

REVIEW

Open Access



# Retrotransposons shape species-specific embryonic stem cell gene expression

Luisa Robbez-Masson and Helen M Rowe\*

## Abstract

Over half of our genome is composed of retrotransposons, which are mobile elements that can readily amplify their copy number by replicating through an RNA intermediate. Most of these elements are no longer mobile but still contain regulatory sequences that can serve as promoters, enhancers or repressors for cellular genes. Despite dominating our genetic content, little is known about the precise functions of retrotransposons, which include both endogenous retroviruses (ERVs) and non-LTR elements like long interspersed nuclear element 1 (LINE-1). However, a few recent cutting-edge publications have illustrated how retrotransposons shape species-specific stem cell gene expression by two opposing mechanisms, involving their recruitment of stem cell-enriched transcription factors (TFs): firstly, they can activate expression of genes linked to naïve pluripotency, and secondly, they can induce repression of proximal genes. The paradox that different retrotransposons are active or silent in embryonic stem cells (ESCs) can be explained by differences between retrotransposon families, between individual copies within the same family, and between subpopulations of ESCs. Since they have coevolved with their host genomes, some of them have been co-opted to perform species-specific beneficial functions, while others have been implicated in genetic disease. In this review, we will discuss retrotransposon functions in ESCs, focusing on recent mechanistic advances of how HERV-H has been adopted to preserve human naïve pluripotency and how particular LINE-1, SVA and ERV family members recruit species-specific transcriptional repressors. This review highlights the fine balance between activation and repression of retrotransposons that exists to harness their ability to drive evolution, while minimizing the risk they pose to genome integrity.

## Part I: Retrotransposons are active in ESCs Retrotransposon activity fluctuates between distinct embryonic cell states

Retrotransposons, which may encompass over two-thirds of the human genome [1] have the potential to cause insertional mutagenesis or transcriptional perturbation prompting their epigenetic silencing during early embryonic development. Retrotransposons are then assumed to remain inactive throughout the organism's adult life through their collective DNA methylation and their aberrant activation has been associated with cancer and autoimmune disorders [2, 3]. However, it has long been known that these elements also play normal roles in development, particularly in the placenta. A

series of papers have now revealed that subsets of retrotransposons are expressed within mouse and human embryonic stem cell (ESC) cultures. The roles that these retrotransposons may play in development relate to their distinct expression patterns between different in vitro subpopulations of ESCs that are described below.

Embryonic stem cells are characterised by their capacity for self-renewal and pluripotency. They constitute a category of primary cells isolated from the inner cell mass (ICM) of blastocysts, which are early-stage pre-implantation embryos, and these cells can form tightly aggregated, three-dimensional colonies in culture. ESCs have been used as model systems to study the pluripotent state and the conditions required for cellular differentiation. Indeed, ESCs possess the remarkable capacity to self-renew and differentiate into cells of all three germ layers of the embryo (endoderm, mesoderm and ectoderm) including the germ line in vitro, or in vivo when

\*Correspondence: h.rowe@ucl.ac.uk  
Division of Infection and Immunity, Medical Research Council Centre for Medical Molecular Virology, University College London, 90 Gower Street, London WC1E 6BT, UK

ESCs are injected into blastocysts leading to the generation of chimeric mice [4].

Cultures of mouse ESCs display a great degree of intercellular heterogeneity, where cells can exhibit different states of pluripotency (as shown by their different gene expression profiles [5]). Cells within these cultures are dynamic, shuttling between a ground state termed “naïve” and a more committed “primed” state [6–8]. Hallmarks of naïve stem cells include high and stable expression of core pluripotency-associated TFs such as NANOG, OCT4 (POU5F1), the active *Oct4* distal enhancer, REX1, and SOX2. Moreover, these cells express key naïve (or ground-state) defining TFs, such as LBP9 and KLF4. Naïve cells also display global DNA hypomethylation, an absence of chromosome X inactivation and a reduced concentration of H3K27me3 repressive histone marks on the main developmental genes [7, 9]. Such ground state ESCs can be maintained in two-inhibitor (2i) medium supplemented with Leukaemia inhibitory factor (LIF). 2i culture conditions involve the addition of two specific small molecule inhibitors directed to the MAPK/ERK and GSK3 pathways. Since these pathways are both stimulated by LIF and can trigger cell differentiation and impair ESCs self-renewal, they both require inhibition [8]. It has recently been shown that a subpopulation of mouse ESCs express the ERV, MERV-L as well as a pool of transcripts normally expressed in two-cell (2-cell)-stage embryos, whereas they do not express classical pluripotency-associated factors like OCT4. Remarkably when isolated, these cells have totipotent potential instead of only pluripotent potential, which means that they can differentiate into extra-embryonic tissues (placenta) as well as into all three germ layers [10]. This is in line with previous data showing MERV-L to be highly expressed in 2-cell stage mouse embryos [11] before being repressed at the blastocyst stage.

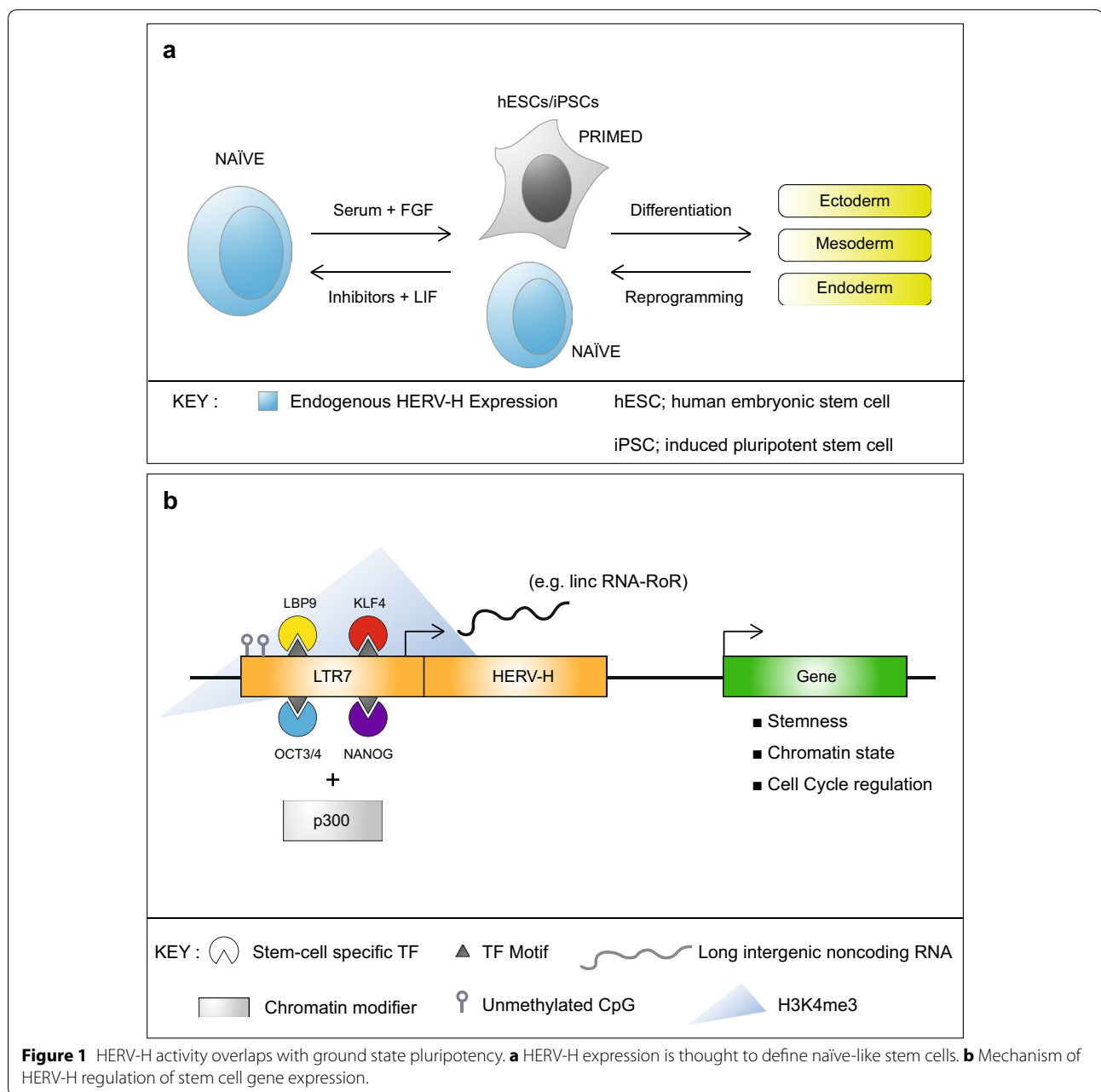
In contrast to mouse ESCs, human ESC lines that have been derived, more closely resemble mouse epiblast stem cells (EpiSCs), a more advanced stage of development of the post-implantation embryo [12]. A lot of studies have focused on trying to “reset” human ESC cultures to a more naïve state of pluripotency in order to study this earlier stage of development. So far, no consensus has been established in terms of the optimum conditions for naïve human pluripotent cell culture, although three main protocols for generating these cells have recently been developed [13–15]. Some strategies require the use of different cocktails of small molecules inhibitors and cytokines to stimulate the expression of naïve pluripotency proteins such as LBP9 [13, 15, 16], whereas other approaches employ the direct introduction of pluripotency factors as transgenes to achieve the same effect

