Mouse germ line mutations due to retrotransposon insertions

Liane Gagnier1, Victoria P. Belancio2 and Dixie L. Mager1*

Abstract

Transposable element (TE) insertions are responsible for a significant fraction of spontaneous germ line mutations reported in inbred mouse strains. This major contribution of TEs to the mutational landscape in mouse contrasts with the situation in human, where their relative contribution as germ line insertional mutagens is much lower. In this focussed review, we provide comprehensive lists of TE-induced mouse mutations, discuss the different TE types involved in these insertional mutations and elaborate on particularly interesting cases. We also discuss differences and similarities between the mutational role of TEs in mice and humans.

Keywords: Endogenous retroviruses, Long terminal repeats, Long interspersed elements, Short interspersed elements, Germ line mutation, Inbred mice, Insertional mutagenesis, Transcriptional interference

Background

The mouse and human genomes harbor similar types of TEs that have been discussed in many reviews, to which we refer the reader for more in depth and general information [1–9]. In general, both human and mouse contain ancient families of DNA transposons, none currently active, which comprise 1–3% of these genomes as well as many families or groups of retrotransposons, which have caused all the TE insertional mutations in these species. As in humans [4], the mouse genome contains active retrotransposon families of long and short interspersed repeats (LINEs and SINEs) that can cause germ line mutations via new insertions but, in contrast to humans, the mouse also contains several groups of retrotranspositionally active endogenous retroviral elements (ERVs) that are responsible for most reported insertional mutations.

ERVs/LTR retrotransposons

ERVs are the result of retroviral infections or retrotranspositions in the germline. The general structure of an ERV is analogous to that of an integrated provirus, with flanking long terminal repeats (LTRs) containing the transcriptional regulatory signals, specifically enhancer, promoter and polyadenylation motifs and often a splice donor site [10, 11]. Sequences of full-length ERVs can encode gag, pol and sometimes env, although groups of LTR retrotransposons with little or no retroviral homology also exist [6–9]. While not the subject of this review, ERV LTRs can often act as cellular enhancers or promoters, creating chimeric transcripts with genes, and have been implicated in other regulatory functions [11–13]. The mouse genome contains many different groups of ERVs and related LTR retrotransposons that together comprise ~10% of the sequenced genome [1] and which have been characterized to varying extents [6, 9, 14, 15]. ERVs in mouse and other vertebrates are generally categorized into three classes. Class I ERVs are most related to the exogenous gamma-retroviral genus, Class II to beta- and alpha-retroviruses and Class III to spuma-retroviruses [6, 9]. The very large non-autonomous MaLR (mammalian apparent LTR retrotransposon) group is also considered Class III but has only very small traces of retroviral homology. Different mammals have distinct collections of ERVs and the mouse is unusual in having a much higher fraction of Class II elements compared to humans or other mammals [1, 6]. For all but very young groups, the majority of ERV loci exist only as solitary LTRs, the product of recombination between the 5′ and 3′ LTRs of integrated proviral forms [16, 17]. Moreover, for ERVs that have not undergone this recombination event, most have...
lost coding competence due to mutational degradation over time.

Unlike human ERVs that are likely no longer capable of autonomous retrotransposition [18, 19], some mouse ERVs are retrotranspositionally active and are significant ongoing genomic mutagens in inbred strains, causing 10–12% of all published germ line mutations via new integration events [1, 20]. The large Intracisternal A-particle (IAP) ERV group is responsible for close to half the reported mutations due to new ERV insertions, with the Early Transposon (ETn)/MusD ERV group also contributing substantially [20](Fig. 1a). These groups and other mutation-causing ERVs will be discussed in more detail in the subsequent relevant sections. The majority of mutagenic ERV insertions occur in introns and disrupt normal transcript processing (e.g. splicing and polyadenylation) to varying degrees, a mechanism well recognized since the 1990s [21–25] and discussed further below.

**Long interspersed repeats (LINEs)**

LINE-1s (L1s) are autonomous non-LTR elements that have accumulated to as many as 500,000 copies in both mouse and human genomes using a copy-and-paste mechanism of amplification [1–3, 26]. Full length L1s are 6–7 kb and contain two open reading frames (ORFs) encoding ORF1p and ORF2p, with the latter having endonuclease and reverse transcriptase activity [27–30]. The number of potentially active L1s (i.e. full-length elements containing intact ORFs) varies significantly between human and mouse. Bioinformatics analyses of the reference genomes have documented 2811 mouse and 146 human L1s that are fully structurally intact [31]. Functional studies have estimated numbers of active L1s to be ~ 3000 for mouse [32] and 80–100 for human [33]. In contrast to the human genome that has had a single subfamily of LINEs active at any given evolutionary time, the mouse genome contains three concurrently active L1 subfamilies (T(F), A, and G(F)) [32, 34] that are insertionally polymorphic among strains [17, 35]. One of the distinguishing features of these subfamilies is the differing 5′ monomer tandem repeats which, when combined with a downstream non-monomeric sequence, form their 5′ UTRs [36]. The 5′ UTR also contains the L1 pol II promoter, which occurs downstream of the transcriptional start site [37, 38], an arrangement common to non-LTR retrotransposons [39], allowing the promoter to be retained in the L1 mRNA.

Mouse and human L1s contain promoters, splice and polyadenylation signals in both sense and antisense directions that are utilized during L1 and host gene transcription, also sometimes leading to the formation of chimeric mRNAs [40–44]. As with ERVs [20, 45], such cis-acting sequences are a likely reason for the negative impact of some intronic L1 insertions on gene expression [43]. De novo L1 inserts can vary in size from just a few bases to those containing a full-length L1 sequence [26], with the vast majority of such inserts being 5′-truncated to varying extents. Although the exact mechanisms underlying this truncation phenomenon remain unclear, there is a positive correlation between the frequency of retrotransposition and insert length [46], and cellular DNA repair interference with L1 integration may play a role [47, 48].

Sporadically, new germ line L1 insertions cause mutations when they land in or near a gene in human [4] or mouse (discussed below), and somatic insertions can also occur, although few of the latter have been shown to exert a significant biological effect [49–51]. Mutagenic L1 inserts can potentially disrupt normal gene function or expression by interfering with it directly or by introducing deletions or complex genomic rearrangements that are sometimes associated with the integration process [3, 52]. In addition to introducing de novo insertions containing L1 sequences, L1 can mobilize flanking
genomic sequences as well. This occurs as a result of their incorporation into the nascent L1 mRNA generated by either inaccurate/upstream transcriptional initiation (5′ transduction) or inefficient transcriptional termination at the L1 3′ polyadenylation site resulting in readthrough and 3′ transduction [3, 53, 54]. Recent analysis of endogenous L1 expression in human cell lines determined that only about a third of expressed L1 loci generate such readthrough transcripts [55] but a similar analysis has not been performed for mouse. The uniqueness of these transduced sequences are often useful in identifying the source L1 element responsible for a newly retrotransposed copy [56].

**Short interspersed repeats (SINEs)**

SINE elements are non-autonomous retrotransposons, as they do not encode proteins involved in their amplification. As with human Alu SINE sequences [57], mouse SINEs have been shown to be retrotransposed by mouse L1 [58]. Only one of the two L1 proteins (ORF2p) is sufficient to drive Alu SINE mobilization in tissue culture [57], although ORF1p enhances the process [59]. Both mouse and human L1s can efficiently mobilize their non-orthologous SINEs, suggesting that such a symbiotic relationship has evolved multiple times [58–62]. There are several SINE classes in the mouse genome that together comprise ~8% of the genome [1]. Among these are B1, B2, B4/RSINE, ID, and MIR. New mutagenic insertions have been documented for B1 and B2 (see below), indicating that at least some copies are still potentially active. B1 (like human Alu) is derived from 7SL RNA, and B2 is derived from tRNA [3]. B1 and B2 SINEs are both present at very high genomic copy numbers: ~560,000 for B1 and ~350,000 for B2 [1]. Like mouse L1s and ERVs, these mouse SINEs are insertional polymorphic in inbred strains [17, 63, 64].

**Cataloging TE-induced mouse mutations**

We assembled lists of mutations caused by TEs by perusing the literature and by querying the Mouse Genome Informatics (MGI) database of mutant alleles [65]. In October 2018 we obtained lists from MGI of all spontaneous mutant alleles that listed “viral”, “transposon” or “insertion” as the cause and extracted all relevant cases through manual curation. To avoid ascertainment bias, we excluded cases where investigators were specifically screening for effects of insertionally polymorphic TEs [35, 66, 67]. While such cases can show effects on gene expression, observable phenotypes due to these insertionally polymorphic TE insertions were not reported in the aforementioned studies. In addition, we excluded cases where the insertion event likely occurred in cultured ES cells used to produce transgenic mice. Nearly all arose spontaneously but two cases of mutations occurring during a chemical mutagenesis experiment, but likely not caused by the chemical mutagen, were also included. This search resulted in a total of 115 TE insertion mutations. Ninety-four of these were caused by insertions of ERVs/LTR retroelements and 21 were L1 or L1-mediated (Fig. 1). In the case of the ERV mutations, the tables shown here are updates of previously published lists [1, 20, 68].

**IAP insertion mutations**

The group of ERVs responsible for the most reported mutations are the IAP elements. IAP sequences are Class II elements and are highly abundant in the mouse [6, 69]. Different estimates for IAP copy number exist in the literature but a recent analysis of all sequences annotated “IAP” by Repeatmasker [70] found ~3000 solitary LTRs and ~2800 full-length or partial full-length elements in the reference C57BL/6 genome [71]. Of the latter, ~1000 have 5′ and 3′ LTRs that are 100% identical, indicative of a very young age, and most of these belong to the IAPLTR1 or 1a subtypes [71]. As expected for such a young ERV group, IAP elements are highly insertionally polymorphic among inbred mouse strains [17, 66, 67, 72]. Although ~200 IAP sequences (IAPE elements) contain an env gene [73], most do not. Loss of env and other specific genetic modifications facilitated adoption of an intracellular retrotranspositional life cycle by IAPs [74] resulting in their accumulation to high copy numbers as genomic “super spreaders” [75]. Besides the lack of env, there are a few common partially deleted proviral forms [69] with the most notable being the 1Δ1 subtype, which has a 1.9 kb deletion removing part of gag and pol, resulting in an ORF encoding a novel gag-pol fusion protein. Although retrotransposition of 1Δ1 proviruses is non-autonomous, requiring gag and pol proteins in trans from other IAPs [76], this subtype is responsible for the great majority of new IAP insertion mutations [20]. Interestingly, it has been shown that the gag-pol fusion protein functions in cis to facilitate retrotransposition [77]. Together with a generally higher level of 1Δ1 transcripts compared to full length IAP mRNAs (see below), this cis effect could explain why most new insertions are of the 1Δ1 subtype.

