REVIEW

Transposable elements in the mammalian germline: a comfortable niche or a deadly trap?

N Zamudio and D Bourc'his

Inserm U934/UMR3215, Institut Curie, Paris, France

Retrotransposable elements comprise around 50% of the mammalian genome. Their activity represents a constant threat to the host and has prompted the development of adaptive control mechanisms to protect genome architecture and function. To ensure their propagation, retrotransposons have to mobilize in cells destined for the next generation. Accordingly, these elements are particularly well suited to transcriptional networks associated with pluripotent and germinal states in mammals. The relaxation of epigenetic control that occurs in the early developing germline constitutes a dangerous window in which retrotransposons can escape from host restraint and massively expand. What could be observed as risky behavior may turn out to be an insidious strategy developed by germ cells to maintain genomic integrity but also certainly promoted the evolution of gene regulation.

sense retrotransposons and hold them back in check. Herein, we review recent insights that have provided a detailed picture of the defense mechanisms that concur toward retrotransposon silencing in mammalian genomes, and in particular in the germline. In this lineage, retrotransposons are hit at multiple stages of their life cycle, through transcriptional repression, RNA degradation and translational control. An organized cross-talk between PIWI-interacting small RNAs (piRNAs) and various nuclear and cytoplasmic accessories provides this potent and multi-layered response to retrotransposon unleashing in early germ cells.

Introduction

Genes have a relatively small contribution to the mass of mammalian genomes, compared with the overwhelming representation of clustered and interspersed repetitive sequences, in particular transposable elements (TEs) (Lander *et al.*, 2001; Waterston *et al.*, 2002). TEs are mobile DNA segments that recruit the cellular machinery for their replication, as well as encode their own specialized proteins to support their nomadic lifestyle. Mammalian genomes host almost three million of these jumping sequences, which have accumulated over evolutionary time and can act as a vehicle for genome size expansion. Although TEs account for <3% of the compact genome of the pufferfish (Fugu), they take over almost half of the genome mass in mice and humans. In addition, contrary to plants and *Drosophila* in which TEs have accumulated in heterochromatin blocks, TEs have mostly colonized euchromatic compartments in mammalian chromosomes, in which they are scattered around and within genes (Goodier and Kazazian, 2008). To tolerate such a level of invasion and the dangerous promiscuity with host genes, genomes have developed a variety of restraining strategies that tame the aggressive nature of TEs. The heterogeneity of TEs has forced the development of flexible and adaptable systems, which

maintain genomic integrity but also certainly promoted the evolution of gene regulation.

TEs are indeed vastly diverse in sequence, but can be categorized into two well-defined classes, according to their structure and movement strategies. Class II elements are made up of DNA transposons, which are mobilized by a 'cut and paste' mechanism. These account for roughly 3% of the human genome and are mostly inert fossils of ancient elements (Lander *et al.*, 2001). Class I elements are retroelements, which account for approximately half of the genome in mammals. A majority of retrotransposons are truncated and mutated, but some active representatives have retained the ability to duplicate themselves, in a 'copy and paste' process involving an RNA intermediate (Craig *et al.*, 2002; Goodier and Kazazian, 2008).

Several families of retrotransposons co-exist in the same genome. Their relative number and level of activity reflects varying evolutionary success in invading and surviving in their host. Competition for vital host factors may also have contributed to the dominance of specific elements (Furano *et al.*, 2004). Retrotransposons can be classified into two types, depending on whether or not they possess long terminal repeats (LTR). LTR-retrotransposons, also known as endogenous retroviruses (ERVs), are potential relics of infectious retroviruses that became *bona fide* endogenous genomic sequences after invading the host germline. In this respect, an ongoing phenomenon of retrovirus endogenization is currently observed in the koala genome (Tarlinton *et al.*, 2006). LTR-retrotransposons produce Gag and Pol proteins that enable them to reverse transcribe a copy of their RNA in the cytoplasm before integration into a new genomic

Correspondence: *Dr D Bourc'his, INSERM U934/UMR 3215, Institut Curie, Unité de Génétique et Biologie du Développement, 26 rue d'Ulm, 75248 Paris Cedex 05, France.*
E-mail: Deborah.Bourchis@curie.fr

location. However, without a functional *envelope* gene, they are defective for horizontal (cell to cell) transmission. LTR-retrotransposons make up approximately 8–10% of the human and mouse genomes (Lander *et al.*, 2001). Although human ERVs do not show signs of recent activity (Moyes *et al.*, 2007), mouse ERVs are still causing *de novo* deleterious mutations and provide a significant source of strain-specific insertional polymorphisms (Zhang *et al.*, 2008). In spite of the scarcity of elements with full coding potential, the retroviral-like intracisternal A particle (IAP) and the MusD/Early Transposon (ETn) families are still responsible for approximately 10% of spontaneous mutations in inbred mice (Maksakova *et al.*, 2006, 2008). Using *ex vivo* assays, a few transposition-competent elements have been shown to efficiently mobilize the impotent members of their own family *in trans*, providing an alternative strategy for *de novo* insertions (Ribet *et al.*, 2004).

Non-LTR elements are the predominant retrotransposons in mammals, and consist of two sub-types, the LINEs (long interspersed nucleotide elements) and the SINEs (short interspersed nucleotide elements). SINEs do not have any coding potential, but LINEs have self-propagating properties rendering them autonomous. Their 5′ UTR region functions as an RNA polymerase II promoter, and they encode two proteins that are necessary and sufficient for their activity, the RNA-chaperone ORF1 and the endonuclease/reverse transcriptase ORF2 protein. LINE retrotransposition relies on a well-described mechanism of 'target-site primed reverse transcription': the RNA is first exported to the cytoplasm to allow the production of LINE proteins, and then retranslocated into the nucleus where it undergoes complementary DNA conversion directly at the site of integration (Ostertag and Kazazian, 2001). The LINE-1 (L1) subclass represents some 17–20% of the total human and mouse genomes, consisting of more than 500 000 copies among which 80–100 elements are still capable of retrotransposition in humans (Brouha *et al.*, 2003). In contrast, several thousand L1 elements are still active in mice. L1 elements are considered as the master retrotransposons of mammalian genomes, not only by their overwhelming success in expansion during evolution, but also by their ability to mobilize *in trans* retrotransposons of the SINE class (Ostertag and Kazazian, 2001; Goodier and Kazazian, 2008). Mammalian SINEs have expanded to one million copies. The most prominent members are Alu repeats in humans and their B1 counterparts in mice, and both are derived from 7SL RNA, a functional component of the signal recognition particle.

In this study, we will review the effect that ancient and modern TEs have on genome evolution and function, and we will examine the nature of the interactions that exist between host genomes and their 'colonists', notably the cellular mechanisms that limit the mobilization and the life cycle of TEs. The germline represents a particularly relevant developmental context in which to study TE biology: TE propagation within a population strictly relies on a vertical mode of inheritance and therefore requires some activity in gamete precursors. A relaxation of epigenetic repression coincides with the acquisition of pluripotency of the early developing germline, which can be observed as an opportunity for TEs to constitute a massive force and invade the host genome. However, this apparent loss of control may rather be an insidious trick played by the germline to spot exuberant and loud TEs and force them to repression.

Good TEs, bad TEs

TEs can modify the landscape and the expression of the genome in a startling number of ways (Goodier and Kazazian, 2008). They can disrupt genes by direct insertional mutagenesis while hopping to a new genomic location. Homologous recombination between non-allelic repeats can cause deletions, duplications and other chromosomal rearrangements. New chimeric mRNA and proteins can be generated by exon shuffling, and by integration of 5′ or 3′ flanking sequences within the TE transcript. Mobilization of cellular mRNAs by the retrotransposon machinery can also generate new genes, known as processed pseudogenes, which may evolve new functions, divergent from the gene from which they originate. These various effects have participated in evolution and speciation. It has been reported that some 8 Mb of primate sequence has been lost in the human lineage merely as the result of target site deletions induced by TE insertions (Xing *et al.*, 2007). Similarly, bursts of pseudogene amplification often coincide with peaks of TE activity during evolution (Ohshima *et al.*, 2003). Beside these structural effects, TEs can also influence the expression pattern of nearby genes. TEs can provide strong constitutive promoters, alternative polyadenylation signals or splicing sites that can modify the cellular transcriptome. Repressive epigenetic marks associated with TEs can spread onto adjacent domains. Accordingly, LINEs have been linked to the acquisition of heterochromatic states, acting as boosting platforms for the spreading of inactivation on one of the X chromosomes in female mammals (Lyon, 2006), or attracting key epigenetic determinants for the emergence of neocentromeres (Chueh *et al.*, 2009).

