

# Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions

Jeffrey S Han

## Abstract

Non-long terminal repeat (non-LTR) retrotransposons are present in most eukaryotic genomes. In some species, such as humans, these elements are the most abundant genome sequence and continue to replicate to this day, creating a source of endogenous mutations and potential genotoxic stress. This review will provide a general outline of the replicative cycle of non-LTR retrotransposons. Recent findings regarding the host regulation of non-LTR retrotransposons will be summarized. Finally, future directions of interest will be discussed.

## Introduction

The biological drive to replicate makes it almost inevitable that selfish genetic elements will populate genomes [1]. Indeed, genome sequencing has revealed that single copy genes are often vastly outnumbered by repetitive transposable elements [2-6]. The abundance and distribution of any particular transposon depends on how aggressive the transposon is (how quickly it can multiply in copy number), where the transposon inserts new copies, and how the host responds. Since these factors can vary greatly, the transposon content of each species is unique and virtually impossible to predict *a priori*. In humans, this number is at least 45% [2].

Transposons can be divided into two broad classes: DNA transposons and retrotransposons. DNA transposons replicate via a cut and paste mechanism [7], whereas retrotransposons replicate using an RNA intermediate. Retrotransposons can be further subdivided into long terminal repeat (LTR) and non-LTR retrotransposons. LTR retrotransposons are retroviral-like in structure and mechanism [8]. Non-LTR retrotransposons (also called LINEs, polyA retrotransposons, or target-primed (TP) retrotransposons), as implied by their name, do not contain LTRs and instead take on the likeness of an integrated mRNA. They are ancient genetic elements that have persisted in eukaryotic genomes for hundreds of

millions of years [9], and are perhaps best known for their enormous success multiplying in the human genome. Although non-LTR retrotransposons can be parasitized by non-autonomous elements (for example, short interspersed transposable elements (SINEs)), this review will focus on autonomous elements, which encode the protein machinery necessary for their self-mobilization.

## Non-LTR retrotransposons: general structure

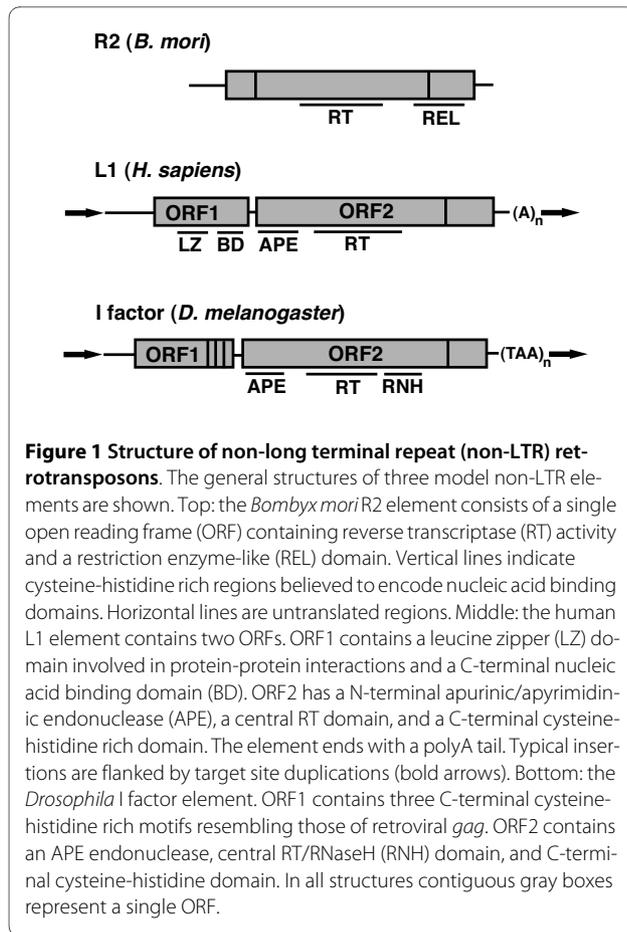
Full length, autonomous non-LTR retrotransposons typically contain one or two open reading frames (ORFs). The general structure of three model examples, the *Bombyx mori* R2 element, the human L1 element, and the *Drosophila melanogaster* I factor, are depicted in Figure 1. Central to retrotransposon mobilization is reverse transcriptase (RT) activity, and thus all autonomous non-LTR retrotransposons contain an RT domain. Also present in virtually all non-LTR retrotransposons is an endonuclease (endo) domain [10-12], although encoded endo activity is not an absolute requirement for non-LTR retrotransposition (see Mechanisms below). A second ORF (ORF1) appears to have been an early evolutionary addition [12]. ORF1 contains RNA binding activity [13-15] and nucleic acid chaperone activity [16], and may play a similar role to the *gag* proteins of retroviruses.

The 5' and 3' untranslated regions (UTRs) of non-LTR retrotransposons are quite variable. Although most characterized 5' UTRs of functional non-LTR retrotransposons contain internal promoter activity [17-25], promoter replacement is frequent [26,27] and therefore promoter

\* Correspondence: han@ciwemb.edu

<sup>1</sup> Department of Embryology, Carnegie Institution of Washington, Baltimore, MD, USA

Full list of author information is available at the end of the article



sequence is not necessarily conserved among elements from different species. Evidence suggests that these promoters are transcribed by RNA polymerase II. There are also elements with no apparent endogenous promoter. For example, the R2 elements, which are found in insect ribosomal DNA loci, do not appear to encode their own promoter, and are thought to be cotranscribed with their host rRNA repeat [28]. The 3' UTR of non-LTR elements usually contains a specific sequence/structure that is recognized by the reverse transcriptase ORF [29-32]. A notable exception is the mammalian L1 element, for which the 3' UTR is dispensable for retrotransposition [33]. The 3' boundary of non-LTR retrotransposons can consist of polydeoxyadenosine (polyA) sequence, short sequence repeats [30,34], or neither. Because of the wide variability in the UTRs of non-LTR retrotransposons, we hesitate to make sweeping generalizations on specific details of non-LTR transcription initiation and termination. Although Figure 1 shows full-length elements, non-LTR retrotransposons in genomes are often 5' truncated due to incomplete reverse transcription [35]. Only a subset of elements are full length and active; for example, out of 500,000 L1s in the human genome, only approximately

7,000 are full length, and of those only 80 to 100 are estimated to be active for retrotransposition [36].