Although transgenic experiments have shown expression of an IAP LTR only in the male germ line [78], endogenous IAP transcription is also detectable in embryogenesis as early as the two cell stage and appears highest in the morula and blastocyst stages [79]. Moreover, at least some IAP elements can be transcribed in normal somatic tissues, particularly in the thymus, where a specific subtype of IAP LTR shows transcriptional activity [80, 81]. Notably, the levels of 1Δ1 5.4 kb IAP transcripts are comparable or often more abundant than full-length IAP transcripts in different tissues or cell
types [69, 80, 82], although the former are present in lower copy numbers [69, 71, 83]. The molecular mechanisms underlying the generally higher transcript levels of 1Δ1 elements are unknown but one possibility is that these elements are more likely to escape the general epigenetic transcriptional repression of IAPs by DNA methylation and repressive histone modifications [84–87].

Table 1 lists mouse germ line mutations caused by insertions of IAPs. Somatic insertions of IAP elements can also occur and cause oncogene or cytokine gene activation in mouse plasmocytomas, myelomas and lymphomas [88–90], likely due to the fact that some IAP LTRs are transcriptionally active in lymphoid tissues [80, 81]. Most of the germ line insertions occur in gene introns and disrupt transcript processing, notably splicing and polyadenylation (Table 1) [20]. However, several IAP-induced mutations involve ectopic gene transcription promoted by an upstream or intronic inserted LTR that is regulated by DNA methylation [20, 91]. In these cases, the IAP is oriented in the opposite transcriptional direction with respect to the gene and it is an antisense promoter within the LTR that is responsible for the ectopic gene transcription. For a number of such cases, including the most well studied Aβ allele of agouti [92], variable establishment of epigenetic repressive marks on the IAP LTR result in variable expressivity of the mutant (IAP) allele in genetically identical mice and have been termed metastable epialleles [91, 93]. Interestingly, a recent genome-wide screen for other IAP metastable epialleles in C57BL/6 mice identified ~100 such loci, with an enrichment of flanking CTCF binding sites as the primary distinguishing feature [94].

### IAP activity in C3H mice

Because high numbers of IAP insertions in C3H mice and high IAP insertional polymorphisms among C3H substrains have been noted before [20, 83], we investigated the strain of origin for all TE-induced mutations. For IAPs, the strain of origin could not be ascertained for three of the 46 cases but, of the remaining 43, a remarkable 84% (36 cases) occurred in a C3H strain or hybrid involving C3H (Table 1, Fig. 1b). This marked skew is not seen for mutations caused by any other retroelements, indicating that ascertainment bias cannot explain the high frequency of IAP-caused mutations in C3H mice. While the date of the mutation is difficult to determine in some cases, IAP retrotranspositions in C3H mice have spanned several decades, with the earliest reported cases in the 1950s and the latest in 2014 (Table 1). This indicates that the unusual IAP activity has been a characteristic of C3H strains for at least 60 years. Indeed, Frankel et al. have shown that at least 26 1Δ1 IAP insertions present in C3H/HeJ are absent from the highly related C3HeB/FeJ substrain [83], again indicative of ongoing activity of IAPs, particularly the 1Δ1 subtype, in this strain.

Although reasons for the numerous IAP insertional mutations in C3H strains are unknown, it is noteworthy that normal spleen, bone marrow and thymus from C3H/HeJ mice have much higher levels of IAP transcripts compared to C57BL/6 and STS/A mice [95], suggesting that transcriptional deregulation may be involved. As well, IAPs are transcriptionally upregulated in radiation-induced acute myeloid leukemia in C3H/HeJ mice, resulting in new insertions in the leukemic cells, most of which are of the 1Δ1 subtype [95, 96]. These observations, coupled with the fact that most new mutations in C3H mice involve the 1Δ1 subtype suggests that this IAP subtype is accumulating in the C3H genome at a faster rate than full length elements.

Two recent reports illustrate the prudence of considering IAP induced mutations whenever working with C3H mice (Fig. 2). In the first case, Frankel et al. found that an IAP insertion in the Pcnx2 gene in C3H/HeJ mice (Pcnx2C3H/HeJ) reduces expression of this gene, which in turn mitigates the effect of an IAP insertion in Gria4 (Gria4pkkw1) which causes seizures [83]. Hence one IAP insertion modifies the effect of another (Fig. 2a). In another intriguing example, Barau et al. conducted a screen in C3HeB/FeJ mice using N-ethyl-N-nitrosourea (ENU) mutagenesis to identify genes involved in retrotransposon silencing in the germ line [97]. They identified several lines with the same mutation, indicating it was not induced by ENU but rather was spontaneous. This mutation was an IAP element inserted in an intron of a gene, annotated as a non-functional pseudogene, that formed as a tandem duplication of Dnmt3B. Barau et al. showed that this gene, now termed Dnmt3C, is indeed a functional DNA methyltransferase responsible for methylating promoters of young retroelements, including L1 elements and IAPs, in the male germ line [97]. Therefore, an IAP insertion facilitated the discovery of a gene involved in its own silencing (Fig. 2b).

### C3H mouse history

The C3H strain was derived by Leonard Strong from a 1920 cross of a Bagg albino female (ancestors to the BALB/c strain) and a male from Little’s strain of “dilute browns” (ancestors to the DBA strain) [98]. One of the original female progeny of this mating developed spontaneous mammary tumors and this trait was selected for or against by subsequent inbreeding to develop the C3H strain (highly susceptible to mammary tumors) and the CBA strain (highly resistant). Mouse mammary tumor virus (MMTV), the transmissible agent responsible for the early onset mammary tumors in C3H [99, 100], was later purged from most C3H related strains by pup fostering or re-derivation. In particular, the most widely
| Mutation | MGI ID | Strain of origin | Year mutation arose | Location & Orientation | Mutational mechanism(s) or effects | References |
|----------|--------|------------------|---------------------|------------------------|----------------------------------|------------|
| Ahy      | 1855945| C3H/HeJ          | ~1994               | 1C, c                  | IAP forms 5′-end of transcript, Regulation by methylation | [181]      |
| Aindy    | 1856403| C57BL/6J x C3H male | 1992              | Intron, –              | IAP forms 5′-end of transcript, Regulation by methylation | [182]      |
| aP       | 1855933| C3H/HeJ          | 1964               | Intron 1, –            | IAP forms 5′-end of transcript, transcript contains internal IAP sequence | [183]      |
| Ay       | 1855930| C3H/HeJ          | 1960               | Exon 1A, –             | IAP forms 5′-end of transcript, Regulation by methylation | [92, 183]  |
| Adams13   | 3579136| unknown, present in several strains | 7 | Intron 23, + | IAP causes premature polyadenylation and protein truncation. Reduced enzyme activity but no obvious phenotype | [184, 185] |
| Ap3d1mhn2 | 1856084| C3H/HeJ          | Before 1988        | Intron 21, +           | Transcript contains internal IAP sequence, Truncated protein | [186]      |
| Atockynyx | 1856898| C3H/HeJ-Tyr-c-a  | 1987               | Intron 1, +            | Transcript contains internal IAP sequence | [187]      |
| Atp2b2fpy | 3664103| DSO, derived from C3H/He-Pw hybrid | 2000 | Intron, + | Reduced mRNA levels, likely by terminating at poly(A) in IAP, also cryptic splicing suggested | [188]      |
| Atrfnoq   | 1856081| C3H– Swiss stock cross | 1950           | Intron 26, –           | Transcript contains internal IAP sequence, Truncated protein, Likely termination in IAP | [189]      |
| Atrfnoq-L | 2156486| C3H/HeJ          | 1981               | Intron 27, +           | Transcript contains internal IAP sequence, Reduced protein levels, Likely termination in IAP | [189]      |
| Axininfu  | 1856035| Bussey stock     | Before 1931        | Intron 6, –            | IAP forms 5′-end of transcript, Regulation by methylation, and transcript contains internal IAP sequence, Truncated protein | [190, 191]|
| Axininfu-ab | 1856037| unknown          | mid 1970s?       | Exon 7, –              | IAP forms 5′-end of transcript, transcript contains internal IAP sequence, Internally deleted protein | [190]      |
| Clcctm1j  | 5618134| C3H/HeSnJ        | After 1947        | Intron 2, +            | Transcripts contains internal IAP sequence, Reduced normal mRNA levels | [192]      |
| Crygeno3  | 3713105| C3H/HeH x 102/E1 F1 | Late 1990s     | Exon 3, +              | Reduced mRNA levels, Transcript contains internal IAP sequence, Truncated protein | [193]      |
| Dob1icm   | 1856801| Dc/Le (Dc arose in an obese stock outcrossed to BALB/c x C3H/He hybrid. One cross to C3H/HeJ, then inbred.) | 1991           | Intron, –              | Transcript contains internal IAP sequence, Truncated protein | [194, 195] |
| Dnmt3cnap | N      | C3H/HeJ/FeJ      | ~2014             | Last intron, –         | IAP provides alternate splice acceptor site, resulting in exclusion of Dnmt3c last exon, chimeric Dnmt3c-IAP mRNA | [97]       |
| Eya1fcr   | 1857803| C3H/HeB/FcJ      | 1984               | Intron 7, +            | Transcript contains internal IAP sequence, Reduced mRNA levels | [25]       |
| Gata3fcr  | 3027100| C3H/HeJ          | 1990s              | Intron 3, –            | Unknown | [196] |
| Gpr17fcrb5 | 5431477| C3H/HeJ          | After 1951        | Intron 1, +            | Drastically reduces gene expression | [197, 198] |
| Gria4skiw | 3580141| C3H/HeJ          | 1950–2002         | Intron 15, +           | Full-length protein expression is significantly reduced | [199]      |
| Gusbmq-2j  | 2152564| C3H/HeOuJ        | mid 1990s?       | Intron 8, +            | Reduced mRNA size and level, No enzyme activity detected | [200]      |
| Hps3fcr6b6 | 1861609| C3H/HeJ          | 1999?              | Exon 10, –             | Transcript contains internal IAP sequence | [201]      |
| Hps1fcr   | 1856712| C3H/HeB/FeJ      | 1957               | 3′-coding exon, –      | Transcript contains internal IAP sequence, Protein contains IAP-encoded amino acids | [202]      |
| Hps6fcr6b6 | 1856410| C3H/HeJ          | 1995               | Unknown, +             | Kidney: loss of expression, Brain: transcript contains IAP sequence | [203]      |
| Kcnq1fcr2b | 2389447| C3H/HeJcr-lf2f1m1Hxr | ~2000          | Exon 2, –              | IAP fusion transcript likely promoted by antisense LTR. IAP provides 5′ end of 31 bp fused in frame to the gene. | [204]      |
| Mutation   | MGI ID   | Strain of origin                      | Year mutation arose | Location & Orientation | Mutational mechanism(s) or effects                                                                 | References |
|------------|----------|---------------------------------------|---------------------|------------------------|-----------------------------------------------------------------------------------------------------|------------|
| LamB3IAP   | 2179716  | C3H (C3Hf/R1 male bred separately from JAX for decades) | ~ 1990              | Intron/exon junction (5′), − | Transcript contains internal IAP sequence, No mRNA or protein expression                           | [205]      |
| Mc1rmpc59H | 5791971  | C3H.Pde6b                             | 2010                | Exon,?                 | Disruption of single exon gene                                                                        | [206]      |
| Mgrn1md   | 1856070  | C3H/HeJ                              | ~ 1960              | Intron 11, +           | Reduced mRNA levels, aberrantly sized transcripts                                                    | [207, 208] |
| Mgrn1md-2J | 1856072  | C3H/HeJ                              | 1978–1993           | Exon 12, +              | Reduced mRNA levels, aberrantly sized transcripts                                                    | [207, 208] |
| Mgrn1md-5J | 1856519  | C3H/HeJ                              | 1978–1993           | Intron 2, +             | Not characterized                                                                                   | [207, 208] |
| Oprm1IAP   | N       | CXBK recomb. Inbred of B6 & Balb     | Before 1984         | Exon 4,?               | Causes truncated protein, LTR may act as antisense promoter?                                         | [209]      |
| PcnxZC3H/HeJ | 6161760  | C3H/HeJ                              | After 1950          | Intron 19, +            | Modifier of Gria4 mutation, reduces Pcnx2 expression                                                  | [83]       |
| Pitpna6   | 1856642  | DBA/2J                               | Early 1960s         | Intron 4, +             | Transcript contains internal IAP sequence, Reduced mRNA & protein                                    | [210]      |
| Pla2g6m1/1 | 4412026  | C3H/HeJ                              | mid 2000s           | Intron 1,?              | Reduces normal gene transcripts by ~ 90%                                                            | [211]      |
| Pld3Nab    | N       | C3H mix                               | ~ 2005              | Intron 2, –             | Causes truncated protein, LTR may act as antisense promoter?                                         | [212]      |
| PofutC3H/HeJ | 3719000  | C3H/HeJ                              | mid 2000s           | Intron 4,               | Reduced mRNA and protein levels                                                                       | [213]      |
| ReflC3H/HeJ | 1857345  | C3H/HeJ                              | ~ 1990              | Exon 36, –              | Exon skipping, Reduced mRNA levels                                                                   | [214]      |
| Spta1C3H/HeJ | 3802578  | C3H/He5n-Rab27a12+/JRos              | 1980–1990s          | Exon 1, –               | IAP fusion transcript likely promoted by antisense LTR. IAP provides 5′ end of 18 bp fused in frame to the gene. | [215]      |
| SptalC3H/HeJ | 2388936  | CcS3/Dem recomb. Con. strain from BALB/cHeA and STS/A | 1991                | Exon 10/ exon 11 junction, + | Exon skipping, Reduced protein expression, Reduced α-β-spectrin dimer and tetramer stability | [216]      |
| TalC3H/HeJ | 1859843  | C57BL/6J x C3HeBe/Fej/Le-a/a)F1.     | 1979                | Intron 4, +             | Promotes overexpression of exons 4 and 5, fusion transcripts found.                                  | [217]      |
| Tnfsf13C3H/HeJ | 2389403 | A/WySnJ                              | Before 1991          | Exon 3, +               | Fusion transcripts, Loss of function mutation                                                        | [218]      |
| TyrQTLC3H/HeJ | 2153728  | C3H/HeF1                              | 1988                | Upstream, –             | IAP does not form 5′-end of transcript, Reduced mRNA expression                                     | [219]      |
| Usp14C3H/HeJ | 1855959 | CBA or kreisler stock                | Early 1950s         | Intron 5, +             | Transcript contains internal IAP sequence, Reduced mRNA levels, Truncated protein                  | [220]      |
| Wnt9bC3H/HeJ | 1856821  | A                                   | 1920s               | Downstream, +           | Produces transcript antisense to Wnt9b.                                                              | [221, 222] |
| Zfp69C3H/HeJ | 2153728  | A/WySnJ                              | Before 1991          | Exon 3, +               | Fusion transcripts, Loss of function mutation                                                        | [218]      |
| Zfp69C3H/HeJ | 1855959 | CBA or kreisler stock                | Early 1950s         | Intron 5, +             | Transcript contains internal IAP sequence, Reduced mRNA levels, Truncated protein                  | [220]      |
| Zfp69C3H/HeJ | 1856821  | A                                   | 1920s               | Downstream, +           | Produces transcript antisense to Wnt9b.                                                              | [221, 222] |
| Zfp69C3H/HeJ | 1855959 | CBA or kreisler stock                | Early 1950s         | Intron 5, +             | Transcript contains internal IAP sequence, Reduced mRNA levels, Truncated protein                  | [220]      |