TE presence and activity can have neutral, deleterious or beneficial outcomes for the host. On a long-term basis and on purifying selection, host genomes have managed to exploit some of their TEs, which are indeed ideally suited to provide raw material for the evolution of new proteins and regulatory sequences (Volff, 2006). TE-encoded proteins have innate roles in nucleic acid biology (binding, copying, breaking, joining, degrading, and so on) and in protein processing and interactions, which can be captured by the host to fulfill useful functions. An example of 'molecular domestication' of a TE-encoded sequence is provided by the RAG1 protein, which has likely emerged from a DNA transposase and is involved in immunity through V(D)J joining (Agrawal *et al.*, 1998). The centromeric protein CENPB is also similar to a transposase, originating from the Pogo superfamily (Smit and Riggs, 1996). Strikingly, a number of placenta-specific genes have appeared in the course of mammalian evolution through the recruitment of existing TE sequences, illustrating their central role in speciation. The placenta-essential *Rtl1* and *Peg10* genes evolved in mice from TE sequences of the Sushi-ichi class that are still present in the Fugu genome but are no longer active in mammals (Ono *et al.*, 2006). Coincidently, syncytin genes have occurred independently in different mammalian lineages from ERVs, and have an essential role in placenta development and function (Heidmann

et al., 2009). Finally, the widespread location and diversity of TEs can also be adopted by mammalian genomes for constituting a repertoire of regulatory sequences, such as binding sites for transcription factors. It is noteworthy that binding sites for the chromatin insulator CTCF are overrepresented in B2 SINE repeats, while motifs for the pluripotency factor Sox2 are found in ERV-Ks in mouse (Wei et al., 2006; Bourque et al., 2008). The weak conservation of transcription factor-binding sites and regulatory networks in mammalian genomes would actually fit with the fact that TE sequence and distribution are highly diverse among species. As a global illustration of the effect of TEs on genic composition and regulation, it was estimated that TE-derived sequences are found in the coding region of 4% of human genes, and are contained in 25% of human promoters (Nekrutenko et al., 2003; van de Lagemaat et al., 2003). TE activity can also indirectly generate beneficial novelties, by hijacking cellular mRNAs and catalyzing the insertion of their complementary DNA in new genomic environments. As an example, an advantageous fusion protein created through L1-induced mobilization of the cyclophilin A gene into the TRIM5 gene explains the resistance of the owl monkey to human immunodeficiency virus-1 infection (Sayah et al., 2004).

Although TE presence and activity can have a useful outcome for the host and provide opportunities for genetic innovation, diversification and speciation during mammalian evolution, it is evident that on a short-term basis, TEs impose a constant threat to the integrity of the genome. This is apparent by the 65 known human diseases solely caused by TE de novo insertions, in which Alu and L1 elements have a major role (Ostertag and Kazazian, 2001; Cordaux and Batzer, 2009). Alu repeats are also frequently involved in major chromosome rearrangements through ectopic recombination between dispersed elements, which have generated almost 50 known diseases. When occurring in the germline, these events can be transmitted to the next generation and lead to hereditary disorders. Since the first report of a case of hemophilia A resulting from an L1 insertion into the blood clotting factor VIII gene, the list of TE-induced congenital pathologies is constantly growing. In the context of somatic cells, cancer can develop, and cases of colorectal cancer have been linked to L1 insertion disrupting the tumor-suppressor gene adenomatous polyposis coli (Miki et al., 1992). A variety of mutant phenotypes caused by TE-induced insertions and rearrangements have also been documented in mouse, and include tumors, infertility and developmental pathologies (Bannert and Kurth, 2004). Independently of their activity, TE toxicity can occur in a retrotransposition-independent manner; overexpression of the L1 ORF2 protein single-handedly induces double-strand breaks and promotes apoptotic and senescence-like responses in cellular assays (Wallace et al., 2008). Finally, relaxation of TE silencing is a typical hallmark of cancerous and aging cells, in which it usually underlies a loss of genome-wide DNA methylation (Barbot et al., 2002; Schulz et al., 2006; Howard et al., 2008). TE hypomethylation and expression may facilitate genetic instability, DNA breaks and chromosome translocations. Whether TE reanimation hastens tumorigenesis and aging processes, or on the contrary triggers cellular responses such as apoptosis, is still under investigation.

Host responses to TEs or how to live with a herd of squatters

To limit the deleterious effects of TEs, mammalian cells have developed a multi-layered response that can affect the various stages of TE life cycle (Table 1). As TEs are greatly diverse in sequence, restricting pathways are quite universal and flexible and can be used for other cellular functions, usually related to general control of gene expression. Active TEs have an RNA-centered mode of replication and accordingly, a majority of defense mechanisms resemble innate immune pathways used for infectious retrovirus restriction. Currently known host responses target (a) TE transcription, (b) post-transcriptional processing of TE RNAs and (c) integration of new TE copies. As a result of these multiple hits, increased transcription and even production of TE accessory proteins may not necessarily convert into higher insertion rate. Although not being directly considered as a host defense strategy, it is also important to mention that cell division could be a strict requirement for TE activity. In vitro reporter assays notably showed that arrests at any stage of the cell cycle or senescent states strongly inhibit L1 retrotransposition (Kubo et al., 2006; Shi et al., 2007).

Transcriptional control: the role of epigenetic modifications

Transcriptional competence results from a combinatorial integration of cis-regulatory sequences, trans-acting factors and epigenetic layouts. To ensure the inheritance of newly retrotransposed elements, TEs have to be expressed in germ cells and their embryonic precursors. As discussed later, regulatory sequences associated with TE 5′LTR or 5′UTR promoters seem to be particularly suited for transcriptional machineries present in early embryos and the germline. However, the sole presence of these factors is not sufficient for TE expression, whose promoters are normally locked by DNA methylation and repressive chromatin states that render them inaccessible for transcription.

In mammals, DNA methylation has a key role in transcriptional silencing and has been suggested to have evolved for the specific purpose of defending the host genome against TE activity (Yoder et al., 1997). Mammalian DNA methylation targets cytosines involved almost exclusively in CpG dinucleotides, and methylated cytosines are in majority contained in TEs. DNA methylation not only causes immediate transcriptional repression of TEs by inducing a local repressive chromatin state, but can also promote permanent inactivation by C→T deamination naturally endured by methylated cytosines. Reflective of their heavily methylated status in the germline, accumulation of TG or TA conversions has eventually led to the immobilization of old TEs (Rollins et al., 2006).

Mutations in genes encoding proteins of the DNA methylation machinery or those facilitating methylation invariably result in TE reanimation and dramatically reduce viability or fertility. Mouse embryos lacking the maintenance DNA-methyltransferase Dnmt1 lose methylation of various types of TEs, massively reactivate IAP elements and die before mid-gestation around 8.5 d.p.c. (days post-coitum) (Walsh et al., 1998; Maksakova et al.,

