic MLV | [126] |
| m | Xpr1 (Rmc1/Sxv) | Xenotropic and polytropic retroviral receptor | Pre-infection | MLV | [127] |
| Rodents | ZAP | Antiviral protein | Viral RNA expression | MLV | [128] |
these repressive marks and an increase in transcription from tandem and interspersed repeats.

**Post-transcriptional gene silencing.** Since transcriptional silencing is unlikely to prevent activity of all TEs, it is essential that some processes act at the level of expressed transcripts. An RNA interference (RNAi)–mediated mechanism, the components of which are discussed elsewhere [51], is involved in post-transcriptional gene silencing of repetitive DNA. High levels of sense and antisense IAP and ERV-L transcripts are expressed concurrently in developing mouse embryos by RNAi-mediated knockdown of Dicer results in a 50% increase in IAP and ERV-L transcripts [52,53]. Dicer knockout mouse ES cells exhibit increased transcription from centromeric repeats, L1s, and IAPs, combined with severe developmental defects [54]. In an analogous example, silencing of the mammalian X chromosome is dependent upon an antisense transcript and shortly after its detection, silencing of the mammalian X chromosome is dependent [54]. In an analogous example, silencing of the mammalian X chromosome is dependent upon an antisense transcript and shortly after its detection, histone 3–lysine 9 and CpG methylation is established at Xist [55], connecting double-stranded RNAs to transcriptional gene silencing. The fact that heterochromatin can be established at homologous loci via short interfering RNAs (siRNAs) is well documented. Examples in model organisms such as fission yeast and Arabidopsis have implicated repeat-derived siRNAs in directing such conformational changes. Fission yeast deleted for RNAi pathway components express centromeric-repeat and integrated transgene transcripts, normally heavily silenced by heterochromatization [56]. Studies in plants show that TEs and tandem repeats specifically become silenced by histone 3-lysine 9 and CpG methylation. These changes are dependent on the chromatin remodeling factor Decrease in DNA Methylation 1 (DDM1) and guided by siRNAs. Indeed, various Arabidopsis genes become subject to RNAi-mediated silencing because of TE proximity to their promoters [57]. Similar results in human cells have demonstrated that non TE-derived siRNAs targeted to the EF1A promoter of a proviral green fluorescent protein reporter inhibits transcription of the transfected EF1A promoter, as well as that of the endogenous copy [58]. Also, siRNAs targeting the E-cadherin promoter induced DNA methylation and heterochromatin [59], but DNA methylation is not a prerequisite, as shown with the CDHI promoter [60].

**Host restriction factors.** Finally, a variety of gene products, some derived from domesticated viral genes, function at various stages of the retroviral life cycle to curtail both exogenous retroviruses and ERVs and have been extensively reviewed recently [61–64] (Table 2). Some particularly relevant examples include Fv1, the Ref/Lv1 family of proteins, APOBEC3G, and Nxf1. Fv1, the “prototypic” retrovirus restriction gene, is an ancient ERV-L gag-like gene that restricts infection by MLV [65]. APOBEC3G encodes a cytidine deaminase that mediates cytosine–uracil transitions when co-packaged with retroviral genomes. It inhibits HIV and MLV replication and also suppresses IAP and MusD/ETn retrotransposition [66]. Nxf1, encoding an mRNA nuclear export factor, has been shown to suppress the hypomorphic effects of intrinsic IAP insertions, presumably by facilitating accurate splicing [67]. The recently described Ref1/Lv1 family of proteins, including TRIM5α, suppresses HIV and MLV [68,69]. However, effects of these proteins on ERVs are unknown.