*Variable phenotype or expression (metastable epiallele)

aID number in Mouse Genome Informatics (MGI) database, "N" indicates not present in MGI

− antisense, + = sense, ? = orientation unknown

Ectopic expression of IAP-driven transcript
used C3H substrain C3H/HeJ was re-derived to be MMTV-free at the Jackson Laboratory (JAX) in 1999 and all C3H substrains carried at JAX have been free of MMTV since that time. Because IAP mutations have continued to occur in C3H/HeJ mice after removal of MMTV (Table 1), it is unlikely that activities of the two retroviral entities are directly related. Various substrains of C3H, including the commonly used C3H/HeJ, were derived in the late 1940s and early 1950s [101].

Interestingly, there is some evidence that C3H/HeJ has a higher spontaneous mutation rate than most other strains. A multi-year study conducted at JAX from 1963 to 1969 examined over 7 million mice derived from 28 inbred strains for spontaneous observable and heritable mutations [102]. C3H/HeJ had marginally the highest overall rate of mutations but not remarkably so [102]. However, this study also documented mutational cases of “irregular inheritance” where the trait was inheritable but showed very poor penetrance. Of the 35 examples of such cases, 16 (46%) arose in C3H/HeJ, even though this strain accounted for only 9.7% of the 7 million mice in the study [102]. It is tempting to speculate that at least some of these unusual cases may involve a new IAP insertion behaving as a metastable epiallele [91, 93].

ETn/MusD insertion mutations
After IAPs, the ETn/MusD group is responsible for the next highest number of germ line mutations, with 31 cases (Fig. 1, Table 2). ETn elements were first described as repetitive sequences expressed highly in early embryogenesis [103]. Subsequent expression analyses showed that ETNs are transcribed in two windows of embryonic development. First during E3.5–7.5 in the inner cell mass and epiblast and second between E8.5–11.5 in various tissues including the neural tube, olfactory/nasal processes and limb buds [103–105]. Although ETNs have LTRs, they have no coding capacity and, hence, their mode of retrotransposition was initially a mystery. Based on traces of retroviral homology in canonical ETNs, we identified anERV group, termed MusD, which is the likely progenitor of ETn [106, 107] and Ribet et al. demonstrated that coding competent MusD elements provide the machinery necessary for ETn elements to retrotranspose [108]. A subsequent phylogenetic analysis of the large betaretrovirus genus classified MusD as belonging to the Class II ERV-β group [14]. One analysis of copy numbers of ETn and MusD in C57BL/6 found ~240 ETn elements, ~100 MusDs and ~500 solitary LTRs [107], and they are highly insertionally polymorphic [17, 66, 109]. As for IAP elements, loss of the env gene and other genetic modifications likely resulted in genomic amplification of MusD (and ETn) elements as intracellular retrotransposons [110]. In another similarity to IAPs, most germ line mutations caused by ETn/MusD are due to insertions of the non-autonomous ETn (Table 2), in particular a specific subtype ETnII-β [20]. Of the 31 cases, only three are documented to be MusD while the rest are ETn (Table 2). The reasons for this are not clear but ETn transcripts are much more abundant than MusD transcripts in embryos and ES cells [107, 111] and there is evidence that MusD is subject to greater levels of epigenetic suppression [111, 112].

ETn/MusD mutations do not show an extreme strain bias as observed for IAP insertions. However, eight mutations have occurred in “A” strain mice (Fig. 1b), such as A/J, and two in the seldom used strain SELH/Bc (Table 2) which has a high incidence of exencephaly [113, 114]. Interestingly, genomic copy number estimates in different mouse strains revealed that, while there are no detectable differences in MusD numbers, A/J, SELH/Bc and CD-1 mice have two to three times more ETnII-β elements compared to C57BL/6 [107]. Transcript levels of MusD and ETnII-β in day 7.5 embryos are also higher in SELH/Bc and CD-1 compared to C57BL/6 [107].

Nearly all of the ETn mutagenic insertions occur in gene introns, in the same transcriptional direction as the gene, and disrupt normal transcript processing through utilization of canonical or cryptic signals within the ETn, notably a specific strong splice acceptor site in the LTR, coupled with either a downstream splice donor or polyadenylation signal [20, 45]. This extreme orientation bias for mutagenic insertions is also observed for the intronic IAP insertions that do not involve IAP promoter activity (Table 1). Such an orientation skew for detrimental
| Mutation          | MGI ID | Strain of origin      | Location & Orientation | Mutational mechanisms or effects                                                                 | References |
|------------------|--------|-----------------------|------------------------|-------------------------------------------------------------------------------------------------|------------|
| Adcy1brl         | 1857405| ICR                   | Intron 14, +           | Transcripts contain ETn sequence, Transcripts terminate within the ETn                           | [226, 227]|
| Bloc1s5nu        | 1856106| Stock-t               | Intron 3, +            | Transcripts contain internal ETn sequence                                                        | [228]      |
| Cacna1f1o0j      | 3605845| AXB6/PgnJ             | Exon 2, +              | Premature termination and alternative splicing within ETn                                       | [229, 230]|
| CaenG2⁵⁰        | 1856548| A/J                   | Intron 2, +            | Transcripts contain ETn sequence, Probable termination within ETn                               | [231, 232]|
| CaenG2⁵⁰-wag     | 1856386| MRL/MpJ-Fat¹pr/J      | Intron 1, +            | Transcripts contain ETn sequence, Probable termination within ETn                               | [231, 232]|
| CaenG2⁵⁰-3J      | 2155891| BALB/cJ               | Intron 2, +            | Transcripts contain internal ETn sequence                                                        | [231, 232]|
| Clcn1adr         | 1856316| A2G                   | Intron 12, +           | Transcripts contain ETn sequence, Termination within ETn                                       | [233]      |
| Dsg4Page         | 3766998| Recom. inbred of MRL/lpr x BXSB | Intron 8, +            | Aberrant splicing in ETn causes in frame additional exon of 63 amino acids                       | [234]      |
| Dysf⁵⁰         | 3055150| A/J                   | Intron 4, +            | ETn fixed in A/J strain. Aberrant splicing, No detectable dysferlin protein                      | [235]      |
| EdarDl-slk       | 1856017| Stock Tyrp1⁰ Dock7⁰ Rb(4.6)2Bnr | unknown                | Transcripts contain ETn sequence, Termination within ETn                                       | [236]      |
| Emfl₁rec       | 5560734| NOR-ICR               | Intron 22, +           | Aberrant splicing and premature termination, transcripts contain ETn sequence                   | [237, 238]|
| Etn3⁵⁰         | 3665247| CD-1                  | 1.6 kb 3’ of Dusp9 gene, – | Upregulates Dusp9 in ES cells, causes phenotype in Polypodia mutant                              | [120]      |
| Fas⁵⁰         | 1856334| MRL/Mp                | Intron 2, +            | Transcripts contain ETn sequence, Termination within ETn                                       | [21, 239, 240]|
| Fbxw4Dac        | 1857833| SM/Ckc                | 10 kb 5’ of exon 1, – | No evident difference in size or abundance of transcript, MusD insertion                       | [121–123]|
| Fbxw4Dac-2J     | 1857834| MRL/MpJ               | Intron 5, +            | Small amounts of normal transcript, MusD insertion disrupts a conserved motif in intron         | [121–123]|
| Fig4⁰          | 3716838| ‘mixed’ (129/Ola, C57BL/6J, C3H & SJL) | Intron 18, +          | Abnormal splicing from exon18 to SA site of EtnII, very low abundance of transcript           | [241]      |
| Fign³           | 1856870| Random bred           | Intron 2, +            | Transcripts contain ETn sequence, Probable termination within ETn                               | [242]      |
| Foxn1nu-Bc      | 1856110| SELH/Bc               | Intron, –              | Transcripts contain internal ETn sequence                                                        | [113]      |
| Gil3⁰           | 1856282| JclICR                | Intron 3, +            | 5 alternatively spliced transcripts contain ETn sequence, three terminate within ETn, two terminate as wild-type | [243]      |
| Hk1dea          | 2151848| A/J                   | Intron 4, +            | Probable aberrant splicing of ETn sequences into mRNA, Decreased hexokinase activity           | [244]      |
| Hsf4ldis1       | 3056560| RIIIS/J               | Intron 9, +            | ETn alters splicing resulting in a chimeric truncated protein, and causes at least 100 fold up-regulation of the transcript | [245]      |
| Lep⁰-2J         | 1858048| SM/Ckc-Fbxw4Dac       | Intron 1, +            | Transcripts contain ETn sequence, Probable termination within ETn, Loss of leptin expression.   | [23]       |
| MipCat-Fr       | 1857104| A/J                   | Intron 3, +            | Transcripts contain ETn sequence, Termination within ETn, LTR sequence translated to protein domain | [246, 247]|