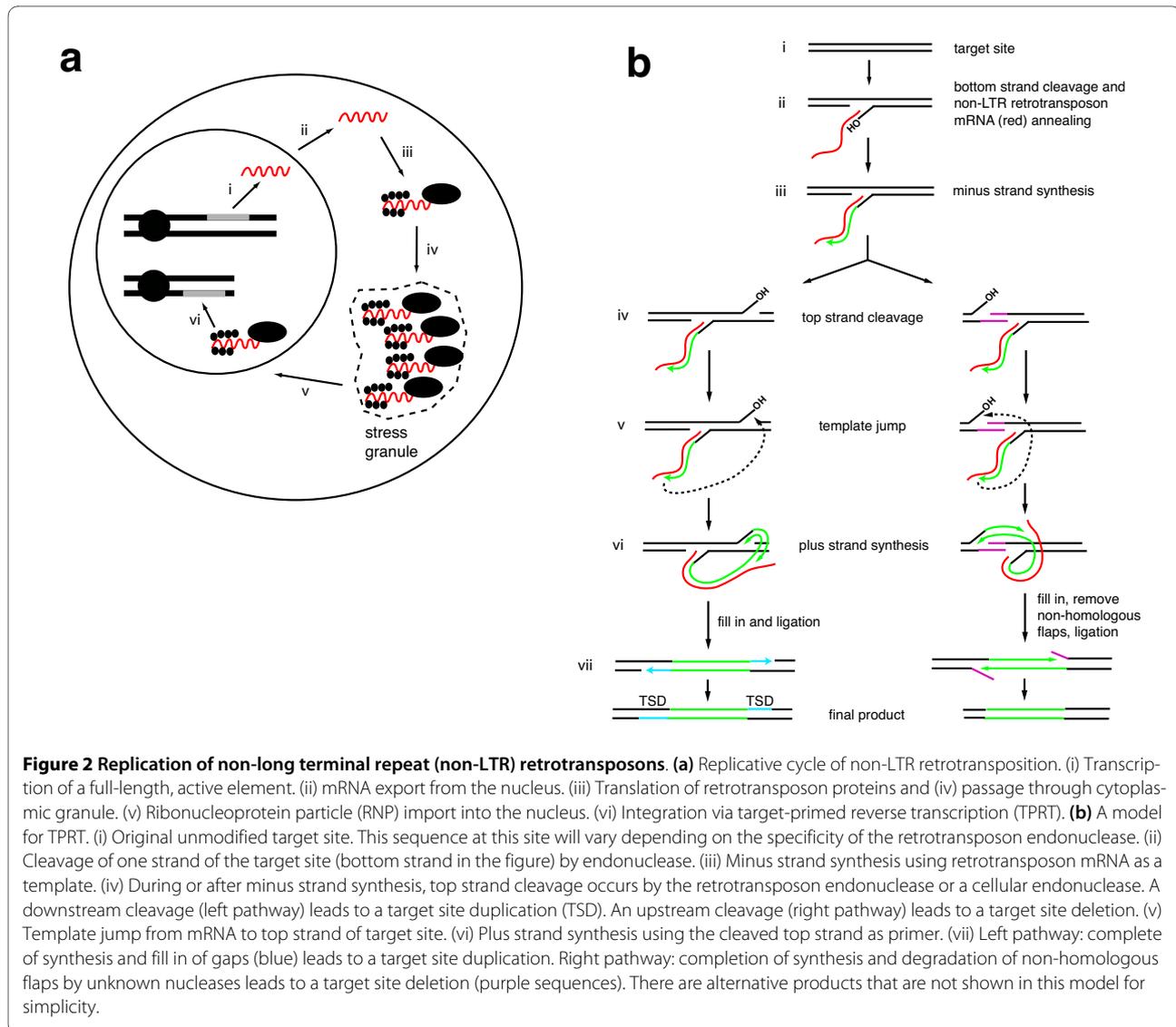
## Non-LTR retrotransposon replication mechanisms

### Outline of life cycle

A general outline of non-LTR retrotransposon replication is shown in Figure 2a. Following transcription and nuclear export, the ORF(s) are translated and assemble to form a ribonucleoprotein particle (RNP) [37-39]. For pol II transcribed elements, the first ORF is most likely translated by the traditional cap recognition and scanning model. This mode of translation (as opposed to an internal ribosomal entry site (IRES)) is supported by experiments in mammals demonstrating that the 5' UTR, 3' UTR, and interORF sequence of L1 are dispensable for retrotransposition ([33]; JSH and Jef Boeke, Johns Hopkins University, unpublished results), the coding region of ORF1 can be extensively mutated without compromising retrotransposition [40], and that translation initiated from the human L1 5' UTR is cap dependent [41]. Elements transcribed by pol I, such as the R1/R2 elements, may be translated via an IRES [28], although this awaits experimental confirmation. In bicistronic transcripts, translation of the second ORF2 relies upon unconventional translation initiation. Studies with the silkworm SART1 element and human L1 element suggest similar translation mechanisms, where ORF2 translation is dependent on ORF1 translation [42,43]. The ribosome that translated ORF1 may reinitiate translation at ORF2, or may recruit another ribosome to serve this purpose.

Using ORF1 as a marker, non-LTR retrotransposon RNPs in most cases are localized predominantly in the cytoplasm [37,39,44-51]. An interesting exception has been reported in rat, where endogenous rat L1 ORF1 protein is concentrated in the nucleus in a chronic myeloid leukemia (CML) cell line [52]. The reason for this anomalous expression pattern is unclear, although one plausible explanation is the production of a mutated version of ORF1 in this particular cell line that escapes cytoplasmic retention [50,52]. Furthermore, a cloned, retrotranspositionally competent rat L1 produced predominantly cytoplasmic RNPs when localized in other cell lines [52]. Other notable examples of nuclear non-LTR retrotransposon RNPs are the telomere targeted elements, which are likely beneficial to the cell [53,54].

Closer examination of non-LTR retrotransposon RNPs in mammals has revealed localization to stress granules, cytoplasmic bodies closely associated with P bodies [50]. Stress granules have been described as a repository for non-translating mRNAs [55]. It is not clear whether stress granule-associated non-LTR RNPs are destined for destruction, or whether transit through the stress granule is an important step in the maturation of a retrotranspo-



son RNP. However, it is noteworthy that loss of P body components in yeast lead to a decrease in Ty1 and Ty3 (LTR retrotransposons) activity, suggesting an important role of P bodies and/or stress granules in the life cycle of retrotransposons [56-60]. The dynamics of non-LTR retrotransposon RNPs have been studied most dramatically with the *Drosophila I* factor [39,47]. *I* factor RNPs appear in nurse cells then are translocated to the oocyte cytoplasm along with other nurse cell components. In the oocyte, the RNPs migrate from the posterior to anterior cortex, following the microtubule organizing center and the nucleus [39,47]. *I* factor RNP migration is reminiscent of *bicoid* mRNA/Staufen protein RNP transport [47,61,62]. The microtubule dependence of this process suggests that non-LTR RNPs may also track along the cytoskeleton. There are also hints that L1 ORF1 protein interacts with intermediate filaments, further bolstering this hypothesis [52]. At some point, a non-LTR re-

trotransposon RNP must have access to a chromosome, and therefore is presumably transported into the nucleus [63,64], where reverse transcription occurs.

#### Target-primed reverse transcription

Once at a suitable chromosomal target site, a non-LTR retrotransposon can begin to copy its genetic information at this new locus. The mechanism for this process has been termed target-primed reverse transcription (TPRT) and the initial stages have been demonstrated most convincingly with *in vitro* studies of the *Bombyx mori* R2 element. Purified R2 protein, when mixed with R2 RNA and a suitable DNA target site, is able to nick the DNA on one strand (the bottom strand in Figure 2b). The resulting free 3' DNA end serves as a primer to reverse transcribe the R2 RNA (minus strand synthesis). A similar system using human L1 ORF2 has also provided evidence for this reaction, although products could not be monitored directly due to their low yield and non-uniformity (L1

ORF2, unlike the R2 ORF, does not nick a rigidly defined target DNA). The target site of non-LTR retrotransposons is dictated by the encoded endonuclease domain. Ancient non-LTR clades, such as CRE, R2, and R4, encode restriction enzyme-like endonucleases [12] that recognize specific sequences. This endo domain was replaced by an abasic endonuclease early in non-LTR evolution [12], leading in most cases to relaxed target site specificity. The importance of the endonuclease domain lies in producing a primer for reverse transcription. If a suitable primer can be created via another mechanism, non-LTR retrotransposition can proceed in an endo-independent manner [65]. Perhaps the most dramatic example of endo-independent non-LTR retrotransposition occurs naturally in *Drosophila*, where the codependent non-LTR retrotransposons TART, HeT-A and TAHRE retrotranspose at chromosome ends to form the telomeres [54,66-68]. A similar reaction has been recapitulated on mammalian chromosome ends with an engineered endo mutant human L1 [69], suggesting the possibility of a 'pre-endo' era of non-LTR retrotransposons that extended and/or repaired free DNA ends.