**Inbred Mice—Out of Balance with Their ERVs?**

The evolution of silencing mechanisms by the host likely, in turn, places pressure on TEs/ERVs to evolve means to escape repression, setting up an “arms race,” not unlike that involving the immune system and infectious agents [1,70,71]. In the case of TEs, waves of amplification are countered by host defenses (Figure 3) and negative selection that quench activity until new variants or “master” elements appear that are capable of instigating further genomic expansions [1,3]. The high rate of ERV germ line and somatic insertional mutations in the laboratory mouse indicates that at least some inbred strains are currently in an active phase of ERV genomic expansion. In contrast, ERV-like elements in humans, while present in comparable overall numbers, have long ago ceased activity [7]. It is interesting to speculate which is the more common situation in modern-day mammals. Without detailed analysis of a variety of mammalian genomes and mutational spectra, it is difficult to answer this question. In mouse, the still active ETn and IAP elements likely amplify via intracellular retrotransposition, thereby avoiding the “front line” defense mechanisms, such as Fv1 and Fv4, in place to inhibit early stages of exogenous infections. In contrast, MLV likely amplifies primarily through rounds of infection of germ line cells, allowing more opportunities for the host to evolve resistance and keep proviral copy number low. We propose that inbred mice represent a relatively transitory state in which host silencing mechanisms have not yet adapted to retrotransposition of new ERV variants. The IAP family nicely illustrates this point. The IAI partly deleted subtype is currently the most active IAP element but is 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 IAI subtype must have arisen recently and, possibly because of specifics of its structure and/or genomic context, has been freed from suppression and allowed to retrotranspose—mainly in the C3H/HeJ strain. A similar scenario is occurring with respect to ETn/MusD elements, where a minor population of ETnII-β elements is causing the bulk of current retrotranspositional activity. This relatively permissive phase of ERV expansion that is ongoing in inbred mice provides a rare opportunity to study how a mammalian host genome responds to new waves of invasion by mobile elements.

**Conclusion**

This review has attempted to highlight the mutational impact that ERVs have had and continue to have on the mouse germ line and to discuss host defenses that have evolved to control these elements. Unlike in human, ERVs in the mouse genome are in an expansion phase, with specific IAP and ETn variants currently playing the dominant role. These elements have accumulated to hundreds of copies in the genome, but evidence indicates that only a few have a high probability of retrotransposing. Identification of their genomic location and/or chromatin state may provide insight into host control mechanisms and why particular elements escape suppression. Genetic factors responsible for variable retrotransposition rates in different strains also await discovery and may reveal new host restriction genes or alleles. Given the propensity for these ERVs to affect gene
expression, it would be interesting to investigate ERV insertions as mediators of phenotypic differences among inbred strains. Indeed, it may be particularly informative to examine genes harboring polymorphic ERV insertions in their introns.

The epigenetic control of mouse and human ERVs is of substantial interest because of their potential effects on adjacent genes. In addition to obvious gene-disruptive effects, mammalian ERVs may also play a role in tissue-specific gene regulation (see Box 4). Some IAP elements act as metastable epialleles [38] with their methylation state determining effects on neighboring genes. The idea that variable silencing of retrotransposons could contribute to gene expression variability in mammals [39] is attractive but, thus far, IAPs are the only type of retroelement shown to display this effect, and it remains to be determined how widespread this phenomenon may be. Functions for RNA-mediated silencing, including potential roles for RNAi [53] and microRNAs [72] in controlling ERVs and exogenous retroviruses, are being rapidly elucidated. A number of questions, however, including the origins of double-stranded RNA necessary for inducing silencing, are currently unanswered [78].

Although ERV insertions are not a source of new mutations in humans, understanding their effects in mice is important for understanding gene regulatory effects of existing human ERVs.LTRs, thousands of which are located within gene borders [74,75], and in elucidating the disruptive effects of therapeutic retroviral vectors. Retroviral activation of protooncogenes has occurred in gene therapy trials, raising major concerns [76]. Therefore, potential long-range promoter or enhancer effects, as displayed by IAP elements, need to be considered and vectors designed to reduce the chances of oncogene activation [77]. The high probability of some retroviruses integrating into introns [78,79] may also limit their usefulness as therapeutic gene delivery systems if aberrant gene splicing and polyadenylation results. Eliminating cryptic splicing and polyadenylation signals within retroviral vectors may be a worthwhile strategy. However, as demonstrated by the mouse ERVs, unique properties and sequence motifs result in distinct mutational mechanisms (see Figure 1), indicating the challenge of attempting to predict a priori the mutagenic behavior of different classes of retroviruses.