Gagnier et al. Mobile DNA (2019) 10:15  Page 8 of 22
insertions is indeed expected, given that fixed/older ERVs have an antisense bias in genes [115, 116], presumably reflecting the fact that such insertions are less likely to be potentially deleterious and selected against.

In an attempt to mechanistically understand these orientation biases, we modeled splicing events involving intronic ERVs (using computationally predicted splice and polyadenylation motifs) and surprisingly found similar predicted frequencies of alternate splicing caused by sense or antisense ERVs [45]. However, actual splicing patterns of human mRNAs with intronic ERVs suggests that suppression of splicing within antisense-oriented ERVs occurs, possibly via steric hindrance due to annealing of sense-oriented ERV mRNAs [45]. This scenario would be analogous to gene therapy approaches where oligonucleotides that anneal to and suppress the use of mutagenic splice sites are used to redirect splicing and restore gene function [117]. Although unproven, such a mechanism could contribute to the general antisense bias for neutral/fixed ERV insertions and the opposite bias for mutagenic insertions.

Unlike for IAPs, there are no documented cases of ETn promoters causing a phenotype by driving ectopic gene expression (Table 2). This is likely due at least in part to the fact that ETn/MusD LTRs are normally only transcriptionally active in embryogenesis, responding to embryonic transcription factors [118, 119], so their promoter/enhancer activity would be silent in somatic tissues where most observable but non-lethal phenotypes manifest themselves. There is, however, at least one case where enhancer effects of an ETn insertion are likely responsible for a mutant phenotype. In this example, an ETn insertion downstream of the \( \text{Dusp9} \) gene upregulates this gene and also causes malformations in Polypodia mice, although a direct link between \( \text{Dusp9} \) deregulation and malformations has not been shown [120].

There is an intriguing but complex story involving two of the three documented MusD insertions [121–123]. Both of these cause the dactylaplasia (\( \text{Dac} \)) embryonic limb malformation phenotype by insertions within (\( \text{Fbxw4}^{\text{Dac-2J}} \)) or upstream (\( \text{Fbxw4}^{\text{Dac}} \)) of the \( \text{Fbxw4} \) gene. Both are full length MusD elements that share 99.6% identity and have occurred in different mouse strains. In the former case (\( \text{Fbxw4}^{\text{Dac-2J}} \)), the intronic, sense oriented MusD severely reduces the amount of normal \( \text{Fbxw4} \) transcripts, likely via typical transcript processing disruption or via physical disruption of a conserved, and hence potentially regulatory, ~ 1.5 kb region within the intron [123], although neither mechanism has been formally demonstrated. In the other \( \text{Dac} \) mutation (\( \text{Fbxw4}^{\text{Dac}} \), also termed \( \text{Dac}^{(J)} \)) the MusD is inserted 10 kb upstream of the \( \text{Fbxw4} \) gene in antisense orientation. However, no effects on the size or abundance of \( \text{Fbxw4} \) transcripts are evident in mice carrying this insertion, so the mechanism by which it causes dactylaplasia remains unclear [121–123].

Interestingly, the \( \text{Dac} \) phenotype is modified by an unlinked polymorphic locus \( \text{mdac} \) (modifier of

| Mutation | MGI IDa | Strain of origin | Location & Orientationb | Mutational mechanisms or effects | References |
|----------|---------|-----------------|------------------------|---------------------------------|------------|
| Rubie\(^{SWR/J}\) | 5568608 | SWR/J | Intron 1, + | Aberrant splicing and premature polyadenylation of the long non-coding RNA transcript Rubie which is proposed to regulate Bmp4 expression | [248] |
| Sgk3\(^{ica}\) | 3032837 | Swiss | Intron 6, + | Aberrant splicing, transcript contains ETn sequence | [249, 250] |
| \( \text{Sil}^{\text{mus}^1} \) | 3055918 | Recomb. inbred CXBS/ByJ | Intron 7, + | Aberrant splicing and premature termination. Transcript contains ETn sequence | [251] |
| S1k6a\(^{pol}^1,1\) | 5086232 | NOD.Cg-Env30\(^D\) Prkdc\(^scid/\)Dvs | Intron 5, + | MusD\(^*\) insertion, Inclusion of 183 bp of MusD sequence in mRNA, complete loss of protein | [252] |
| Ttc\(^{fln}\) | 1856879 | A/J | Intron 14, + | Transcripts contain internal ETn sequence | [253] |
| \( \text{Tgg}\) | 1857760 | A/J | splice donor of exon 7, – | 8 alternatively spliced transcripts, 4 contain ETn sequence, 1 terminates within ETn, 7 terminate as wild-type, 5 have exon skipping | [254, 255] |
| Try\(^{bc3}\) | 1856303 | SELH/Bc | Exon 1, – | Transcripts probably contain ETn sequence, Probable termination in ETn | [113] |
| Zha\(^{BAlb/c}\) | 3623885 | BALB/cJ | Intron 1, + | Transcripts contain ETn sequence, Termination within ETn | [256, 257] |

*aID number in Mouse Genome Informatics (MGI) database, “N” indicates not present in MGI

*b+ = antisense, + = sense, ? = orientation unknown

*MusD insertions are bolded. All others are ETns
dactylaplasia) [124]. In strains homozygous for the mdac allele (eg. BALB/c and A/J), the dactylaplasia phenotype is observed if the mice carry either dac mutation. However, in strains carrying the other allele Mdac (eg. CBA, C3H or C57BL), the phenotypic effects of the dac mutations are not observed [122, 124]. Although the identity of mdac is still unknown, it could be a gene involved in epigenetic regulation of MusD. In mdac/mdac mice, the 5′ LTR of the Dac1J MusD element is unmethylated and enriched in active histone marks whereas this LTR is heavily methylated and enriched in repressive histone marks in mice carrying the Mdac allele [122]. Moreover, ectopic MusD transcript expression is observed in embryos and limb buds of dactylaplasia mdac/mdac mice, but not in wildtype mdac/mdac mice, suggesting that the increased MusD expression is due to transcription of the Dac1J MusD element itself, rather than general up-regulation of MusDs in the genome [122].

### Table 3 MLV Insertions

| Mutation | MGI ID | Strain of origin | ERV location, orientation | Mutational mechanisms or effects | References |
|----------|--------|------------------|---------------------------|---------------------------------|------------|
| Abcb1αdms | 304239 | CF-1 | Solitary MLV LTR, Intron 22; − | Transcripts contain viral sequence, Exon skipping, Disruption of gene function. | [258, 259] |
| Aifm1Hq | 1861097 | CF-1 | Intron 1; + | 80% decrease in transcript and protein levels, Aberrant splicing | [260] |
| HrPV | 1856075 | HRS/J | Intron 6; + | Transcripts contain viral sequence, Probable termination within LTR | [24, 137] |
| Lamc2<sup>E1</sup> | 3609880 | 129X1/SvJ | Solitary MLV LTR, intron 18; + | Aberrant splicing and premature termination, low level of normal transcripts | [261] |
| Lmf1<sup>E1</sup> | 1856200 | M. m. musculus | Intron 7; ? | Transcripts contain ERV sequence, Termination within ERV | [262] |
| Myo5ad | 1856004 | Fancy mice | Intron; + | Shortened and abnormally spliced transcripts that vary among tissues | [135] |
| Pde6e<sup>E1</sup> | 1856373 | unknown | Intron 1; − | Associated with nonsense mutation in gene but effect of ERV is unclear | [263] |

