Transposable elements and the epigenetic regulation of the genome

R. Keith Slotkin and Robert Martienssen

Abstract | Overlapping epigenetic mechanisms have evolved in eukaryotic cells to silence the expression and mobility of transposable elements (TEs). Owing to their ability to recruit the silencing machinery, TEs have served as building blocks for epigenetic phenomena, both at the level of single genes and across larger chromosomal regions. Important progress has been made recently in understanding these silencing mechanisms. In addition, new insights have been gained into how this silencing has been co-opted to serve essential functions in 'host' cells, highlighting the importance of TEs in the epigenetic regulation of the genome.

Transposable elements
All mobile DNA segments in the genome, regardless of their mechanism of transposition.

Transpose
The movement of a genetic element from one location of the genome to another.

Autonomous element
A transposable element that produces all the proteins that are required for transposition.

Cryptic element
A transposable element that is epigenetically inactivated for a period of time and has lost the ability to mobilize.

Epigenetic
A heritable change that is not caused by a genetic mutation.

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. Correspondence to R.M. martiens@cshl.edu
doi:10.1038/nrg2072

Genome sequencing has revealed that transposable elements (TEs) of various classes constitute a large fraction of most eukaryotic genomes, including nearly 50% of our own[1]. Active TEs are highly mutagenic, often targeting protein-coding genes for insertion, and also causing chromosome breakage, illegitimate recombination and genome rearrangement. TEs can also influence neighbouring genes by altering splicing and polyadenylation patterns, or by functioning as enhancers or promoters[2].

TEs are often considered as 'selfish' or 'parasitic' elements, because their success (that is, an increase in copy number) is negatively correlated with the fitness of the host. However, most TEs are not actively transposing or duplicating. Although in most cases this is due to mutations and deletions that abolish transposition, some full-length autonomous TEs remain intact but silent in host genomes, in the form of cryptic elements. To combat the potentially harmful effects of active TEs, the genome has evolved epigenetic 'defense' mechanisms to suppress their activity. An epigenetically inactive TE retains the coding potential to mobilize itself but does not produce the necessary proteins owing to a repressive chromatin environment.

It has long been known that cryptic and non-autonomous TEs accumulate in heterochromatic regions of the genome. However, recent discoveries concerning the molecular mechanisms by which TEs are epigenetically silenced are increasing our understanding of their contribution to the function of heterochromatic regions that flank centromeres and telomeres. It is becoming clear that TEs have become part of the regulatory toolkit of the genome, with important roles in directing gene expression. Recent studies have implicated TEs in imprinting and X-chromosome inactivation, as well as in the developmental regulation of gene expression, a role that was first described by McClintock shortly after her discovery of TEs[3]. Here we bring together the current understanding of TEs as epigenetic regulators of the genome, drawing on studies from yeast, plants and animals.

Silencing mechanisms that suppress TEs

Genome-wide screens for genes that are required for the regulation of class I TE activity in budding yeast[4] and Class II TE activity in *Caenorhabditis elegans*[5] (see BOX 1 for an overview of the different types of TE) have uncovered more than 100 genes with various functions. Most of the genes that were identified did not overlap between the two screens, suggesting that many more genes might alter TE activity. Here we focus on epigenetic mechanisms of TE silencing at the transcriptional and post-transcriptional levels, which involve a host of factors that regulate TE mobility and carry out important cellular functions.

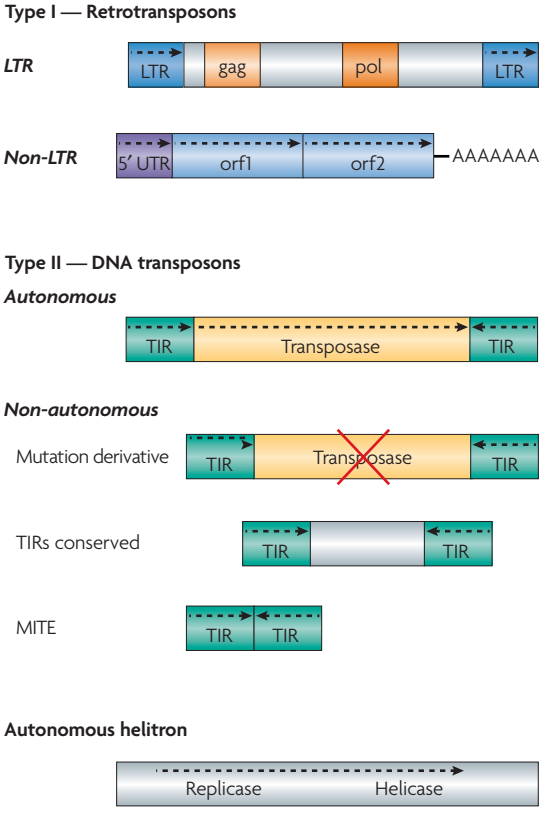
Post-transcriptional silencing of TEs by RNAi. RNAi is the mechanism by which dsRNA is cleaved by members of the dicer family of proteins into short 21–30-nucleotide small interfering RNAs (siRNAs) that guide RNA-degrading complexes to a complementary transcript (FIG. 1a). Argonaute proteins constitute the catalytic component of the siRNA-guided transcript-cleavage complex, which is known as RISC. RNAi is the main mechanism of TE silencing in *C. elegans*, in which genes that are required for RNAi are essential for silencing the

Box 1 | **Types and structures of transposable elements**

The diverse array of transposable elements (TEs) can be classified using several criteria, such as the requirement for a reverse-transcription step to transpose. TEs that require a reverse-transcription step are called retrotransposons, or type I elements. They can be divided into two types, on the basis of the presence or absence of direct repeats at the ends of the element called long terminal repeats (LTRs). LTR retrotransposon proteins pol and gag are closely related to retroviral proteins; however, these elements lack the envelope protein that is required to exit the cell. Retrotransposons undergo duplicative transposition, as their total number increases after each transposition with the potential to expand genomes. For example, one non-LTR retrotransposon family, LINE1, constitutes 17% of the total human genome[1].

Type II TEs, known as DNA transposons, do not require a reverse-transcription step to integrate into the genome. Instead, a transposon-encoded protein called a transposase recognizes the terminal inverted repeats (TIRs) that flank the TE, excises the TE out of the donor position, and then integrates the transposon into the new acceptor site. The gap that is left at the donor position can then be either repaired without element replacement in cut-and-paste transposition, or filled with a copy of the transposon by gap repair. Helitrons are newly identified DNA transposons that duplicate differently, using a rolling-circle mechanism. Autonomous helitrons contain a DNA helicase protein, as well as a replicase protein similar to replicon protein A (RPA).

TEs can also be classified by their self-sufficiency. Both retrotransposon and DNA-transposon families have autonomous elements and non-autonomous elements. Non-autonomous elements are often mutated relics of autonomous family members, but sometimes have only limited sequence similarity to their autonomous counterparts. Non-autonomous DNA transposons often consist of a pair of TIRs surrounding non-transposon DNA, which is frequently copied from an unrelated protein-coding gene. Additionally, some non-autonomous elements, called miniature inverted-repeat transposable elements (MITES), consist of two TIRs that are connected in a tail-to-tail orientation. The number of non-autonomous elements in a genome can greatly outnumber the autonomous elements. For example, there are an estimated 1.2 million copies of the *Alu* repeat in the human genome (reviewed in REF. 122), which is a non-autonomous non-LTR retrotransposon (also known as a short interspersed nuclear element, or SINE).



Non-autonomous element
A transposable element that does not produce the proteins required for transposition. These elements are dependent on the proteins produced by autonomous elements of the same element family to transpose.

Heterochromatin
The portion of interphase chromosomes that remains densely stained and condensed after cell division. Heterochromatic regions are generally gene poor, replicate late, and have a low level of transcription.

Imprinting
A gene is expressed from only one allele in a diploid organism, depending on the parent of origin.

Duplicative transposition
A mechanism of transposition that results in a copy of the element at both the excision and acceptor site. This can occur even after excision of the element by the process of gap repair.

Tc1 DNA transposon in the germ line[6,7]. In addition, mutations in both argonaute- and dicer-family proteins cause the reactivation of TEs in many eukaryotic species. TEs give rise to numerous siRNAs[8] in most species, including the recently reported TE siRNAs in humans[9], and TE-siRNA levels are correlated with element activity[7,10]. Although our understanding of how TEs are specifically targeted by the RNAi pathway is incomplete, structural features of TEs might help to distinguish their transcripts from those produced by host genes. For example, transcripts containing the terminal inverted repeats (TIRs) of DNA transposons might fold back to produce a hairpin dsRNA structure. Other mechanisms that are responsible for triggering RNAi are explored in BOX 2.

Chromatin modifications. A range of chromatin modifications suppress TE transcription, including modifications of histone tails, DNA methylation and alterations in chromatin packing and condensation. Modifications of histone amino (N)-terminal tails alter the binding of protein factors and relay information to transcription factors. Nucleosomes that are associated with TEs are enriched for methylation of histone H3 at lysine 9 (H3K9), which is a signal for transcriptionally repressive and inactive chromatin[11,12]. Mutations in genes required for repressive histone tail modifications lead to TE reactivation; for example, mutations in the histone H3K9 methyltransferase gene *Suv39* result in modest upregulation of TE transcripts in mouse embryonic stem (ES) cells[12].

Although it is not present in all eukaryotes, DNA methylation on cytosine residues is another important signal that represses TE transcription. In both plants and mammals, cytosine residues can be methylated in a symmetrical context (CpG), and this methylation is copied to the new DNA strand upon DNA replication, providing a mechanism for inheritance of TE silencing. In the mouse, the DNMT1 DNA methyltransferase is responsible for this maintenance of DNA methylation. Elevated intracisternal A-particle (IAP) retrotransposon transcript levels have been reported in *Dnmt1*-deficient embryos, which die after 9–10 days[13], but not in undifferentiated ES cells[12]. Asymmetrical DNA methylation also targets TEs for silencing in plants and, to a lesser extent, in mammals[14]. However, this methylation must be targeted to the TE *de novo* with each generation, and does not require partially methylated substrates as a guide. In the mouse, DNMT3 methyltransferase carries out *de novo* (mostly CpG) DNA methylation, and the non-catalytic homologue DNMT3L is required for IAP retrotransposon methylation and silencing in premeiotic male germ cells[15]. These findings suggest that DNA methylation might be required for the epigenetic silencing of TEs at specific stages of mammalian development (see below). The mechanism that is responsible for targeting *de novo* and asymmetrical DNA methylation is discussed later with respect to RNA-directed DNA methylation (particularly in the model plant *Arabidopsis thaliana*).

Cut-and-paste transposition
A mechanism of transposition in which no copy of the element remains at the excision site.