(NANOG and KLF2 transgene expression coupled to ERK inhibition) [14].

However, it has recently been shown that a natural population of naïve-like cells exists in human ESC cultures and transiently during reprogramming, and these cells can be isolated simply by using a GFP reporter driven by a HERV-H promoter, since they express this primate-specific retrovirus ([17–19] and see Figure 1a). This would obviate the need for complicated and expensive culture protocols to induce naïve ESCs. Retroviruses such as MERV-L in mouse and HERV-H in human are therefore powerful tools to naturally select distinct populations of primitive pluripotent cells in order to study early development and offer new perspectives in reprogramming strategies and in the production of stem cell medicines. It is clear that retrotransposons impact on cell fate and reprogramming but how do they do this? Their distinct functions can be divided into three categories detailed below, which are (a) the recruitment of TFs to activate cellular genes, (b) the production of noncoding RNAs and finally, (c) potential roles of viral proteins and particles.

(a) *Retrotransposon DNAs bind core pluripotency factors in a species-specific manner* Retrotransposons have likely evolved to recruit activating TFs expressed early in development to ensure their propagation through the germ line and overall persistence in the genome. This has been beneficial to the host, providing us with a pool of regulatory sequences that can act as novel enhancers and promoters [20]. The creation of new TF binding sites within retrotransposons has occurred similarly in both mice and humans, although since many retrotransposon types and locations are distinct between these two species, this has led to unique gene regulatory circuits.

Several retrotransposons are abundantly expressed in human ESCs such as long interspersed nuclear element 1 (LINE-1) [21] and HERV-H [18]. This tissue-specificity is determined in part by TFs. For example, one LINE-1 family member, L1TD1 that has lost enzymatic activity is specifically expressed in ESCs because its promoter recruits NANOG, OCT4 and SOX2 [22]. HERV-H recruits the TF, LBP9 [19], and since this TF is essential for ground-state pluripotency [23], HERV-H, as one of its targets in human cells, is an integral component of naïve cells. LBP9 is a recent example that fits with a previous broad observation that core pluripotency TFs exert highly species-specific binding profiles. Indeed, it was established that up to 25% of pluripotency-associated TF binding sites were contributed from transposable elements, and that these TF-bound loci are lineage or species-specific because of the often random nature of the retrotransposition process [24]. Not all TFs are recruited to retrotransposon sequences and it appears that



pluripotency associated TFs are enriched at these sites, although interestingly, the transcription insulator CTCF that is ubiquitously expressed also binds retrotransposons, but in this case B2 SINES (short-interspersed nuclear elements) [25]. Certain classes of retrotransposons are enriched for particular TFs: for instance, OCT4/SOX2 binding is enriched on ERVK repeats in mouse and ERV1 in humans and TP53 is predominantly found on Mer61-type ERV1 repeats in human [26]. Sequence

patterns that resemble binding site motifs of some TFs (ESR1, TP53, OCT4/SOX2, and CTCF) embedded within different transposable elements can therefore predispose them to becoming mammalian TF binding sites.

The majority of TF binding sites occur in the long terminal repeats (LTRs) of retrotransposons, which flank open reading frames (ORFs), or on UTRs of non-LTR retrotransposons such as LINE and SINE elements. LTRs and UTRs therefore serve as platforms for the

recruitment of pluripotency-associated TFs and co-factors (such as chromatin modifiers) to moderate expression of cellular genes. The influence of retrotransposons on gene expression is significant because local epigenetic modifications can exert long-range effects by remodelling large chromatin domains and higher order chromatin structure, and LTR enhancers may loop to cellular genes in *cis* or even in *trans*, for example through CTCF [25]. In human and mouse ESCs, it was recently shown that LTR-derived enhancers affected numerous genes involved in chromatin organization, cell cycle and stemness [27]. We have also shown that repressed LTRs can become enhancers when they escape silencing, showing that silenced LTRs contain intrinsic enhancer activity and may act as temporal enhancers [28]. Similarly, LTRs also exert significant promoter effects that can be switched on during critical developmental stages. For instance, waves of stage-specific retrotransposon activation occur during pre-implantation embryo development. Indeed, by investigating the transcriptomes of human pre-implantation embryos, Goke and colleagues [29] have documented transient ERV activation taking place between the oocyte and blastocyst stages. Their analysis of different ERV families revealed that each stage of the embryo expresses distinct ERV classes, for example, HERVK14 is only expressed between the pronucleus and 4-cell stage, whereas THE1A is restricted to the 8-cell and Morula stages. Interestingly, both elements are not expressed in human ESCs, reinforcing the idea that cultured human ESCs only represent a snapshot of development and are primarily not naïve ESCs [29]. Another example of a stage-specific ERV is MERV-L [30], whose promoters drive expression of hundreds of 2-cell stage expressed genes [10]. While retrotransposons can act as promoters or enhancers for cellular genes, their enhancer function is particularly fascinating due to the recently documented rapid evolution of enhancer modules across species, which includes enhancers derived from retrotransposon sequences [31].