<sup>a</sup>ID number in Mouse Genome Informatics (MGI) database

<sup>b</sup>− = antisense, + = sense, ? = orientation unknown

### Table 4 Other ERV Insertions

| Mutation | MGI ID | Strain of origin | ERV, Location, Orientation | Mutational mechanisms or effects | References |
|----------|--------|------------------|---------------------------|---------------------------------|------------|
| Agtbp1<sup>β-2/Δ</sup> | 1856536 | SM/J | ERV-β2/ETnERV3, Intron 13; + | Greatly reduced full length gene transcripts | [264] |
| Etn2<sup>β</sup> | 1857746 | Danforth’s posterior duplication stock | ERV-β2/ETnERV3, 12 kb upstream; + | Overexpression of Ptf1a and two neighboring genes, acts as an enhancer | [140–142] |
| Nax3<sup>β</sup> | 1856606 | GL/Le | Presumed ERV-β2/ETnERV3, Intron 12; + | Transcripts show aberrant splicing from Nax3 into retroviral element | [139, 265] |
| Prph2<sup>β</sup> | 1856523 | O20/A | ERV-β2/ETnERV3, Exon 2; − | Transcripts contain entire ERV, coding sequence disruption | [266, 267] |
| Ednrb<sup>α</sup> | 1856148 | Fancy mice | ERV-β-4, Intron 1; + | Aberrant splicing and premature polyadenylation | [268] |
| a | 1855937 | Fancy mice | VL30, Intron 1; −; ERV-β-4 within VL30; + | Smaller gene mRNA and levels 8 fold lower than in wild-type. Aberrant splicing and premature termination within the ERV-β-4 | [22, 143] |
| Dock<sup>α</sup> | 1856946 | JAX Dilute Brown | MERV-L, Exon 18; + | Frameshift and premature termination | [269, 270] |
| Fgfl<sup>β</sup> | 6147688 | ICR | Partial MTA MaLR, Intron 2; − | 498 bp insertion is combined with 9.3 kb deletion of exon 3 and flanking sequence, no detectable transcripts | [158] |
| Grm1<sup>α</sup> | 3604783 | BALB/c/Pas | MERV-L solitary LTR, intron 4; + | Aberrant splicing and premature termination, transcripts contain LTR sequence | [271] |
| Npc1<sup>α</sup> | 1857409 | BALB/c | Partial MERV-L, Exon 9 | 824 bp partial and rearranged MERV-L combined with 703 bp deletion. Transcripts contain MERV-L sequence, decreased expression, premature truncation | [159] |

<sup>a</sup>ID number in Mouse Genome Informatics (MGI) database

<sup>b</sup>− = antisense, + = sense, ? = orientation unknown
locus has been mapped to a 9.4 Mb region between markers D13Mit310 and D13Mit113 on chromosome 13 [122, 124]. Interestingly, this region includes a cluster of KRAB-ZFP (zinc finger protein) transcription factor genes. KRAB-ZFP genes are found in multiple clusters in the genome, are rapidly evolving and highly polymorphic in mice [125, 126] and some are involved in epigenetic silencing of ERVs [126]. Hence, it is tempting to speculate that the identity of mdac is such a gene.

MLV insertion mutations
The murine leukemia virus (MLV or MuLV) group is the most well characterized ERV group in the mouse and has caused seven documented spontaneous mutations (Fig. 1a,Table 3). MLV is also likely responsible for retrotransposing the non-autonomous VL30 ERV involved in the non-agouti mutation that will be discussed in the next section. MLVs are Class I elements, belonging to the gamma retrovirus genus, entered the mouse genome less than 1.5 million years ago and still contains infectious members [127]. MLV loci are highly insertionally polymorphic among strains [128, 129] with copy numbers of ~20 for xenotropic MLV and ~40 for polytropic MLV [9]. Ecotropic copies, i.e. those able to infect only mouse cells (and not those of other species) based on env protein recognition of a cellular receptor, are present in very few copies in various strains [127]. New germ line insertions appear to occur primarily through oocyte reinitiation, rather than intracellular retrotransposition [130], which has likely kept MLV copy numbers low. Ever since it was first reported that exogenous MLV can integrate into the germ line [131], MLV and MLV-based vectors have been widely used for many applications including insertional mutagenesis screens, gene therapy and oncogene discovery [132–134].

All of the MLV mutation-causing insertions occur in gene introns and affect normal gene transcript processing to varying degrees (Table 3). The very first ERV-induced mutation to be described, over 35 years ago, was an MLV insertion causing the dilute coat color mutation (Myo5adr) in DBA/2J mice [135]. This mutation can revert due to homologous recombination between the 5’ and 3’ LTR of the full length provirus, leaving a solitary LTR at the locus [136]. Phenotypic reversion by this mechanism also occurs for the hairless mutation (Hrhr), another of the first documented cases caused by an MLV insertion [137].

Insertional mutations by other class II ERVs
In addition to the ERVs discussed above, members of five other ERV groups have caused mouse mutations (Table 4). Like the IAP and ETn/MusD groups, two of the groups, ERV-β2 and ERV-β4, belong to Class II or the betaretrovirus genus as defined by pol homology [14]. Both of these groups are heterogeneous and relatively low in copy number. The ERV-β2 group includes mouse mammary tumor virus (MMTV) but the ERVs responsible for the four cases of mutations belong to a different ERV-β2 cluster which has internal sequences annotated in Repbase [138] primarily as “ETnERV3” with LTRs annotated as “RLTR13A” [14]. The full ERV was not sequenced for the Nox3het mutation but we presume it to be an ERV-β2 since the limited LTR sequence provided matched RLTR13A or RLTR13B [139]. For the other three ERV-β2 cases in Table 4, their full sequences have been published and they are 96–99% identical to each other with the major differences being internal deletions in the Agtpbp1prd3l and Prph2sd2 elements with respect to the longer Etn2sd ERV insertion (D. Mager, unpublished observations).

The above cases highlight the continual difficulties and confusion with ERV annotation. As an example, the ERV insertion causing the allele termed “Etn2sd”, where the ERV likely acts as an enhancer, was reported to be an “ETn” element [140–142]. However, as discussed above, this is misleading since “ETnERV3” is a separate entity compared to the more well-known ETn/MusD group, an important distinction but likely generally overlooked. Interestingly, when the reference C57BL/6 genome was analyzed in 2004, less than 15 ERV loci falling into the ERV-β2 group were found and none were fully coding competent [14]. Moreover, all of the ERV-β2s discussed above also lack full open reading frames. Nonetheless, the presence of these elements at sites of new mutations in other strains suggests such strains have or had coding-competent members to provide proteins in trans, allowing retrotransposition of defective elements. The strains in which the ERV-β2 mutations arose (Table 4) do not share close relationships so the origin of any active autonomous copies is unknown.

The ERV-β4 group [14] has been involved in two known mutations and both occurred in old “fancy mice” (Table 4). One of these mutations (Ednrbb) was caused by insertion of a 5 kb non-coding competent element whose internal sequence is classified as “ERV-β4_1B-I (internal)” in Repbase [138] but half of the sequence in the middle of the element actually lacks homology to retroviruses (unpublished observations). Fifteen to 20 sequences closely related to the Ednrbb element exist in the C57BL/6 reference genome and, since they contain LTRs and parts of the 5’ and 3’ internal sequences highly similar to the ERV-β4 element discussed below, it is likely that this small non-autonomous group has amplified using retroviral proteins provided by coding competent ERV-β4 elements.

The other mutation case involving an ERV-β4 is complex. The a (non-agouti) allele of the agouti gene is one of many agouti alleles affecting coat color [143], including four caused by IAP insertions (Table 1). The a allele
yses by Bulman et al. showed that the white-bellied agouti (\textit{agouti}) gene involving ERV insertions in the first intron of the \textit{agouti} gene was fixed in the reference strain C57BL/6. It involves a VL30 ERV and an ERV-\(\beta\)4 inserted within it. Partial phenotypic reversion of \textit{non-agouti} occurs frequently. The \(\alpha\) (black and tan) allele results from recombination between the LTRs of the ERV-\(\beta\)4. The \(\alpha'\) (white-bellied agouti) allele results from recombination between the VL30 LTRs. See text for references. Gene structure in black is shown to very rough scale. Green arrows and lines are the LTRs and internal VL30 sequences. Purple arrows and line depict the ERV-\(\beta\)4.

![Fig. 3 Three alleles of the agouti gene involving ERV insertions. The \(\alpha\) (non-agouti) mutant allele is fixed in the reference strain C57BL/6. It involves a VL30 ERV and an ERV-\(\beta\)4 inserted within it. Partial phenotypic reversion of non-agouti occurs frequently. The \(\alpha'\) (black and tan) allele results from recombination between the LTRs of the ERV-\(\beta\)4. The \(\alpha''\) (white-bellied agouti) allele results from recombination between the VL30 LTRs. See text for references. Gene structure in black is shown to very rough scale. Green arrows and lines are the LTRs and internal VL30 sequences. Purple arrows and line depict the ERV-\(\beta\)4.](image)

is fixed in the reference strain C57BL/6 and is responsible for its black coat color. Molecular characterization of \textit{non-agouti} in the early 1990s revealed that it was caused by a 5.5 kb VL30 ERV insertion in the first intron of the \textit{agouti} gene with another reported ~5.5 kb segment flanked by 526 bp direct repeats found within the VL30 [22, 143]. Our perusal of the fully sequenced reference C57BL/6 genome shows that the sequence within the VL30 is ~9.3 kb. The mutation is reported to be caused by a VL30, which belongs to a well-studied medium repetitive non-autonomous Class I ERV group that is co-packaged with MLV, allowing its retrotransposition [144, 145]. Although VL30 is insertionally polymorphic among inbred strains [17], this is the only reported VL30-caused mutation. The nature of the insertion within the VL30 was not known at the time of analysis, but the C57BL/6 sequence shows it to be an ERV-\(\beta\)4 (coordinates of the full ~14.7 kb VL30/ERV-\(\beta\)4 insertion are chr2:155014951–155,029,651, GRCm38/mm10). Hence two ERV insertion events contributed to the \textit{non-agouti} mutation, a VL30 insertion followed by insertion of an ERV-\(\beta\)4 within it (Fig. 3). The \textit{non-agouti} \(\alpha\) allele reverts at a high frequency to “black and tan” (\(\alpha'\)) or white-bellied agouti (\(\alpha''\)) [22, 143]. Molecular analyses by Bulman et al. showed that the \(\alpha''\) allele contains the VL30 element with a single ERV-\(\beta\)4 LTR and the \(\alpha''\) allele contain just one VL30 LTR [22](Fig. 3). Therefore, normal \textit{agouti} gene expression can be partly restored by homologous recombination between the LTRs of the VL30 or the ERV-\(\beta\)4, as has also been observed for MLV mutations (discussed above). Notably, the ERV-\(\beta\)4 element involved in the \textit{non-agouti} \(\alpha\) allele is the only fully coding competent ERV-\(\beta\)4 copy in the C57BL/6 genome [14].

**Insertions by MERV-L/MaLR elements**

The Class III MERV-L LTR retrotransposon group has also caused a few mutations (lower part of Table 4). MERV-L is a large, recently amplified group in the mouse with coding competent members but lacking an \textit{env} gene [146–148]. These retrotransposons are highly expressed in the 2-cell embryo [79, 149], create viral-like particles [150] and ~700 full length or near full length elements exist in the reference C57BL/6 genome [148]. Therefore, the fact that there are only three reported germ line mutations caused by MERV-L insertions is somewhat paradoxical. Despite the high transcript level and particle formation by MERV-L at the two cell stage, it appears that any fully retrotranspositionally competent members are very rare or effectively blocked from completing retrotransposition by host defense mechanisms. Indeed, MERV-L elements amplified in two major bursts in mouse evolution, approximately 2 and 10 million years ago [147] and it is possible that host genetic adaptations as a result of a host-virus “arms race” [151] have effectively repressed further MERV-L expansion. Interestingly, MERV-L and associated MT MaLR LTRs have been co-opted to drive expression of genes and other transcripts involved in early embryogenesis and zygotic genome activation [79, 152–154] and there is evidence that MERV-L expression is important for embryonic development [155].