The remaining steps of non-LTR integration remain somewhat murky but are hypothesized to occur as follows (see Figure 2b). After or during minus strand synthesis, a second strand nick of the target site occurs (on the top strand of Figure 2b). Depending on the specific non-LTR retrotransposon involved, the second strand nick can occur downstream, upstream, or in line with the bottom strand nick to generate target site duplications (TSDs), target site deletions, or blunt insertions. The non-LTR endo domain is presumed although not proven to also make this second nick, and evidence suggests that at least one non-LTR RT/endo ORF works as a dimer, with each subunit catalyzing one of the nicks [70,71]. How the position of the second strand nick is selected is not entirely clear and probably differs between different elements. In some cases, the second strand nick is likely predefined by the retrotransposon endo domain. For mammalian elements, there is some evidence that microhomology to the newly synthesized bottom strand dictates second strand cleavage [72,73]. In most cases, however, we simply do not know how the second strand cleavage occurs.

Second strand cleavage creates a primer for plus strand synthesis. The newly synthesized minus strand jumps templates from the retrotransposon RNA to the target site DNA top strand. This strand exchange is possibly facilitated by the addition of non-templated nucleotides [74] and/or the assistance of the ORF1 nucleic acid chaperone activity [16]. Plus strand synthesis proceeds from the second nick. Some non-LTR clades have acquired an RNase H domain, which likely removes mRNA template as minus strand synthesis occurs [12,75]. Others have no

evidence of an RNase H domain, and the mRNA may simply be displaced by second strand synthesis [76]. In the case of a downstream second strand nick, gaps are filled in to produce TSDs (Figure 2b, left pathway). In the case of an upstream second strand nick (Figure 2b, right pathway), non-homologous flaps are removed by unknown factors to generate target site deletions. After DNA synthesis is complete, the remaining nicks are ligated to complete the insertion.

#### Host-encoded factors facilitating non-LTR retrotransposons

It is generally believed that host-encoded factors are required for a complete integration (TPRT) reaction [77]. There are multiple reasons for this line of thought. No non-LTR retrotransposon has ever been shown (biochemically or bioinformatically) to encode proteins with all the expected abilities required to complete such a reaction. For example, the nicks generated by TPRT need to be sealed after DNA synthesis, and encoded ligase activity has not been found within non-LTR elements. In the case of target site deletions, unidentified nucleases eliminate unpaired DNA (Figure 2b, vii, right pathway). In addition, the complete retrotransposition event has never been recapitulated *in vitro* with purified substrate and non-LTR proteins. This of course could simply be due to the complex multistep nature of non-LTR integration and our inability to find the appropriate test tube conditions. However, sequence analysis of genomic integration events suggests the involvement of host DNA repair machinery in some instances of non-LTR integration [72,78]. Finally, experimental data suggests a dependence of non-LTR retrotransposons on host DNA repair pathways. Transfection of a human L1 plasmid into mammalian cells leads to the generation of endodependent DNA double strand breaks (DSBs), and successful L1 retrotransposition in this tissue culture system depends on the ATM kinase involved in homologous recombination and non-homologous end joining (NHEJ) [79]. More recently, cells deficient in the NHEJ factors Ku70, Artemis, and LigIV showed partial inhibition of retrotransposition product formation [80]. Similar studies of the *Lactococcus lactis* L1.LtrB group II intron have revealed multiple *Escherichia coli* DNA repair factors required for intron retrohoming [81]. These bacterial introns share a common ancestry with non-LTR retrotransposons and thus are highly relevant to understanding non-LTR mechanisms [9]. Overall, these data implicate host factors in the resolution of TPRT; however, the limitations of the assays used allow alternative explanations (see Lingering questions below). Non-LTR retrotransposons are not unique in this regard, as other transposon classes also require host factors for resolution of transposition intermediates and/or repair of the donor site.

### Host regulatory mechanisms limiting non-LTR retrotransposons

From the perspective of a transposon, it makes sense to be most active in the germ line or sexual phase of the host organism's life cycle; after all, these are the cells that will give rise to the next generation. In contrast, transposition events in somatic cells will be evolutionarily non-productive, since those new transposition events are lost when the host organism dies. As would be predicted, non-LTR retrotransposons are expressed and retrotranspose in germ cells [45-47,82]. While transposon 'success' depends on transposition in the germ line, the genomic insult of excess transposition can have negative effects on the host. Therefore, host cells generally repress transposon activity. This implies that germ cells are a major evolutionary battleground between transposons and the host, where loss of transposon repression can have dramatic consequences. For example, hybrid dysgenesis is a long known phenomenon in *Drosophila* where the progeny of transposon-naïve females and active transposon-containing males are subject to a range of syndromes including mutation, chromosomal rearrangements, and sterility [83,84]. This syndrome is due to the host's inability to control transposon activity. More recently, the devastating effect of unleashing excess transposition in the mammalian germ line has become apparent, where upregulation of L1 and IAP elements are postulated to lead to sterility [85-88].

#### Small RNA pathways

The discovery of RNA interference (RNAi) [89] led to the hypothesis that small RNAs are involved in silencing transposons via DNA methylation and degradation of transposon RNA. Forward genetic screens revealed that RNAi mutants can indeed lead to general transposon derepression [90,91]. We now know that cells express various classes of small RNAs: microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs), and Piwi interacting RNAs (piRNAs) are among the most studied. A significant fraction of endo-siRNAs and piRNAs correspond to transposon sequence, implying a sequence-based recognition and repression mechanism [92-107]. These RNAs interact with Argonaute protein family members in order to carry out their regulatory functions [108]. piRNAs are germ cell specific, while endo-siRNAs are ubiquitous.

A model for piRNA-mediated silencing of transposons in the germ line has emerged from elegant experiments in *Drosophila* and mice. In this model, two classes of piRNAs associate with different Argonaute family members. A class of piRNAs corresponding to the sense orientation of transposon transcripts associates with Piwi and Ago3 (MILI in mice) [85,94], and a class of piRNAs antisense with respect to transposon mRNA associates with Aubergine (MILI2 in mice) [94,109]. These classes of piRNAs

are amplified by a 'ping-pong' mechanism whereby the sense strand piRNA directs cleavage and formation of an antisense strand, and *vice versa* [94]. The origin of the antisense piRNAs, in many cases, can be traced to long piRNA precursors transcribed from a small number of discrete piRNA clusters in the genome [94], some of which are already known to be 'master regulators' of transposon control [110-116].