Table 1 Known host repressors of TE activity in the mouse and human genomes

Gene	Biochemical Function /Role	Functional evidence for TE repression	TE target (level of reactivation)	References
Dnmt1	DNA methyltransferase / maintenance methylation	KO mouse Embryonic lethality	IAP+++	Walsh et al .(1998)
Dnmt3A	DNA methyltransferase / de novo methylation	KO mouse Male sterility/Neonatal lethality	L1+++, IAP+++	Okano et al. (1999) Kato et al. (2007)
Dnmt3L	Co-factor of Dnmt3A / de novo methylation	KO mouse Male sterility	L1+++, IAP+++	Bourc'his and Bestor (2004)
Lsh	Helicase / assists de novo methylation	KO mouse Embryonic lethality	IAP++	De la Fuente et al. (2006)
ESET	Histone methyltransferase / H3K9 methylation	KO mouse ES cells	IAP+++, MusD+++, L1+	Matsui et al. (2010)
Suv39	Histone methyltransferase / H3K9 methylation	KO mouse ES cells	IAP+++, L1+, SINEs+	Martens et al. (2005)
TRIM28/KAP1	Binds to zing finger proteins / transcriptional repressor	KO mouse and ES cells Embryonic lethality	IAP+++, L1+	Rowe et al. (2010)
PRC1/ PRC2	Polycomb repressor complex / histone modifications	Double KO mouse ES cells	IAP+++, MLV+++	Leeb et al. (2010)
Rb	Binds to proteins / tumor suppressor	KO mouse fibroblast cells	L1+++	Montoya-Durango et al. (2009)
Uhrf1	Binds to hemimethylated DNA / assists maintenance methylation	KO mouse Embryonic lethality	IAP+++	Sharif et al. (2007)
Miwi2	RNaseH (Piwi family) / piRNA biogenesis	KO mouse Male sterility	L1+++, IAP++	Carmell et al. (2007)
Mili	RNaseH (Piwi family) / piRNA biogenesis	KO mouse Male sterility	L1+++, IAP++	Aravin et al. (2008)
Maelstrom	Unknown / supports piRNA pathway	KO mouse Male sterility	L1+++, IAP+	Soper et al. (2008)
Tdrd1	Tudor protein, binds to MILI / supports piRNA pathway	KO mouse Male sterility	L1+++	Reuter et al. (2009)
Tdrd9	Tudor protein, binds to MIWI2 / supports piRNA pathway	KO mouse Male sterility	L1+++	Shoji et al. (2009)
GASZ	Unknown / supports piRNA biogenesis	KO mouse Male sterility	L1 +++, IAP+	Ma et al. (2009)
ADAR	Adenosine deaminases / RNA editing	Cellular assays	Alu, L1, SINEs	Nishikura (2006)
APOBEC3	Cytosine deaminases / RNA editing	Cellular assays	IAP, Alu, L1, MusD, HERV-K	Bogerd et al. (2006) Esnault et al. (2008)
AID	Cytosine deaminase / RNA editing	Cellular assays	L1, MusD	MacDuff et al. (2009)
ERCC/XPG	Endonuclease / nucleotide excision pathway	Cellular assays	L1	Gasior et al. (2008)
Tex19.1	Unknown / unknown	KO mouse Male sterility	MMERVK10C+++	Ollinger et al. (2008)

Abbreviations: HERV, human endogenous retrovirus; IAP, intracisternal A particle; KO, knockout; L1, long interspersed nucleotide element-1, MLV, murine leukemia virus; piRNA, PIWI-interacting RNA; SINEs, short interspersed nucleotide elements; TE, transposable element. Green = genes repressing TEs at the transcriptional level. Purple = genes involved in TE repression at both transcriptional and post-transcriptional levels. Blue = genes involved in repression at post-transcriptional level. '+' indicates the level of reactivation in various mouse mutants.

2008). Combined inactivation of *de novo* DNA-methyltransferases Dnmt3A and Dnmt3B, which co-operate in the establishment of methylation patterns in early embryos, phenotypically mimics the *Dnmt1* knockout phenotype (Okano *et al.*, 1999). In the male germline, Dnmt3A strictly requires its co-factor Dnmt3L to methylate L1 and IAP repeats. Failure to achieve this process leads to a high level of expression of these elements in germ cells and complete sterility of *Dnmt3L* or *Dnmt3A* mutant males (Bourc'his and Bestor, 2004; Kato *et al.*, 2007). Although catalytically inactive, biochemical assays have revealed that Dnmt3L facilitates *de novo* methylation by stabilizing the active conformation of Dnmt3A, this therefore allows a more efficient transfer of methyl groups onto target sequences (Chedin *et al.*, 2002; Jia *et al.*, 2007). It should be noted that Dnmt3L is not only functionally but also evolutionarily linked to the protection of the germline against TEs: Dnmt3L specifically emerged in eutherian mammals, some 150 million years ago, coinciding with an important TE expansion in

mammalian genomes (Yokomine *et al.*, 2006; Warren *et al.*, 2008).

In addition to members of the DNA-methyltransferase family, proteins that assist the DNA methylation reaction also have a role in TE transcriptional repression. Among them, Lsh (lymphoid-specific helicase) is a member of the SNF2 family of chromatin remodeling ATPases that facilitates the access of Dnmts to the DNA substrate (Zhu *et al.*, 2006). Inactivation of *Lsh* in mice results in decreased methylation and increased expression of IAP elements in female germ cells and in embryonic tissues, which leads to early post-natal lethality (Huang *et al.*, 2004; De La Fuente *et al.*, 2006). To ensure the clonal propagation of methylation patterns upon cellular divisions, the maintenance DNA-methyltransferase Dnmt1 requires UHRF1 to load onto hemi-methylated DNA strands generated by replication (Sharif *et al.*, 2007). *Uhrf1* mutant mouse embryos phenocopy the *Dnmt1* mutation, showing hypomethylation at various TEs and reactivation of IAP elements.

DNA methylation and chromatin remodeling concur in the formation of condensed heterochromatic states, and specific repressive histone tail modifications have been associated with TE promoters. Alu repeats are enriched in H3K9 methylation in human somatic cells (Kondo and Issa, 2003), and IAP elements are targeted by both H3K9 and H4K20 trimethylation in mouse ES cells (Martens et al., 2005; Mikkelsen et al., 2007). As a result of the diversity and redundancy of histone-modifying enzymes, the importance of these marks is difficult to assess functionally. Decreases in H3K9 methylation levels through inactivation of H3K9 methyltransferases have variable outcomes on TE silencing in mouse ES cells: inactivation of ESET leads to a robust reactivation of a range of LTR-retrotransposons (Matsui et al., 2010), Suv39h deficiency induces a modest IAP reactivation (Martens et al., 2005), while inactivation of G9a has no effect (Dong et al., 2008). Reduction in H4K20 methylation through Suv420h1 and Suv420h2 deficiency does not affect TE expression in ES cells (Matsui et al., 2010). Combined loss of H3K9 and H4K20 trimethylation can induce the reactivation of L1 elements, as was observed in mouse fibroblasts lacking the heterochromatin-associated retinoblastoma protein (Montoya-Durango et al., 2009). Finally, the repressive mark H3K27 trimethylation probably has a role in silencing TEs, as combined inactivation of the polycomb complexes PRC1 and PRC2 leads to an upregulation of LTR-sequences in ES cells (Leeb et al., 2010).

Post-transcriptional silencing: playing with the RNA
Altering the genetic information through RNA editing: The term RNA editing refers to molecular processes that modify the information content of an RNA molecule. RNA-editing enzymes can target a variety of nucleosides in RNA transcripts, most often by deamination. By altering cellular RNAs and therefore the amino-acid sequence of encoded proteins, RNA editing has a large role in expanding the genome capacity, as in the case of immunoglobulin class switching. RNA-editing proteins can also act against invading RNA particles, and in mammals, these proteins are involved in innate immune responses against infectious RNA viruses. Accordingly, mutant mouse models often show immunity issues and increased susceptibility to viral infection (Muramatsu et al., 2000). RNA editases are also potent inhibitors of endogenous TE activity when overexpressed in cellular assays, but presently, very little is known about this role during development in vivo.

The ADAR family of RNA-editing enzymes converts adenosine residues into inosines in regions of double-stranded RNAs. Analysis of the human transcriptome has revealed that ADARs target double-stranded RNAs that are formed from inverted Alu and L1 repeats. In mouse, SINEs are undergoing the same conversion (Kim et al., 2004; Eisenberg et al., 2005; Nishikura, 2006). The APOBEC proteins form another family that catalyzes the deamination of cytosine residues into uracils and have greatly expanded in the primate lineage. APO-BEC3G notably reduces the replication of human immunodeficiency virus, by inducing the accumulation of uracil mutations on the nascent retroviral complementary DNA strand and subsequently inactivating the newly integrated copy (Bishop et al., 2004). Retrotransposition assays have shown that APOBEC3A, 3B, 3C and 3F enzymes are potent restrictors of different classes of LTR- and non-LTR retrotransposons, such as L1, IAP, Alu, human ERV-K and MusD elements in human and mouse cells (Bogerd et al., 2006; Esnault et al., 2006, 2008). AID, another cytosine deaminase with a wider phylogenetic distribution, also represses L1 and MusD retrotransposition in mouse cells (MacDuff et al., 2009). Unexpectedly, TE restriction triggered by AID and some of the APOBEC3 proteins does not involve C to U hypermutation. It has been hypothesized that these enzymes mediate cytoplasmic sequestration of L1 RNA and/or L1-encoded proteins, or directly inhibit L1 ORFs (Stenglein and Harris, 2006; Beauregard et al., 2008). Considering the master role of L1 in retrotransposition biology, this would not only affect L1 activity but could also render the L1 machinery inaccessible to non-autonomous TEs of other classes. Alternatively, AID-like enzymes may also regulate TE activity at the nuclear level (Popp et al., 2010).