Supporting Information

Table S1. Germ Line Mouse Mutations Caused by ERV Insertions

Found at DOI: 10.1371/journal.pgen.0020002.s001 (233 KB DOC).

Acknowledgments

Related work in our laboratory was supported by a grant from the Canadian Institutes of Health Research with core support provided by the BC Cancer Agency. We thank Patrick Medstrand and anonymous reviewers for helpful suggestions and Christine Kelly for manuscript preparation. We also thank Susan Mc Clintchy at The Jackson Laboratory for help with the Mouse Genome Informatics database. We regret that space limitations prevented the citing of all relevant work by colleagues.

References

1. Brookfield JFY (2005) The ecology of the genome—Mobile DNA elements and their hosts. Nat Rev Genet 6: 128–136.
2. Kazazian HH Jr (2004) Mobile elements: Drivers of genome evolution. Science 303: 1626–1632.
3. Kidwell MG, Lisch DR (2001) Perspective: transposable elements, parasitic DNA, and genome evolution. Evolution Int J Org Evolution 55: 1–24.
4. Bartolome C, Maside X, Charlesworth B (2002) On the abundance and distribution of transposable elements in the genome of Drosophila melanogaster. Mol Biol Evol 19: 926–937.
5. Kaminker J, Bergman C, Kronmiller B, Carlson J, Svirskas R, et al. (2002) The transposable elements of the Drosophila melanogaster euchromatin: A genomics perspective. Genome Biol 3: RESEARCH0084.
6. Eickhobb TH, Furano AV (2009) Fruit flies and humans respond differently to retrotransposons. Curr Opin Genet Dev 12: 697–674.
7. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409: 860–921.
8. Kazazian HH Jr (1998) Mobile elements and disease. Curr Opin Genet Dev 8: 345–350.
9. International Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420: 5280–5282.
10. Bannert N, Kurth R (2004) Retroelements and the human genome: New perspectives on an old problem. Proc Natl Acad Sci U S A 101: 14572–14579.
11. Methow DL, Medstrand P (2005) Retroviral repeat sequences. In: Cooper D, editor. Nature encyclopedia of the human genome. Hampshire (United Kingdom): Macmillan Publishers. pp. 57–63.
12. Smit AF (1999) Interpersed repeats and other remnants of transposable elements in mammalian genomes. Curr Opin Genet Dev 9: 657–663.
13. Boeke JD, Stoye JP (1999) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In: Coffin JM, Hughes SH, Varmus H, editors. Retroviruses. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press. pp. 345–453.
14. Katzonakis A, Tzistan M (2005) Phylogeny of human endogenous and exogenous retroviruses. In: Sverdlov ED, editor. Retroviruses and primate genome evolution. Georgetown (Texas): Landes Bioscience. pp. 186–203.
15. Deininger PL, Moran JV, Batzer MA, Kazazian J, Haig H (2005) Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 15: 651–658.
16. Jurka J (2004) Evolutionary impact of human Alu repetitive elements. Curr Opin Genet Dev 14: 605–608.
17. Chen JM, Stenson PD, Cooper DN, Ferrer C (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet 117: 411–427.
18. Brohia B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, et al. (2003) Hot Linds for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A 100: 5280–5285.
19. Kazazian HH Jr (1999) An estimated frequency of endogenous insertion mutations in humans. Nat Genet 22: 130.
20. Li X, Scarcinge WA, Hill KA, Roberts S, Menges A, et al. (2001) Frequency of recent retrotransposition events in the human factor IX gene. Hum Mutat 17: 511–519.
21. Deininger PL, Batzer MA (1999) Alu repeats and human disease. Mol Genet Metab 67: 183–193.