Mutation of endo-siRNA and piRNA Argonaute partners leads to transposon upregulation [85-87,117-119]. Transposon upregulation is associated with defects in germ cell development, emphasizing the importance of transposon silencing [86-88]. In germ cells, ping-pong amplification is presumed to not only cleave transposon mRNA leading to post-transcriptional repression, but also serves to generate sequence-specific substrates to guide DNA methylation. It is well known that transposons are associated with DNA methylation, and this methylation plays an important role in long-term transposon silencing [120,121]. Mutations in MILI and MIWI2 lead not only to piRNA defects but also failure of proper transposon DNA methylation, providing a genetic link between piRNA production and the eventual methylation of transposon DNA [87]. It is assumed that the piRNA effector complexes somehow recruit the *de novo* DNA methyltransferases DMNT3A, DMNT3B, and their associated protein DMNT3L to methylate transposon loci. How this is accomplished is not yet clear, and the molecular mechanisms underlying these processes are currently an area of active investigation [122-124]. It is likely that this process will resemble well studied pathways of heterochromatin formation in various organisms (reviewed in [125,126]). We should also mention that these mechanisms are not specific to non-LTR retrotransposons but apply to all transposons.

#### Intracellular defense by the APOBEC protein family

The antiviral activity of the APOBEC3G cytidine deaminase was first recognized for its ability to inhibit the HIV retrovirus [127,128]. APOBEC3G cytosine deamination of the newly synthesized HIV cDNA strand results in G to A hypermutation in the viral coding sequence. Since retroelements synthesize a similar nascent cDNA strand, this finding led to the hypothesis that APOBEC family members similarly inhibit retrotransposons. Indeed, all APOBEC3 family members have been shown to inhibit retrotransposon activity [129-135] using a cell line-based assay, although there are inconsistencies in the literature as to whether APOBEC3G inhibits L1 elements. These inconsistencies may be due to varying levels of endogenous APOBEC3 proteins expressed in laboratory cell lines [135]. In addition, the *activation-induced deaminase* (AID) gene, which may share common ancestry with the APOBEC3 proteins, also inhibits L1 in the same cell line-based assay [136].

Surprisingly, hypermutation of the integrated retrotransposition product does not occur upon APOBEC3 member coexpression with non-LTR retrotransposons, and cytidine deaminase activity was shown to be dispensable for the antiretrotransposon effect in most but not all cases (a separation of function of antiviral activity from cytidine deaminase activity has also been shown for HIV [137]). Thus, there are likely at least two distinct mechanisms for non-LTR retrotransposon inhibition by APOBEC family members: deaminase dependent and deaminase independent. The lack of coding strand G to A hypermutation in the deaminase-dependent mechanism could be due to the quick degradation of nascent cDNA minus strands containing uracil, preventing the resolution of deaminase-modified retrotransposition intermediates. The mechanism of deaminase-independent APOBEC3 inhibition of non-LTR retrotransposons is currently unclear, although localization of the specific proteins may give clues to possible function. APOBEC3A, B, C, and H, inhibitors of human L1, localize at least partially to the nucleus [131,135], suggesting a possible function in blocking L1 integration. APOBEC3DE, F, and G appear predominantly in the cytoplasm. They may serve a role in cytoplasmic sequestration of RNPs, although a fraction of these APOBECs may also shuttle to and perform a function in the nucleus [131]. Recently APOBEC3G has been shown to localize to P granules and stress granules [138]. Given the presence of L1 RNPs in stress granules [50], APOBEC3G may sequester L1 in these granules. Coexpression of ORF1 protein with Alu (a non-autonomous non-LTR parasite of L1) shifts a fraction of Alu transcripts from the nucleus to ORF1p-containing stress granules [51]. Interestingly, APOBEC3G also inhibits Alu elements [139,140], and the colocalization of APOBEC3G, L1, and Alu suggest a common (but poorly understood) mechanism.

It should be noted that APOBEC3 proteins have recently been implicated in the clearance of foreign DNA introduced into cultured cells, irrespective of whether the foreign DNA encodes a transposable element [141]. This restriction of foreign DNA also has both deaminase-dependent and deaminase-independent pathways, and perhaps casts some doubt on the specificity of APOBEC retroelement control. However, these data are somewhat at odds with controls performed in prior studies [132,142]. Ultimately, definitive proof for biologically relevant regulation of non-LTR retrotransposons by APOBECs will likely require the demonstration of an APOBEC3-related effect on endogenous retroelements in an organism.

#### **Lingering questions**

There are still large gaps in our knowledge of non-LTR retrotransposon biology. Here I will discuss open ques-

tions that I feel are of outstanding interest. Since we are humans, and the non-LTR L1 element is that only known autonomous transposon in the human genome, these questions are framed in the context of mammalian L1 biology. However, some these questions can be addressed in various model systems, as cell biology between species is remarkably conserved.

#### ***What is the natural frequency of human retrotransposition?***

'Back of an envelope' calculations have been used to estimate that, among every 10 to 20 new humans born, there is on average one new L1 retrotransposition event [143,144]. These calculations were based on multiplying the fraction of known human mutations caused by retrotransposition by an estimated total mutation rate per generation. Possible sources for error in these calculations are the underestimation of retrotransposition events due to detection bias and the reliance on phenotypes to find mutations, which only samples a subset of genes and genomic regions. Human retrotransposition frequency has also been estimated at 1 new insertion per 2 to 33 individuals, based on the summed retrotransposition frequency of tagged active L1s in a tissue culture or transgenic mouse retrotransposition assay [33,36,145]. Although consistent with previous estimates, these assays carry the caveat that they are performed in non-natural environments for a human retrotransposon (somatic cells and live mice). Current projects to catalog retrotransposon polymorphisms in humans will provide another way to estimate retrotransposition rates based on allele frequencies. However, the advent of next generation sequencing [146] and rapidly decreasing per base sequencing costs should make it possible in the near future to directly determine retrotransposition frequency by sequencing and comparing the genomes of successive family generations. This would not only give an average frequency of retrotransposition, but would also be able to highlight the potential wide variability of retrotransposition frequency between human individuals. This individual variability is due to variation of the number of 'hot' (highly active) L1s in a particular genome, as well as mutation of 'hot' L1s to give 'cool' alleles [147,148].

#### ***What pathways are at the LINE-host interfaces?***

Previously, technical limitations have precluded laboratories from performing large systematic screens for factors that regulate LINES. The studies leading to the recent exciting findings that small RNAs and APOBECs inhibit retroelements actually originated from non-retrotransposon laboratories. However, these serendipitous findings hint that careful examination will yield even more host pathways that interact with L1-like elements. In the past decade, the development of technology to knock down mammalian gene expression with short interfering RNAs [149] has made it possible to assess candidate gene contributions to retrotransposition without the time con-

suming step of manipulating mammalian chromosomes. This should allow the eventual screening of many, if not all, mammalian genes for effects on retrotransposition in standard tissue culture [33]. Simple yeast assays are also now available for LINE retrotransposition [150,151]. The fast and low cost methods of yeast genetics provide an attractive complementary system to probe the non-LTR/host interface.