Degrading TE transcripts through RNA interference: RNA interference represents another post-transcriptional mechanism of TE suppression (Obbard et al., 2009). In this case, small RNAs operate through homology-based recognition to induce the degradation of complementary TE transcripts, through the recruitment of the RNA-induced silencing complex. The catalytic component of RNA-induced silencing complex is made up of Argonaute proteins that can bind single-stranded nucleic acids and slice them through their RNaseH-like activity. One class of TE-associated small RNAs are the small interfering RNAs (siRNAs), which are 21–23 nt long double-stranded RNAs produced by the DICER-dependent cleavage of hybrid molecules naturally formed by sense and antisense retrotranscripts. The siRNA pathway is a potent mechanism of transposon silencing in Plants, fungi and Drosophila (Brodersen and Voinnet, 2006). In human cell lines, transfected exogenous siRNAs can limit L1 expression (Soifer et al., 2005). But while TE-associated endogenous siRNAs have been reported in mammals, their formal involvement in host defense against TE invasion is still lacking. Their presence might be incidental, reflecting by-products of DICER activity, which is strictly required for the biogenesis of other small RNA effectors, the microRNAs. Accordingly, Dicer knock-out greatly impairs microRNA-directed gene silencing in oocytes, but does not lead to TE upregulation (Hayashi et al., 2008). Nevertheless, the oocyte actively produces siRNAs containing L1, IAP and Mouse Transcripts (MT) sequences (Tam et al., 2008; Watanabe et al., 2008). Interestingly, these siRNAs seem to be mostly required for degrading cellular genes bearing TE repeats in their 3′ UTR regions but not necessarily TEs themselves (Murchison et al., 2007). This suggests that the female germline may use the siRNA pathway to regulate the expression level of a particular class of genes through their TE-related sequences. Abundant LTR transcripts are also specific to early embryos and ES cells, and are found in both sense and antisense orientation (Peaston et al., 2004; Svoboda et al., 2004). These may potentially form double-stranded substrates for DICER and trigger an RNA interference response.

Another class of small RNAs has recently been linked to RNA interference-based TE repression. The piRNAs (PIWI-interacting RNAs) are single-stranded RNAs that are processed independently of DICER and are slightly longer than siRNAs (24–30 nt). They are loaded onto specific members of the Argonaute proteins, the PIWI proteins. PIWI proteins are strictly restricted to the germline and have a wider phylogenetic distribution than conventional Argonautes (Grimson *et al.*, 2008). As for the siRNA/Argonaute complex, the piRNA/PIWI interaction forms a recognition and processing machinery for degrading target transcripts. The mouse genome encodes three PIWI proteins (four in humans) and among them, MILI and MIWI2 are involved in host defense. *Mili* or *Miwi2* mouse mutant males are deficient in TE-associated piRNAs and consequently fail to degrade L1 and IAP transcripts that accumulate in their germ cells, thus leading to sterility (Carmell *et al.*, 2007; Aravin *et al.*, 2007b). The striking similarity with the *Dnmt3L* mutant phenotype, including a failure to properly methylate TE sequences, indicates that the piRNA pathway not only acts as a post-transcriptional suppressor but also controls the transcriptional output of TE elements (Aravin *et al.*, 2008; Kuramochi-Miyagawa *et al.*, 2008). This dual role is reminiscent of RNA-directed DNA methylation originally described in plants (Brodersen and Voinnet, 2006). A number of recent studies have greatly expanded our understanding of the interplay between homology-dependent piRNA and DNA methylation pathways in the mammalian germline and have provided a scenario of events that we will discuss later.

Blocking the integration

Finally, the last stage of the TE life cycle, the integration of the complementary DNA copy into the genome, is also subject to host defense. The ERCC1/XPF heterodimer complex has endonuclease activity and is involved in DNA repair mostly through the nucleotide excision repair pathway. Reduction of XPF in human cell lines increases L1 retrotransposition, suggesting that intermediates of the target-site primed reverse transcription process may be cleaved by this enzyme (Gasior *et al.*, 2008). Interestingly, other DNA repair enzymes have an inverse effect on retrotransposition: the double-strand break repair protein ATM is indeed facilitating L1 integration (Gasior *et al.*, 2006), indicating that various DNA repair pathways may be able to recognize and process TE integration intermediates, in a positive or negative manner.

Learning from infectious retroviruses

The recent characterization of TRIpartite Motif (TRIM) proteins as important mediators of innate immunity against retroviruses has opened up a new field of potential suppressors of endogenous retrotransposons. TRIM proteins can form high molecular-mass complexes that localize to specific subcellular compartments present in the cytoplasm and the nucleus and can target retroviruses at different stages of their cycle (Ozato *et al.*, 2008). TRIM proteins have greatly expanded in the mammalian lineage and have evolved species-specific functions particularly well adapted to defend the host against both exogenous and intragenomic parasites.

TRIM28/KAP1 (KRAB-associated protein1) is a transcriptional repressor responsible for the intrinsic resistance of ES cells to infectious retroviruses, in particular against Moloney murine leukemia virus. The zinc-finger protein ZFP809 bridges TRIM28 to a specific sequence used by the integrated proviral DNA to initiate its synthesis, named the primer-binding site (Wolf and Goff, 2009). TRIM28 induces the heterochromatinization of this sequence by recruiting H3K9 di- and trimethylation, HP1 and the NuRD histone deacetylase complex (Sripathy *et al.*, 2006). It was recently shown that TRIM28 also binds to the 5′ sequence of endogenous LTR-retrotransposons in ES cells and early embryos, and induces their repression in conjunction with ESET-dependent H3K9 trimethylation (Matsui *et al.*, 2010; Rowe *et al.*, 2010). Interestingly, although reactivated IAP sequences lose H3K9 trimethylation and gain H4 acetylation, they still harbor a normal level of DNA methylation in *TRIM28*-deficient cells. This would suggest that TRIM28 either acts downstream of DNA methylation or functions in a DNA-methylation-independent manner. Another member, TRIM22, probably functions to reduce Gag protein stability and could also target TE-encoded proteins (Ozato *et al.*, 2008). A last potential candidate for TE restriction is the zinc-finger anti-viral protein (ZAP), which does not belong to the TRIM family. ZAP promotes the degradation of viral mRNAs by interacting with the exosome, and may also prevent the accumulation of TE ribonucleoparticles in the cytoplasm (Zhu and Gao, 2008). Future studies should be directed toward understanding the importance of these various anti-retroviral proteins in suppressing endogenous TEs, and in particular in the germline in which TE-induced genetic modifications can irreversibly shape the host genome.

TE silencing in the germline

The male germline as a biological niche for TE biology

TE survival in the host organism requires their repression in somatic cells not to risk harming the host, while their obligate mode of dissemination through vertical transmission requires some level of activity and mobilization in the germline. In mammals, retrotransposons are particularly active in germ cells and in early embryos before the emergence of the germline, suggesting an adaptive strategy to transcriptional networks associated with germinal and pluripotent states. The MT family of LTR retrotransposons accounts for 13% of the transcriptome of the mature mouse oocyte, although these sequences comprise <5% of the genome (Peaston *et al.*, 2004). ETn elements were originally identified by their specific expression during early embryogenesis (Maksakova *et al.*, 2008). L1 transcripts and proteins are naturally found in male germ cells entering meiosis and in preimplantation embryos, while they appear to be completely repressed in differentiated somatic tissues (Branciforte and Martin, 1994). Moreover, L1 transcripts may be competent for integration in the early mouse embryo after being carried over from the gametes through fertilization (Kano *et al.*, 2009). It is in fact increasingly suspected that most of L1 mobilization may occur during early embryonic development in humans, as evidenced by the

retrotransposition of a transgenic L1 element observed in human embryonic stem cells (Garcia-Perez *et al.*, 2007). Somatic and germinal mosaicisms of L1 insertions observed in human pathology also support this notion (van den Hurk *et al.*, 2007).