Proteins that modify chromatin structure are also involved in TE silencing, and several of these proteins are implicated in chromatin packaging and condensation. In plants, SWI/SNF chromatin-remodelling proteins, which use the energy of ATP hydrolysis to remodel nucleosomes, are required for TE silencing. In *A. thaliana*, the SWI2/SNF2-like protein DDM1 is specifically required for TE silencing and chromatin condensation[16]. Interestingly, TEs remain activated, or 'preset' (BOX 3), when passed from a *ddm1* mutant to a wild-type background[17], and do not regain DNA and H3K9 methylation, both of which are lost in *ddm1* mutants[16]. In the mouse, the *ddm1* homologue *Lsh1* also remodels chromatin and suppresses TEs, which lose DNA methylation in *Lsh1*$^{-/-}$ embryos[18]. Transcriptional analysis indicates that, although TE transcripts are strongly elevated in *ddm1* and *Lsh1* mutants, most genes are unaffected, suggesting that the remodelling activity is specifically targeted to TEs[19].

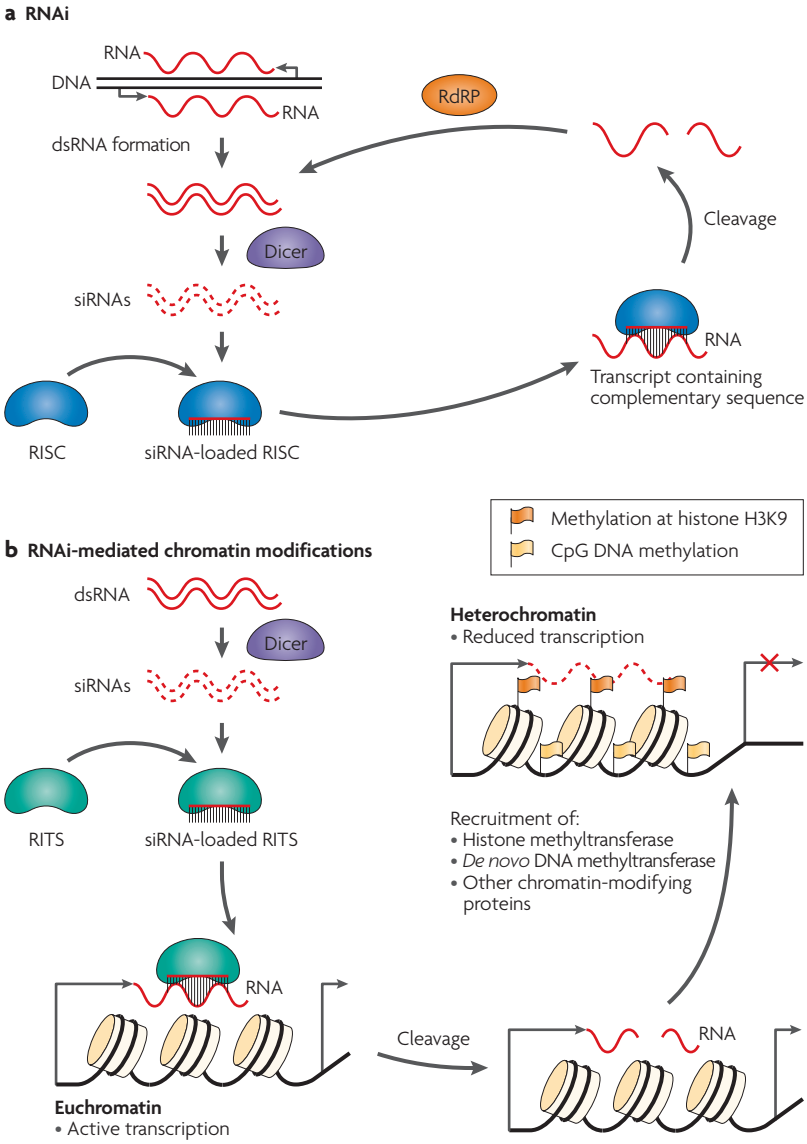


Figure 1 | **Mechanisms of transposable element silencing. a** | In the post-transcriptional RNAi pathway, dsRNA is cleaved into small interfering RNAs (siRNAs) by a dicer-family protein. These siRNAs are incorporated into the RISC complex, which contains an argonaute-family protein. The siRNA-loaded RISC complex then cleaves transcripts that are complementary to the siRNA sequence. In this example, dsRNA is initially formed by overlapping antiparallel transcripts. RdRP can additionally produce dsRNA from the cleaved RNA products (BOX 2). **b** | siRNAs can feed into an alternate pathway by which a different complex containing an argonaute-family protein (called RITS in *Schizosaccharomyces pombe*) guides the cleavage of nascent transcripts, which are still attached to RNA polymerase II and the DNA strand. Cleavage of this nascent transcript targets this region of chromatin for modification by the recruitment of a histone H3 lysine 9 (H3K9) methyltransferase and other proteins. Modifications of heterochromatin include the methylation of the H3K9 and possibly methylation of the cytosine bases of DNA in organisms in which DNA is methylated. Modifications to transposable element chromatin result in heterochromatin that is condensed and inaccessible to transcription.

RNAi-mediated chromatin modifications. In a number of eukaryotes, RNAi components are required for chromatin modifications. Although the maintenance of some repressive chromatin modifications does not require RNAi, it is not known whether these modifications are initially established by RNAi-based mechanisms. It is currently thought that the maintenance of TE silencing at the chromatin level involves redundant functions of both RNAi-dependent and RNAi-independent pathways, although our understanding of RNAi-mediated chromatin modifications is incomplete.

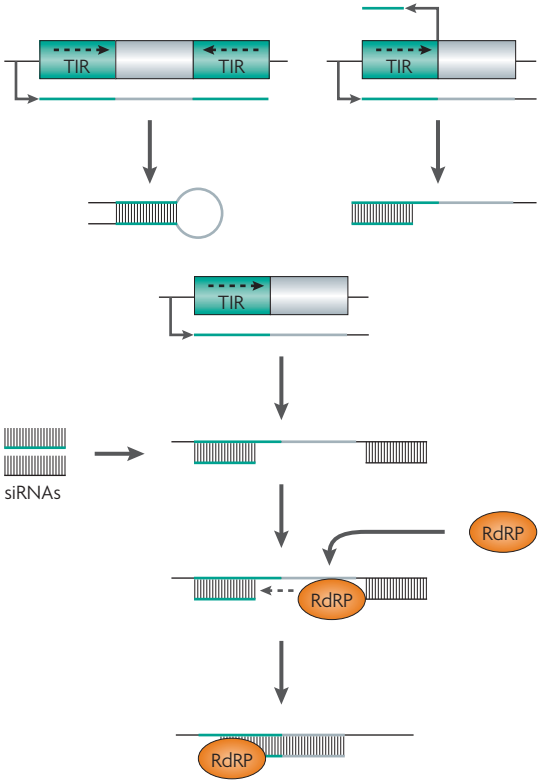
In the fission yeast, *Schizosaccharomyces pombe*, heterochromatic silencing depends on transcription by RNA Polymerase II and on the RNAi machinery (reviewed in REF. 20). Pericentromeric heterochromatin in *S. pombe* has provided a model for RNAi-mediated chromatin modification, which parallels TE silencing in using a similar set of siRNA triggers and silencing factors. In *S. pombe*, the RNAi machinery works on nascent RNAs, which are still bound to the chromosome and are processed co-transcriptionally into siRNAs[21] (FIG. 1b). These siRNAs are used by an argonaute protein (Ago1) to cleave nascent RNAs, and this cleavage recruits the histone H3K9 methyltransferase Clr4 to the chromosome[22] by an unknown mechanism. RNAi is specifically required for spreading of H3K9 methylation from repeats into reporter genes that are embedded into heterochromatic transcripts[22], similar perhaps to the spreading of heterochromatic silencing that is observed in position effect variegation (PEV) in *Drosophila melanogaster* (see below). In *S. pombe*, both pericentromeric repeats and *Tf2* LTR retrotransposons are under the control of histone modifications, but the role of RNAi in *Tf2* silencing is modest by comparison[23].

In the ciliated protozoa *Tetrahymena* and *Paramecium*, small RNAs target regions of the genome for chromatin modification, but these target regions subsequently undergo developmentally regulated DNA deletion (reviewed in REF. 24). Argonaute, dicer and histone methyltransferases are all required for this process, in which siRNA from one nucleus are somehow compared with transcripts from another nucleus to select sequences for elimination. In many ways, these internal eliminated sequences (IESs) resemble DNA transposons, but, unlike transposons in other eukaryotes, they are silenced and excised at the same time. This interesting system is likely to yield future insights into the epigenetic regulation of TEs.

Box 2 | **The origin of dsRNAs**

Small RNAs can initiate silencing at the post-transcriptional level or lead to chromatin modifications (FIG. 1). Although the process by which small interfering RNAs (siRNAs) are generated from dsRNA is well understood, the origin of the dsRNA from transposable elements (TEs) remains mysterious. Owing to their inverted-repeat structure (BOX 1), DNA transposons might form a hairpin upon read-through transcription of the entire element from external promoters. Such read-through has been detected in *Caenorhabditis elegans* for the *Tc1* DNA transposon, and it was proposed that this process initiates siRNA production from the *Tc1* terminal inverted repeats[7]. Likewise, read-through of a defective inverted-repeat transposon derivative is a dominant repressor of a maize DNA transposon family because it works as an siRNA trigger[123]. By this mechanism, the TE is continuously transcribed while silenced, because the element itself does not initiate transcription. dsRNA can also be generated from two overlapping antiparallel transcripts that base pair together. For example, bidirectional transcription from the LINE1 5′ UTR in human cells accounts for siRNA production from this element[9]. Insertion of a small piece of a TE into a bidirectionally transcribed region of the genome might also initiate transposon silencing by producing dsRNA. For example, in *Drosophila melanogaster*, insertions of TEs into heterochromatic loci result in the Aubergine-dependent germline silencing of homologous TEs elsewhere in the genome[124]. Similarly, in fission yeast, reporter genes that are inserted into heterochromatin are silenced by read-through transcription and RNAi[22].

In plants, *C. elegans* and fission yeast, dsRNA is also produced by a family of RNA-dependent RNA polymerases (RdRPs), which copy ssRNA templates into dsRNA. RdRP proteins participate in TE-mediated heterochromatin production[25]. How RdRP proteins select transcripts to feed into the RNAi and heterochromatin pathway is not currently understood. It might be that RdRPs are drawn to the RNA template by an siRNA-loaded complex, or by processing of sliced transcripts, suggesting that RdRPs might be involved only in producing secondary siRNAs along the length of the transcript template, and not in the production of the primary siRNA that is needed to initiate silencing. In plants, a recent report suggested that transcripts with two (or more) complementary microRNA sites might preferentially trigger RdRP[125]. (TIR, terminal inverted repeat).



Gap repair
A process whereby, after excision of a DNA transposon, the element can be copied back into the excision site from the template of the sister chromatid or homologous chromosome.

DNA transposon
Transposable elements that do not use a reverse-transcription step to integrate copies into the genome.

Terminal inverted repeat
Repeats that flank most DNA transposons and lie in an inverted orientation.

Retrotransposon
Transposable elements that use a reverse-transcription step to integrate copies into the genome. Also known as retroposons.

Presetting
Describes the inheritance of transposable-element-mediated gene control that is independent of the transposable element. Usually confined to the first generation.