Insertion of a partial MTA MaLR element, belonging to a large young group of non-autonomous retrotransposons related to MERV-L [15, 156], and also highly expressed in early embryogenesis [153, 157], has contributed to a mutation in the \textit{Fgf5} gene [158]. However, this case and the MERV-L insertion causing the \textit{Npc1}m\textsubscript{IN} mutation [159] are both partial elements and are coupled with genomic deletions, so the order of events resulting in these mutations is unclear. It is noteworthy that two of the four cases associated with Class III MERV-L/MaLR mutagenic insertions involve rearrangements of the ERV itself as well as genomic deletions. Interestingly, MaLR elements are associated with formation of independent hypervariable minisatellite sequence arrays in both human and mouse [160, 161], suggesting that these elements may foster genomic recombination and rearrangements.

**LINE1 insertion mutations**

Our literature and MGI database search resulted in a list of 12 germ line mutations caused by L1 insertions (Table 5, Fig. 1). Of the 11 where the length and/or sequence of the insertion was published, five are full length (or nearly full length) and six are partial elements, with the shortest being only 81 bp. All five full length insertions belong to the L1MdTF family, subtypes I or II, which are among the youngest L1 subfamilies, each with
over 1000 full length elements in C57BL/6 [34]. (Note that some revisions and updates to L1 subfamily classification and nomenclature have occurred [34]). In two cases, the source L1 element could be identified due to inclusion of flanking transduced sequence at the new insertion site. In the \( \text{Nr2e3rd7} \) mutant allele, the L1 insertion includes 28 bp of 5’ transduced sequence, allowing the source element to be traced to the L1 at chr4:21650298–21656544 (GRCm38/mm10) [162]. The other case (\( \text{Lama2}^{\text{dy-Pas}} \)) is interesting in that it involves an IAP LTR and an L1 [163]. While not reported as an L1 3’ transduction event in the original paper [163], our

| Mutation | MGI ID | Strain of origin | L1 length, location and orientation | Mutational mechanisms or effects | References |
|----------|--------|------------------|-------------------------------------|----------------------------------|------------|
| Atp7aMo-ca | 2387450 | unknown | 81 bp L1, exon 10, – | Aberrant splicing causing in-frame loss of 10 amino acids | [272] |
| Dab1yot | 1857750 | Chimera of 129 and C57BL/6 | Rearranged partial L1MdTf, intron | Partial gene deletion and atypical 962 bp L1 insertion | [273] |
| Frem1yot | 1856897 | AKRJ | Unknown length L1, exon 17, + | Transcript and protein truncation | [274] |
| Gbrbyot | 1856363 | Random-bred stock | Full length L1MdTf_J intron 6, – | Reduced expression of normal transcript and exon skipping | [275–277] |
| Lama2dy-Pas | 6220701 | Non-inbred Agouti | Nearly full length L1MdTf_I and 3’ transduction, intron 34, + | Transcript contains internal IAP sequence, truncated protein. L1-mediated retrotransposition event involving an L1 and 3’ transduced IAP LTR | [163] |
| Lyc2 | 1855968 | C3H/RI X 101/R1 (from radiation exp.) | 1.1 kb 5’ truncated L1, intron, + | Two alternatively spliced transcripts containing L1 sequence results in two truncated forms of the protein | [278, 279] |
| MifFybyot | 1856089 | C3H | Full length L1MdTf_I, intron 3, + | Decreased expression and exon skipping of two isoforms and abolished expression of the third isoform | [280] |
| Nr2e3rod | 1859180 | 77-2C2a-special- JAX | Full length L1MdTf_J with 28 bp 5’ transduction, exon 5, – | Accumulation of incompletely spliced transcripts. | [162] |
| PdeoCort | 2657247 | Recom. inbred CXB1/ByJ | 1.5 kb insertion of truncated L1 & likely 3’ transduction, Intron 4, – | Aberrant splicing within inserted sequence causes inclusion of extra 116 bp exon and frame shift. | [164] |
| RelyCort | 1856416 | BALB/c | Full length L1MdTf_J exon 59, + | 220 bp deletion in mRNA due to skipping of exon containing L1 | [281] |
| Scn8amed | 1856078 | PCT | 180 bp 5’ truncated L1, Exon 2, – | Skipping of exon containing L1, loss of functional protein | [282] |
| Ttnnda | 1856953 | C57BL/6J | 2.4 kb L1, exon 3, + | 779 bp gene deletion and L1 insertion causes exon loss or chimeric transcripts | [283] |

*ID number in Mouse Genome Informatics (MGI) database
\(-=\) antisense, + = sense, ? = orientation unknown

**Fig. 4** **a** Transduction of IAP LTR by an L1. A full length L1MdTf element interrupted by an IAP ERV exists in intron 3 of the Akr1c14 gene on chromosome 13. This L1 is the source element responsible for the \( \text{Lama2}^{\text{dy-Pas}} \) mutation, with the newly inserted sequence polyadenylated in the IAP LTR. Thick orange lines are L1 genomic sequences and thin orange lines represent L1 RNA. The IAP LTRs and internal sequences are in green. Genes and number of exons are not to scale. **b** B2 insertion causing gene upregulation. The \( \text{Tnf}^{\text{Rep/Rep}} \) mutation is a B2 insertion (in yellow) in the 3’ LTR of Tnf, causing Tnf upregulation due to polyadenylation within the B2 which removes the negative regulatory ARE (AU rich element) from the Tnf mRNA. Mice with this mutation have heart disease and arthritis due to overexpression of TNF. B2 is yellow and thicker black boxes are coding sequences.
perusal of the inserted sequence (Genbank accession AJ277888) revealed that the L1 has transduced the IAP LTR, with the inserted sequence polyadenylated within the 5’ LTR (Fig. 4a). The source L1 has a 3.7 kb partly deleted IAP element inserted within it, so that ~700 bp of the 3’ end of the L1 occurs on the other side of the IAP (coordinates of the source L1/IAP are chr13:4065522–4,076,041, GRCm38/mm10). Another L1 insertion (Pde6c<sup>ref</sup>), which occurred in a recombinant inbred strain established from a C57BL/6 and BALB/c intercross, has the classical molecular structure of a 3’ transduction event [164]. However, there is no L1 element in either the sequenced C57BL/6 or BALB/c genomes at the original location of the transduced sequence (unpublished observations), which occurs in an intron of the Diaph2 gene [164]. Therefore, the simplest explanation is that an L1 inserted in the Diaph2 gene in the particular mouse colony being used and then retrotransposed again, creating the Pde6c<sup>ref</sup> allele.

L1 insertions have occurred in a variety of genetic backgrounds, with no evident strain bias. The mutational effects of these insertions are as expected, with intronic L1s affecting splicing and exonic cases physically disrupting the coding sequence. Interestingly of the 12 L1 cases, half occur in gene exons and half in introns (Table 5), which is more skewed toward exons compared to the ERV insertions discussed above (Tables 1-4). It is a reasonable assumption that truncated (and hence shorter) L1 insertions might be less likely to affect transcript processing if inserted in an intron. (See also discussion of SINE insertions below). Indeed, the two shortest L1 insertions of 81 and 180 bp both occur in exons (Table 5). However, two of the five full length L1s, which are similar to size to ERVs, also occur in exons.

**Table 6** SINEs and other L1-mediated insertion mutations

| Mutation | MGI ID<sup>a</sup> | Strain of origin | Insertion type, location, and orientation<sup>b</sup> | Mutational mechanisms or effects | References |
|----------|-----------------|-----------------|------------------------------------------|-------------------------------|-------------|
| Atco2<sup>f</sup> | 1856574 | Bagg albino | B1, exon 4, + | B1 insertion causes protein truncation and rare exon skipping | [187, 284] |
| Comt<sup>At</sup> | 4819952 | Strain variant, likely origin in Lathrop stock | B2, 3’ UTR, + | Premature polyadenylation leads to shortened 3’UTR and increased protein expression. Associated with behavioral difference. | [170, 171] |
| Gsdma<sup>3rif</sup> | 2385837 | BALB/c | B2, exon 7, + | B2 causes premature termination, transcript includes B2 sequence | [285] |
| Ndufs4<sup>At</sup> | 4942335 | Male Dnmt3L +/− x female transgenic RIP-mtDNA C57BL/6 | B2, exon 3, + | B2 insertion causes exon skipping and premature termination | [286] |
| Ncam<sup>111</sup> | 5444298 | B6.129P2-Cnr2tm1Dgen/J | B2, exon 26, − | Aberrant splicing and premature termination | [287] |
| Poc1a<sup>v</sup> | 2135614 | DBA, gene cDNA, exon 8, − | L1 mediated insertion of Cenpw cDNA in exon 8. Exon 8 is skipped, reading frame preserved | [165] |
| Prtn6<sup>Av</sup> | 4947257 | BALB/c | B2, exon 6, + | Replacement of exon 6 sequence with B2 sequence, reading frame preserved | [288] |
| Slc27a4<sup>At</sup> | 445864 | B129X1/SvJ x 129S1/Sv | B2, exon 3, − | Coding sequence disruption, no mRNA or protein detected | [289] |
| Tnf<sup>b-prom</sup> | 579894 | C57BL/6 | B2, 3’UTR, + | Premature polyadenylation leads to shortened 3’UTR and gene overexpression | [172] |

<sup>a</sup>ID number in Mouse Genome Informatics (MGI) database  
<sup>b</sup>− = antisense, + = sense, ? = orientation unknown
Unlike the ERV mutation-causing insertions, where most cases occur in introns (Tables 1-4), all such mouse SINE insertions have occurred in exons (Table 6), which represent a much smaller genomic space. The marked bias toward exonic insertions also occurs for disease-causing Alus [4]. This could simply be due to the fact that SINEs are shorter and therefore new insertions are much less likely to significantly disrupt gene expression if inserted into an intron. Indeed, although SINEs, particularly Alus, can cause alternative splicing and exonization [168], both human and mouse SINEs are relatively enriched in introns [169] and show less evidence of selection against intronic insertions compared to ERVs or L1s [68].

As is the case for mutation-causing human Alu insertions [4], most of the mouse SINE insertions directly disrupt the gene’s coding sequence, causing exon skipping, protein ablation, truncations or amino acid replacements (Table 6). However, in the ComtB2i allele, which is a strain variant present in C57BL/6 and a few other strains [170, 171] and in the Tnfβpsm1 mutation [172], a B2 element inserted into the 3’ UTR causes gene upregulation, which modifies the phenotype. This effect is due to a shortened 3’ UTR caused by premature polyadenylation within the B2 and a resultant replacement or disruption of negative regulatory motifs within the UTR, which has been directly shown for Tnfβpsm1 [172] (Fig. 4b).