The propensity of the germline and early embryos to express TEs is not only linked to the availability of key transcription factors, but also to a relaxation of epigenetic control in these cells. Indeed, genome-wide loss of DNA methylation accompanies the acquisition of pluripotent states in primordial germ cells and preimplantation embryos, which opens a window of opportunity for TEs to escape from host restraint (Rougier *et al.*, 1998; Hajkova *et al.*, 2002). The struggle between TEs and their host is expected to be the most obvious in these natural sites, and indeed, recent studies have highlighted the defense strategies specifically developed by the germline to hold TEs back in check. However, the sensitivity to TE repression is subject to a very strong sexual dimorphism, which likely reflects the different developmental kinetics of the male and female germlines.

In males, individual inactivation of components of the germline arsenal against TEs invariably leads to sterility (Figure 1). Mutant animals do not produce any mature gametes, a condition known as azoospermia. TE reactivation is associated with a high rate of illegitimate pairing between non-homologous chromosomes at meiosis, which triggers an apoptotic checkpoint and ends spermatogenesis. TE reactivation also perturbs the self-renewal program of spermatogonial stem cells, which results in spermatogenic tubules completely devoid of any germ cells within a few weeks. This phenotype was originally described for the methylation-deficient *Dnmt3L* mutant males (Bourc'his and Bestor, 2004), observed again in *PIWI* mutants (Carmell *et al.*, 2007; Aravin *et al.*, 2007b) and more recently recapitulated in mutations for proteins supporting the piRNA/PIWI pathway, such as the Tudor domain-containing proteins TDRD1 and TDRD9 (Ollinger *et al.*, 2008; Soper *et al.*, 2008; Ma *et al.*, 2009; Reuter *et al.*, 2009; Shoji *et al.*, 2009). Tudor domain-containing proteins assemble specialized RNA processing platforms in the cytoplasm of germ cells, to which they recruit PIWI proteins through their ability to bind symmetrically dimethylated arginines, a post-translational modification undergone by both MILI and MIWI2 (Aravin *et al.*, 2009; Vagin *et al.*, 2009; Wang *et al.*, 2009; Shoji *et al.*, 2009). As a whole, directed mutagenesis in mice has allowed the identification of nine proteins with a key role in safe guarding the male germline against TEs (Table 1 and Figure 1): Dnmt3L, Dnmt3A, MIWI2, MILI, Maelstrom (MAEL), TDRD1, TDRD9, GASZ and Tex19.1 (Testis expressed 19.1) (Ollinger *et al.*, 2008; Soper *et al.*, 2008; Ma *et al.*, 2009). All of these proteins, except Tex19.1, have been involved in the same TE restricting pathway, initiated by
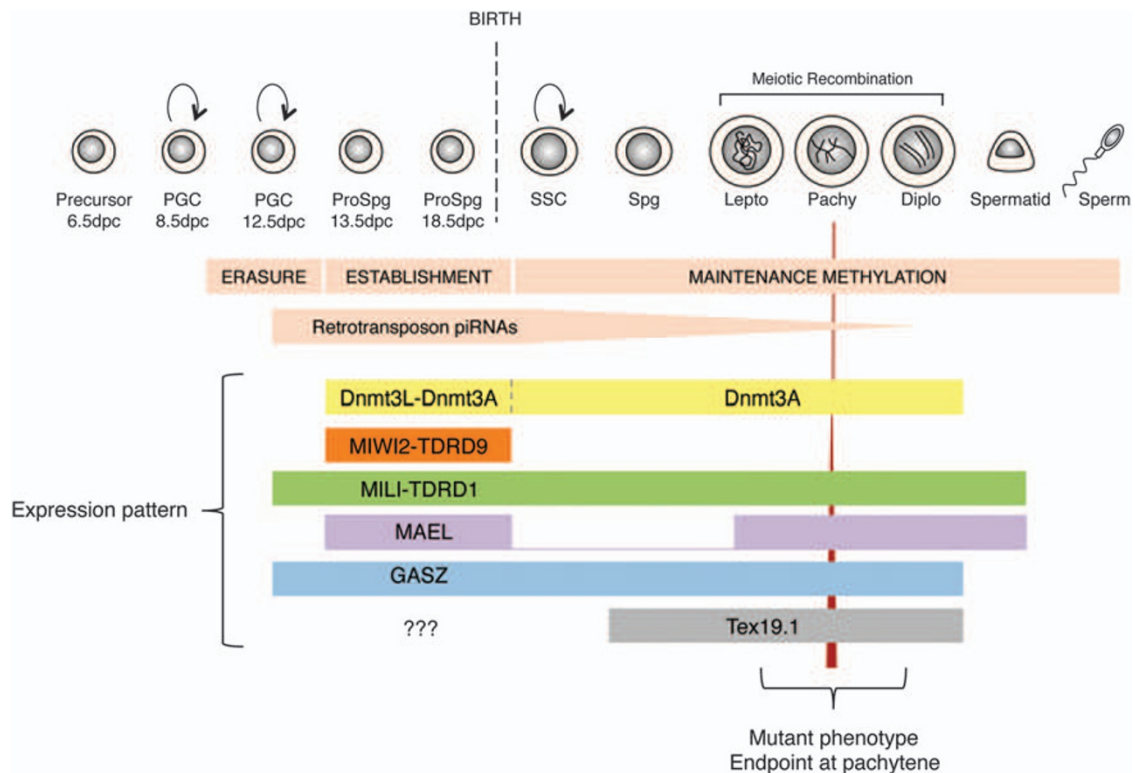

Figure 1 Expression pattern and mutant phenotype of proteins involved in TE repression in the developing male germline in mice. The germline genome gets heavily demethylated as primordial germ cells (PGCs) colonize the genital ridges around 10.5 d.p.c. After sex determination, male PGCs become prospermatogonia (ProSpg) and get remethylated from 13.5 d.p.c. to birth, at the formation of spermatogonial stem cells (SSC). This period coincides with an accumulation of TE-derived piRNAs and a peak of expression of *de novo* DNA-methyltransferases Dnmt3L and Dnmt3A, PIWI proteins MILI and MIWI2 and their respective associated Tudor proteins TDRD1 and TDRD9. Other proteins with an important role in supporting piRNA biogenesis are also present in ProSpg, Maelstrom (MAEL) and GASZ. Some of these proteins are expressed at later stages of spermatogenesis. Mutations in these genes results in male sterility, with male germ cells ending their progression at the pachytene stage.

PIWI-dependent TE transcript cleavage and ending with DNA methylation-induced transcriptional repression. These proteins not only protect fertility in the short term, but also prevent reduction of fitness in the long term, by limiting the accumulation of heritable TE-induced germline mutations.

The situation in the female germline differs greatly to that in males. Females lacking any of the components mentioned above do not significantly reactivate TEs in their germline. It is noteworthy that *Mili* and *Miwi2* mutant females are fertile even when harboring combined mutations (Carmell *et al.*, 2007; Watanabe *et al.*, 2008; Shoji *et al.*, 2009). So what controls TE expression in oocytes? As discussed earlier, the DICER-dependent siRNAs may constitute an alternative or redundant degradation system to the piRNA pathway for controlling TE expression in oocytes (Tam *et al.*, 2008; Watanabe *et al.*, 2008). However, *Dicer* deficiency barely affects TE expression levels (Murchison *et al.*, 2007; Watanabe *et al.*, 2008). Moreover, TEs are naturally hypomethylated and expressed in mature oocytes (Peaston *et al.*, 2004; Lucifero *et al.*, 2007), suggesting that these cells are apt to tolerate the presence of TE transcripts. Active cell division is an important requirement for TE transposition, and the long meiotic arrest endured by oocytes may constitute an innate barrier to the completion of the full TE cycle (Shi *et al.*, 2007). The stringency of TE control is therefore not as primordial as for the male germline, which relies on life-long dividing cells, the spermatogonial stem cells (Aravin and Bourc'his, 2008). This hypothesis is strongly supported by the absence of such a sexual dimorphism in *Drosophila*, a species in which both male and female germlines rely on a pool of dividing stem cells. In this case, the two sexes are equally susceptible to TE reactivation in their germlines and share common defense mechanisms (Brennecke *et al.*, 2007; Aravin *et al.*, 2007a). The only known case of potential TE reactivation in mammalian oocytes is reported for female mice bearing mutations in the *Lsh* helicase gene, a regulator of DNA methylation (De La Fuente *et al.*, 2006). However, this effect could be secondary to a more general perturbation of embryonic development and the effect on fertility is not known.