22. Druker R, Whitleaf E (2004) Retrotransposon-derived elements in the mammalian genome: A potential source of disease. J Inherit Metab Dis 27: 319–330.
23. Kazazian HH Jr, Moran JV (1998) The impact of L1 retrotransposons on the human genome. Nat Genet 19: 19–24.
24. Jenkins NA, Copeland NG (1995) High frequency germline acquisition of ecotropic MuLV proviruses in SWRJ-RFJ hybrid mice. Cell 43: 811–819.
25. Spencer SE, Gilbert DJ, Swing DA, Copeland NG, Jenkins NA (1990) Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Sve lines of mice. Mol Cell Biol 9: 177–184.
26. Szabo C, Kim YK, Mark WH (1995) The endogenous ecotropic murine retroviruses Env-16 and Env-17 are both capable of producing new proviral insertions in the mouse genome. J Virol 67: 5704–5708.
27. Rosenberg N, Jolicoeur P (1997) Retroviral pathogenesis. In: Coffin JM, Hughes SH, Varmus H, editors. Retroviruses. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press. pp. 457–589.
28. Dudley JP (2003) Tag, you're it: Retroviral insertions identify genes involved in cancer. Trends Mol Med 9: 43–45.
29. Suzuki T, Shen H, Akagi K, Morse HC, Malley JD, et al. (2002) New genes involved in cancer identified by retroviral tagging. Nat Genet 32: 166–174.
30. Wang XY, Steelman LS, McCubrey JA (1997) Abnormal activation of cytokine gene expression by intracisternal type A particle transposition: Effects of mutations that result in autocrine growth stimulation and cytokine gene expression by intracisternal type A particle transposition: Effects of mutations that result in autocrine growth stimulation and malignant transformation. Cytokines Cell Mol Ther 3: 3–19.
31. Baust C, Baille GJ, Mager DL (2002) Insertional polymorphisms of ETn retrotransposons include a disruption of the wz gene in C57BL/6 mice. Mammm Genome 15: 425–428.
32. Ishihara H, Tanaka I, Wan H, Nojima K, Yoshida K (2004) Retrotransposition of limited deletion type of intracisternal A-particle elements in the myeloid leukemia Clls of C3HHe mice. J Radiat Res (Tokyo) 45: 25–32.
33. Rakyan VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, et al. (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. Proc Natl Acad Sci U S A 100: 2538–2543.
34. Dewannieux M, Dupressoir A, Harper F, Pietron G, Heidmann T (2004)
Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. Nat Genet 56: 534–539.
35. Baust C, Gagnier L, Baille JG, Harris MJ, Juriloff DM, et al. (2003) Structure and expression of mobile TEfI retroelements and their contribution to Hash relatives in the mouse. J Virol 77: 11448–11458.
36. Ribo D, Dewannieux M, Heidmann T (2004) An active murine transposon family: Retrotransposition of "master" MusD copies and ETn transmobilization. Genome Res 14: 2261–2267.
37. Kuff FL, Luedke CR (1988) The intracisternal A particle gene family: Structure and functional aspects. Adv Cancer Res 51: 183–276.
38. Ravan K, Blevitt ME, Drucker R, Preis J, Whitelaw E (2002) Metastable epialleles in mammals. Trends Genet 18: 548–551.
39. Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroelements is constrained by cytosine methylation. Nat Genet 20: 116–117.
40. Goll GM, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem 74: 481–514.
41. Lorincz MC, Dickerson DR, Schmitt M, Grouin M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11: 1066–1072.
42. Reisler R, Horowitz Sc, MusD (2005) Multiple groups of transposable elements in mice. Gene Expr Patterns 4: 467–471.
44. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases (DNMTs) and gene silencing by short interfering RNAs in human cells. Nature 431: 211–217.