**Concluding remarks**

This review has provided a comprehensive catalog and discussion of mouse mutations caused by insertions of ERVs, LINEs and SINEs. It is clear that, among these TE types, ERV insertion mutations are the most prevalent (Fig. 1a). Through an accounting of all independent spontaneous mutant alleles in mouse, it was previously estimated that ERV insertions comprise 10–12% of all published spontaneous mutations [1, 20]. Another previous report estimated that L1 insertions account for 2–3% of mouse mutations [173], suggesting a relative ratio of ERV to L1 insertion mutations of 4 to 6. Our updated numbers (94 ERV cases and 12 L1 cases) reveal a somewhat higher ratio of approximately eight. If the nine SINE insertion cases reported here are included, the ratio of ERV to L1 insertion mutations is ~4.5. Since both human and mouse have active L1s, we can attempt to compare relative L1 recent “activity” based solely on the number of documented mutations due to L1 insertions. Both bioinformatics and functional studies [31–33] suggest that the typical inbred mouse genome harbors roughly 20–30 times more retrotranspositionally competent L1s compared to human (~3000 versus ~100–150). All else being equal, one might then expect the frequency of L1 insertional mutations to be 20–30 times higher in mouse. Recent reviews on retrotransposons in human disease report 22 cases of L1 insertions causing heritable mutations/diseases [4, 174]. To put these numbers in context, it should be remembered that many more mutations have been described in human compared to mouse. The Human Gene Mutation Database [175], lists ~240,000 entries as of January 2019. In contrast, the MGI database [65], lists only ~2100 spontaneous mutant alleles as of the same date, and many of these are non-independent entries or revertant cases. While comparing such overall numbers is fraught with caveats, they are however still useful to illustrate the point that the mouse “mutational space” is vastly understudied compared to human. Hence, the relatively low number of 12 mouse L1 mutations (when compared to the number of human L1 mutations) is not unexpected but rather simply appears low when viewed against the high numbers of ERV mutations. Indeed this number is approximately in line with expectations when compared to human, given the much higher number of active L1s but much lower numbers of all characterized mutations in mouse.

In considering L1-mediated insertion mutations as a fraction of all mutations, the numbers reported here suggest a frequency of 3–5% in mouse, building on the previous L1 estimate of 2–3% [173] and including the SINE cases. There have been various estimates for the frequency of L1-mediated mutations in human, with an early estimate of 1 in 600 (0.16%) reported by Kazazian [176]. A more recent study of the spectrum of mutations in a single gene found that TE insertions caused 0.4% of all mutations in NF-I [177], although it is unclear if this figure can be extrapolated to all genes. In any case, these estimates suggest that the contribution of L1 activity to overall mutational burden is at least 10 fold higher in mouse.

Concerning mouse ERVs, there are several distinct ERV groups currently able to retrotranspose at least in some strains, including the low copy number and poorly characterized ERV-β2 and ERV-β4 groups [14], previously not known to be active. Unpublished transcriptome analysis indicates that expression of both these groups is readily detectable in early embryonic stages (Julie Brind’Amour and Matt Lorincz, personal communication) but little else is known about them. The fact that new insertions have been found for such low copy number ERV groups indicates they are still mutagenic in some strains and worthy of further investigation.

Another point worth emphasizing is that, although IAP ERVs are young and have accumulated to high copy numbers in inbred strains, they perhaps do not deserve the often used designation as the currently “most active” group of mouse ERVs. This is likely true only in C3H mice and, if this strain is removed from consideration, a modest seven IAP-caused mutations can be documented...
to have occurred in strains unrelated to C3H (Table 1, Fig. 1b). This number of mutations places IAP recent “activity” more on a par with the low copy number MLV and ERV-β2 groups and suggests that the genomic expansion of IAPs in most strains has largely ceased, likely due to host defense mechanisms [86, 151, 178–180] gaining the upper hand. Exclusive of the C3H strain, the ETn/MusD group accounts for the most mutagenic ERV insertions. One possible reason for the high IAP-induced mutations in C3H mice could be a slight relaxation of repression in the germ line, so it would seem prudent for investigators to consider including this strain in studies to investigate the regulation of IAPs. This extreme strain bias for IAP activity also illustrates the difficulty in attempting to compare de novo TE insertion mutation rates in the “outbred” human population with those in the artificial environment of inbred mice. Nonetheless, the primary difference between human and mouse in terms of TE-induced insertional mutations is clearly the lack of ongoing ERV activity in modern humans.

Abbreviations
Dac: Dactylaplasia; ERV: Endogenous retrovirus; ETn: Early transposon; IAP: Intracisternal A type particle; JAX: The Jackson Laboratory; L1: LINE-1 family; LINE: Long interspersed element; LTR: Long terminal repeat; MalR: Mammalian apparent LTR retrotransposon; MLV: Murine leukemia virus; ORF: Open reading frame; SINE: Short interspersed element; TE: Transposable element

Acknowledgements
We thank Rita Rebollo for help with figures and for making suggestions on this review. We also thank Diana Juriloff for advice on mouse history and Jonathan Stoye for reading the MLV section. We apologize to authors if we failed to mention relevant work.

Funding
Relevant work in the Mager lab is currently funded by the Natural Sciences and Engineering Council of Canada. V. Belancio is funded by NIH grants R21AG055387 and R01AG057597 and the Brown Foundation.

Availability of data and materials
This is a review article. Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Authors’ contributions
LG and DLM compiled data and VLP and DLM wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable. This is a review article on mouse.

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1Terry Fox Laboratory, BC Cancer and Department of Medical Genetics, University of British Columbia, V5Z1L3, Vancouver, BC, Canada. 2Department of Structural and Cellular Biology, Tulane University School of Medicine, Tulane Cancer Center, Tulane Center for Aging, New Orleans, LA 70112, USA.
80. Metz JA, F ewell JW, Kulf EL. Selective activation of a discrete family of endogenous proviral elements in normal BALB/c lymphocytes. Mol Cell Biol. 1992;12(1):220–8.
81. Kulf EL, F ewell JW. Intracisternal A-particle gene expression in normal mouse thymus tissue: gene products and strain-related variability. Mol Cell Biol. 1985;5(10):474–83.
82. Ludeser KS, F ewell JW, Morozov VE, Kulf EL. Selective expression of intracisternal A-particle genes in established mouse plasmacytomas. Mol Cell Biol. 1993;13(12):7439–46.
83. Frankel WN, Mahaffey CL, McCarr TC, Beyer BJ, Letts VA. Unraveling genetic modifiers in the gria4 mouse model of absence epilepsy. PLoS Genet. 2014;10(7):e1004544.
84. Walsh CP, Chillett JR, Restor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 1998;20(2):116–7.
85. Maksaka JA, Mager DL, Reiss DI. Keeping active endogenous retroviral-like sequences in check: the epigenetic perspective. Cellular and Molecular Life Sciences (CMLS). 2006;63(21):3329–47.
86. Leung DC, Lorio NC. Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci. 2012;37(4):127–33.
87. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virolology. 2011;412(1):273–87.
88. Givol D. Activation of oncogenes by transposable elements. Biochem Soc Trans. 1985;82(7):2054–63.
89. Wang XY, Steelman LS, McCubrey JA. Abnormal activation of 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. 1997;3(1):3–19.
90. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene. 2008;27(3):404–5.
91. Rakytnik VK, Blewitt ME, Drucker R, Preis JI, Whitelaw E. Metastable epialleles in mammals. Trends Genet. 2002;18(7):348 –51.
92. Morgan HO, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet. 1999;23(3):314–8.
93. Dolinoy DC, Das R, Weidman JR, Little RJ. Metastable Epialleles, imprinting, and the fetal origins of adult diseases. Pediatr Res. 2007;61:30R.
94. Kazekchenko A, Bertozzi TM, Sjoberg-Herrera MK, Walker N, Gardner J, Gunnin R, et al. Identification, Characterization, and Heritability of Mutant Metastable Epialleles: Implications for Non-genetic Inheritance. Cell. 2018; 175(1):1259–71 e13.
95. Ishihara H, Tanaka I, Furuse K, Tsunekura K. Increased expression of intracisternal A-particle RNA in regenerated myeloid cells after X irradiation in C3H/Hej mice. J Radiat Res (Tokyo). 2004;45(1):25–32.
96. Ishihara H, Tanaka I, Wan H, Nojima K, Yoshida K. Retrotransposition of limited deletion type of intracisternal A-particle elements in the myeloid leukemia cells of C3H/He mice. J Radiat Res (Tokyo). 2004;45(1):25–32.
97. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science. 2016;354(6314):907–12.
98. Strong LC. The establishment of the C(3)H inbred strain of mice for the study of spontaneous carcinoma of the mammary gland. Genetics. 1935;20(6):586–91.
99. Comstand PW, Patrei KM, Owens KM, Villareal LM, Galbraith K, et al. Novel intracisternal A-particle RNA in regenerated myeloid cells after X irradiation in C3H/Hej mice. J Radiat Res (Tokyo). 2004;45(1):25–32.
100. Ishihara H, Tanaka I, Wan H, Nojima K, Yoshida K. Retrotransposition of limited deletion type of intracisternal A-particle elements in the myeloid leukemia cells of C3H/He mice. J Radiat Res (Tokyo). 2004;45(1):25–32.
101. Akeson EC, Donahue LR, Beamer WG, Shultz KL, Ackert-Bicknell C, Rosen CJ, et al. Chromosomal inversion discovered in C3H/HeJ mice. J Radiat Res (Tokyo). 2004;45(1):25–32.
102. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, et al. The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science. 2016;354(6314):907–12.
103. Krebs CJ, Larks KK, Khan SM, Robbins DM. Expansion and diversification of KRAB zinc-finger genes within a cluster including regulator of sex-limitation box zinc finger protein family. Mob DNA. 2015;6:17.
104. Steye JP, Coffin JM. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. J Virol. 1988;62(1):168–75.
105. Frankel WN, Stoye JP, Taylor BA, Coffin JM. A linkage map of endogenous murine leukemia proviruses. Genetics. 1990;124(2):211–36.
221. Juriloff DM, Harris MJ, Dewell SL, Brown CJ, Mager DL, Gagner L, et al. Investigations of the genomic region that contains the C1 mutation, a causal gene in multifocal cleft lip and palate in mice. Birth Defects Res A Clin Mol Teratol. 2005;73(2):103–13.

222. Juriloff DM, Harris MJ, Mager DL, Gagner L. Epigenetic mechanism causes Wnt9b deficiency and nonsyndromic cleft lip and palate in the a/Wysn mouse strain. Birth Defects Res A Clin Mol Teratol. 2014;100(10):772–88.

223. Schenecke S, Nesterling M, Vogel H, Blüher M, Block-M D, Díaz MB, et al. Positional cloning of zinc finger domain transcription factor Zpf69, a candidate gene for obesity-associated diabetes contributed by mouse locus Nidd/SJL. PLoS Genet. 2009;5(7):e1000541.