Developmental scenario of TE regulation in the male germline

In mammals, the germline is set aside from the rest of the embryo at around 6.5 d.p.c. in mice. After a phase of migration, the primordial germ cells colonize the genital ridges around 10.5 d.p.c. Unknown signals emanating from the future gonads trigger a genome-wide erasure of methylation patterns (Figure 1), as part of the epigenetic reprogramming leading to primordial germ cell pluripotency and resetting of genomic imprinting (Trasler, 2006). L1 elements get severely demethylated at this stage, while IAP elements conserve some residual methylation (Hajkova *et al.*, 2002). This lack of methylation-dependent control lasts for a few days in male germ cells. Commitment to a male-specific program occurs at 12.5 d.p.c. in response to the expression of the sex-determining *Sry* gene. Subsequent establishment of male-specific methylation patterns starts at 13.5 d.p.c. to be fully completed at birth (20.5 d.p.c.), at the time of

appearance of spermatogonial stem cells (Kato *et al.*, 2007). This window of TE *de novo* methylation coincides with a quiescent period of development, in which germ cells are in G0/G1 arrest and take the name of prospermatogonia (Figure 1). Expression of Dnmt3L, Dnmt3A, the PIWI proteins MILI and MIWI2, as well as TDRD, MAEL and GASZ proteins specifically peaks during this period.

Although epigenetic relaxation of TE control may primarily be observed as suicidal for the host, TE reanimation is actually leading to their own fall in the male germline. Methylated and repressed TEs live undercover in the host genome. Their demethylation in germ cell precursors reveals their existence to the host, which can deploy an ancestral PIWI-based recognition and destruction machinery to maintain them in hold. The piRNA pathway is deeply conserved among metazoans, and is invariably involved, from invertebrates to vertebrates, in protecting genomic integrity in the germline by restricting TE transcripts (Aravin *et al.*, 2007a). Contrary to flies and worms, which rely on the piRNA pathway but are devoid of genomic methylation, the mammalian germline combines both piRNA-induced degradation and DNA methylation to silence TEs. Their epistatic relationship explains why mutations in genes required for piRNA production, such as *Mili*, *Miwi2*, *Tdrd1*, *Tdrd9* and *Gasz*, also affect the process of *de novo* methylation of TE elements in prospermatogonia (Carmell *et al.*, 2007; Aravin *et al.*, 2007b; Ma *et al.*, 2009; Reuter *et al.*, 2009; Shoji *et al.*, 2009).

Cellular, biochemical and molecular analyses have evidenced the elaborate spatio-temporal control of TE expression occurring in the developing male germline (Figures 1 and 2). This mammalian-specific scenario is slightly different from the piRNA-based suppressing pathway described in *Drosophila* (Brennecke *et al.*, 2007). After methylation erasure in primordial germ cells, TE retrotranscripts are potentially expressed, sensed by PIWI proteins and cleaved into piRNAs. As a result of their presence in opposite orientations in the genome, TEs can generate both sense and antisense retrotranscripts that will be respectively cleaved into primary sense and secondary antisense piRNAs. What initiates this process is not fully elucidated yet, but primary and secondary piRNAs are sorted into distinct cytoplasmic compartments. Primary sense piRNAs are preferentially loaded onto MILI proteins that are supported by TDRD1 Tudor proteins and reside in specific organelles called the 'pi-bodies' (Kojima *et al.*, 2009; Reuter *et al.*, 2009) (Figure 2). Pi-bodies were recently shown to contain another important player of fetal piRNA biogenesis, the GASZ protein whose biochemical function is unknown (Germ cell protein with Ankyrin repeats, Sterile alpha motifs and Leucine Zipper) (Ma *et al.*, 2009). Secondary antisense piRNAs are more prone to associate with MIWI2 proteins, which interact with other specialized Tudor proteins, TDRD9, and localize to distinct compartments called 'piP-bodies' (Aravin *et al.*, 2009; Shoji *et al.*, 2009). Exchange of sense and antisense piRNAs between pi- and piP-bodies intensifies and accelerates the degradation of TE transcripts, which eventually leads to the accumulation of TE-associated piRNAs in prospermatogonia. Mammals, similarly to *Drosophila*, are devoid of an RNA-dependent RNA polymerase activity. This feed-forward loop provides an alternative way for boosting the production of small effector RNAs.
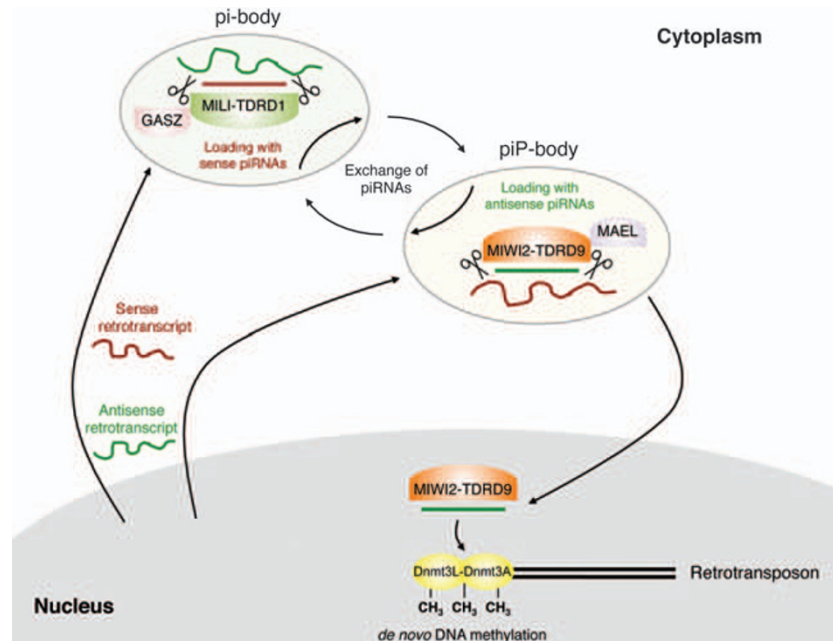
Figure 2 Subcellular compartimentalization of the piRNA pathway and interaction with the DNA methylation machinery in the male germline. Sense (red line) and antisense (green line) retrotranscripts are produced from TE sequences within the nucleus. These transcripts get cleaved into sense and antisense piRNAs, respectively. Sense piRNAs associate with MILI-TDRD1 complexes within pi-bodies, in which GASZ also resides. Antisense piRNAs associate with MIWI2-TDRD9 complexes and localize to piP-bodies, in which Maelstrom (MAEL) is also found. Amplification of post-transcriptional degradation of retrotranscripts is likely to occur in these cytoplasmic compartments, through an exchange of sense and antisense piRNAs. MIWI2-TDTD9 modules and their associated antisense piRNAs can also feedback into the nucleus to promote a stable transcriptional repression through the recruitment of the Dnmt3L/Dnmt3A machinery involved in *de novo* DNA methylation of TEs in the germline.

Any cellular mRNA expressed around 12.5 d.p.c. is likely to be subject to PIWI-directed cleavage. However, a robust post-transcriptional repression will be fostered only on parallel processing of homologous sense and antisense transcripts. Transcripts originating from single-copy and single-oriented genes fail to engage into the amplification cycle. Exon-derived piRNAs are only found in a sense orientation in association with MILI and they form <3% of the total piRNA population present at 16.5 d.p.c. (Aravin *et al.*, 2008). In contrast, TE-derived sense and antisense piRNAs account for almost 50% of piRNAs. Interestingly, MILI and MIWI2 not only have strand orientation preference but also bind piRNAs of different sizes, with MIWI2-bound piRNAs being slightly longer than MILI ones (average of 28.5 versus 26 nt). Nucleotide differences also reflect specific slicing signatures between MILI and MIWI2. How sense and antisense TEs are directed toward MILI or MIWI2 is unknown, but MIWI2 loading with antisense piRNAs is strictly dependent on MILI slicing activity, while MILI loading occurs without the need of MIWI2 (Aravin *et al.*, 2008).