RNA-directed DNA methylation and TE silencing have been investigated extensively in *A. thaliana*. The *A. thaliana* RNAi gene families have duplicated and diversified to participate in post-transcriptional regulation as well as siRNA-mediated chromatin modifications[25]. For example, *A. thaliana* contains four different dicer proteins, each with specific yet overlapping functions. Likewise, two distinct size classes of siRNA have been identified in this species[26]. The smaller class of 21–22-nucleotide siRNAs are involved in post-transcriptional regulation (RNAi and microRNAs), whereas the larger class of 24–26-nucleotide siRNAs are involved in RNA-dependent DNA methylation, and are derived predominantly from TEs and tandem repeats. The *DICER-LIKE 3* (*DCL3*) gene generates the larger 24–26-nucleotide siRNA size class from TEs, and one of the ten argonaute proteins, AGO4, binds a subset of these siRNAs. The siRNA-loaded AGO4 complex seems to have both catalytic and non-catalytic roles in silencing different TEs by targeting asymmetrical DNA methylation[27].

Heterochromatin formation targeted by siRNAs raises an interesting question: how do TEs that are silenced at the chromatin level produce transcripts that are to be cleaved into siRNAs? In *A. thaliana*, this is accomplished by the plant-specific RNA polymerase IV (REFS 28,29). Loss of Pol IV can transcriptionally reactivate several silenced TEs, as well as isolated solo long terminal repeats (LTRs)[30]. There are two forms of *A. thaliana* Pol IV (Pol IVa and Pol IVb), which are distinguished by their unique largest subunits. Both forms have a specific role in RNA-directed DNA methylation: Pol IVa in generating siRNAs, and Pol IVb in transferring the siRNA signal to produce the silencing outcome[31,32]. Pol IVb is probably guided to the correct location by its carboxy (C)-terminal domain, which is significantly longer than that of Pol IVa. This C-terminal domain interacts directly with AGO4 (REF. 33) in specific nuclear bodies in an RNA-dependent manner[34], and might serve as a platform for different protein factors. It is possible that Pol IV-generated 'trigger' transcripts that are produced from full-length or cryptic TE elements are targeted co-transcriptionally as they are in *S. pombe*. It is also possible that Pol IV acts directly on an RNA template (rather than DNA), generating dsRNA. Mutant screens for genes that are involved in RNA-directed DNA methylation have recovered both Pol IV and a SWI/SNF chromatin-remodelling gene, *DRD1* (REF. 35). Along with *DDM1*, *DRD1* might cooperate with Pol IV in transcribing heterochromatic DNA (or even producing dsRNA), and subsequent targeting of these regions for DNA methylation.

Although symmetrical DNA methylation is heritable, asymmetrical DNA methylation must be replaced at each cell division by a targeted mechanism, perhaps involving siRNAs. The *A. thaliana Dnmt3* homologue, *DRM2*, cooperates with another methyltransferase, *CMT3*, to impose asymmetrical DNA methylation, and is required for RNA-dependent DNA methylation (reviewed in REF. 36). Asymmetrical methylation of TEs in *A. thaliana* requires *DRM2* and *CMT3*, but transgenes and TEs are only selectively reactivated in

Box 3 | Transposable element cycling versus presetting

In her seminal research on transposable elements (TEs), Barbara McClintock described cycling and presetting of the *Spm* class of DNA transposons. Autonomous TEs cycled from active to inactive epigenetic states and back again in a single generation, or in multiple generations. By contrast, when a non-autonomous TE-generated epiallele responded to the activity of an autonomous element in *trans*, the epiallele sometimes continued to respond even after the active element segregated away. McClintock argued that these epialleles had predetermined, or 'preset', patterns of gene activity that were conditioned by the autonomous element[3], although this response was mostly lost in subsequent generations (erasure). Recent studies on this transposon family have shown that a single active autonomous element can transactivate a second, previously silenced autonomous element, requiring a *trans*-acting protein that is produced by *Spm*[126]. Today, presetting is thought to be due to multiple *trans*-reactivations and modifications of previously silenced autonomous elements that continue to condition the epiallele.

these mutants if the symmetrical CpG maintenance methyltransferase *MET1* (the homologue of *Dnmt1*) is still present[37]. This indicates that asymmetrical DNA methylation is much less important than symmetrical CpG methylation in maintaining TE silencing[10]. The observation that only a subset of *A. thaliana* TEs are regulated by RNAi in the presence of *MET1* and *DDM1* reinforces the importance of chromatin remodelling and CpG methylation in maintaining TE silencing.

The mechanisms involved in RNAi-mediated chromatin modifications are not as well understood in mammals. It is known that the DNA methyltransferase *Dnmt3a* binds to artificially introduced siRNAs[38], and artificial siRNA can direct DNA methylation, which is consistent with a requirement for this enzyme downstream of RNAi[36].

Germline silencing and the Piwi/Aubergine pathway. In animals, the germ line is sequestered early in embryogenesis, making this tissue uniquely sensitive to the heritability of TE activity. In *D. melanogaster*, crosses between males with *P* DNA transposons and females that are devoid of *P* elements result in sterility associated with high rates of germline transposition in the progeny, phenotypes that are collectively referred to as hybrid dysgenesis. *P* elements in the maternal parent suppress hybrid dysgenesis, suggesting the involvement of a cytoplasmic factor, or cytotype, that suppresses the activity of these elements (reviewed in REF. 39). *P* elements that are inserted near the X-chromosome telomere strongly induce the repressive nature of the maternal cytotype, and this location is a hot spot for *P*-element transposition. Recently, silencing mediated by these telomeric *P* elements was found to depend on the Polycomb group gene *Enhancer of zeste* and the argonaute protein Aubergine[40]. These results suggest that RNAi and heterochromatin might be responsible for the maternal repression of subtelomeric *P* elements. One idea is that heterochromatin-producing non-coding RNAs that are transcribed from telomere-associated repeats might extend into *P* elements that are inserted within these repeats, resulting in siRNA-mediated *trans* silencing of active *P* elements genome-wide[41].

Similar to *P* elements, non-LTR retrotransposons are also responsible for hybrid dysgenesis in *D. melanogaster*. Transformation of an *I* non-LTR retrotransposon into a genome that is devoid of *I* elements results in an increase in element copy number before the entire family of elements is eventually silenced. Silencing of the *I* element can be induced by introducing short pieces of the *I* element before transformation with the full-length element. This silencing requires transcription (but not translation) and is orientation-independent[42]. Similar to *P* cytotype, *I* activity depends on Aubergine[43], and the repressive signal from specific *I* elements that are located in heterochromatin is passed from females to their offspring. Therefore, two different heterochromatic TEs (*P* and *I*) produce signals that are transmitted in the maternal cytoplasm to protect the progeny from the detrimental effects of element mobility and hybrid dysgenesis. These signals are likely to be small RNAs, and they are important for genome integrity. Similarly, although they do not contribute to hybrid dysgenesis, *Gypsy* LTR retrotransposons are regulated in *trans* by the *Flamenco* locus, which contains *Gypsy* and other LTR retrotransposon fragments[44]. Silencing by *Flamenco* also depends on the argonaute protein Piwi (reviewed in REF. 41).

P, *I* and *Gypsy* elements are all silenced in the *D. melanogaster* germ line, precisely where the argonaute proteins Piwi and Aubergine are expressed. This might account for the activation of several classes of *D. melanogaster* TEs in cultures of somatic cells that lack *piwi* and *aubergine* expression[45]. Consistent with such a role, the Piwi/Aubergine subclade of argonaute-family genes is essential for germline development, not only in *Drosophila*, but also in mammals and *C. elegans*[46]. *piwi/aubergine* mutants have phenotypes that suggest that they are involved in regeneration of germ cells, and a recent report provides evidence that *aubergine* developmental phenotypes might result from a triggering of the DNA-damage pathway caused by excessive dsDNA breaks[47], perhaps generated by TEs. Interestingly, the Piwi subclade is not found in plants, which do not sequester germline cells early in development.

Piwi proteins bind special small RNAs (termed piRNAs) that are larger (26–30 nucleotides) than most small RNAs[48,49,50]. In mammals, many piRNAs in the male germ line are derived from TEs, but most (80%) are derived from approximately 140 loci that correspond to long non-coding RNAs, which are not known to be TEs[49,50]. piRNA clusters in mice and humans occur in syntenic locations, suggesting a conserved function in germline development; however, they are not related in sequence. These clusters are strand-specific, suggesting that the piRNA precursors are transcribed from only one strand. In *D. melanogaster*, piRNAs are referred to as repeat-associated siRNAs, or rasiRNAs[51], because they are generated from repeats and TEs[43] that are clustered in the genome. *Drosophila melanogaster* rasiRNAs have 3′ modifications that differ from other small RNAs[43], but it is not clear whether piRNAs in other species (such as mammals) carry these same modifications. Aubergine

Position effect variegation
Unstable gene-expression patterns that are determined by the location of a gene.

LTR retrotransposon
Long terminal repeat retrotransposons are a type of retrotransposon that has repeats in a direct orientation at either end, in contrast with non-LTR retrotransposons.

Transposable element cycling
The transformation of a transposable element from an epigenetically inactive state to an active state, and back to an epigenetically inactive state, over the course of development or over multiple generations.

Polycomb group
A class of proteins, originally described in *Drosophila melanogaster*, the function of which is to maintain stable and heritable epigenetic repression.

non-LTR retrotransposon
A retrotransposon that is not flanked by long terminal repeats.

is involved in silencing *P* and *I* TEs (reviewed in REF. 41), but it is not clear whether rasiRNAs regulate all TEs in the germ line, or only those with TE remnants in piRNA clusters.

Intriguingly, rasiRNA production and TE silencing are independent of Dicer in *D. melanogaster*[43], but at least three helicases — Homeless/Spindle E, Lighten up and Armitage — are required, as well as Aubergine and Piwi[41]. Given the unusual single-stranded nature of piRNA, it is likely that these helicases and Argonaute homologues cooperate in piRNA biogenesis. Interestingly, similar helicases are found in the chromatin-altering RNA-dependent RNA polymerase (RdRP) silencing complex in fission yeast[52], suggesting a role for piRNAs in heterochromatin formation. In support of this idea, Piwi and Aubergine are required for some types of PEV (see below) that result in chromatin modifications.

TEs contribute to chromosome form and function

Centromeres, telomeres and knobs. Constitutive heterochromatin is typically found at centromeres and telomeres. Both structures are essential for chromosome function and genome integrity — the centromere for trafficking chromosomes at cell division, and the telomere for preventing chromosome shortening following replication. In most eukaryotic species, the centromere varies in size from several kilobases to several megabases, and consists of long tandem arrays of simple sequence repeats (satellite repeats) surrounded by a peripheral region that is rich in TEs. The satellite regions of human centromeres are generally free of TEs, but pericentromeric regions are composed of long blocks of LINE (long interspersed nuclear element) and SINE (short interspersed nuclear element) retrotransposons[53]. In centromeres from *A. thaliana*, the inner satellite arrays are interspersed with retrotransposons, whereas the outer pericentromeric region is enriched for DNA transposons[54]. The DNA composition of centromeres differs widely in both length and primary nucleotide sequence between species, and it is the specific epigenetic context that confers centromeric function (reviewed in REF. 55). The presence of TEs, and the epigenetic silencing apparatus that they recruit, might therefore have a role in centromere function. It is currently unknown whether TEs specifically target heterochromatic regions for transposition, or whether they merely congregate at this location owing to reduced recombination and silencing, although recent data support the second theory[17].