Even with these technical advances, the limitations of current retrotransposition assays must be kept in mind when interpreting the results of a genetic screen. TPRT by definition involves breakage of DNA strands, and is therefore potentially genotoxic. The rate of L1 RNP formation, DNA damage, and stress (virtually every cell expressing active L1) far exceeds the rate of successful retrotransposition in a typical human cell culture assay (1 in 20 to 1 in 10,000) [79,152-154]. Therefore, one can easily imagine that any mutation that modifies the ability to tolerate DNA damage or stress will alter the apparent retrotransposition frequency, even if the true retrotransposition frequency remains unchanged. The retrotransposition assays are also not yet adapted for genetic screens in the natural L1 environment (germ cells). Although screens in embryonic stem cells [155] or sporulating yeast may come closer to approximating the natural context of LINE retrotransposition, the germ cell environment is quite unique and may contain non-somatic processes that stimulate or inhibit retrotransposition. Finally, unraveling the contribution of potential 'hits' to retrotransposition will be a daunting task, as the assay gives no indication as to what step in the poorly understood L1 life cycle is affected, and we still do not have a way to directly monitor which cells are undergoing active retrotransposition. This is perhaps not such a bad thing, as it will keep us occupied for years to come.

#### **Localization of L1 proteins: what are we really looking at?**

The vast excess of L1-containing cytoplasmic granules as compared to the actual number of successful retrotransposition events leads us to our next question. If TPRT occurs in the nucleus, why are RNPs overwhelmingly in the cytoplasm? Are we simply looking at non-functionally relevant, retrotransposon trash? An answer to this question is suggested by looking at the distribution of non-LTR retrotransposon RNPs in *Drosophila*. *Drosophila* has 20 families of non-LTR retrotransposons [156]. Rashkova *et al.* examined the location of five non-LTR element ORF1s, which presumably track with RNPs [48]. The *Het-A* and *TART* elements, which retrotranspose onto chromosome ends to form telomeres, are believed to perform an important cellular function for the host since *Drosophila* lacks telomerase enzyme. The RNPs for *Het-A* and *TART* were found to be efficiently transported into the nucleus. In contrast, RNPs for the euchromatic-targeted non-LTR retrotransposons *jockey*, *doc*, and *I* factor were

predominantly sequestered in the cytoplasm. Thus it seems that host cells have a surveillance system that is able to distinguish 'friendly' from 'unfriendly' RNPs, and target only the latter for inhibition. It remains an open question whether these sequestered RNPs are still active and ready for retrotransposition, or marked for certain death. Is retrotransposition the result of a lucky RNP that escapes the stress granule, or is there an as of yet undetected 'true' retrotransposition pathway that bypasses these granules altogether? Is there a rare activating signal that can induce retrotransposition? Genetics, improvements in retrotransposition frequency, and the development of sensitive tools to visualize RNP transport and action in the nucleus will go a long way towards answering these questions.

#### **Do we need retrotransposon activity?**

A glance through genomes of higher eukaryotes will show something akin to 'genes floating on a sea of retrotransposons' [157]. How can these repeated sequences multiply to such high numbers if they do not provide some function for the host? A common mistake is to think of transposons as being under typical Mendelian constraints. This is not the case. Transposons can replicate and increase in number at a rate faster than single copy genes in the genome. In principle, an aggressively replicating transposon can rapidly sweep through a sexual population without conferring any fitness benefit [158]. The increase of transposon presence with each successive generation must simply outweigh any negative fitness effects.

Even if not conferring a net fitness benefit, transposons provide a source of endogenous mutation. Rather than simple point mutations, deletions and rearrangements, transposon mutations have the advantage of dispersing prebuilt functional units, such as transcription regulatory sequences, protein-binding DNA, and open reading frames encoding nucleic acid binding, cleavage, and synthesis activities. Thus, it is not surprising that after millions of years of evolution some of the many transposon bits and pieces littered among genomes have been put to use by the host cell. For example, genome rearrangements such as V(D)J recombination and programmed DNA deletion in ciliates are performed by proteins that share ancestry with transposases [159,160]. Retrotransposon sequences have also been incorporated into the coding sequence of many human proteins [161]. In addition, the silencing of retrotransposons can lead to epigenetic effects on adjacent genes [162]. The abundance of L1 elements on the X chromosome has even led to the theory that retrotransposons are involved in X inactivation spreading [163]. It is important, however, to note that host utilization of transposon remnants does not necessitate or even imply that active (retro)transposition is an important functional component of host cell biology. The

only case of probable host requirement for retrotransposon activity that I am aware of is the aforementioned telomere maintenance of some insects, which is performed by non-LTR retrotransposons. Overall evidence suggests that transposons do not 'come in peace', and only after host adaptation are they subdued and in rare cases turned from the 'dark side'. Nevertheless, the presence of at least one case of a 'good' retrotransposon hints that there may be other interesting examples of symbiotic active retrotransposons awaiting discovery.

## Conclusions

The past 20 years has seen remarkable progress in our knowledge of non-LTR retrotransposon biology. Non-LTR retrotransposons are major components of eukaryotic genomes and our cells have evolved elaborate mechanisms to deal with these selfish elements. Since the L1 is the only active autonomous transposon in humans, and has directly or indirectly produced over one-third of our genome sequence, non-LTR retrotransposon biology is particularly important for understanding human genome evolution. The genotoxic effects of retrotransposons in mice and tissue culture also suggest that further investigation into the mechanisms of non-LTR retrotransposition will allow us to manipulate retrotransposon activity to confer potential health benefits. Finally, insight into host-retrotransposon interactions will give us a clearer picture on how cells recognize and silence DNA repeats; a question relevant to all of biology.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