*(b) Non-coding RNAs derived from LTRs play a role in stem cell identity* Pluripotency TF binding to retrotransposons leads to transcription not only of cellular genes but also of retrotransposon RNAs, some of which are coding and many of which are short or long non-coding RNAs (lncRNAs) of largely unknown function. Interestingly, some non-coding RNA molecules are chimeric because they originate from splicing events between cellular and viral transcripts. Of particular interest, rare sub-populations of human ESCs (termed naïve cells, introduced above) have an active chromatin configuration at LTR7 sites in the genome (hypomethylated DNA, active histone marks and bound NANOG, OCT4, KLF4 and LBP9) and show elevated expression of their linked

HERV-H transcripts, as well as of HERV-K transcripts, ([19, 27, 32] and see Figure 1b). High-throughput transcriptional profiling of mouse and human stem cells has revealed a large pool of species-specific chimeric and lncRNAs, including several pluripotency-associated lncRNAs, such as lin-ROR [33] and linc00458 [34].

Increased expression of some members of the HERV-H family is also observed in human induced pluripotent stem cells (iPSCs) [17, 35]. Depletion of some HERV-H expressed loci or of lin-ROR in human ESCs, using RNA interference, results in a drastic change in cellular morphology, with cells adopting a more differentiated phenotype (fibroblast-like) [32]. lncRNAs are implicated in many cellular processes including chromatin remodelling, control of promoter activity, X-chromosome inactivation, imprinting and nuclear import (reviewed in [36]). However, the specific role of most lncRNAs remains unclear. In the context of stem cells, retrotransposon-derived transcript expression levels closely mirrors the expression patterns of core pluripotency factors such as OCT4, NANOG and SOX2, suggesting that they might be essential to the pluripotent state. Importantly, HERV-H must be silenced to guarantee successful cell differentiation [34]. Reminiscent of pluripotent cells, LTR7-induced transcripts (including HERV-H lincRNA-RoR [33], LINE-1 [37] and HERV-K [35]) are transiently activated during reprogramming to iPSCs, indicating that their role in ESCs may parallel their role in restoring pluripotency in differentiated cells. However, they need to be subsequently repressed for successful reprogramming [17]. In general, high HERV-H RNA levels define naïve ESCs, concomitant with a complete loss of repressive chromatin marks such as condensed H3K27me3, whereas HERV-H is only lowly detected in primed ESCs. Of note, a similar enrichment of ERV derived transcripts has also been described in trophoblast stem cells and placenta [38].

However, what do HERV-H-lncRNAs do? Recent work has shown evidence that these RNA molecules can recruit transcriptional co-activators and other proteins into DNA-binding regulatory complexes. A specific function of long intergenic non-coding RNAs (lincRNA) was first demonstrated in human fibroblasts and their derivative iPSCs where a lincRNA specific microarray analysis was performed [33]. Additionally, RNA cross-linking experiments in human ESCs show that HERV-H-lncRNAs act as a scaffold unit that recruits the co-activators CBP, p300, MED6 and MED12, to enhancer regions [32]. These lncRNAs are also associated with OCT4, and are thought to play an essential role in LTR-specific enhancer activity (Figure 1b). For example, one of these co-activators is the histone acetyltransferase p300, which was showed to be essential for the recruitment of the

NANOG/OCT4/SOX2 complex and regulates transcription via chromatin remodelling [39].

In sum, the involvement of retrotransposon lncRNAs in the control of pluripotency in early development and in reprogramming is a common mechanism in mammals, likely acting through RNA-recruited co-activators but operating via species-specific transposable elements.

*(c) Viral proteins, and particles that bud from embryos may function in development* Some of the first electron microscopy images of chicken embryos revealed mysterious virus-like particles (VLPs) of unknown function, which were mainly extracellular [40]. Likewise, in mouse embryos, ERVs, including of the IAP and MERV-L classes can be observed budding into the endoplasmic reticulum, particularly at the 2-cell stage [41]. The potential function of retroviral particles is unknown, although they may serve an antiviral role. In contrast, it is well established that certain retroviral proteins serve vital functions in reproduction and development. The best example of this is the *syncytin* family: *syncytins* 1 and 2 are essential placental genes derived from retroviral envelopes. These proteins emerged in mammals on at least six occasions independently and were retained each time by natural selection to carry out the same function. Syncytins are responsible for the formation of the syncytiotrophoblast, the multicellular element of the placenta responsible for nutrient exchange and shielding the embryo from the mother's immune system [42].

Another documented example of retrotransposon protein expression is for LINE-1. It was recently demonstrated that mammals have evolved to use LINE-1 retrotransposon activity as a way to assess the quality of gametes. The massive loss of oocytes (two-thirds are lost in mice and around 80% in humans) that takes place during their maturation serves as a key quality-control checkpoint. For example, a recent study demonstrated that levels of the LINE-1 protein, L1ORF1p, which is essential for retrotransposition, acted as a marker to govern oocyte fate. Apoptosis is triggered only in oocytes with high L1ORF1p levels, ensuring that aberrant LINE-1 activation during epigenetic reprogramming of the genome remained as low as possible in the surviving oocytes and potential offspring [43]. An analogous mechanism likely exists in the male germ line. Conversely, the L1TD1 gene is an interesting example of a LINE-1 protein that has been positively selected in both primates and mice, due to beneficial roles it is thought to play in both genome defence and pluripotency [44]. Of note, active retrotransposition of LINE-1 has been reported in neural progenitor cells and brains of rodents and humans [45–48], although the potential function of this is unknown.

Although it is largely unknown how viral proteins and particles might contribute to genome defence and maintain pluripotent states, an exciting recent study on

HERV-K provides new insight into these questions [49]. The authors reveal firstly that OCT4 drives expression of human-specific HERV-K proviruses by binding to their promoters (LTR5HS), leading to the production of GAG proteins and VLPs during early human development. Secondly, the HERV-K accessory protein, Rec binds to a subset of cellular mRNAs and can influence their translation, and finally HERV-K may serve to combat exogenous viral infections because it is shown to upregulate classical virus restriction factors such as IFITM1. Of note, data from this paper also suggests that HERV-K may be a more accurate marker of naive human ESCs than HERV-H because it is expressed in naive but not primed human ESCs.

## Part II: Retrotransposons are repressed in ESCs

### KAP1 retrotransposon repression shapes gene regulation

The very retrotransposons that are active and have been exapted to serve useful gene regulatory functions are often the same families that our genomes have evolved to repress. This is because these families contain elements with intact regulatory sequences that could interfere with gene expression and/or functional open reading frames that could lead to retrotransposition events. One example of this is MERV-L, which, as mentioned above, is highly abundant at the 2-cell stage of development (contributing to 3% of mRNAs) but repressed by the blastocyst stage [30].