During cell division, the condensed structure of the centromere is required to properly traffic the chromosome. Reactivation of epigenetically silenced TEs and satellite repeats is accompanied by chromosome segregation and meiotic defects in the mouse[15,56,57], loss of sister chromatid cohesion in fission yeast (reviewed in REF. 58), and loss of centromere condensation in *A. thaliana* (reviewed in REF. 59). This suggests that TE silencing mechanisms have a vital role in maintaining centromeric constitutive heterochromatin, and therefore in centromere function. In plants, retrotransposons help to silence centromeric satellite arrays (and centromeric satellite arrays help to silence retrotranspons)[60]. Additionally, highly conserved retrotransposons are found exclusively at the centromere, indicating a conserved function (reviewed in REF. 61). In mammals, inactivation of dicer leads to the accumulation of centromeric transcripts[62], presumably because they are no longer processed into siRNAs. These mutants also have centromeric condensation and differentiation defects[63], suggesting that RNAi has an important role in centromere function and heterochromatin formation through the regulation of satellite-repeat and TE transcripts.

TEs might have given rise to the centromeric satellite tandem repeats themselves. Centromere protein B (CENPB) helps to regulate centromeric heterochromatin by binding a short motif in most satellite repeats[64], and resembles the transposase of *Tc1/mariner* DNA transposons. Additionally, centromeric satellite repeats from many organisms have sequence homology to known TEs[65], although they are not mobile and no longer resemble TEs in structure. Tandem repeats such as centromeric satellite repeats are a major source of siRNAs, and it has been suggested that tandem duplications have been retained in the genome to efficiently maintain epigenetic silencing owing to their ability to replenish the siRNA pool[66].

Telomeres are composed of short tandem repeats that, in most species, are added to the ends of chromosomes by the reverse transcriptase telomerase, which uses an RNA template in a mechanism that resembles non-LTR retrotransposon integration[67]. Subtelomeric heterochromatin that lies adjacent to the telomeric repeats resembles pericentromeric heterochromatin, and is composed of full-length and fragmented TEs. *Drosophila* species lack a telomerase homologue, and non-LTR retrotransposons have taken over its function by specifically transposing to the chromosome ends in a developmentally regulated and orientation-specific manner (reviewed in REF. 68). These non-LTR retrotransposons (named *HeT-A* and *TART*) accomplish this insertion-site specificity by transporting their Gag proteins (see BOX 1) into the nucleus, where they recruit more copies to the chromosome ends. Recent studies have shown that these retrotransposons are regulated by the same epigenetic mechanisms that govern TE activity, including RNAi (reviewed in REF. 69). Specifically, mutations in the RNA helicase gene *spnE* and the argonaute gene *aubergine* cause an increase in *HeT-A* and *TART* transcript levels, reduced siRNA abundance and more frequent transpositions to chromosome ends[70]. This RNAi control of *HeT-A* and *TART* was found to occur specifically in *Drosophila* ovaries and oocytes, suggesting that the transposition and subsequent chromosome elongation occurs at premeiotic stages of the *Drosophila* germ line. It is currently unclear whether TEs contribute to telomere-length regulation in mammals. However, recent studies show that the epigenetic status (predominantly DNA methylation) of telomeric and sub-telomeric repeats has a role in regulating telomere elongation in mammalian cells[71].

In contrast to the centromere and telomere, a chromosomal knob is a fragment of constitutive heterochromatin that has no known function, although some knobs have been associated with neo-centromeric activity.

Constitutive heterochromatin
Heterochromatin that is heritable and is found at a particular locus at all times.

Analysis of an *A. thaliana* knob, composed almost entirely of full-length and fragmented TEs (many of which are inserted into each other), showed that heterochromatin was dependent on the epigenetic mechanisms of TE silencing, and that the jumbled TE islands that are located in these knobs potentially lead to read-through transcription from existing elements that trigger RNAi-mediated chromatin modifications[16]. This finding, and others, support the view that TE-containing constitutive heterochromatin is governed by the epigenetic regulators that are recruited by TEs.

PEV and chromatin insulators. Muller first described PEV in *D. melanogaster* in the 1930s as a variegated gene-silencing phenotype associated with the spreading of heterochromatin into adjacent genes. Transgene insertions in heterochromatin might confer variegated phenotypes (reviewed in REF. 72), in some cases depending on proximity to TEs (FIG. 2a), indicating that TEs are nucleation centres for the formation of this type of facultative heterochromatin[73]. The basis of the variegated phenotype that is seen in PEV is currently enigmatic, although this feature suggests that TE silencing must vary both spatially and temporally. Genetic screens to identify suppressors and enhancers of PEV suggest that the silencing of reporter transgenes is dependent on the same factors that are responsible for silencing TEs, including histone methyltransferases and argonaute genes, indicating that RNAi-mediated chromatin modifications are involved[74,75]. Genes must be in close proximity to a TE to be influenced by its regulation[16], and analysis of *cis*-acting deletions or duplications in *D. melanogaster* indicate that facultative heterochromatin of this sort can spread linearly along the chromosome (for up to 10 kb) until it encounters a boundary before the next actively expressed gene[73].

Chromatin insulators, such as boundary elements, set up territories of gene expression along the chromosome by establishing domains of chromatin structure. Insulators are thought to have a role in organizing chromatin into nuclear compartments that are dependent on transcriptional states (reviewed in REF. 76). TEs themselves can function as insulators. For example, TEs in constitutive heterochromatin might function to distinguish and functionally separate euchromatic chromosome arms from telomeric repeats or from the satellite core of the centromere. In *D. melanogaster*, the *Gypsy* retrotransposon carries an insulator element that can block a promoter's interactions with a distal enhancer element, or buffer a transgene from the influence of its genomic context. It remains unresolved whether other TEs besides *Gypsy* function as insulators. A recent report has shown that, like PEV, *Gypsy* insulator elements are regulated and controlled by the same set of genes that are responsible for suppression of TEs, including the argonaute family[77]. However, the function of the *Gypsy* insulator, along with examples of PEV that also require the activity of Piwi and Aubergine, are manifest long after these genes are expressed in the germ line, and must therefore involve chromosomal memory and chromatin modification (reviewed in REF. 41).

Facultative heterochromatin
Heterochromatin that switches to euchromatin at points during development.

X inactivation. Female mammals must compensate for their double dosage of X-chromosome genes compared with XY males. This compensation is achieved by the inactivation and heterochromatinization of one of the two X chromosomes. Inactivation is initiated at the X-chromosome inactivation center (XIC), and then spreads outwards from the XIC to the rest of the chromosome. However, some regions of the X chromosome are less efficiently silenced. According to one hypothesis, LINE retrotransposons might help as 'boosters' for the efficient spread of the silencing away from the XIC[78]. In support of this idea, the X chromosomes of many mammals, including humans, are rich in LINE elements, except in regions that are prone to escaping X inactivation[79]. Translocations between the X chromosome and autosomes (which have a much lower LINE-element density) show that the spread of silencing away from the XIC and into the autosome cannot be maintained (reviewed in REF. 78). It is still unknown whether LINEs on the X chromosome function in the spread of heterochromatin. However, a recent study has shown that short tandem repeats with ancient homology to retrotransposons regulate X-chromosome inactivation by producing bidirectional transcripts in differentiating mouse ES cells, providing indirect evidence that TEs function in both the initiation and spread of X inactivation[80].

TE-induced epigenetic regulation of specific genes

TEs are known to produce genetic and phenotypic variation between individuals because of their polymorphic locations. They can also produce variation within the cells and tissues of an individual by subjecting nearby genes to the epigenetic regulation that is targeted to TEs. Epialleles are heritable but reversible epigenetic changes in a gene's expression that are often under the control of a nearby TE or tandem repeat. Epialleles are metastable by nature, having variable expression in genetically identical cells (reviewed in REF. 81), which is manifested as variegation and mosaicism (FIG. 2b) that resembles PEV. As 25% of human promoter regions contain TE-derived sequences, including experimentally verified *cis*-regulatory elements (reviewed in REF. 82), cryptic TEs might contribute to the regulatory regions of many genes. For example, a recent report showed the transcriptional regulation of the inhibitor of apoptosis protein family in humans, mice and rats were all regulated by domesticated LTR retrotransposons. Interestingly, these LTR elements were independently acquired in these species and show no similarity, suggesting that retrotransposons have a regulatory role for orthologous genes and might differentially modulate apoptotic responses between species[83].

Epiallele production and phenotypic variation. TE promoters can influence the transcription of nearby genes. In the simplest case, transcripts from full-length TEs can read through the end of the element into a neighbouring gene. Read-through transcripts can then be subject to post-transcriptional and transcriptional control, depending on the activity of the TE. Spurious transcription can also initiate from fractured and incomplete elements that retain promoters but have lost termination