Once loaded with antisense piRNAs produced by MILI-dependent cleavage, the MIWI2/TDRD9 complex can translocate from its cytoplasmic compartment into the nucleus (Aravin *et al.*, 2009; Shoji *et al.*, 2009) (Figure 2). MIWI2-bound antisense piRNAs will there promote *de novo* methylation of complementary genomic copies of TE elements. The piRNA pathway therefore converts a post-transcriptional response to TE reactivation into a mitotically stable mode of transcriptional repression (Aravin and Bourc'his, 2008; Aravin *et al.*,

2008; Kuramochi-Miyagawa *et al.*, 2008). Modalities of piRNA-dependent recruitment of the Dnmt3L/Dnmt3A complex are currently unknown and represent the next challenge in understanding RNA-directed DNA methylation pathways in mammals. The genome of *Drosophila* also encodes a PIWI protein with a nuclear localization, which suggests that chromatin modifications are the ancestral mechanism of piRNA-induced transcriptional repression (Brennecke *et al.*, 2007). Nuclear antisense piRNAs could work in RNA:RNA hybrids, by recognizing nascent TE transcripts and promote in *cis* the formation of repressive chromatin states at target TE loci. A similar mechanism operated by siRNAs is involved in heterochromatin formation in fission yeast (Zofall and Grewal, 2006), and could be consolidated in mammals by recruiting the Dnmt3L/Dnmt3A methylation complex (Ooi *et al.*, 2007; Aravin and Bourc'his, 2008). Alternatively, antisense piRNAs could also act in RNA:DNA hybrids. Invasion of double-stranded TE genomic segments with single-stranded piRNAs may generate alternative secondary structures such as DNA:RNA triplexes, which could attract directly or indirectly the DNA-methyltransferases (Bestor and Tycko, 1996).

The piRNA pathway is a major fertility protector in mammals, having a dual role in restricting TE activity through post-transcriptional degradation and DNA methylation-dependent transcriptional silencing. Last but not least, some evidence suggests that it may also affect translational control of TE-encoded proteins. MILI/TDRD1-containing pi-bodies associate with mitochondria and are gathering sites for multiple RNA

processing events. These include not only piRNA-mediated degradation and nonsense-mediated decay, but also translational repression. MIWI2/TDRD9-containing piP-bodies correspond to germline analogs of processing bodies, which are enriched in translationally inactive mRNAs. One key player of germ cell piP-bodies is MAEL, an enigmatic protein that contains a HMG box and a domain homologous to DnaQ-H 3′-5′ exonuclease (Soper et al., 2008; Aravin et al., 2009). MAEL mutants fail to properly assemble MIWI2 and TDRD9 in piP-bodies and reactivate L1-encoded proteins. However, piRNA biogenesis and DNA methylation are not affected, suggesting an alternative role of MAEL on TE protein translation. This observation opens up the possibility that piRNA-based defense mechanisms may target TEs at multiple stages of their life cycle, by recruiting specialized nuclear and cytoplasmic machineries available in germ cells. By interacting with specialized proteins, piRNA-directed mechanisms can also be adapted to specific TE classes. Although Dnmt3L, MILI and MIWI2 restrict L1 and LTR- elements, TDRD1 and TDRD9 show an exclusive preference for L1 suppression (Table 1) (Reuter et al., 2009; Shoji et al., 2009). Although its mode of action is currently unknown, the Tex19.1 protein is involved in limiting the expression of LTR elements of the MMERVK10C family during spermatogenesis (Ollinger et al., 2008). Identification of new partners of the piRNA machinery will provide us with a clearer picture of these specialized sub-pathways of TE suppression. Given the evolutionary relationship between infectious retroviruses and ERVs, anti-retroviral proteins such as TRIM proteins or zinc-finger anti-viral protein are promising candidates for LTR-specific defense routes.

Conclusion

The genome-wide loss of DNA methylation that characterizes the early stages of germ cell development in mammals renders the host genome vulnerable to TE invasion. However, male germ cells are extremely well prepared to face the burst of TE expression. Their piRNA-centered counteracting strategies involve many protagonists and cellular compartments that converge toward building an adaptive and multi-layered response. Moreover, germ cells undergoing demethylation rapidly stop dividing, a process that may limit retrotransposition efficiency (Shi et al., 2007). Rather than providing a comfortable niche, loosening of epigenetic repression may be observed as a trap programmed by the germline genome to sense and repress its resident TEs. Although not fully identical in detail, this detection strategy is also used by Arabidopsis thaliana, pointing to an interesting case of convergent evolution between mammals and plants (Slotkin et al., 2009). However, plants function in a safer way, by demethylating and reactivating TEs in surrounding cells of the pollen rather than in sperm cells per se. Populations of TE-associated siRNAs are then fed to the sperm to promote RNA-directed DNA methylation and correct any DNA methylation defect before fertilization.

Pluripotent cells of the mammalian embryo also undergo a genome-wide loss of DNA methylation followed by a TE remethylation at the time of implantation (Rougier et al., 1998; Hajkova et al., 2002). The piRNA/Dnmt3L pathway is apparently not active during this period (Bourc'his and Bestor, 2004; Carmell et al., 2007; Aravin et al., 2007b). Alternative RNA-directed DNA methylation or innate immunity-related pathways may be operant in sensing TEs and inducing their repression during early embryonic development.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

References

Agrawal A, Eastman QM, Schatz DG (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394: 744–751.

Aravin AA, Bourc'his D (2008). Small RNA guides for de novo DNA methylation in mammalian germ cells. Genes Dev 22: 970–975.

Aravin AA, Hannon GJ, Brennecke J (2007a). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318: 761–764.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF et al. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol Cell 31: 785–799.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ (2007b). Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316: 744–747.

Aravin AA, van der Heijden GW, Castaneda J, Vagin VV, Hannon GJ, Bortvin A (2009). Cytoplasmic compartmentalization of the fetal piRNA pathway in mice. PLoS Genet 5: e1000764.

Bannert N, Kurth R (2004). Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci USA 101(Suppl 2): 14572–14579.

Barbot W, Dupressoir A, Lazar V, Heidmann T (2002). Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. Nucleic Acids Res 30: 2365–2373.

Beauregard A, Curcio MJ, Belfort M (2008). The take and give between retrotransposable elements and their hosts. Annu Rev Genet 42: 587–617.

Bestor TH, Tycko B (1996). Creation of genomic methylation patterns. Nat Genet 12: 363–367.

Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH (2004). Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol 14: 1392–1396.

Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV et al. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci USA 103: 8780–8785.

Bourc'his D, Bestor TH (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 431: 96–99.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 18: 1752–1762.

Branciforte D, Martin SL (1994). Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol* **14**: 2584–2592.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R et al. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089–1103.

Brodersen P, Voinnet O (2006). The diversity of RNA silencing pathways in plants. *Trends Genet* **22**: 268–280.

Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* **100**: 5280–5285.

Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG et al. (2007). MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* **12**: 503–514.

Chedin F, Lieber MR, Hsieh CL (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci USA* **99**: 16916–16921.

Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KH, Wong LH (2009). LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. *PLoS Genet* **5**: e1000354.

Cordaux R, Batzer MA (2009). The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**: 691–703.

Craig NL, Craigie R, Gellert M, Lambowitz AM (2002). *Mobile DNA II*. American Society for Microbiology Press: Washington, DC.

De La Fuente R, Baumann C, Fan T, Schmidtmann A, Dobrinski I, Muegge K (2006). Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nat Cell Biol* **8**: 1448–1454.

Dong KB, Maksakova IA, Mohn F, Leung D, Appanah R, Lee S et al. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J* **27**: 2691–2701.

Eisenberg E, Nemzer S, Kinar Y, Sorek R, Rechavi G, Levanon EY (2005). Is abundant A-to-I RNA editing primate-specific? *Trends Genet* **21**: 77–81.

Esnault C, Millet J, Schwartz O, Heidmann T (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res* **34**: 1522–1531.

Esnault C, Priet S, Ribet D, Heidmann O, Heidmann T (2008). Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: *ex vivo* effects and *in vivo* 'traces' on the murine IAPE and human HERV-K elements. *Retrovirology* **5**: 75.

Furano AV, Duvernell DD, Boissinot S (2004). L1 (LINE-1) retrotransposon diversity differs dramatically between mammals and fish. *Trends Genet* **20**: 9–14.

Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS et al. (2007). LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet* **16**: 1569–1577.

Gasior SL, Roy-Engel AM, Deininger PL (2008). ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst)* **7**: 983–989.