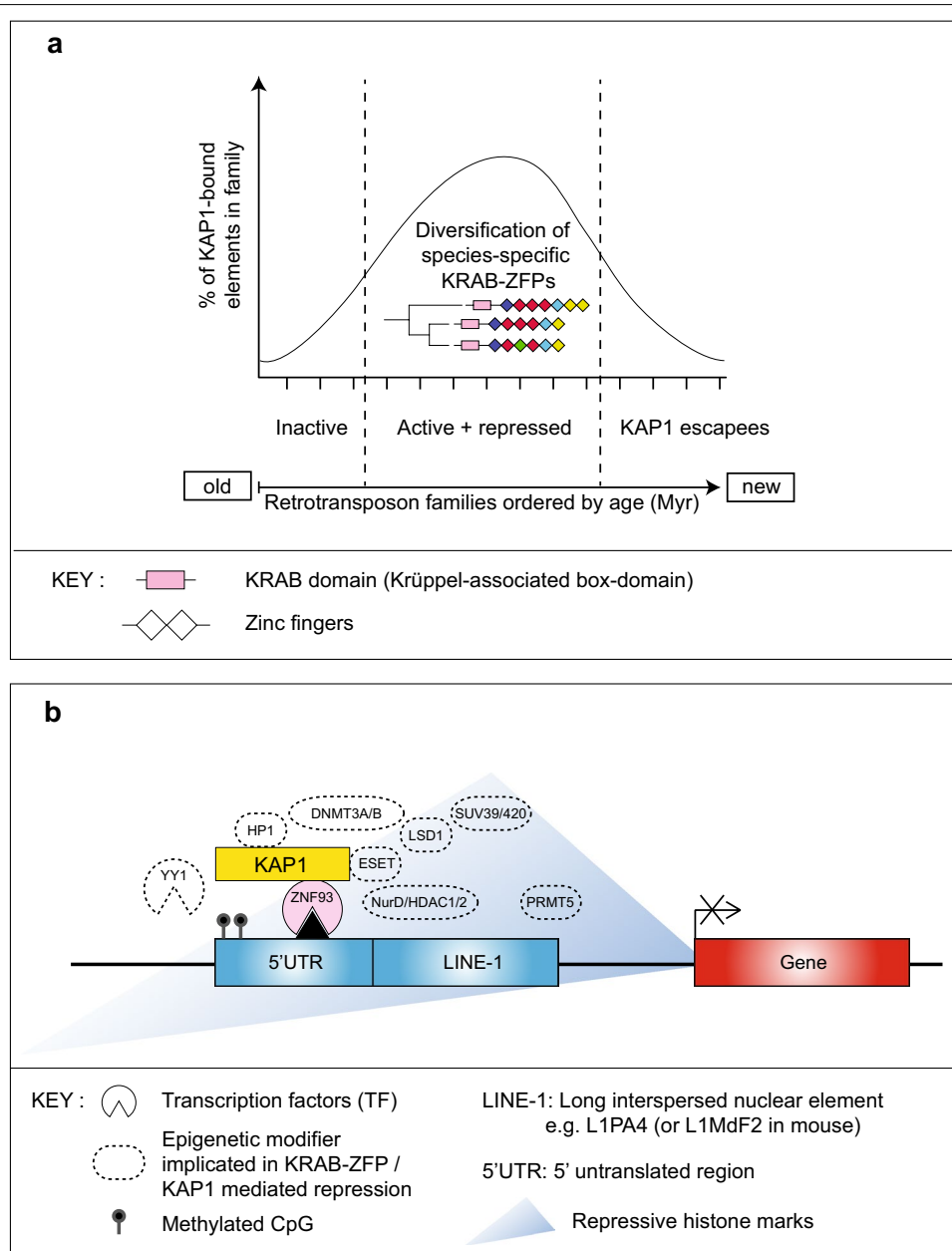
Active retrotransposon families are targeted for epigenetic silencing early in development and during reprogramming [17, 35, 50]. One important repression pathway we and others have uncovered is the KAP1 (TRIM28) pathway that operates in ESCs and early embryos: KAP1 is recruited to repetitive sequences through site-specific krüppel-associated box domain-containing-zinc finger proteins (KRAB-ZFPs) and represses them through the histone methyltransferase ESET/SETDB1 [51–56] (and reviewed in [57, 58]), leading to their subsequent DNA methylation [59]. ERV silencing also leads to the repression of nearby genes due to the spreading of epigenetic marks, suggesting that this mechanism may have been co-opted for the fine-tuning of gene expression; certain genes are not switched off but maintained in a lowly expressed state in early development [28, 60, 61].

KAP1 repression is sequence-specific *in vitro* and *in vivo* [59] and retrotransposons that have more recently invaded the genome escape KAP1 through subtle changes in their nucleotide content, presumably because the KRAB-ZFP system has not yet adapted to repress them. This is true for both mouse (for example for the IAP class [53]) and human retrotransposons (LINE-1 [51]), although it is best illustrated with the LINE-1 family, due to the recent classification of LINE families based on their relative ages [62]. The most ancient LINE-1 families are neither KAP1-bound nor DNA methylated,



presumably because they are dead by mutation, whereas the newer ones are KAP1-bound, repressed and highly DNA methylated, and finally the most recent families escape KAP1-repression, but they are regulated through DNA methylation, which may be deposited through one or more small RNA pathways ([51] and see Figure 2a). The KAP1-ERV repression pathway operates in mouse

and human ESCs but is not required in mouse embryonic fibroblasts [53, 54], presumably because DNA methylation takes over as the dominant silencing mechanism later in development, a hypothesis we and others have provided evidence for [59, 63, 64]. Of interest, KAP1 repression of ERVs is still detected in mouse neural progenitor cells [65].



**Figure 2** Adaptive evolution of retrotransposon repression in ESCs. **a** Co-evolution of retrotransposons and KRAB-ZFPs. **b** Mechanism of KAP1 repression of retrotransposons.

KAP1 repression of retroviruses was initially discovered in the context of murine leukaemia virus (MLV) [55], which led on from original observations that MLV was restricted in embryonic cells [66] through its primer binding site Pro (PBS-pro) sequence that binds proline tRNA [67, 68]. This sequence was later discovered to recruit KAP1 through the mouse KRAB-ZFP, Zfp809 [56]. MLV still serves as a practical model to explore the KAP1 repression pathway and it was recently uncovered that YY1 and EBP1 contribute to an MLV silencing complex [69–71], factors that may also repress ERVs with a PBS-pro and/or ERVs with unrelated PBS sites (reviewed in [72]). Elegant work has just revealed, through Zfp809 knockout mice and genetic and biochemical experiments in ESCs, that Zfp809 not only restricts exogenous MLV but also several ERVs that contain PBS-pro sequences, as predicted [73]. Strikingly, in Zfp809 knockout mice, disruption of silent chromatin marks normally established early in development at VL30-(virus-like 30) type PBS-pro ERVs leads to their overexpression in differentiated tissues, together with nearby genes. Of note, VL30 elements lack coding regions, illustrating how KAP1 represses not only coding but also non-coding ERVs that remain a threat because of their regulatory sequences. This new study, therefore, provides conclusive evidence that the KAP1/KRAB-ZFP pathway is necessary to repress retrotransposons and linked genes *in vivo*.

#### **Human retrotransposons are targeted by human-adapted KRAB-ZFPs**

The finding that KAP1 represses multiple classes of ERVs (mainly IAPs and MERVK in mouse [53] or HERVK and LINE-1 in humans [54], most of which do not operate through a PBS-pro or even have mutated PBS sequences or no PBS sequence (LINE-1s), led to the concept that retrotransposons may recruit KAP1 through a multitude of different site-specific KRAB-ZFPs. This would ensure that even retrotransposons that cannot reverse transcribe are maintained inactive so as not to affect cellular genes through their potentially active enhancer/promoter sequences [28, 60]. This is supported by the diverse KRAB-ZFPs that our genomes encode, many of which are species-specific and rapidly evolving with largely unknown functions, which suggests their participation in genetic conflict with viral sequences that are also rapidly evolving [74–78]. One example of KRAB-ZFP adaptive evolution is the ZNF91 subfamily that has expanded across primate lineages [79].