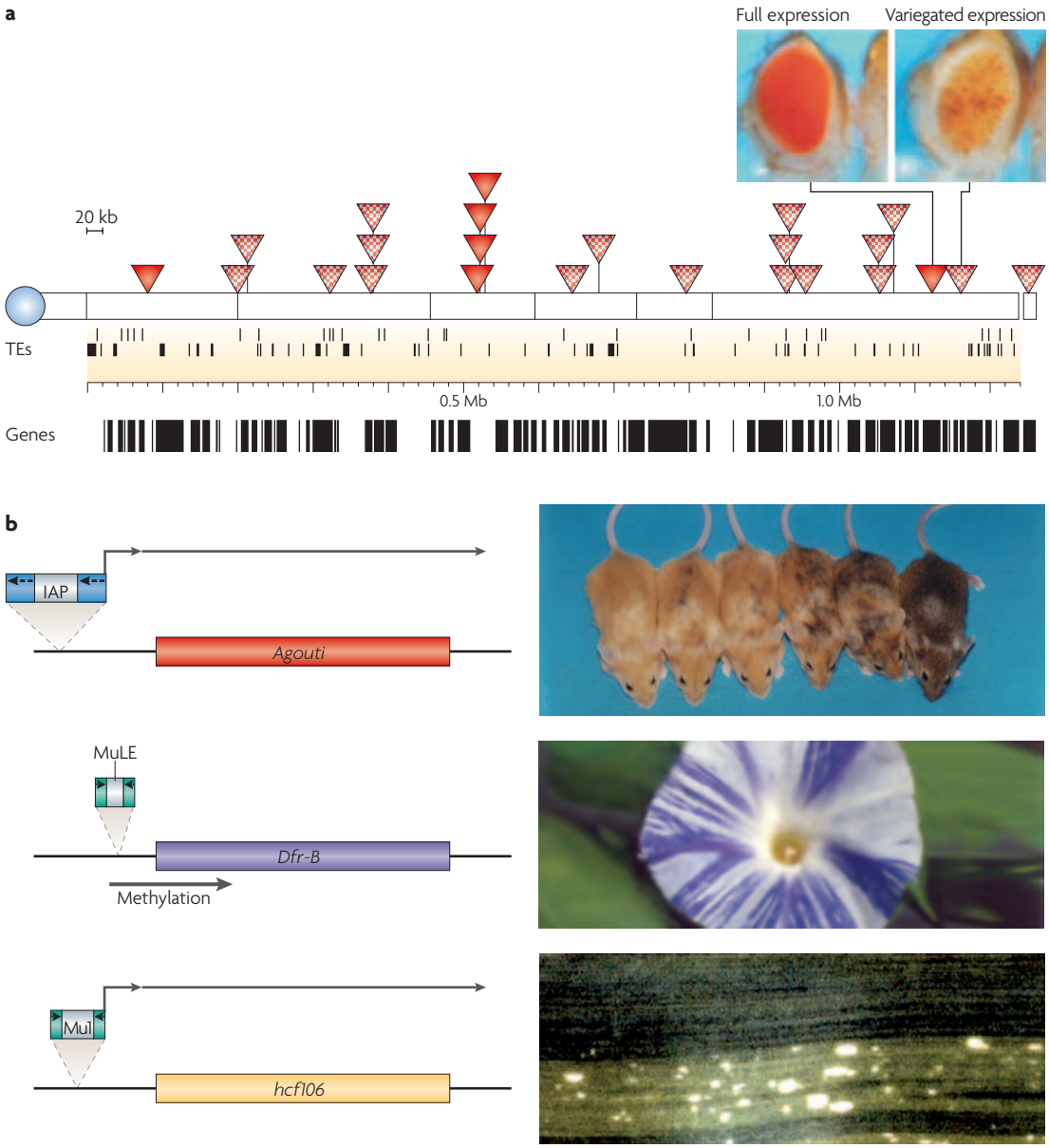


Figure 2 | **The influence of transposable elements on gene expression. a** | Position effect variegation (PEV) on *Drosophila melanogaster* chromosome 4 correlates with the proximity of a transgene to a transposable element (TE)[73]. In this example, expression of the *white* transgene results in a red eye. Transgene insertions (indicated by triangles) along the chromosome can be near TEs or genes. The phenotype of the eye for each transgene insertion is marked by a red or 'variegated' triangle. Insertions of the transgene into active regions of the genome (such as genes) result in transgene expression and a red eye. Transgene insertions into or near TEs have variable expression resulting in variegated eyes. These insertions demonstrate that variegated transgene expression correlates with proximity to a TE, non-variegated (red-eye) producing transgenes are clustered in gene islands, and the majority of this *D. melanogaster* chromosome arm is heterochromatic. **b** | Three examples of variegation induced by TE-generated epialleles. In mice, the intracisternal A-particle (IAP) retrotransposon produces an outward-reading transcript that extends into the agouti coat-colour gene. The level of agouti transcript, and the colour of the coat, is subject to the epigenetic status of the retrotransposon and is heritable[87]. In morning glory flowers, DNA methylation of a non-autonomous MuLE transposon can spread to the promoter of a flower-colour gene (*Dfr-B*), creating petal-colour streaks[143]. In maize, the activity of one TE family regulates two epialleles. In sectors where the *Mutator* transposon family is active, the mutant phenotypes of both epialleles occur, generating single sectors of pale green (*hcf106* mutant) and necrotic spotted (*les28* mutant) tissue on leaves[92]. *hcf106* transcripts are initiated by a non-autonomous (Mu1) transposon only when the autonomous transposon elsewhere in the genome is inactive. The molecular nature of the *les28* epiallele has not been elucidated. Image in panel **a** reproduced with permission from REF. 73 © (2004) American Society for Microbiology. Top image in panel **b** reproduced with permission from *Nature* REF. 87 © (1999) Macmillan Publishers Ltd. Middle image in panel **b** reproduced with permission from REF. 143 © (2004) Elsevier Science. Bottom image in panel **b** reproduced with permission from REF. 144 © (2001) American Association for the Advancement of Science.

Box 4 | **Genome stress, transposable elements and the adaptive response**

McClintock first suggested that transposable elements (TEs) were activated in response to challenges to the genome[127] and, since then, the release of the epigenetic silencing of TEs has been described in response to UV exposure, temperature, radiation, wounding, cell culture, pathogen infection and polyploidization (reviewed in REF. 128). Stress-reactivated TEs might generate the raw diversity that a species requires over evolutionary time to survive the specific stress. This adaptive response is a long-term strategy to increase variability for selection, but might not necessarily need to be genetic, as TE-induced epialleles would also be affected if the control of TEs were lost.

The response of TEs to stress can occur through one of two mechanisms (reviewed in REF. 128). First, the stress could directly activate TEs and their mutagenic activity. For example, the *Tnt1* retrotransposon in tobacco is reactivated by infection, a process that is mediated through the *Tnt1* LTR promoter, which has regions that are similar to pathogen defense genes that respond the stress-response hormone salicylic acid. Likewise, some TEs in *Drosophila melanogaster* that respond to heat stress also contain the same regulatory motifs as heat-shock inducible promoters. Second, stress might inhibit gene-silencing mechanisms in the genome, indirectly resulting in the reactivation of TEs. For example, position effect variegation (PEV) is temperature sensitive in fission yeast[129] as well as in *D. melanogaster*, as are *P* element cytotype and hybrid dysgenesis[130]. Additionally, in *Drosophila simulans*, temperature influences the rate of transposition and, in wild populations, changes in copy number follow a minimum temperature cline[131]. In *Schizosaccaromyces pombe*, the genes and TEs that are activated in abiotic stress conditions are similar to those that are reactivated in histone deacetylation mutants[23], and stress-response factors cooperate with RNAi in heterochromatic silencing[132]. These examples demonstrate how the environment can influence the epigenetic regulation of TEs. Once reactivated, these elements can then further alter the genome randomly in response to stress.

In one final striking example, dietary supplementation with methyl-group donors for DNA and histone methylation was shown to cause a shift in coat colour in the offspring of A^{vy} mice[133]. Mice that had supplemented diets had a statistically significant shift in the colour of their offspring's coats compared with non-supplemented control mice. The supplementation altered the level of DNA methylation of the IAP LTR at the A^{vy} locus[134,135]. Genome-wide, it is not known to what extent diet and other aspects of the environment effect epigenetic gene expression that is mediated through TEs, and this remains an important question to be resolved.

signals. For example, retrotransposon 5′ LTRs contain the *cis* promoter elements that are required to initiate transcription, which are also found in 3′ LTRs (BOX 1). Therefore, outward-reading transcripts from retrotransposons and solo-LTRs can express downstream genes, a process that was first described for *Ty1* in budding yeast (reviewed in REF. 84).

In addition, DNA transposons, LTR and non-LTR retrotransposons express outward-reading antisense transcripts from their ends. These outward-reading transcripts can result in ectopic expression of neighbouring genes that is dependent on the activity state of the TE[85], as illustrated by the *P*-cytotype-dependent *vestigial* locus of *D. melanogaster*[86]. The mouse IAP LTR retrotransposon produces an outward-reading antisense transcript from its 5′ LTR and reads into neighbouring genes, altering their regulation. In one example, a naturally occurring IAP insertion has generated the A^{vy} allele at the mouse agouti locus (FIG. 2b), which confers a different coat colour from the wild-type allele, in addition to increased susceptibility to obesity, diabetes and tumours[87]. In some ways resembling presetting (see BOX 3), coat-colour changes that are induced by the IAP element are heritable through the female germ line[88], and depend on *trans*-acting modifiers that regulate chromatin and DNA methylation[89].

The human LINE1 element also has an outward-reading promoter in its 5′ end, and many human genes use this as their primary promoter[85]. The chimeric transcripts can be tissue specific and potentially expand the expression pattern of the neighbouring gene[90]. Retrotransposons are expressed in mouse cleavage-stage embryos and mature oocytes, providing alternative promoters to a subset of host genes[91]. A recent study showed that the LINE1 outward (antisense) transcript, which overlaps the inward sense transcript, results in dsRNA production and is subject to RNAi, which might bring genes that neighbour the LINE1 element under its epigenetic control[9]. It is also possible that a single TE family can regulate multiple epialleles at the same time[92]. This can result in coordinate epigenetic regulation of several unrelated genes (FIG. 2b), a phenomenon that was first described by McClintock. This might suggest that the activity state of a TE, which is subject to regulation by the environment and stress (BOX 4), can potentially regulate a set of genes that have related functions. Furthermore, this highlights the differential regulation of autonomous and non-autonomous elements; although epigenetically silenced, non-autonomous elements are often just responding to the activity of the master autonomous element of the same family.

TE-induced epialleles have been proposed to account for phenotypic variability in mammals[93], especially differences between genetically identical twins. Variegation and mosaic patterns must begin with a single cell either losing or gaining gene silencing. How the early epigenetic expression patterns of TEs and epialleles are established differently in identical cells remains a major question in the field. However, because TEs comprise a large fraction of the eukaryotic genome and are the targets of the cell's gene-silencing machinery, their contribution to epigenetic gene regulation and to human disease could be large.

Paramutation. Some epialleles undergo paramutation — an interaction between alleles that results in a heritable expression change of one allele. The allele with reduced expression (or silenced allele) is then able to initiate silencing of other active alleles in subsequent generations. In maize, several genes in the anthocyanin pigmentation pathway have paramutable alleles, which are associated with and dependent on TEs or tandem repeats. At the *R* gene, DNA methylation differs between active and silenced alleles, and is localized to a *Doppia* DNA transposon fragment in the *R* gene promoter[94]. Intriguingly, in an allele in which most of the transposon fragment had been deleted, cytosine methylation was not present and the ability to change paramutation states was compromised, suggesting that this transposon fragment is responsible for carrying the epigenetic mark. The *R* homologue *B* can also be regulated by paramutation, but the regions that are required to control *B* paramutation were mapped to tandem repeats that lie 100 kb upstream from the start of transcription of the *B* gene[95]. Alleles of *B* that do not participate in paramutation have only one copy of this repeat, whereas paramutable alleles