Gasior SL, Wakeman TP, Xu B, Deininger PL (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* **357**: 1383–1393.

Goodier JL, Kazazian Jr HH (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135**: 23–35.

Grimson A, Srivastava M, Fahey B, Woodcroft BJ, Chiang HR, King N et al. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**: 1193–1197.

Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W et al. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* **117**: 15–23.

Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, Lao K et al. (2008). MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. *PLoS One* **3**: e1738.

Heidmann O, Vernochet C, Dupressoir A, Heidmann T (2009). Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new 'syncytin' in a third order of mammals. *Retrovirology* **6**: 107.

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A (2008). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* **27**: 404–408.

Huang J, Fan T, Yan Q, Zhu H, Fox S, Issaq HJ et al. (2004). Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res* **32**: 5019–5028.

Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449**: 248–251.

Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM et al. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev* **23**: 1303–1312.

Kato Y, Kaneda M, Hata K, Kumaki K, Hisano M, Kohara Y et al. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum Mol Genet* **16**: 2272–2280.

Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S et al. (2004). Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* **14**: 1719–1725.

Kojima K, Kuramochi-Miyagawa S, Chuma S, Tanaka T, Nakatsuji N, Kimura T et al. (2009). Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. *Genes Cells* **14**: 1155–1165.

Kondo Y, Issa JP (2003). Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J Biol Chem* **278**: 27658–27662.

Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian Jr HH et al. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci USA* **103**: 8036–8041.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**: 908–917.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* **24**: 265–276.

Lucifero D, La Salle S, Bourc'his D, Martel J, Bestor TH, Trasler JM (2007). Coordinate regulation of DNA methyltransferase expression during oogenesis. *BMC Dev Biol* **7**: 36.

Lyon MF (2006). Do LINEs have a role in X-chromosome inactivation? *J Biomed Biotechnol* **2006**: 59746.

Ma L, Buchold GM, Greenbaum MP, Roy A, Burns KH, Zhu H et al. (2009). GASZ is essential for male meiosis and suppression of retrotransposon expression in the male germline. *PLoS Genet* **5**: e1000635.

MacDuff DA, Demorest ZL, Harris RS (2009). AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* **37**: 1854–1867.

Maksakova IA, Mager DL, Reiss D (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell Mol Life Sci* **65**: 3329–3347.

Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet* **2**: e2.

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P et al. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J 24: 800–812.

Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H et al. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 464: 927–931.

Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 52: 643–645.

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560.

Montoya-Durango DE, Liu Y, Teneng I, Kalbfleisch T, Lacy ME, Steffen MC et al. (2009). Epigenetic control of mammalian LINE-1 retrotransposon by retinoblastoma proteins. Mutat Res 665: 20–28.

Moyes D, Griffiths DJ, Venables PJ (2007). Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. Trends Genet 23: 326–333.

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102: 553–563.

Murchison EP, Stein P, Xuan Z, Pan H, Zhang MQ, Schultz RM et al. (2007). Critical roles for Dicer in the female germline. Genes Dev 21: 682–693.

Nekrutenko A, Chung WY, Li WH (2003). An evolutionary approach reveals a high protein-coding capacity of the human genome. Trends Genet 19: 306–310.

Nishikura K (2006). Editor meets silencer: crosstalk between RNA editing and RNA interference. Nat Rev Mol Cell Biol 7: 919–931.

Obbard DJ, Gordon KH, Buck AH, Jiggins FM (2009). The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci 364: 99–115.

Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol 4: R74.

Okano M, Takebayashi S, Okumura K, Li E (1999). Assignment of cytosine-5 DNA methyltransferases Dnmt3a and Dnmt3b to mouse chromosome bands 12A2-A3 and 2H1 by in situ hybridization. Cytogenet Cell Genet 86: 333–334.

Ollinger R, Childs AJ, Burgess HM, Speed RM, Lundegaard PR, Reynolds N et al. (2008). Deletion of the pluripotency-associated Tex19.1 gene causes activation of endogenous retroviruses and defective spermatogenesis in mice. PLoS Genet 4: e1000199.

Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N et al. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat Genet 38: 101–106.

Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448: 714–717.

Ostertag EM, Kazazian Jr HH (2001). Biology of mammalian L1 retrotransposons. Annu Rev Genet 35: 501–538.

Ozato K, Shin DM, Chang TH, Morse III HC (2008). TRIM family proteins and their emerging roles in innate immunity. Nat Rev Immunol 8: 849–860.

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7: 597–606.

Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M et al. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature 463: 1101–1105.

Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol 16: 639–646.

Ribet D, Dewannieux M, Heidmann T (2004). An active murine transposon family pair: retrotransposition of 'master' MusD copies and ETn trans-mobilization. Genome Res 14: 2261–2267.

Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J et al. (2006). Large-scale structure of genomic methylation patterns. Genome Res 16: 157–163.

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Paldi A et al. (1998). Chromosome methylation patterns during mammalian preimplantation development. Genes Dev 12: 2108–2113.

Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463: 237–240.

Sayah DM, Sokolskaja E, Berthoux L, Luban J (2004). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature 430: 569–573.

Schulz WA, Steinhoff C, Florl AR (2006). Methylation of endogenous human retroelements in health and disease. Curr Top Microbiol Immunol 310: 211–250.

Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450: 908–912.

Shi X, Seluanov A, Gorbunova V (2007). Cell divisions are required for L1 retrotransposition. Mol Cell Biol 27: 1264–1270.

Shoji M, Tanaka T, Hosokawa M, Reuter M, Stark A, Kato Y et al. (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. Dev Cell 17: 775–787.

Slotkin RK, Vaughn M, Borges F, Tanurdzic M, Becker JD, Feijo JA et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell 136: 461–472.

Smit AF, Riggs AD (1996). Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 93: 1443–1448.

Soifer HS, Zaragoza A, Peyvan M, Behlke MA, Rossi JJ (2005). A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. Nucleic Acids Res 33: 846–856.

Soper SF, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P et al. (2008). Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. Dev Cell 15: 285–297.

Sripathy SP, Stevens J, Schultz DC (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. Mol Cell Biol 26: 8623–8638.

Stenglein MD, Harris RS (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. J Biol Chem 281: 16837–16841.

Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev Biol 269: 276–285.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453: 534–538.

Tarlinton RE, Meers J, Young PR (2006). Retroviral invasion of the koala genome. Nature 442: 79–81.

Trasler JM (2006). Gamete imprinting: setting epigenetic patterns for the next generation. Reprod Fertil Dev 18: 63–69.

104

Vagin VV, Wohlschlegel J, Qu J, Jonsson Z, Huang X, Chuma S et al. (2009). Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev* **23**: 1749–1762.

van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* **19**: 530–536.

van den Hurk JA, Meij IC, Seleme MC, Kano H, Nikopoulos K, Hoefsloot LH et al. (2007). L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet* **16**: 1587–1592.

Volff JN (2006). Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**: 913–922.

Wallace NA, Belancio VP, Deininger PL (2008). L1 mobile element expression causes multiple types of toxicity. *Gene* **419**: 75–81.

Walsh CP, Chaillet JR, Bestor TH (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**: 116–117.

Wang J, Saxe JP, Tanaka T, Chuma S, Lin H (2009). Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr Biol* **19**: 640–644.

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F et al. (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**: 175–183.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* **453**: 539–543.

Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.

Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T et al. (2006). A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**: 207–219.

Wolf D, Goff SP (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**: 1201–1204.

Xing J, Witherspoon DJ, Ray DA, Batzer MA, Jorde LB (2007). Mobile DNA elements in primate and human evolution. *Am J Phys Anthropol* (Suppl 45): 2–19.

Yoder JA, Walsh CP, Bestor TH (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**: 335–340.

Yokomine T, Hata K, Tsudzuki M, Sasaki H (2006). Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting. *Cytogenet Genome Res* **113**: 75–80.

Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL (2008). Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. *PLoS Genet* **4**: e1000007.

Zhu H, Geiman TM, Xi S, Jiang Q, Schmidtmann A, Chen T et al. (2006). Lsh is involved in *de novo* methylation of DNA. *EMBO J* **25**: 335–345.

Zhu Y, Gao G (2008). ZAP-mediated mRNA degradation. *RNA Biol* **5**: 65–67.

Zofall M, Grewal SI (2006). RNAi-mediated heterochromatin assembly in fission yeast. *Cold Spring Harb Symp Quant Biol* **71**: 487–496.