However, while previously only a model, it has not been until now that exciting work has illustrated that indeed our genomes do encode a repertoire of KRAB-ZFPs adapted to recognize and target species-specific retrotransposons. Specifically, the human proteins ZNF91 and

ZNF93 bind to and repress SVA and LINE-1 retrotransposons respectively, in the human genome [80]. SVA elements are a newly emerged retrotransposon class that invaded great ape genomes 18–25 million years (myr) ago. They are composite retrotransposons that contain an Alu-like fragment that is joined by a variable number tandem repeat (VNTR) domain to a SINE region that contains 3'LTR sequences similar to the ERV, HERV-K10 [81]. ZNF91 underwent structural changes between 8 and 12 myr ago to restrict SVAs, including the addition of seven zinc fingers. The authors nicely link structure to function showing that while macaque ZNF91 is unable to repress a human SVA reporter plasmid in mouse ESCs, (which do not express endogenous ZNF91), transfected human ZNF91 with its seven newly evolved zinc fingers induces strong SVA repression, as expected [80].

ZNF93 is another interesting example of a host-retrotransposon interaction, particularly because LINE-1s exert a unique pattern of evolution with a single L1PA subfamily active at one time in a genome before being replaced by a new subfamily, allowing their approximate ageing [62]. For example, in the human genome, L1PA4 (18 myr old) was replaced by L1PA3 (15.8 myr old), which was replaced by L1PA2 (7.6 myr old). KRAB-ZFP evolution relates to the activity of LINE-1 subfamilies because ZNF93 targets a sequence present in the L1PA4 UTRs and some L1PA3 UTRs but which is deleted in L1PA2 elements leading to their escape from ZNF93 repression in mouse ESCs when LINE-1 reporters and ZNFs are co-transfected [80]. ZNF93 underwent zinc finger deletions and other structural adaptations to repress human LINE-1s. As such, human ZNF93 but not macaque ZNF93 is able to repress an L1PA4 reporter construct in mouse ESCs. In the case of SVAs, nearby genes were also repressed through ZNF91, which supports previous findings that retrotransposon repression can lead to species-specific fine-tuning of gene circuits *in vitro* [28, 60] and *in vivo* [73]. Species-specificity is driven not only by LINEs and SVAs, some of which are distinct in the human genome but also by the TFs they recruit, which are also species-specific, as they have undergone adaptation between primate lineages. These new findings lead us to a summary model of KAP1 repression of retrotransposons, although the exact enzymes required are still unclear (See Figure 2b).

Of note, still only a handful of KRAB-ZFPs that recognise repetitive DNA have been characterized to date. This includes the two human KRAB-ZFPs discussed above and the mouse KRAB-ZFP Zfp809, already mentioned [55, 56, 73]. Apart from these, an additional mouse KRAB-ZFP, Zfp819 has been implicated in modulating expression of IAP ERVs and LINEs through an unknown sequence, which impacts on the balance between

pluripotency and differentiation [82] and the new mouse KRAB-ZFP *Gm6871* that was previously only a predicted gene, targets a subset of LINES (mainly of the L1MdF2 family), again through a 5'UTR sequence [51]. A mouse KRAB-ZFP *Ssm1b* has also been implicated in DNA methylation of foreign DNA [83]. Many questions persist concerning how KRAB-ZFPs exert their functions, their patterns of evolution, how many of them recognise repetitive sequences, where and when they act, how they impact on cellular genes and how they relate to disease settings. However, a recent paper provides evidence that most KRAB-ZFPs may target repetitive sequences since 16 out of 18 human KRAB-ZFPs sampled at random bound repeated elements, including LINE-1, HERVs and SVAs, as determined by chromatin immunoprecipitation [84]. Of note, the authors also used high throughput binding data to create an improved ZFP recognition code predictor.

#### Other retrotransposon repression pathways

While a discussion of all other potential retrotransposon repression pathways acting within ESCs is beyond the scope of this review, we mention a few of the main ones below and refer to other reviews [72, 85]. These can be divided into transcriptional repression pathways, which together with KAP1 act as the first line of defence against retrotransposons, and post-transcriptional repression pathways that are crucial at later stages of the retrotransposon life cycle to prevent new retrotransposition events.

Transcriptional repression pathways involving histone deacetylases and histone methyltransferases/demethylases are fundamental to retrotransposon repression in ESCs because in these cells, there is a layer of repression additional to DNA methylation, highlighted by the finding that triple knockout of DNMT1, DNMT3a and DNMT3b is not sufficient to significantly reactivate retrotransposons such as IAPs [63, 86]. DNA methylation-independent mechanisms of repression are presumably required in development in the face of global re-setting of the epigenome. Key enzymes implicated in retrotransposon repression in ESCs include HDACs [87, 88], ESET/SETDB1 [52, 59, 89], which likely has KAP1-independent as well as dependent targets, LSD1/KDM1a [90], G9a [91], Suv420H1/2 that mediates H4K20me3 [28, 52], polycomb complexes [92] and Suv39h [93]. Other less studied enzymes such as the arginine methyltransferase PRMT5 may also play a role, since it has recently been uncovered to interact with KAP1 [94] and repress some retrotransposons in primordial germ cells and preimplantation embryos during DNA methylation reprogramming [95]. The role of PRMT5 in retrotransposon repression was also assessed in *prmt5* deleted and control 2i- cultured ESCs (see ESC culture protocols in part I above),

which display DNA hypomethylation, but only 1.8-fold of IAP-Gag upregulation was observed in this context, perhaps due to PRMT5 compensation by PRMT7. Other co-factors that have been implicated in retrotransposon transcriptional repression include HP1 family members that interact with KAP1 [96–98] and may be involved in long-range repression [99, 100], DNMT3L [101, 102], hnRNPK and MCAF1 [103], REX1 [104, 105] and SIRT6 [106]. Exactly which enzymes are directed to certain retrotransposons and how and when KAP1 is involved is unknown. It is possible that non-coding RNAs play a targeting role in a similar way to how they tether co-activators to HERV-H (see part I above).

Post-transcriptional repression pathways include intrinsic factors that block later stages of the retrotransposon lifecycle such as SAMHD1 ([107] and see [108] for a review). A multitude of small RNA pathways are crucial to retrotransposon regulation in development and the intricate details of these pathways are only now being unravelled. Some small RNAs like piRNAs (small non-coding piwi-interacting RNAs) can even induce the silent histone mark H3K9me3 and de novo DNA methylation at least in the germ line, adding to transcriptional silencing [50, 109–111] and the piRNA pathway may play a role in ESCs [112]. Interestingly, small RNAs derived from LINE-1 have been implicated in transcriptional activation of LINE-1 at the 2-cell stage of embryo development [113], whereas in mouse ESCs, there is a role for small RNAs in LINE-1 restriction because there is an increase in LINE-1 transcripts in Dicer knockout ESCs that is rescued by ectopic Dicer expression [114]. Still much is unknown about small RNA transposon silencing in ESCs, particularly as small RNAs detected (in this case from LINE-1) include both Dicer-dependent and independent classes [114].