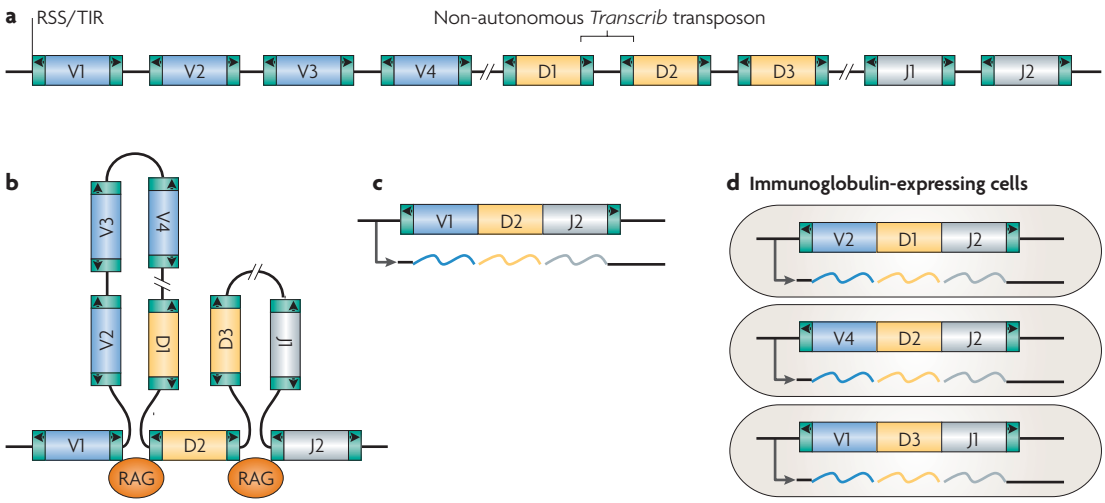


Figure 3 | **The transposable element origins of V(D)J recombination.** To generate a diverse set of antibodies, B and T cells of jawed vertebrates undergo V(D)J recombination to create specific DNA rearrangements. **a** | Immunoglobulin genes are arranged with multiple variable (V), diversity (D) and joining (J) regions, from which genetic diversity is generated upon rearrangement of these regions. Each segment (or exon) is flanked by a short inverted repeat called a recombinational signal sequence (RSS), which is analogous to a DNA transposon's terminal inverted repeats (TIRs) (BOX 1). The introns that are flanked by RSSs can be thought of as non-autonomous DNA transposons. The two genes that are responsible for the DNA rearrangements, *Rag1* and *Rag2*, are immobilized autonomous versions of the *Transcrib* family of DNA transposons[145]. These proteins are essentially transposase proteins (BOX 1) that recognize the RSS inverted repeat sequences that are located within the immunoglobulin gene at the site of the coding-sequence boundaries. **b** | The RAG proteins bind different combinations of RSS sequences in each B or T cell precursor and excise the internal region, which can be thought of as excision of a non-autonomous transposon. **c** | After excision by the RAG proteins, an immunoglobulin gene is produced with a coding region that has been rearranged compared with the unmodified genome. **d** | The RSS sequences that interact with the RAG transposase are variable in each cell, giving rise to different excision events and individual immunoglobulin cells that carry different coding potentials. The site of element excision is now a rearranged immunoglobulin gene that can be transcribed and translated into an antibody.

have seven. DNA methylation and nuclease sensitivity differ at the repeats between active and silenced alleles, and are altered after undergoing paramutation. Recently, an RdRP gene (see BOX 2) was found to be required for paramutation[96], as well as TEs[97] and transgene silencing[98]. This implicates the RdRP-dependent amplification of non-coding tandem repeat RNAs in the regulation of paramutation. However, paramutation-like phenomena also occur in the mouse[99], which is not thought to have RdRPs.

Imprinting. As on the mammalian X chromosome, TEs are specifically associated with monoallelic expressed genes that are located on autosomes, raising the possibility that TEs are involved in the regulation of imprinted genes. In mammals, the IAP and LINE1 retrotransposons are hypomethylated in the female germ line, and constitute a high proportion of mRNA in oocytes, which are deposited in the embryo[91]. By contrast, non-autonomous SINE elements are hypomethylated in sperm[100]. This raises the possibility that different classes of TEs might contribute to the differential expression of paternally and maternally imprinted genes. Consistent with this, there is an excess of LINE1 elements that are associated with paternally expressed autosomal imprinted genes[101], whereas SINE elements are excluded from the promoters of these same genes[102]. X inactivation (see above) in the mouse is also imprinted, such that the paternal X chromosome is preferentially inactivated in the placenta (reviewed in REF. 103), and presumably this could again be due to differential TE activation in the germ line.

The effect of TE-mediated imprinting in mammals might be mediated through the action of DNMT3L. This *de novo* DNA methyltransferase gene is expressed exclusively in female germ line cells and is responsible for DNA methylation and establishment of some genomic imprints[104]. Currently, it is not known how this non-catalytic methyltransferase homologue is guided to TEs, however it might be that stage-specific siRNAs or piRNAs guide DNMT3L to TEs such as IAP and LINE1.

Other types of TEs have also given rise to imprinted genes in mammals. For example, the imprinted genes retrotransposon-like 1 (*Rtl1*) and paternally expressed 10 (*Peg10*) are immobilized *Sushi*/*Gypsy* class retrotransposons that have undergone positive selection to acquire new and essential functions[105]. Only the paternal allele of *Rtl1* is expressed[106], and two maternally imprinted microRNAs are encoded on the opposite strand that guide the cleavage of the *Rtl1* mRNA[107].

The *FWA* gene of *A. thaliana* provides an excellent example of how TE-generated tandem repeats are involved in epiallele formation and imprinting. The first two exons of the *FWA* gene are composed of tandem

Box 5 | **Genome invasion and expansion**

When a transposable element (TE) enters a genome for the first time (through cross-fertilization, horizontal transfer or another mechanism), the TE can often duplicate freely before becoming epigenetically silenced. For example, engineered versions of *Tc1/mariner* DNA transposons from frogs and fish[136,137], or a LINE1 retrotransposon from humans[138], all had a high transposition frequency when inserted into the mouse genome. Studies in *Drosophila melanogaster* using the *I*, *hobo* and *P* TEs have provided insight into this phenomenon (reviewed in REF. 139). Older and more fragmented TEs are located primarily in heterochromatin, where they are generally not selected against. It is in this fragmented and epigenetically regulated environment that they are likely to produce the silencing signals (such as small RNAs) that function to silence the entire TE family in *trans*[41]. As a naive genome does not have these silencing triggers, incoming TE families are able to remain active until they (perhaps by chance) produce an element that functions as a source of the silencing signal. This can occur quickly for TEs that contain inverted repeats, but might also eventually occur due to tandem insertions or abortive transpositions. Using these mechanisms, TEs themselves have been shown to be responsible for their initial epigenetic silencing because of their propensity to generate defective copies upon transposition[123].

TEs also go through bursts of activity during evolution, whenever the restraints on element copy number are released[140]. The nature of these constraints is not known, but epigenetic suppression of TEs probably has a role. Bursts in TE activity can lead to genome size increases, as can genome duplications and polyploidization. For example, large genomes, such as those of humans and maize, consist of ~45% and ~75% TEs, respectively (reviewed in REF. 122). Interestingly, particular families of reactivated retrotransposons provide the bulk of genome size increases in many species. One family of *Gypsy*-like retrotransposons is responsible for most of the threefold difference in genome size among several *Gossypium* species[141]. Similarly, as few as three retrotransposon families have doubled the size of some rice genomes[142]. In cases in which only one or a few TEs contribute to genome expansion, the expansion might not be due to total loss of epigenetic control but, instead, loss of control of only one element or family, a process that is similar to what occurs when a TE enters a new genome (see above). For example, the trigger that is responsible for initiating silencing of one TE family might be lost due to deletion or drift, resulting in reactivation of only that family. How particular individual TE families escape silencing or reactivate whereas others do not remains an important question to be resolved.

repeats from a fragmented SINE TE. *FWA* is normally expressed in the seed, and transcription starts in the upstream tandem repeat (which is required for ectopic expression), consistent with a SINE antisense outward-reading promoter. The tandem repeats produce siRNAs that regulate the *FWA* promoter, and are targeted for DNA methylation[16]. A recent study showed that these tandem repeats recruit RNA-dependent DNA methylation to unmethylated *FWA* epialleles in *trans*, reminiscent of paramutation[108]. Normally, *FWA* is imprinted and expressed from only the maternal allele in the endosperm, in which the tandem repeats (and therefore the SINE) are demethylated[109]. This might reflect the activity of SINE TEs in this tissue; however, TE expression that is dependent on parent of origin has not yet been reported in plants.

Mammalian IAP, LINE and SINE retrotransposons show clear parent-of-origin effects (reviewed in REF. 110). However, the extent to which imprinted TEs control imprinted genes remains to be determined. Evidence from plants and animals suggest that the expression patterns of genes that are responsible for epigenetic gene silencing are responsible for generating imprinting patterns. This gene silencing might be mediated through the presence of TEs in or near the imprinted genes.

V(D)J Recombination. To generate the enormous diversity of antibodies that are needed for a successful immune response, the B and T cells of jawed vertebrates have co-opted a TE family to create cell-type-specific DNA rearrangements, called V(D)J recombination (FIG. 3). Antibody production and V(D)J recombination are subject to epigenetic regulation, but the extent to which TEs contribute to this regulation is unclear. Because the action of V(D)J recombination must be tissue- and stage-specific, there is tight regulation at the level of recombination signal sequence (RSS) chromatin accessibility (reviewed in REF. 111). Non-functional immunoglobulin genes are positioned in the nuclear space close to constitutive heterochromatin, and their position and levels of condensation change upon activation. Histone tail modifications, including methylated H3K9 and chromatin-remodelling by proteins in the *SWI2/SNF2* family, regulate the accessibility of the template to V(D)J recombination. Additionally, the process of DNA rearrangement might be under post-transcriptional regulation. Intergenic antisense transcripts have been described that temporally correlate with V(D)J recombination[112]. Therefore, the epigenetic mechanisms that are responsible for regulating V(D)J recombination are reminiscent of those that regulate TE activity. Although it is intriguing that the origin of this rearrangement mechanism is a TE system, whether this origin contributed to the epigenetic regulation of these rearrangements remains unknown.

Evolutionary implications

TEs have the ability to mutate genes, alter gene regulation and generate new genes, each providing fuel for evolution (reviewed in REF. 113). However, this is presumably not the goal of transposition; TEs are likely to be sexually transmitted, self-replicating parasites that do not aim to aid evolution[114,115]. Because the raw variation that is induced by TEs is dependent on their activity, much of the evolutionary potential of TEs is governed by their control and epigenetic regulation. Epigenetic TE regulation might account for some key aspects of genome evolution, including the stress-induced reactivation of TEs (BOX 4) and the invasion of a new genome by TEs (BOX 5).

TEs also have a major role in generating intraspecies variation. Intergenic regions in different inbred strains of maize, for example, are essentially unrelated owing to differential insertion of LTR-retrotransposons and helitrons[116]. Helitrons are a class of DNA transposon that frequently duplicate passenger genes. In human populations, although active full-length LINE1 elements are undergoing negative selection[117], they significantly contribute to levels of human genetic diversity[118]. The *Alu* family of human SINE elements might also increase intraspecies genetic variability through insertions and non-homologous recombination events between elements, leading to various chromosomal rearrangements, duplications and deletions (reviewed in REF. 119). These *Alu* elements also show epigenetic variability at the level of cytosine methylation of individual elements[100,120], producing intraspecies variation.