224. Jao Y, Jin X, Yan J, Jiao F, Li X, Roe BA, et al. An insertion of intracisternal A particle retrotransposon in a novel member of the phosphoglycerate mutase family in the lew allele of mutant mice. Genes Genet Syst. 2009;84(5):327–34.

225. Wang L, Jiao Y, Sun S, Jarrett HW, Sun D, Gu W. Gene network of a phosphoglycerate mutase in muscle wasting in cells. Cell Biol Int. 2015;39(6):666–77.

226. Abdel-Majid RM, Leong WL, Dobson MJ, Logsdon JJM, Abdel-Majid RM, Schalkwyk LC, et al. Gene expression in the mouse 6-day-old embryo at 21 days postconception. Genome Res. 2005;15(1):179–103.

227. Zhang G, Liu J, Wang L, Jiao Y, Sun S, Martin D, et al. Identification of an endogenous retrovirus in an apoptosis gene. J Exp Med. 1993;178(2):723.

228. Croquelois A, Giuliani F, Savary C, Kielar M, Amiot C, Schenk F, et al. The nob2 mouse, a null mutation in Cacna1f: anatomical and functional abnormalities in the outer retina and their consequences on retinal cell visual responses. Vis Neurosci. 2006;23(1):11–24.

229. Doering CJ, Rehak R, Bonfield S, Peloquin JB, Stell WK, Mema SC, et al. The mouse Clc1/myotonia gene: ETn insertion, a variable AATC repeat, Mutation in mice, connection between myotonia congenita and neurodegeneration in the pale tremor mouse and patients with CMT4J. Nature. 2007;448(7149):68–70.

230. Wu J, Zhou T, He J, Zou J, Mairitz JD. Autoimmune disease in mice due to integration of an endogenous retrovirus in an apoptosis gene. J Exp Med. 1993;178(2):461–8.

231. Chow CY, Zhang Y, Dowling JJ, Jin N, Adamska M, Shiga K, et al. Mutation of FkG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4A. Nature. 2007;448(7149):68–72.

232. Shiels A, Bassnett S. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. Nat Genet. 1996;12(2):212–5.

233. Shiels A, Mackay D, Bassnett S, Al-ghoul K, Kuszak J. Disruption of lens fiber cell architecture in mice expressing a chimeric AQP0-LTR protein. FASEB J. 2000;14(14):2207–12.

234. Roberts KA, Abraira VE, Tucker AF, Goodrich LV, Andrews NC. Mutation of Rubie, a novel long non-coding RNA located upstream of Bmp4, causes vestibular malformation in mice. PLoS One. 2012;7(1):e24945.

235. Mecklenburg L, Tobin DJ, Cilarn MV, Lucian C, Paus R. Premature termination of hair follicle morphogenesis and accelerated hair follicle cycling in lasi congenital ataxia (fis) mice points to fuzzy as a key element of hair cycle control. Exp Dermatol. 2005;14(8):561–70.

236. Campagna DR, Custodio ADO, Antochos BB, Cilarn MV, Fleming MD. Mutations in the serum/glucocorticoid regulated kinase 3 (Skg3) are responsible for the mouse fuzzy (fz) hair phenotype. J Invest Dermatol. 2008;128(3):730–2.

237. Zhao L, Longo-Guess C, Harris BS, Lee J-W, Ackerman SL. Protein accumulation and neurodegeneration in the wozzy mutant mouse is caused by disruption of SliL1, a cochaperone of BiP. Nat Genet. 2005;37:974–7.

238. Bogdanik LP, Chapman HD, Miers KE, Sereze DV, Burgess RW. A MusD retrotransposon insertion in the mouse Scol5A5 gene causes alterations in neuromuscular junction maturation and behavioral phenotypes. PLoS One. 2012;7(1):e20217.

239. White RA, McNulty SG, Nsunu NN, Boydston LA, Brewer BP, Shimizu K. Positional cloning of the Ttr7 gene required for normal iron homeostasis and mutated in hea and fsn anemia mice. Genomics. 2005;85(3):330–7.

240. Herrmann BG, Labeit S, Poutoua K, King TR, Lehrach H. Cloning of the T gene required for mesoderm formation in the mouse. Nature. 1990;346(629):67–72.

241. Goldin SN, Papaiakovou VE. Unusual misregulation of RNA splicing caused by insertion of a transposable element into the Tr (Brachyury) locus. BMC Genomics. 2005;6(1):14.

242. Perincheri S, Dingle RWC, Peterson ML, Spear BT. Hereditary persistence of -fetoprotein and H19 expression in liver of BALB/CJ mice is due to a retroviral insertion in the Zhs2 gene. Proc Natl Acad Sci. 2004;101(2):396–401.

243. Perincheri S, Peyton DK, Glenn M, Peterson ML, Spear BT. Characterization of the Entra-endoegenous retroviral element in the BALB/CJ Zhs2 Afl1 allele. Mamm Genome. 2007;19(1):26–31.

244. Jun K, Lee S-B, Shin H-S. Insertion of a retroviral solo long terminal repeat in BDR-3 locus disrupts mRNA splicing in mice. Incorporating Mouse Genome. 2000;11(10):843–8.

245. Pipert TR, Umbenhauer DR. The subpopulation of CF-1 mice deficient in P-glycoprotein contains a murine retroviral insertion in the Brachyury locus. BMC Genomics. 2005;6(1):14.

246. Pipped TR, Umbenhauer DR. The subpopulation of CF-1 mice deficient in P-glycoprotein contains a murine retroviral insertion in the Brachyury locus. BMC Genomics. 2005;6(1):14.

247. Pipped TR, Umbenhauer DR. The subpopulation of CF-1 mice deficient in P-glycoprotein contains a murine retroviral insertion in the Brachyury locus. BMC Genomics. 2005;6(1):14.
263. Bowes C, Li T, Frankel WN, Danciger M, Coffin JM, Applebury ML, et al. Localization of a retroviral element within the rd gene coding for the beta subunit of cGMP phosphodiesterase. Proc Natl Acad Sci U S A. 1993;90(7):2955–9.

264. Fernandez-Gonzalez A, Spada ARL, Treadaway J, Higdon JC, Harris BS, Sidran RL, et al. Purkinje cell degeneration (pcd) phenotypes caused by mutations in the Axotomy-induced gene, Nna1. Science. 2002;295(5561):1904–6.

265. Paffenholz R, Bergstrom RA, Pasutto F, Wabnitz P, Munroe RJ, Jagla W, et al. Vestibular defects in head-tilt mice result from mutations in Nox3, encoding an NADPH oxidase. Genes Dev. 2004;18(5):486–91.

266. Travis GH, Brennan MB, Danielson FE, Kozak CA, Sutcliffe JG. Identification of a photoreceptor-specific mRNA encoded by the gene responsible for retinal degeneration slow (rd8). Nature. 1989;338(6261):70–3.

267. Ma J, Norton JC, Allen AC, Burns JB, Hasel KW, Burns JL, et al. Retinal degeneration slow (rd8) in mouse results from simple insertion of a t haplotype-specific element into protein-coding exon II. Genomics. 1995;28(2):212–9.

268. Yamada T, Ohtani S, Sakurai T, Tsuji T, Kunieda T, Yanagisawa M. Reduced expression of the endothelin receptor type B gene in piebald mice caused by insertion of a retroprotein-like element in intron 1. J Biol Chem. 2006;281(16):10799–807.

269. Woolley GW. “misty” a new coat color dilution in the mouse, Mus musculus. J Biol Chem. 2012;287(37):31185–94.

270. Perou CM, Pryor RJ, Naas TP, Kaplan J. The bg allele mutation is due to a long interspersed nuclear element insertion. Nat Genet. 1994;7(2):136–40.

271. Kojima T, Nakajima K, Mikoshiba K. The disabled 1 gene is disrupted by a photoreceptor-specific mRNA encoded by the gene responsible for retinal degeneration slow (rd8). J Biol Chem. 2006;281(16):10799–807.

272. Cunliffe P, Reed V, Boyd Y. Intragenic deletions at Atp7a in mouse models of the spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA is associated with normal function. Proc Natl Acad Sci. 2009;106(8):2706–11.

273. Conti V, Aghaie A, Cilli M, Martin N, Catidi G, Musante L, et al. cvn4, a mouse model for human ataxia associated with kYPHOSCOLIOSIS caused by an mRNA splicing mutation of the metabotropic glutamate receptor 1 (Gmr1). Int J Mol Med. 2006;18(4):593–600.

274. Cunliffe P, Reed V, Boyd Y. Intragenic deletions at Atp7a in mouse models for Menkes disease. Genomics. 2001;74(2):155–62.

275. Kojima T, Nakajima K, Mikoshiba K. The disabled 1 gene is disrupted by a replacement with L1 fragment in yotari mice. Mol Brain Res. 2000;75(1):121–7.

276. Smyth I, Du X, Taylor MS, Justice MJ, Beutler B, Jackson IJ. The extracellular matrix gene Fenn1 is essential for the normal adhesion of the embryonic epidermis. Proc Natl Acad Sci U S A. 2004;101(37):13560–5.

277. Kingsmore SF, Giros B, Suh D, Bieniarz M, Caron MG, Seldin MF. Glycine receptor β-subunit gene mutation in spastic mouse associated with LINE-1 element insertion. Nat Genet. 1994;7(2):136–40.

278. Mulhardt C, Fischer M, Gass P, Simon-Chazottes D, Guenet JL, Kuhse J, et al. The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intrinsic insertion of L1 element. Neuron. 1994;13(4):1003–15.

279. Becker K, Braune M, Benderska N, Buratti E, Baralle F, Villmann C, et al. A Retroelement modifies pre-mRNA splicing: THE Murine Mediation of a retrotransposition event. Proc Natl Acad Sci. 2011;108(17):6974–9.

280. Becker K, Braune M, Benderska N, Buratti E, Baralle F, Villmann C, et al. A Retroelement modifies pre-mRNA splicing: THE Murine Mediation of a retrotransposition event. Proc Natl Acad Sci. 2011;108(17):6974–9.

281. Morelli KH, Seburn KL, Schroeder DG, Spaulding EL, Dionne LA, Cox GA, et al. Severity of demyelinating and axonal neuropathy mouse models is modified by genes affecting structure and function of peripheral nodes. Cell Rep. 2017;18(1):3178–91.

282. Nesterovitch AB, Szanto S, Gonda A, Bardos T, Kis-Toth K, Adarichev VA, et al. Spontaneous insertion of a B2 element in the ptpn6 gene drives a systemic autoinflammatory disease in mice resembling neutrophilic dermatosis in humans. Am J Pathol. 2011;178(4):1701–14.

283. Moulson CJ, Martin DR, Lugus JJ, Schaffer JE, Lind AC, Minner JR. Cloning of wrinkly-free, a previously uncharacterized mouse mutation, reveals crucial roles for fatty acid transport protein 4 in skin and hair development. Proc Natl Acad Sci U S A. 2003;100(9):5274–9.