#### Discussion and perspectives

As discussed, retrotransposons can confer regulatory complexity to gene networks early in development and in ESCs by serving as enhancers and/or promoters of key developmental genes. They provide new regulatory sequences that have been integrated into novel gene networks and they are particularly important in the maintenance of the naïve pluripotent cell state, likely through their adaptation to bind TFs expressed early in development (see [72, 85, 115, 116] for additional reviews on this topic).

Many questions remain about the extent to which a family of retrotransposons can be repressed or activated in certain tissues or at particular developmental stages, and which factors coordinate switches in activation status. One interesting perspective is that the very KRAB-ZFP pathways that repress retrotransposons may actually



only induce temporal or tissue-specific repression, and there is some evidence for KRAB-ZFPs even being activators in some cases [76]. It would make sense for the host to evolve to restrict retrotransposons in situations where they pose heritable threats to the germ line such as in germ cells or early embryos where many KRAB-ZFPs are enriched. This may have led to retrotransposons becoming ideal lineage-specific enhancers that get switched on during later stages of development, a hypothesis worth exploring. Another interesting perspective is that since KRAB-ZFP gene clusters are heavily intermingled with retrotransposons, some of them may have been reverse transcribed along with retrotransposons, which may have contributed to their rapid evolution and increase in copy number.

In summary, retrotransposons are, unlike classical viruses, essential to the evolution of our genomes and have contributed to genome plasticity and the creation of new genes. We can use them as tools to direct cell fate and reprogramming but studying their intricate pathways of expression is necessary to understand gene regulation in development, and to develop safe stem cell medicines. Research into this topic is particularly relevant to understanding and treating genetic diseases and cancers in which retrotransposons have been implicated.

#### Authors' contributions

LR and HMR jointly planned, drafted and wrote the manuscript, and prepared the figures. Both authors read and approved the final manuscript.

#### Acknowledgements

The authors thank Steen K Ooi and Yasu Takeuchi for reading the manuscript. LR and HMR are supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and Royal Society (Grant number 101200/Z/13/Z) awarded to HMR.

#### Compliance with ethical guidelines

#### Competing interests

The authors declare that they have no competing interests.

Received: 26 February 2015 Accepted: 7 May 2015

Published online: 29 May 2015

#### References

- de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7(12):e1002384. doi:10.1371/journal.pgen.1002384
- Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenz D et al (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med* 16(5):571–579. doi:10.1038/nm.2129 (1p following 9)
- Nexo BA, Christensen T, Frederiksen J, Moller-Larsen A, Oturai AB, Villesen P et al (2011) The etiology of multiple sclerosis: genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. *PLoS One* 6(2):e16652. doi:10.1371/journal.pone.0016652
- Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292(5819):154–156
- Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X et al (2010) Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6(5):468–478. doi:10.1016/j.stem.2010.03.015
- Marks H, Kalkan T, Menafrá R, Denissov S, Jones K, Hofmeister H et al (2012) The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149(3):590–604. doi:10.1016/j.cell.2012.03.026
- Nichols J, Smith A (2009) Naive and primed pluripotent states. *Cell Stem Cell* 4(6):487–492. doi:10.1016/j.stem.2009.05.015
- Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J et al (2008) The ground state of embryonic stem cell self-renewal. *Nature* 453(7194):519–523. doi:10.1038/nature06968
- Nichols J, Smith A (2011) The origin and identity of embryonic stem cells. *Development* 138(1):3–8. doi:10.1242/dev.050831
- Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D et al (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487(7405):57–63. doi:10.1038/nature11244
- Kigami D, Minami N, Takayama H, Imai H (2003) MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod* 68(2):651–654
- Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL et al (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* 448(7150):196–199. doi:10.1038/nature05972
- Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D et al (2013) Derivation of novel human ground state naive pluripotent stem cells. *Nature* 504(7479):282–286. doi:10.1038/nature12745
- Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F et al (2014) Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* 158(6):1254–1269. doi:10.1016/j.cell.2014.08.029
- Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J et al (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell* 15(4):471–487. doi:10.1016/j.stem.2014.07.002
- Hanna J, Cheng AW, Saha K, Kim J, Lengner CJ, Soldner F et al (2010) Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *Proc Natl Acad Sci* 107(20):9222–9227. doi:10.1073/pnas.1004584107
- Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M et al (2014) Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci* 111(34):12426–12431. doi:10.1073/pnas.1413299111
- Santoni FA, Guerra J, Luban J (2012) HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* 9:111. doi:10.1186/1742-4690-9-111
- Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A et al (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 516(7531):405–409. doi:10.1038/nature13804
- Jacques PE, Jeyakani J, Bourque G (2013) The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* 9(5):e1003504. doi:10.1371/journal.pgen.1003504
- García-Pérez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS et al (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet* 16(13):1569–1577. doi:10.1093/hmg/ddm105
- Narva E, Rahkonen N, Emani MR, Lund R, Pursiheimo JP, Nasti J et al (2012) RNA-binding protein L1TD1 interacts with LIN28 via RNA and is required for human embryonic stem cell self-renewal and cancer cell proliferation. *Stem Cells* 30(3):452–460. doi:10.1002/stem.1013
- Martello G, Bertone P, Smith A (2013) Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. *EMBO J* 32(19):2561–2574. doi:10.1038/emboj.2013.177
- Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS et al (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* 42(7):631–634. doi:10.1038/ng.600

25. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C et al (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148(1–2):335–348. doi:[10.1016/j.cell.2011.11.058](https://doi.org/10.1016/j.cell.2011.11.058)
26. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG et al (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* 18(11):1752–1762. doi:[10.1101/gr.080663.108](https://doi.org/10.1101/gr.080663.108)
27. Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A et al (2014) Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* 46(6):558–566. doi:[10.1038/ng.2965](https://doi.org/10.1038/ng.2965)
28. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y et al (2013) TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* 23(3):452–461. doi:[10.1101/gr.147678.112](https://doi.org/10.1101/gr.147678.112)
29. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F et al (2015) Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16(2):135–141. doi:[10.1016/j.stem.2015.01.005](https://doi.org/10.1016/j.stem.2015.01.005)
30. Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM et al (2004) RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* 269(1):276–285. doi:[10.1016/j.ydbio.2004.01.028](https://doi.org/10.1016/j.ydbio.2004.01.028)
31. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M et al (2015) Enhancer evolution across 20 mammalian species. *Cell* 160(3):554–566. doi:[10.1016/j.cell.2015.01.006](https://doi.org/10.1016/j.cell.2015.01.006)
32. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G et al (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* 21(4):423–425. doi:[10.1038/nsmb.2799](https://doi.org/10.1038/nsmb.2799)
33. Loewer S, Cabili MN, Guttman M, Loh YH, Thomas K, Park IH et al (2010) Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* 42(12):1113–1117. doi:[10.1038/ng.710](https://doi.org/10.1038/ng.710)
34. Ng SY, Johnson R, Stanton LW (2012) Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* 31(3):522–533. doi:[10.1038/emboj.2011.459](https://doi.org/10.1038/emboj.2011.459)
35. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Diaz N, Rowe HM et al (2014) Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* 24(8):1251–1259. doi:[10.1101/gr.172809.114](https://doi.org/10.1101/gr.172809.114)
36. Geisler S, Collier J (2013) RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14(11):699–712. doi:[10.1038/nrm3679](https://doi.org/10.1038/nrm3679)
37. Wissing S, Munoz-Lopez M, Macia A, Yang Z, Montano M, Collins W et al (2012) Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet* 21(1):208–218. doi:[10.1093/hmg/ddr455](https://doi.org/10.1093/hmg/ddr455)
38. Chuong EB, Rumi MA, Soares MJ, Baker JC (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* 45(3):325–329. doi:[10.1038/ng.2553](https://doi.org/10.1038/ng.2553)
39. Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB et al (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133(6):1106–1117. doi:[10.1016/j.cell.2008.04.043](https://doi.org/10.1016/j.cell.2008.04.043)
40. Cromack AS (1968) An electron microscope study of virus-like particles in chick embryo and L cell cultures. *J Gen Virol* 2(1):195–198
41. Calarco PG (1979) Intracisternal A particles in preimplantation embryos of feral mice (*Mus musculus*). *Intervirology* 11(6):321–325
42. Cornelis G, Vernochet C, Carradec Q, Souquere S, Mulot B, Catzeflis F et al (2015) Retroviral envelope gene captures and syncytium exaptation for placentation in marsupials. *Proc Natl Acad Sci* 112(5):E487–E496. doi:[10.1073/pnas.1417000112](https://doi.org/10.1073/pnas.1417000112)
43. Malki S, van der Heijden GW, O'Donnell KA, Martin SL, Bortvin A (2014) A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Dev Cell* 29(5):521–533. doi:[10.1016/j.devcel.2014.04.027](https://doi.org/10.1016/j.devcel.2014.04.027)
44. McLaughlin RN Jr, Young JM, Yang L, Neme R, Wichman HA, Malik HS et al (2014) Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency. *PLoS Genet* 10(9):e1004531. doi:[10.1371/journal.pgen.1004531](https://doi.org/10.1371/journal.pgen.1004531)
45. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F et al (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479(7374):534–537. doi:[10.1038/nature10531](https://doi.org/10.1038/nature10531)
46. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT et al (2009) L1 retrotransposition in human neural progenitor cells. *Nature* 460(7259):1127–1131. doi:[10.1038/nature08248](https://doi.org/10.1038/nature08248)
47. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS et al (2012) Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151(3):483–496. doi:[10.1016/j.cell.2012.09.035](https://doi.org/10.1016/j.cell.2012.09.035)
48. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH et al (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435(7044):903–910. doi:[10.1038/nature03663](https://doi.org/10.1038/nature03663)
49. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ et al (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature*. doi:[10.1038/nature14308](https://doi.org/10.1038/nature14308)
50. Crichton JH, Dunican DS, MacLennan M, Meehan RR, Adams IR (2014) Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell Mol Life Sci* 71(9):1581–1605. doi:[10.1007/s00018-013-1468-0](https://doi.org/10.1007/s00018-013-1468-0)
51. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M et al (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev* 28(13):1397–1409. doi:[10.1101/gad.241661.114](https://doi.org/10.1101/gad.241661.114)
52. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H et al (2010) Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464(7290):927–931. doi:[10.1038/nature08858](https://doi.org/10.1038/nature08858)
53. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T et al (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463(7278):237–240. doi:[10.1038/nature08674](https://doi.org/10.1038/nature08674)
54. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J et al (2014) Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* 24(8):1260–1270. doi:[10.1101/gr.172833.114](https://doi.org/10.1101/gr.172833.114)
55. Wolf D, Goff SP (2007) TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* 131(1):46–57. doi:[10.1016/j.cell.2007.07.026](https://doi.org/10.1016/j.cell.2007.07.026)
56. Wolf D, Goff SP (2009) Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 458(7242):1201–1204. doi:[10.1038/nature07844](https://doi.org/10.1038/nature07844)
57. Leung DC, Lorincz MC (2011) Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem Sci* 37(4):127–133. doi:[10.1016/j.tibs.2011.11.006](https://doi.org/10.1016/j.tibs.2011.11.006)
58. Rowe HM, Trono D (2011) Dynamic control of endogenous retroviruses during development. *Virology* 411(2):273–287. doi:[10.1016/j.virol.2010.12.007](https://doi.org/10.1016/j.virol.2010.12.007)
59. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J et al (2013) De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* 140(3):519–529. doi:[10.1242/dev.087585](https://doi.org/10.1242/dev.087585)
60. Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX et al (2011) DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 8(6):676–687. doi:[10.1016/j.stem.2011.04.004](https://doi.org/10.1016/j.stem.2011.04.004)
61. Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y et al (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* 7(9):e1002301. doi:[10.1371/journal.pgen.1002301](https://doi.org/10.1371/journal.pgen.1002301)
62. Konkel MK, Batzer MA (2010) A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. *Semin Cancer Biol* 20(4):211–221. doi:[10.1016/j.semcancer.2010.03.001](https://doi.org/10.1016/j.semcancer.2010.03.001)
63. Hutnick LK, Huang X, Loo TC, Ma Z, Fan G (2010) Repression of retrotransposal elements in mouse embryonic stem cells is primarily