Likewise, in plants, meiotically stable epigenetic polymorphisms between individuals of the same species are often generated by TEs[121]. This intraspecies variation produced by the epigenetic regulation of TEs generates phenotypic variation, and could potentially contribute to speciation itself, through mechanisms such as hybrid vigour or hybrid dysgenesis.

Conclusions

TEs are a common component in many epigenetic mechanisms, from imprinting and X inactivation to PEV. They have also contributed significantly to the evolution of linear chromosomes, constituting the bulk of pericentromeric and telomeric heterochromatin in most genomes. Likewise, epigenetic regulation of gene expression and chromosome function is strikingly similar to that of TEs, suggesting that they were the original target of this epigenetic regulation. The discovery of transcriptional and post-transcriptional gene-silencing processes that are mediated by RNAi has provided a sequence-specific mechanism that can distinguish TEs from genes. Environmental factors that influence this epigenetic regulation might also have a vital role in disease, natural variation and evolution.

Many questions remain. For example, what role do the fractured and jumbled TEs that compose constitutive heterochromatin have in regulating the full-length elements that lie elsewhere in the genome? Additionally, do antiparallel transcripts, such as those that regulate the *S. pombe* pericentromere, regulate TEs? And how are early expression patterns of TEs and epialleles established differently in identical cells, resulting in variegation? These questions are important and, as we understand more about the epigenetic regulation of TEs, many other epigenetic phenomena are likely to be elucidated because of their common origin.

1. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Girard, L. & Freeling, M. Regulatory changes as a consequence of transposon insertion. *Dev. Genet.* **25**, 291–296 (1999).
3. McClintock, B. Components of action of the regulators *Spm* and *Ac*. *Carnegie Inst. Wash. Year Book* **64**, 527–536 (1965).
4. Griffith, J. L. *et al.* Functional genomics reveals relationships between the retrovirus-like *Ty1* element and its host *Saccharomyces cerevisiae*. *Genetics* **164**, 867–879 (2003).
5. Vastenhouw, N. L. *et al.* A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Curr. Biol.* **13**, 1311–1316 (2003).
6. Tabara, H. *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* **99**, 123–132 (1999).
7. Sijen, T. & Plasterk, R. H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**, 310–314 (2003).
This report demonstrates that the endogenous RNAi system functions to repress TE activity in a tissue- and developmentally specific manner.
8. Watanabe, T. *et al.* Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **20**, 1732–1743 (2006).
This report confirms the previous findings in non-mammalian eukaryotes that TEs are an important source and target of siRNAs.
9. Yang, N. & Kazazian, H. H. Jr. *L1* retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13**, 763–771 (2006).
This publication demonstrated that siRNAs are generated from TEs in human cells, as well as showing that the LINE1 antisense promoter is essential for the production of these siRNAs.
10. Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, e67 (2003).
11. Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871–1873 (2002).
12. Martens, J. H. *et al.* The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
13. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20**, 116–117 (1998).
14. Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P. & Warren, W. D. Asymmetric methylation in the hypermethylated CpG promoter region of the human *L1* retrotransposon. *J. Biol. Chem.* **272**, 7810–7816 (1997).
15. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking *Dnmt3L*. *Nature* **431**, 96–99 (2004).
This study provides evidence of the tissue-specific *de novo* methylation of retrotransposons in mammals, as well as demonstrating that this methylation is essential for viability.
16. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
This study describes the microarray analysis of an entire chromosomal knob, its TE composition and its dependency on DNA methylation and chromatin remodelling.
17. Kato, M., Takashima, K. & Kakutani, T. Epigenetic control of *CACTA* transposon mobility in *Arabidopsis thaliana*. *Genetics* **168**, 961–969 (2004).
18. Yan, Q., Cho, E., Lockett, S. & Muegge, K. Association of *Lsh*, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. *Mol. Cell. Biol.* **23**, 8416–8428 (2003).
19. Huang, J. *et al. Lsh*, an epigenetic guardian of repetitive elements. *Nucleic Acids Res.* **32**, 5019–5028 (2004).
This work demonstrates that the *SWI/SNF Lsh1* gene regulates only repetitive DNA such as TEs, as well as showing the large amount of the genome that is mis-expressed when TE silencing is lost.
20. Martienssen, R. A., Zaratiegui, M. & Goto, D. B. RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends Genet.* **21**, 450–456 (2005).
21. Buhler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).
22. Irvine, D. V. *et al.* Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**, 1134–1137 (2006).
This report demonstrates that the spread of chromatin modifications is dependent on read-through transcription and the catalytic activity of argonaute.
23. Hansen, K. R. *et al.* Global effects on gene expression in fission yeast by silencing and RNA interference machineries. *Mol. Cell. Biol.* **25**, 590–601 (2005).
24. Mochizuki, K. & Gorovsky, M. A. Small RNAs in genome rearrangement in *Tetrahymena*. *Curr. Opin. Genet. Dev.* **14**, 181–187 (2004).
25. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
26. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**, 4671–4679 (2002).
This is the first correlation between the TE-produced longer class of siRNAs in plants and their role in DNA methylation.
27. Qi, Y. *et al.* Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**, 1008–1012 (2006).
28. Herr, A. J., Jensen, M. B., Dalmay, T. & Baulcombe, D. C. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**, 118–120 (2005).
29. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
30. Huettel, B. *et al.* Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* **25**, 2828–2836 (2006).
31. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
32. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nature Genet.* **37**, 761–765 (2005).
33. Li, C. F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106 (2006).
34. Pontes, O. *et al.* The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92 (2006).
35. Kanno, T. *et al.* A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep.* **6**, 649–655 (2005).
36. Matzke, M. A. & Birchler, J. A. RNAi-mediated pathways in the nucleus. *Nature Rev. Genet.* **6**, 24–35 (2005).
37. Cao, X. & Jacobsen, S. E. Role of the *Arabidopsis* DRM methyltransferases in *de novo* DNA methylation and gene silencing. *Curr. Biol.* **12**, 1138–1144 (2002).
38. Weinberg, M. S. *et al.* The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* **12**, 256–262 (2006).
39. Castro, J. P. & Carareto, C. M. *Drosophila melanogaster* P transposable elements: mechanisms of transposition and regulation. *Genetica* **121**, 107–118 (2004).
40. Reiss, D., Josse, T., Anxolabehere, D. & Ronsseray, S. *aubergine* mutations in *Drosophila melanogaster* impair P cytotype determination by telomeric *P* elements inserted in heterochromatin. *Mol. Genet. Genomics* **272**, 336–343 (2004).
This report was the first to link an argonaute gene and RNA-based silencing to the *trans* silencing signal in TE cytotype control.
41. Kavi, H. H., Fernandez, H. R., Xie, W. & Birchler, J. A. RNA silencing in *Drosophila*. *FEBS Lett.* **579**, 5940–5949 (2005).
42. Jensen, S., Gassama, M. P. & Heidmann, T. Taming of transposable elements by homology-dependent gene silencing. *Nature Genet.* **21**, 209–212 (1999).
This important report shows that a short, non-protein-encoding fragment of a TE can suppress hybrid dysgenesis and transposition activity when the full-length TE enters a naïve genome.