JSH is supported by the Carnegie Institution of Washington

## Author Details

Department of Embryology, Carnegie Institution of Washington, Baltimore, MD, USA

Received: 1 February 2010 Accepted: 12 May 2010

Published: 12 May 2010

## References

1. Dawkins R: *The Selfish Gene* UK Oxford University Press; 1976.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrum J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, et al: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
3. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, et al: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
4. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reilly AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, et al: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**:1112-1115.
5. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, et al: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*).** *Science* 2002, **296**:92-100.
6. Wicker T, Robertson JS, Schulze SR, Feltus FA, Magrini V, Morrison JA, Mardis ER, Wilson RK, Peterson DG, Paterson AH, Ivarie R: **The repetitive landscape of the chicken genome.** *Genome Res* 2005, **15**:126-136.
7. Craig NL, Craigie R, Gellert M, Lambowitz AM: *Mobile DNA II* Washington DC USA: ASM Press; 2002.
8. Boeke JD, Stoye JP: **Retrotransposons, endogenous retroviruses, and the evolution of retroelements.** In *Retroviruses* Edited by: Coffin JM, Hughes SH, Varmus HE. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press; 1997:343-435.
9. Eickbush TH, Malik HS: **Origins and evolution of retrotransposons.** In *Mobile DNA II* Edited by: Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington, DC: ASM Press; 2002:1111-1144.
10. Yang J, Malik HS, Eickbush TH: **Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements.** *Proc Natl Acad Sci USA* 1999, **96**:7847-7852.
11. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD: **Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.** *Cell* 1996, **87**:905-916.
12. Malik HS, Burke WD, Eickbush TH: **The age and evolution of non-LTR retrotransposable elements.** *Mol Biol Evol* 1999, **16**:793-805.
13. Hohjoh H, Singer MF: **Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon.** *EMBO J* 1997, **16**:6034-6043.
14. Dawson A, Hartwood E, Paterson T, Finnegan DJ: **A LINE-like transposable element in *Drosophila*, the I factor, encodes a protein with properties similar to those of retroviral nucleocapsids.** *EMBO J* 1997, **16**:4448-4455.
15. Kolosha VO, Martin SL: **High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1).** *J Biol Chem* 2003, **278**:8112-8117.
16. Martin SL, Bushman FD: **Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon.** *Mol Cell Biol* 2001, **21**:467-475.
17. Mizrokhi LJ, Georgieva SG, Ilyin YV: **Jockey, a mobile *Drosophila* element similar to mammalian LINES, is transcribed from the internal promoter by RNA polymerase II.** *Cell* 1988, **54**:685-691.
18. McLean C, Bucheton A, Finnegan DJ: **The 5' untranslated region of the I factor, a long interspersed nuclear element-like retrotransposon of *Drosophila melanogaster*, contains an internal promoter and sequences that regulate expression.** *Mol Cell Biol* 1993, **13**:1042-1050.
19. Minchiotti G, Di Nocera PP: **Convergent transcription initiates from oppositely oriented promoters within the 5' end regions of *Drosophila melanogaster* F elements.** *Mol Cell Biol* 1991, **11**:5171-5180.
20. Contursi C, Minchiotti G, Di Nocera PP: **Identification of sequences which regulate the expression of *Drosophila melanogaster* Doc elements.** *J Biol Chem* 1995, **270**:26570-26576.
21. Swergold GD: **Identification, characterization, and cell specificity of a human LINE-1 promoter.** *Mol Cell Biol* 1990, **10**:6718-6729.
22. Minkami R, Kurose K, Etoh K, Furuhashi Y, Hattori M, Sakaki Y: **Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element.** *Nucleic Acids Res* 1992, **20**:3139-3145.
23. Adey NB, Tollefsbol TO, Sparks AB, Edgell MH, Hutchison CA: **Molecular resurrection of an extinct ancestral promoter for mouse L1.** *Proc Natl Acad Sci USA* 1994, **91**:1569-1573.
24. DeBerardinis RJ, Kazazian HH Jr: **Analysis of the promoter from an expanding mouse retrotransposon subfamily.** *Genomics* 1999, **56**:317-323.
25. Schumann G, Zundorf I, Hofmann J, Marschalek R, Dingermann T: **Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the**

- Dictyostelium* repetitive element, from *Dictyostelium discoideum*. *Mol Cell Biol* 1994, **14**:3074-3084.
26. Haas NB, Grabowski JM, North J, Moran JV, Kazazian HH, Burch JB: Subfamilies of CR1 non-LTR retrotransposons have different 5'UTR sequences but are otherwise conserved. *Gene* 2001, **265**:175-183.
  27. Khan H, Smit A, Boissinot S: Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* 2006, **16**:78-87.
  28. George JA, Eickbush TH: Conserved features at the 5 end of *Drosophila* R2 retrotransposable elements: implications for transcription and translation. *Insect Mol Biol* 1999, **8**:3-10.
  29. Luan DD, Eickbush TH: RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. *Mol Cell Biol* 1995, **15**:3882-3891.
  30. Kajikawa M, Okada N: LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 2002, **111**:433-444.
  31. Osanai M, Takahashi H, Kojima KK, Hamada M, Fujiwara H: Essential motifs in the 3' untranslated region required for retrotransposition and the precise start of reverse transcription in non-long-terminal-repeat retrotransposon SART1. *Mol Cell Biol* 2004, **24**:7902-7913.
  32. Anzai T, Osanai M, Hamada M, Fujiwara H: Functional roles of 3'-terminal structures of template RNA during *in vivo* retrotransposition of non-LTR retrotransposon, R1Bm. *Nucleic Acids Res* 2005, **33**:1993-2002.
  33. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr: High frequency retrotransposition in cultured mammalian cells. *Cell* 1996, **87**:917-927.
  34. Chambeyron S, Bucheton A, Busseau I: Tandem UAA repeats at the 3'-end of the transcript are essential for the precise initiation of reverse transcription of the I factor in *Drosophila melanogaster*. *J Biol Chem* 2002, **277**:17877-17882.
  35. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D, Boeke JD: Molecular archeology of L1 insertions in the human genome. *Genome Biol* 2002, **3**:research0052.
  36. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr: Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* 2003, **100**:5280-5285.
  37. Hohjoh H, Singer MF: Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J* 1996, **15**:630-639.
  38. Martin SL: Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol* 1991, **11**:4804-4807.
  39. Selem MC, Disson O, Robin S, Brun C, Teninges D, Bucheton A: *In vivo* RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. *Nucleic Acids Res* 2005, **33**:776-785.
  40. Han JS, Boeke JD: A highly active synthetic mammalian retrotransposon. *Nature* 2004, **429**:314-318.
  41. Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC, Shatsky IN: Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol Cell Biol* 2007, **27**:4685-4697.
  42. Kojima KK, Matsumoto T, Fujiwara H: Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol Cell Biol* 2005, **25**:7675-7686.
  43. Alish RS, Garcia-Perez JL, Muotri AR, Gage FH, Moran JV: Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev* 2006, **20**:210-224.
  44. Lachaux P, Bouhidel K, Mesure M, Pinon H: Spatial and temporal expression of the I factor during oogenesis in *Drosophila melanogaster*. *Development* 1992, **115**:729-735.
  45. Branciforte D, Martin SL: Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol* 1994, **14**:2584-2592.
  46. Trelogan SA, Martin SL: Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc Natl Acad Sci USA* 1995, **92**:1520-1524.
  47. Selem MC, Busseau I, Malinsky S, Bucheton A, Teninges D: High-frequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics* 1999, **151**:761-771.
  48. Rashkova S, Karam SE, Pardue ML: Element-specific localization of *Drosophila* retrotransposon Gag proteins occurs in both nucleus and cytoplasm. *Proc Natl Acad Sci USA* 2002, **99**:3621-3626.
  49. Goodier JL, Ostertag EM, Engleka KA, Selem MC, Kazazian HH Jr: A potential role for the nucleolus in L1 retrotransposition. *Hum Mol Genet* 2004.
  50. Goodier JL, Zhang L, Vetter MR, Kazazian HH Jr: LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol* 2007, **27**:6469-6483.
  51. Goodier JL, Mandal PK, Zhang L, Kazazian HH Jr: Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum Mol Genet* 2010, **19**:1712-1725.
  52. Kirilyuk A, Tolstonog GV, Damert A, Held U, Hahn S, Lower R, Buschmann C, Horn AV, Traub P, Schumann GG: Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. *Nucleic Acids Res* 2008, **36**:648-665.
  53. Matsumoto T, Takahashi H, Fujiwara H: Targeted nuclear import of open reading frame 1 protein is required for *in vivo* retrotransposition of a telomere-specific non-long terminal repeat retrotransposon, SART1. *Mol Cell Biol* 2004, **24**:105-122.
  54. Pardue ML, Rashkova S, Casacuberta E, DeBaryshe PG, George JA, Traverse KL: Two retrotransposons maintain telomeres in *Drosophila*. *Chromosome Res* 2005, **13**:443-453.
  55. Balagopal V, Parker R: Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. *Curr Opin Cell Biol* 2009, **21**:403-408.
  56. Irwin B, Aye M, Baldi P, Beliakova-Bethell N, Cheng H, Dou Y, Liou W, Sandmeyer S: Retroviruses and yeast retrotransposons use overlapping sets of host genes. *Genome Res* 2005, **15**:641-654.
  57. Aye M, Irwin B, Beliakova-Bethell N, Chen E, Garrus J, Sandmeyer S: Host factors that affect Ty3 retrotransposition in *Saccharomyces cerevisiae*. *Genetics* 2004, **168**:1159-1176.
  58. Griffith JL, Coleman LE, Raymond AS, Goodson SG, Pittard WS, Tsui C, Devine SE: Functional genomics reveals relationships between the retrovirus-like Ty1 element and its host *Saccharomyces cerevisiae*. *Genetics* 2003, **164**:867-879.
  59. Nyswaner KM, Checkley MA, Yi M, Stephens RM, Garfinkel DJ: Chromatin-associated genes protect the yeast genome from Ty1 insertional mutagenesis. *Genetics* 2008, **178**:197-214.
  60. Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM, Garfinkel DJ: P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. *Mol Cell Biol* 2010, **30**:382-398.
  61. Berleth T, Burri M, Thoma G, Bopp D, Riechstein S, Frigerio G, Noll M, Nusslein-Volhard C: The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J* 1988, **7**:1749-1756.
  62. Ferrandon D, Elphick L, Nusslein-Volhard C, St Johnston D: Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* 1994, **79**:1221-1232.
  63. Kinsey JA: Tad, a LINE-like transposable element of *Neurospora*, can transpose between nuclei in heterokaryons. *Genetics* 1990, **126**:317-323.
  64. Kubo S, Selem MC, Soifer HS, Perez JL, Moran JV, Kazazian HH Jr, Kasahara N: L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci USA* 2006, **103**:8036-8041.
  65. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV: DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 2002, **31**:159-165.
  66. Biessmann H, Valgeirsdottir K, Lofsky A, Chin C, Ginther B, Levis RW, Pardue ML: HeT-A, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol Cell Biol* 1992, **12**:3910-3918.
  67. Levis RW, Ganesan R, Houtchens K, Tolar LA, Sheen FM: Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 1993, **75**:1083-1093.
  68. Sheen FM, Levis RW: Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc Natl Acad Sci USA* 1994, **91**:12510-12514.
  69. Morrish TA, Garcia-Perez JL, Stamato TD, Taccioli GE, Sekiguchi J, Moran JV: Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 2007, **446**:208-212.