- mediated by a DNA methylation-independent mechanism. *J Biol Chem* 285(27):21082–21091. doi:10.1074/jbc.M110.125674
64. Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20(2):116–117. doi:10.1038/2413
  65. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jonsson ME, Manne C et al (2015) TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep*. 10(1):20–28. doi:10.1016/j.celrep.2014.12.004
  66. Teich NM, Weiss RA, Martin GR, Lowy DR (1977) Virus infection of murine teratocarcinoma stem cell lines. *Cell* 12(4):973–982 (pii:0092-8674(77)90162-3)
  67. Barklis E, Mulligan RC, Jaenisch R (1986) Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 47(3):391–399 (pii:0092-8674(86)90596-9)
  68. Colicelli J, Goff SP (1986) Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. *J Virol* 57(1):37–45
  69. Schlesinger S, Goff SP (2013) Silencing of proviruses in embryonic cells: efficiency, stability and chromatin modifications. *EMBO Rep* 14(1):73–79. doi:10.1038/embor.2012.182
  70. Schlesinger S, Lee AH, Wang GZ, Green L, Goff SP (2013) Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell Rep*. 4(1):50–58. doi:10.1016/j.celrep.2013.06.003
  71. Wang GZ, Wolf D, Goff SP (2014) EBP1, a novel host factor involved in primer binding site-dependent restriction of moloney murine leukemia virus in embryonic cells. *J Virol* 88(3):1825–1829. doi:10.1128/JVI.02578-13
  72. Schlesinger S, Goff SP (2015) Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol Cell Biol* 35(5):770–777. doi:10.1128/MCB.01293-14
  73. Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, Park C et al (2015) The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev* 29(5):538–554. doi:10.1101/gad.252767.114
  74. Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM et al (2014) Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS One* 8(2):e56721. doi:10.1371/journal.pone.0056721
  75. Emerson RO, Thomas JH (2009) Adaptive evolution in zinc finger transcription factors. *PLoS Genet* 5(1):e1000325. doi:10.1371/journal.pgen.1000325
  76. Liu H, Chang LH, Sun Y, Lu X, Stubbs L (2014) Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol* 6(3):510–525. doi:10.1093/gbe/evu030
  77. Lukic S, Nicolas JC, Levine AJ (2014) The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ* 21(3):381–387. doi:10.1038/cdd.2013.150
  78. Thomas JH, Schneider S (2011) Coevolution of retroelements and tandem zinc finger genes. *Genome Res* 21(11):1800–1812. doi:10.1101/gr.121749.111
  79. Hamilton AT, Huntley S, Tran-Gyamfi M, Baggott DM, Gordon L, Stubbs L (2006) Evolutionary expansion and divergence in the ZNF91 subfamily of primate-specific zinc finger genes. *Genome Res* 16(5):584–594. doi:10.1101/gr.4843906
  80. Jacobs FM, Greenberg D, Nguyen N, Haessler M, Ewing AD, Katzman S et al (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516(7530):242–245. doi:10.1038/nature13760
  81. Hancks DC, Kazazian HH Jr (2010) SVA retrotransposons: evolution and genetic instability. *Semin Cancer Biol* 20(4):234–245. doi:10.1016/j.semcancer.2010.04.001
  82. Tan X, Xu X, Elkenani M, Smorag L, Zechner U, Nolte J et al (2013) Zfp819, a novel KRAB-zinc finger protein, interacts with KAP1 and functions in genomic integrity maintenance of mouse embryonic stem cells. *Stem Cell Res*. 11(3):1045–1059. doi:10.1016/j.scr.2013.07.006
  83. Ratnam S, Engler P, Bozek G, Mao L, Podlutzky A, Austad S et al (2014) Identification of Ssm1b, a novel modifier of DNA methylation, and its expression during mouse embryogenesis. *Development*. 141(10):2024–2034. doi:10.1242/dev.105726
  84. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A et al (2015) C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol*. doi:10.1038/nbt.3128
  85. Gifford WD, Pfaff SL, Macfarlan TS (2013) Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol* 23(5):218–226. doi:10.1016/j.tcb.2013.01.001
  86. Tsumura A, Hayakawa T, Kumaki Y, Takebayashi S, Sakaue M, Matsuoka C et al (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11(7):805–814. doi:10.1111/j.1365-2443.2006.00984.x
  87. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC et al (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 466(7307):769–773. doi:10.1038/nature09209
  88. Reichmann J, Crichton JH, Madej MJ, Taggart M, Gautier P, Garcia-Perez JL et al (2012) Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of retrotransposon expression in mouse embryonic stem cells. *PLoS Comput Biol* 8(4):e1002486. doi:10.1371/journal.pcbi.1002486
  89. Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P et al (2014) Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc Natl Acad Sci* 111(18):6690–6695. doi:10.1073/pnas.1322273111
  90. Macfarlan TS, Agarwal S, Driscoll S, Lettieri K, Wang J et al (2011) Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25(6):594–607. doi:10.1101/gad.2008511
  91. Leung DC, Dong KB, Maksakova IA, Goyal P, Appanah R, Lee S et al (2011) Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc Natl Acad Sci* 108(14):5718–5723. doi:10.1073/pnas.1014660108
  92. Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A et al (2010) Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* 24(3):265–276. doi:10.1101/gad.544410
  93. Bulut-Karslioglu A, De La Rosa-Velazquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J et al (2014) Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* 55(2):277–290. doi:10.1016/j.molcel.2014.05.029
  94. di Caprio R, Ciano M, Montano G, Costanzo P, Cesaro E (2015) KAP1 is a novel substrate for the arginine methyltransferase PRMT5. *Biology* 4(1):41–49. doi:10.3390/biology4010041
  95. Kim S, Gunesdogan U, Zyllicz JJ, Hackett JA, Cougot D, Bao S et al (2014) PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* 56(4):564–579. doi:10.1016/j.molcel.2014.10.003
  96. Cammas F, Herzog M, Lerouge T, Chambon P, Losson R (2004) Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev* 18(17):2147–2160. doi:10.1101/gad.30290418/17/2147
  97. Lechner MS, Begg GE, Speicher DW, Rauscher FJ 3rd (2000) Molecular determinants for targeting heterochromatin protein 1-mediated gene silencing: direct chromoshadow domain-KAP-1 corepressor interaction is essential. *Mol Cell Biol* 20(17):6449–6465
  98. Wolf D, Cammas F, Losson R, Goff SP (2008) Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *J Virol* 82(9):4675–4679. doi:10.1128/JVI.02445-07
  99. Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G, Denervaud N et al (2010) KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* 6(3):e1000869. doi:10.1371/journal.pgen.1000869
  100. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410(6824):116–120. doi:10.1038/3506513235065132
  101. Kao TH, Liao HF, Wolf D, Tai KY, Chuang CY, Lee HS et al (2014) Ectopic DNMT3L triggers assembly of a repressive complex for retroviral silencing in somatic cells. *J Virol* 88(18):10680–10695. doi:10.1128/JVI.01176-14

102. Vlachogiannis G, Niederhuth CE, Tuna S, Stathopoulou A, Viiri K, de Rooij DG et al (2015) The Dnmt3L ADD domain controls cytosine methylation establishment during spermatogenesis. *Cell Rep*. doi:10.1016/j.celrep.2015.01.021
103. Thompson PJ, Dulberg V, Moon KM, Foster LJ, Chen C, Karimi MM et al (2015) hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. *PLoS Genet* 11(1):e1004933. doi:10.1371/journal.pgen.1004933
104. Guallar D, Perez-Palacios R, Climent M, Martinez-Abadia I, Larraga A, Fernandez-Juan M et al (2012) Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. *Nucleic Acids Res* 40(18):8993–9007. doi:10.1093/nar/gks686
105. Schoorlemmer J, Perez-Palacios R, Climent M, Guallar D, Muniesa P (2014) Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. *Front Oncol*. 4:14. doi:10.3389/fonc.2014.00014
106. Van Meter M, Kashyap M, Rezazadeh S, Geneva AJ, Morello TD, Seluanov A et al (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun*. 5:5011. doi:10.1038/ncomms6011
107. Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X et al (2013) Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep*. 4(6):1108–1115. doi:10.1016/j.celrep.2013.08.019
108. Dewannieux M, Heidmann T (2013) Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Curr Opin Virol* 3(6):646–656. doi:10.1016/j.coviro.2013.08.005
109. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF et al (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 31(6):785–799. doi:10.1016/j.molcel.2008.09.003
110. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M et al (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 22(7):908–917. doi:10.1101/gad.1640708
111. Pezic D, Manakov SA, Sachidanandam R, Aravin AA (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev* 28(13):1410–1428. doi:10.1101/gad.240895.114
112. Marchetto MC, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL et al (2013) Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* 503(7477):525–529. doi:10.1038/nature12686
113. Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E et al (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* 20(3):332–338. doi:10.1038/nsmb.2495
114. Ciaudo C, Jay F, Okamoto I, Chen CJ, Sarazin A, Servant N et al (2013) RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. *PLoS Genet* 9(11):e1003791. doi:10.1371/journal.pgen.1003791
115. Feschotte C, Gilbert C (2012) Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 13(4):283–296. doi:10.1038/nrg3199
116. Weiss RA, Stoye JP (2013) Virology. Our viral inheritance. *Science* 340(6134):820–821. doi:10.1126/science.1235148

**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