43. Vagin, V. V. *et al.* A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**, 320–324 (2006).
This recent report shows that rasiRNAs are structurally different from siRNAs, and they associate with Piwi to silence TEs in *D. melanogaster*.
44. Sarot, E., Payen-Groschene, G., Bucheton, A. & Pelisson, A. Evidence for a Piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* **166**, 1313–1321 (2004).
45. Rehwinkel, J. *et al.* Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **26**, 2965–2975 (2006).
46. Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* **12**, 3715–3727 (1998).
47. Klattenhoff, C. *et al. Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12**, 45–55 (2007).
48. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442**, 203–207 (2006).
49. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714 (2006).
50. Girard, A., Sachidanandam, R., Hannon, G. J. & Carmell, M. A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **442**, 199–202 (2006).
51. Aravin, A. A. *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell.* **5**, 337–350 (2003).
52. Motamedi, M. R. *et al.* Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**, 789–802 (2004).
53. Schueler, M. G. & Sullivan, B. A. Structural and functional dynamics of human centromeric chromatin. *Annu. Rev. Genomics Hum. Genet.* **7**, 301–313 (2006).
54. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
55. Dawe, R. K. & Henikoff, S. Centromeres put epigenetics in the driver's seat. *Trends Biochem. Sci.* **31**, 662–669 (2006).
56. Peters, A. H. *et al.* Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**, 323–337 (2001).
57. De La Fuente, R. *et al.* Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nature Cell Biol.* **8**, 1448–1454 (2006).
58. Pidoux, A. L. & Allshire, R. C. The role of heterochromatin in centromere function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 569–579 (2005).
59. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364–370 (2004).
60. May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L. & Martienssen, R. A. Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genet.* **1**, e79 (2005).
61. Jiang, J., Birchler, J. A., Parrott, W. A. & Dawe, R. K. A molecular view of plant centromeres. *Trends Plant Sci.* **8**, 570–575 (2003).
62. Fukagawa, T. *et al.* Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nature Cell Biol.* **6**, 784–791 (2004).
63. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19**, 489–501 (2005).
This report confirms that, as in lower eukaryotes, dicer and small-RNA-based mechanisms are responsible for centromere condensation in mammals.
64. Ohzeki, J., Nakano, M., Okada, T. & Masumoto, H. CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA. *J. Cell Biol.* **159**, 765–775 (2002).
65. Wong, L. H. & Choo, K. H. Evolutionary dynamics of transposable elements at the centromere. *Trends Genet.* **20**, 611–616 (2004).
66. Martienssen, R. A. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genet.* **35**, 213–214 (2003).
67. Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955–959 (1997).
68. Pardue, M. L. *et al.* Two retrotransposons maintain telomeres in *Drosophila*. *Chromosome Res.* **13**, 443–453 (2005).
69. Casacuberta, E. & Pardue, M. L. RNA interference has a role in regulating *Drosophila* telomeres. *Genome Biol.* **7**, 220 (2006).
70. Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. & Gvozdev, V. Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes Dev.* **20**, 345–354 (2006).
71. Gonzalo, S. *et al.* DNA methyltransferases control telomere length and telomere recombination in mammalian cells. *Nature Cell Biol.* **8**, 416–424 (2006).
72. Schotta, G., Ebert, A., Dorn, R. & Reuter, G. Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin. Cell Dev. Biol.* **14**, 67–75 (2003).
73. Sun, F. L. *et al. cis*-acting determinants of heterochromatin formation on *Drosophila melanogaster* chromosome four. *Mol. Cell. Biol.* **24**, 8210–8220 (2004).
74. Haynes, K. A., Caudy, A. A., Collins, L. & Elgin, S. C. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr. Biol.* **16**, 2222–2227 (2006).
75. Schotta, G., Ebert, A. & Reuter, G. SU(VAR)3–9 is a conserved key function in heterochromatic gene silencing. *Genetica* **117**, 149–158 (2003).
76. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Rev. Genet.* **7**, 703–713 (2006).
77. Lei, E. P. & Corces, V. G. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nature Genet.* **38**, 936–941 (2006).
78. Lyon, M. F. Do LINEs have a role in X-chromosome inactivation? *J. Biomed. Biotechnol.* **2006**, 59746 (2006).
79. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
80. Cohen, D. E. *et al.* The DXPas34 repeat regulates random and imprinted X inactivation. *Dev. Cell* **12**, 57–71 (2007).
81. Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
82. Jordan, I. K., Rogozin, I. B., Glazko, G. V. & Koonin, E. V. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19**, 68–72 (2003).
83. Romanish, M. T., Lock, W. M., de Lagemaat, L. N., Dunn, C. A. & Mager, D. L. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. *PLoS Genet.* **3**, e10 (2007).
84. Garfinkel, D. J. Genome evolution mediated by *Ty* elements in *Saccharomyces*. *Cytogenet. Genome Res.* **110**, 63–69 (2005).
85. Nigumann, P., Redik, K., Matlik, K. & Speek, M. Many human genes are transcribed from the antisense promoter of *L1* retrotransposon. *Genomics* **79**, 628–634 (2002).
This paper reports the discovery that tissue-specific transcripts that are generated from the outward-reading antisense promoter of LINE1 express neighbouring genes in large numbers.
86. Hodgetts, R. B. & O'Keefe, S. L. The mutant phenotype associated with *P*-element alleles of the vestigial locus in *Drosophila melanogaster* may be caused by a readthrough transcript initiated at the *P*-element promoter. *Genetics* **157**, 1665–1672 (2001).
87. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23**, 314–318 (1999).
This landmark report demonstrates that a mosaic coat-colour pattern is defined by the epigenetic state of a nearby TE, and that these epigenetic patterns are not fully erased when passed through meiosis, creating heritable patterns of epigenetic gene expression.
88. Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. & Whitelaw, E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* **2**, e49 (2006).
89. Gaudet, F. *et al. Dnmt1* expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Mol. Cell. Biol.* **24**, 1640–1648 (2004).
90. Matlik, K., Redik, K. & Speek, M. *L1* antisense promoter drives tissue-specific transcription of human genes. *J. Biomed. Biotechnol.* **2006**, 71753 (2006).
91. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
92. Martienssen, R. & Baron, A. Coordinate suppression of mutations caused by Robertson's mutator transposons in maize. *Genetics* **136**, 1157–1170 (1994).
93. Whitelaw, E. & Martin, D. I. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nature Genet.* **27**, 361–365 (2001).
In this provocative commentary, it is suggested that the differential metastable epigenetic silencing of TEs that are scattered throughout the genome produce intraspecies diversity.
94. Walker, E. L. Paramutation of the *r1* locus of maize is associated with increased cytosine methylation. *Genetics* **148**, 1973–1981 (1998).
95. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev.* **16**, 1906–1918 (2002).
96. Alleman, M. *et al.* An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**, 295–298 (2006).
97. Woodhouse, M. R., Freeling, M. & Lisch, D. Initiation, establishment, and maintenance of heritable MuDR transposon silencing in maize are mediated by distinct factors. *PLoS Biol.* **4**, e339 (2006).
98. McGinnis, K. M., Springer, C., Lin, Y., Carey, C. C. & Chandler, V. Transcriptionally silenced transgenes in maize are activated by three mutations defective in paramutation. *Genetics* **173**, 1637–1647 (2006).
99. Rassoulzadegan, M. *et al.* RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**, 469–474 (2006).
100. Rubin, C. M., VandeVoort, C. A., Teplitz, R. L. & Schmid, C. W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res.* **22**, 5121–5127 (1994).
101. Allen, E. *et al.* High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc. Natl Acad. Sci. USA* **100**, 9940–9945 (2003).
In this unique survey of the genome, LINE elements were found to be significantly associated with imprinted genes.
102. Greally, J. M. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl Acad. Sci. USA* **99**, 327–332 (2002).
103. Reik, W. & Lewis, A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nature Rev. Genet.* **6**, 403–410 (2005).
104. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294**, 2536–2539 (2001).
105. Youngson, N. A., Kocialkowski, S., Peel, N. & Ferguson-Smith, A. C. A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. *J. Mol. Evol.* **61**, 481–490 (2005).
106. Ono, R. *et al.* A retrotransposon-derived gene, *PEG10*, is a novel imprinted gene located on human chromosome 7q21. *Genomics* **73**, 232–237 (2001).
107. Davis, E. *et al.* RNAi-mediated allelic *trans*-interaction at the imprinted *Rtl1/Peg11* locus. *Curr. Biol.* **15**, 743–749 (2005).
108. Chan, S. W., Zhang, X., Bernatavichute, Y. V. & Jacobsen, S. E. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol.* **4**, e363 (2006).
109. Kinoshita, T. *et al.* One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303**, 521–523 (2004).
110. Walter, J., Hutter, B., Khare, T. & Paulsen, M. Repetitive elements in imprinted genes. *Cytogenet. Genome Res.* **113**, 109–115 (2006).
111. Sen, R. & Oltz, E. Genetic and epigenetic regulation of *IgH* gene assembly. *Curr. Opin. Immunol.* **18**, 237–242 (2006).
112. Bolland, D. J. *et al.* Antisense intergenic transcription in V(D)J recombination. *Nature Immunol.* **5**, 630–637 (2004).
113. Kidwell, M. G. & Lisch, D. R. Transposable elements and host genome evolution. *Trends Ecol. Evol.* **15**, 95–99 (2000).

114. Hickey, D. A. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics* **101**, 519–531 (1982).
115. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).
116. Wang, Q. & Dooner, H. K. Eukaryotic transposable elements and genome evolution special feature: remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc. Natl Acad. Sci. USA* **103**, 17644–17649 (2006).
117. Boissinot, S., Davis, J., Entezam, A., Petrov, D. & Furano, A. V. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl Acad. Sci. USA* **103**, 9590–9594 (2006).
118. Seleme Mdel, C. *et al.* Extensive individual variation in *L1* retrotransposition capability contributes to human genetic diversity. *Proc. Natl Acad. Sci. USA* **103**, 6611–6616 (2006).
This report demonstrates that there is a surprisingly high level of individual variation and TE-activity potential between humans.
119. Deininger, P. L., Moran, J. V., Batzer, M. A. & Kazazian, H. H. Jr. Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* **13**, 651–658 (2003).
120. Sandovici, I. *et al.* Interindividual variability and parent of origin DNA methylation differences at specific human *Alu* elements. *Hum. Mol. Genet.* **14**, 2135–2143 (2005).
121. Rangwala, S. H. *et al.* Meiotically stable natural epialleles of *Sadhu*, a novel *Arabidopsis* retroposon. *PLoS Genet.* **2**, e36 (2006).
122. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
123. Slotkin, R. K., Freeling, M. & Lisch, D. Heritable transposon silencing initiated by a naturally occurring transposon inverted duplication. *Nature Genet.* **37**, 641–644 (2005).
124. Ronsseray, S., Josse, T., Boivin, A. & Anxolabehere, D. Telomeric transgenes and *trans*-silencing in *Drosophila*. *Genetica* **117**, 327–335 (2003).
125. Axtell, M. J., Jan, C., Rajagopalan, R. & Bartel, D. P. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**, 565–577 (2006).
126. Schlappi, M., Raina, R. & Fedoroff, N. Epigenetic regulation of the maize *Spm* transposable element: novel activation of a methylated promoter by *TnpA*. *Cell* **77**, 427–437 (1994).
127. McClintock, B. The significance of responses of the genome to challenge. *Science* **226**, 792–801 (1984).
128. Capy, P., Gasperi, G., Biemont, C. & Bazin, C. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85**, 101–106 (2000).
129. Allshire, R. C., Javerzat, J. P., Redhead, N. J. & Cranston, G. Position effect variegation at fission yeast centromeres. *Cell* **76**, 157–169 (1994).
130. Kidwell, M. G. Reciprocal differences in female recombination associated with hybrid dysgenesis in *Drosophila melanogaster*. *Genet. Res.* **30**, 77–88 (1977).
131. Vieira, C., Aubry, P., Lepetit, D. & Biemont, C. A temperature cline in copy number for 412 but not *roo/B104* retrotransposons in populations of *Drosophila simulans*. *Proc. Biol. Sci.* **265**, 1161–1165 (1998).
132. Jia, S., Noma, K. & Grewal, S. I. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. *Science* **304**, 1971–1976 (2004).
133. Wolff, G. L., Kodell, R. L., Moore, S. R. & Cooney, C. A. Maternal epigenetics and methyl supplements affect agouti gene expression in A^{vy}/a mice. *FASEB J.* **12**, 949–957 (1998).
134. Cooney, C. A., Dave, A. A. & Wolff, G. L. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J. Nutr.* **132**, 2393S–2400S (2002).
This study demonstrates the surprising result that dietary supplementation of a methylation donor molecule heritably alters a TE-induced epiallele that confers coat colour in mice.
135. Waterland, R. A. & Jirtle, R. L. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).
136. Ivics, Z., Hackett, P. B., Plasterk, R. H. & Izsvak, Z. Molecular reconstruction of Sleeping Beauty, a *Tc1*-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501–510 (1997).
137. Miskey, C., Izsvak, Z., Plasterk, R. H. & Ivics, Z. The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells. *Nucleic Acids Res.* **31**, 6873–6881 (2003).
138. An, W. *et al.* Active retrotransposition by a synthetic *L1* element in mice. *Proc. Natl Acad. Sci. USA* **103**, 18662–18667 (2006).
139. Biemont, C. & Vieira, C. What transposable elements tell us about genome organization and evolution: the case of *Drosophila*. *Cytogenet. Genome Res.* **110**, 25–34 (2005).
140. Kidwell, M. G. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49–63 (2002).
141. Hawkins, J. S., Kim, H., Nason, J. D., Wing, R. A. & Wendel, J. F. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **16**, 1252–1261 (2006).
142. Piegu, B. *et al.* Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262–1269 (2006).
143. Iida, S., Morita, Y., Choi, J. D., Park, K. I. & Hoshino, A. Genetics and epigenetics in flower pigmentation associated with transposable elements in morning glories. *Adv. Biophys.* **38**, 141–159 (2004).
144. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074 (2001).
145. Kapitonov, V. V. & Jurka, J. RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons. *PLoS Biol.* **3**, e181 (2005).
This study provides the missing evidence of the TE-family origin of the RAG transposase proteins.

Acknowledgements
The authors apologize if relevant publications were not cited due to space constraints. R.K.S is supported by a postdoctoral fellowship from the US National Institutes of Health. Work in the authors' laboratory is supported by a grant from the National Institutes of Health to R.M.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
The following terms in this article are linked online to:
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
AGO4 | agouti | Argonaute | *CMT3* | *DCL3* | *Dnmt1* | *Dnmt3a* | *DRD1* | *DRM2* | *FWA* | *MET1* | *Peg10* | *Rag1* | *Rag2* | *Rtl1* | *vestigial*
UniProtKB: http://ca.expasy.org/sprot
Aubergine | CENPB | Clr4 | DDM1 | dicer | DNMT1 | Piwi
Access to this links box is available online.