70. Yang J, Eickbush TH: RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. *Mol Cell Biol* 1998, **18**:3455-3465.
71. Christensen SM, Ye J, Eickbush TH: RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc Natl Acad Sci USA* 2006, **103**:17602-17607.
72. Zingler N, Willhoeft U, Brose HP, Schoder V, Jahns T, Hanschmann KM, Morrish TA, Lower J, Schumann GG: Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res* 2005, **15**:780-789.
73. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD: Human L1 retrotransposition is associated with genetic instability *in vivo*. *Cell* 2002, **110**:327-338.
74. Bibillo A, Eickbush TH: End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem* 2004, **279**:14945-14953.
75. Blesa D, Martinez-Sebastian MJ: bilbo, a non-LTR retrotransposon of *Drosophila subobscura*: a clue to the evolution of LINE-like elements in *Drosophila*. *Mol Biol Evol* 1997, **14**:1145-1153.
76. Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A, Eickbush TH: DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol* 2007, **374**:322-333.
77. Moran JV, Gilbert N: Mammalian LINE-1 retrotransposons and related elements. In *Mobile DNA II* Edited by: Craig NL, Craigie R, Gellert M, Lambowitz AM. Washington, DC, USA: ASM Press; 2002:836-869.
78. Ichijyanagi K, Nakajima R, Kajikawa M, Okada N: Novel retrotransposon analysis reveals multiple mobility pathways dictated by hosts. *Genome Res* 2007, **17**:33-41.
79. Gasior SL, Wakeman TP, Xu B, Deininger PL: The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 2006, **357**:1383-1393.
80. Suzuki J, Yamaguchi K, Kajikawa M, Ichijyanagi K, Adachi N, Koyama H, Takeda S, Okada N: Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet* 2009, **5**:e1000461.
81. Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M: Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohomology. *Genes Dev* 2005, **19**:2477-2487.
82. Kidd SJ, Glover DM: *Drosophila melanogaster* ribosomal DNA containing type II insertions is variably transcribed in different strains and tissues. *J Mol Biol* 1981, **151**:645-662.
83. Bucheton A, Lavigne JM, Picard G, L'Heritier P: Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variations in the efficiency of inducer and reactive strains. *Heredity* 1976, **36**:305-314.
84. Kidwell MG, Kidwell JF, Sved JA: Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 1977, **86**:813-833.
85. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ: Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 2007, **316**:744-747.
86. Carmell MA, Girard A, Kant HJ van de, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ: MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 2007, **12**:503-514.
87. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T: DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 2008, **22**:908-917.
88. Soper SF, Heijden GW van der, Hardiman TC, Goodheart M, Martin SL, de Boer P, Bortvin A: Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* 2008, **15**:285-297.
89. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998, **391**:806-811.
90. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello CC: The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 1999, **99**:123-132.
91. Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH: Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 1999, **99**:133-141.
92. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, Gaasterland T, Meyer J, Tuschl T: The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 2003, **5**:337-350.
93. Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC: Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev* 2006, **20**:2214-2222.
94. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 2007, **128**:1089-1103.
95. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC: A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 2007, **315**:1587-1590.
96. Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC: Gene silencing mechanisms mediated by Aubergine piRNA complexes in *Drosophila* male gonad. *Rna* 2007, **13**:1911-1922.
97. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, Chien M, Russo JJ, Ju J, Sheridan R, Sander C, Zavolan M, Tuschl T: A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 2006, **442**:203-207.
98. Girard A, Sachidanandam R, Hannon GJ, Carmell MA: A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 2006, **442**:199-202.
99. Grivna ST, Beyret E, Wang Z, Lin H: A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev* 2006, **20**:1709-1714.
100. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H: 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* 2006, **20**:1732-1743.
101. Houwing S, Kamminga LM, Berezikov E, Cronenbold D, Girard A, Elst H van den, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RH, Hannon GJ, Draper BW, Ketting RF: A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 2007, **129**:69-82.
102. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R: Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev* 2008, **22**:2773-2785.
103. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC: The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 2008, **453**:803-806.
104. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H: *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 2008, **453**:793-797.
105. Czech B, Malone CD, Zhou R, Stark A, Schlingehayde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J: An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 2008, **453**:798-802.
106. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD: Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 2008, **320**:1077-1081.
107. Chung WJ, Okamura K, Martin R, Lai EC: Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol* 2008, **18**:795-802.
108. Hutvagner G, Simard MJ: Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 2008, **9**:22-32.
109. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ: A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 2008, **31**:785-799.
110. Pelissou A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG: Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J* 1994, **13**:4401-4411.

111. Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A: **Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*.** *Genetics* 1995, **139**:697-711.
112. Desset S, Meignin C, Dastugue B, Vaury C: **COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*.** *Genetics* 2003, **164**:501-509.
113. Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabehere D, Ronsseray S: **P-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy.** *Genetics* 2000, **155**:1841-1854.
114. Ronsseray S, Lehmann M, Anxolabehere D: **The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome.** *Genetics* 1991, **129**:501-512.
115. Stuart JR, Haley KJ, Swedzinski D, Lockner S, Kocian PE, Merriman PJ, Simmons MJ: **Telomeric P elements associated with cytotypic regulation of the P transposon family in *Drosophila melanogaster*.** *Genetics* 2002, **162**:1641-1654.
116. Reiss D, Josse T, Anxolabehere D, Ronsseray S: **aubergine mutations in *Drosophila melanogaster* impair P cytotypic determination by telomeric P elements inserted in heterochromatin.** *Mol Genet Genomics* 2004, **272**:336-343.
117. Kalmykova AI, Klenov MS, Gvozdev VA: **Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline.** *Nucleic Acids Res* 2005, **33**:2052-2059.
118. 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* 2004, **166**:1313-1321.
119. Vagin VV, Klenov MS, Kalmykova AI, Stolyarenko AD, Kotelnikov RN, Gvozdev VA: **The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*.** *RNA Biol* 2004, **1**:54-58.
120. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 1997, **13**:335-340.
121. Bourc'his D, Bestor TH: **Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.** *Nature* 2004, **431**:96-99.
122. Aravin AA, Hannon GJ: **Small RNA silencing pathways in germ and stem cells.** *Cold Spring Harb Symp Quant Biol* 2008, **73**:283-290.
123. Halic M, Moazed D: **Transposon silencing by piRNAs.** *Cell* 2009, **138**:1058-1060.
124. Girard A, Hannon GJ: **Conserved themes in small-RNA-mediated transposon control.** *Trends Cell Biol* 2008, **18**:136-148.
125. Grewal SI, Elgin SC: **Transcription and RNA interference in the formation of heterochromatin.** *Nature* 2007, **447**:399-406.
126. Henderson IR, Jacobsen SE: **Epigenetic inheritance in plants.** *Nature* 2007, **447**:418-424.
127. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D: **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.** *Nature* 2003, **424**:99-103.
128. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L: **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** *Nature* 2003, **424**:94-98.
129. Dutko JA, Schafer A, Kenny AE, Cullen BR, Curcio MJ: **Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases.** *Curr Biol* 2005, **15**:661-666.
130. Schumacher AJ, Nissley DV, Harris RS: **APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast.** *Proc Natl Acad Sci USA* 2005, **102**:9854-9859.
131. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: **Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.** *Proc Natl Acad Sci USA* 2006, **103**:8780-8785.
132. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C: **APOBEC3 proteins inhibit human LINE-1 retrotransposition.** *J Biol Chem* 2006, **281**:22161-22172.
133. Stenglein MD, Harris RS: **APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism.** *J Biol Chem* 2006, **281**:16837-16841.
134. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD: **APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.** *Curr Biol* 2006, **16**:480-485.
135. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K: **All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.** *Nucleic Acids Res* 2007, **35**:2955-2964.
136. MacDuff DA, Demorest ZL, Harris RS: **AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity.** *Nucleic Acids Res* 2009, **37**:1854-1867.
137. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM: **Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.** *Curr Biol* 2005, **15**:166-170.
138. Gallois-Montbrun S, Kramer B, Swanson CM, Byers H, Lynham S, Ward M, Malim MH: **Antiviral protein APOBEC3G localizes to ribonucleoprotein complexes found in P bodies and stress granules.** *J Virol* 2007, **81**:2165-2178.
139. Chiu YL, Witkowska HE, Hall SC, Santiago M, Soros VB, Esnault C, Heidmann T, Greene WC: **High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition.** *Proc Natl Acad Sci USA* 2006, **103**:15588-15593.
140. Hulme AE, Bogerd HP, Cullen BR, Moran JV: **Selective inhibition of Alu retrotransposition by APOBEC3G.** *Gene* 2007, **390**:199-205.
141. Stenglein MD, Burns MB, Li M, Lengyel J, Harris RS: **APOBEC3 proteins mediate the clearance of foreign DNA from human cells.** *Nat Struct Mol Biol* 2010, **17**:222-229.
142. Bogerd HP, Wiegand HL, Doeble BP, Lueders KK, Cullen BR: **APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells.** *Nucleic Acids Res* 2006, **34**:89-95.
143. Kazazian HH Jr: **An estimated frequency of endogenous insertional mutations in humans.** *Nat Genet* 1999, **22**:130.
144. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS: **Frequency of recent retrotransposition events in the human factor IX gene.** *Hum Mutat* 2001, **17**:511-519.
145. Osterstag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH Jr: **A mouse model of human L1 retrotransposition.** *Nat Genet* 2002, **32**:655-660.
146. Metzker ML: **Sequencing technologies - the next generation.** *Nat Rev Genet* 2010, **11**:31-46.
147. Seleme MC, Vetter MR, Cordaux R, Bastone L, Batzer MA, Kazazian HH Jr: **Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity.** *Proc Natl Acad Sci USA* 2006, **103**:6611-6616.
148. Lutz SM, Vincent BJ, Kazazian HH Jr, Batzer MA, Moran JV: **Allelic heterogeneity in LINE-1 retrotransposition activity.** *Am J Hum Genet* 2003, **73**:1431-1437.
149. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.** *Nature* 2001, **411**:494-498.
150. Goodwin TJ, Busby JN, Poulter RT: **A yeast model for target-primed (non-LTR) retrotransposition.** *BMC Genomics* 2007, **8**:263.
151. Dong C, Poulter RT, Han JS: **LINE-like retrotransposition in *Saccharomyces cerevisiae*.** *Genetics* 2009, **181**:301-311.
152. Haoudi A, Semmes OJ, Mason JM, Cannon RE: **Retrotransposition-competent Human LINE-1 induces apoptosis in cancer cells with intact p53.** *J Biomed Biotechnol* 2004, **2004**:185-194.
153. Belgnaoui SM, Gosden RG, Semmes OJ, Haoudi A: **Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells.** *Cancer Cell Int* 2006, **6**:13.
154. Wallace NA, Belancio VP, Deininger PL: **L1 mobile element expression causes multiple types of toxicity.** *Gene* 2008, **419**:75-81.
155. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS, Moran JV: **LINE-1 retrotransposition in human embryonic stem cells.** *Hum Mol Genet* 2007, **16**:1569-1577.
156. Eickbush TH, Furano AV: **Fruit flies and humans respond differently to retrotransposons.** *Curr Opin Genet Dev* 2002, **12**:669-674.
157. Bushman F: *Lateral DNA Transfer* Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press; 2002.
158. Hickey DA: **Selfish DNA: a sexually-transmitted nuclear parasite.** *Genetics* 1982, **101**:519-531.
159. Kapitonov VV, Jurka J: **RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.** *PLoS Biol* 2005, **3**:e181.

160. Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Betermier M: **PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*.** *Genes Dev* 2009, **23**:2478-2483.
161. Li WH, Gu Z, Wang H, Nekrutenko A: **Evolutionary analyses of the human genome.** *Nature* 2001, **409**:847-849.
162. Whitelaw E, Martin DI: **Retrotransposons as epigenetic mediators of phenotypic variation in mammals.** *Nat Genet* 2001, **27**:361-365.
163. Lyon MF: **X-chromosome inactivation: a repeat hypothesis.** *Cytogenet Cell Genet* 1998, **80**:133-137.

doi: [10.1186/1759-8753-1-15](https://doi.org/10.1186/1759-8753-1-15)

**Cite this article as:** Han, Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions *Mobile DNA* 2010, **1**:15

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