45. Heard E, Rougeulle C, Arnaud D, Avner P, Allis CD, et al. (2001) Role of transposable elements in heterochromatin and epigenetic control. Rev Biochem 74: 481–514.
46. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements: A family of endogenous retrovirus-like elements active in mammalian LTR-transposons. Nucleic Acids Res 21: 1863–1872.
47. Lorincz MC, Dickerson DR, Schmitt M, Grouin M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11: 1066–1072.
48. Mizutani T, Ito T, Nishina M, Yamamichi N, Watanabe A, et al. (2002) DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Genet 37: 906–910.
49. Lorincz MC, Dickerson DR, Schmitt M, Grouin M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11: 1066–1072.
50. Baillie GJ, van de Lagemaat LN, Mager DL (2004) Multiple groups of transposable elements in mice. Gene Expr Patterns 4: 467–471.
51. Hannon GJ (2002) RNA interference. Nature 418: 244–251.
52. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, et al. (2005) The profile of repeat-associated histone lysine methylation states in preimplantation embryos. Dev Cell 7: 597–606.
53. Saavedra P, Stein P, Anger M, Bernstein E, Hannon GJ, et al. (2004) RNAi and expression of retrotransposons MusERV-L and IAP in preimplantation mouse embryos. Dev Biol 269: 276–285.
54. Baillie GJ, van de Lagemaat LN, Mager DL (2004) Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. J Virol 78: 5784–5798.
55. Reus F, Frankel W, Moriwaki K, Shiroishi T, Coffin J (1996) Genetics of intracisternal-A-particle-related envelope-encoding proviral elements in mice. J Virol 70: 6450–6454.
56. Bru ˆ let P, Condamine H, Jacob F (1985) Spatial distribution of transcripts associated with intracisternal A particle transcripts. Genes Dev 8: 481–490.
57. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, et al. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297: 1855–1857.
58. Tohyama Y, Tsuchihashi M, Kume M, Kano T, Kuroda K, et al. (2004) Silencing of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.
59. McCormick J, Coffin J (1988) Polymorphism of murine endogenous proviruses revealed by using virus-class specific oligonucleotide probes. J Virol 62: 168–175.
60. Baillie GJ, van de Lagemaat LN, Baus C, Mager DL (2004) Multiple groups of endogenous betaretroviruses in mice, rats, and other mammals. J Virol 78: 5784–5798.
61. Best S, Le Tissier PR, Stoye JP (1997) Endogenous retroviruses and the evolution of resistance to retroviral infection. Trends Microbiol 5: 313–318.
62. Bieniasz PD (2003) Restriction factors: A defense against retroviral infection. Trends Microbiol 11: 286–291.
63. Tsuchihashi M, Kume M, Kano T, Kuroda K, et al. (2004) Silencing of transposable elements in heterochromatin and epigenetic control. Nature 430: 471–476.
64. McCormick J, Coffin J (1988) Polymorphism of murine endogenous proviruses revealed by using virus-class specific oligonucleotide probes. J Virol 62: 168–175.
65. Satin (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti dilute coat-color mutation of DBA mice. Proc Natl Acad Sci U S A 85: 189–192.
66. Lorincz MC, Dickerson DR, Schmitt M, Grouin M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nat Struct Mol Biol 11: 1066–1072.
67. Hatzioannou T, Perez-Caballero D, Yang A, Cowan S, Biasniz PD (2004) Retrovirus resistance factors RefI and LvI are species-specific variants of TRIM5alpha. Proc Natl Acad Sci U S A 101: 10774–10779.
68. Mack MJ, Hatziioannou T, Perez-Caballero D, Yang A, Cowan S, Biasniz PD (2004) Retrovirus resistance factors RefI and LvI are species-specific variants of TRIM5alpha. Proc Natl Acad Sci U S A 101: 10774–10779.
