**REVIEW**

Open Access

# Transposable elements in the mammalian embryo: pioneers surviving through stealth and service

Patricia Gerdes[1], Sandra R. Richardson[1], Dixie L. Mager[2*] and Geoffrey J. Faulkner[1,3*]

## Abstract

Transposable elements (TEs) are notable drivers of genetic innovation. Over evolutionary time, TE insertions can supply new promoter, enhancer, and insulator elements to protein-coding genes and establish novel, species-specific gene regulatory networks. Conversely, ongoing TE-driven insertional mutagenesis, nonhomologous recombination, and other potentially deleterious processes can cause sporadic disease by disrupting genome integrity or inducing abrupt gene expression changes. Here, we discuss recent evidence suggesting that TEs may contribute regulatory innovation to mammalian embryonic and pluripotent states as a means to ward off complete repression by their host genome.

## Background

Mammalian embryonic development is governed by a complex set of genetic and epigenetic instructions. This genomic blueprint undergoes evolutionary selection and, as such, the fundamental order of development is well conserved among mammals. At fertilization, sperm and egg unite to form the zygote, which undergoes successive cleavage divisions, yielding two-, four-, and eight-cell embryonic stages [1, 2]. Initially, the zygotic genome is transcriptionally inactive, with maternally inherited factors regulating embryonic metabolism and development. Embryonic genome activation occurs at around the eight-cell stage in humans and the two-cell stage in mice [3] and is accompanied in each species by

epigenome-wide remodeling [4]. The zygote and its daughter cells are totipotent; that is, they have the potential to differentiate into all embryonic and extraembryonic cell types. During development, the differentiation potential of embryonic cells becomes progressively more restricted. At the blastocyst stage, the cells of the inner cell mass (ICM) are pluripotent, meaning that while they cannot give rise to extraembryonic tissues, they can generate all cell lineages and are able to self-renew. Hence, early development involves rapid cellular diversification driven by myriad, and largely still undefined, transcriptional and epigenetic programs (Box 1).

Pluripotent states arising embryonically in vivo, or achieved in vitro by cellular reprogramming, are associated with epigenetic derepression and transcriptional activation of transposable elements (TEs) [4–6]. These mobile genetic elements are found in every eukaryotic genome sequenced to date and account for at least half of mammalian DNA [7–9]. In most mammals, retrotransposons are the predominant TEs. These can be divided into long terminal repeat (LTR) retrotransposons, including endogenous retroviruses (ERVs), and non-LTR retrotransposons such as long interspersed elements (LINEs) and short interspersed elements (SINEs) (Fig. 1a) [10–12]. LINE-1 (L1; Box 2), and ERV families are the only autonomous retrotransposons identified in the human and mouse genomes, though, importantly, human ERVs (HERVs) are all likely now retrotransposition incompetent (Box 3).

All retrotransposons mobilize via a "copy-and-paste" mechanism involving a transcribed RNA intermediate that is reverse transcribed and integrated as a nascent cDNA into genomic DNA. However, there are essential differences in the retrotransposition mechanisms used by LTR and non-LTR retrotransposons (Fig. 1b, c). L1 mRNA transcription relies on an internal 5′ promoter, whereas ERV proviruses utilize a 5′ LTR promoter for transcription initiation (Fig. 1a). Crucially, most new L1

* Correspondence: dmager@bccrc.ca; faulknergj@gmail.com
[2]Department of Medical Genetics, Terry Fox Laboratory, British Columbia Cancer Agency, University of British Columbia, Vancouver, BC V5Z 1L3, Canada
[1]Mater Research Institute, University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia
Full list of author information is available at the end of the article

Gerdes *et al. Genome Biology* (2016) 17:100

Page 2 of 17

---

**Box 1. Regulatory networks controlling pluripotency**

Programmed shifts in transcriptional and epigenetic states during embryogenesis have been studied primarily using in vitro systems. Embryonic stem cells (ESCs) are pluripotent cells derived from the blastocyst inner cell mass. Cultured ESCs are intensively used to study pluripotency, particularly in humans. Over the past decade, a core regulatory circuit incorporating the transcription factors Oct4 (also known as Pou5f1), Sox2, and Nanog [126–128] has been revealed to regulate ESC pluripotency [129]. This circuit activates pluripotency-associated factors and represses lineage-specific genes [130]. Pluripotent cells can also be derived in vitro via somatic cell reprogramming. Induced pluripotent stem cells (iPSCs) were initially produced by forced expression of Oct4, Sox2, Klf4, and c-*Myc* using retroviral vectors [131, 132]. Numerous methods have since been developed to improve reprogramming efficiency and iPSC safety [133]. As for ESCs, iPSCs provide a powerful system to understand the pluripotent state and can differentiate to all cell types of the body [131, 132]

---

insertions are 5′ truncated and therefore lack the core L1 regulatory sequence. Of 500,000 human L1 copies, only about 7000 retain the canonical 5′ promoter [7, 13]. By contrast, about 90 % of HERVs exist in the genome as solitary LTRs due to recombination of proviral 5′ and 3′ LTRs [11, 14]. Many of these LTRs maintain, or restore through acquired mutations, their natural transcriptional and regulatory signatures, which can perturb the expression of nearby genes [15]. While the regulatory capacity of older LTRs will tend to diminish over time, the approximately 440,000 identifiable LTRs in the human genome [7] still carry enormous potential to regulate genes and gene networks [14–17]. Therefore, compared with L1, ERVs are arguably a much greater source of regulatory innovation (Fig. 2).

Recent studies have revealed a complex and somewhat paradoxical interplay between retrotransposons and their host genome in pluripotent cells. On one hand, retrotransposons have long been regarded as fundamentally selfish genetic elements [18] that, to ensure their survival, must evade host genome surveillance and mobilize in cells that provide opportunities for germline transmission. Transcriptional reactivation of retrotransposons in the early mammalian embryo aligns with this evolutionary imperative, despite retrotransposition posing a threat to genome integrity. Indeed, cells employ numerous mechanisms to restrict retrotransposition at this stage [19–23]. On the other hand, transcription from ERV promoters drives the expression of cellular genes as well as ERV-derived sequences and appears to be a fundamental

characteristic of the pluripotent state [16, 24–31]. LTRs may be permitted to thrive in this environment due to the materials they provide to the host genome for regulatory network innovation (Fig. 3). Indeed, as well as providing alternative promoters to pluripotency genes [28], ERVs can serve as long-range enhancers [26], produce regulatory noncoding RNAs [27, 30], and may, in some cases, express their own viral proteins [29, 31]. Hence, transcribed products arising from ERVs may promote, or even be required for, the pluripotent state [24–33]. Finally, reports of L1 retrotransposition in somatic cells have fueled speculation that TE-derived mosaicism may lead to functional innovation during development [34–37].

Here, we review the restraint and activity of TEs in embryonic cells and later in development, as well as the unexpected promotion of pluripotent states by ERVs. We further appraise the convergent contributions to embryogenesis made by ERVs in distinct mammalian clades as evidence of an evolved strategy to avoid, or at least delay, host genome repression.

## ERV-driven transcription in the early embryo
### ERV regulation of protein-coding genes
Although there are spectacular examples of TE proteins underpinning functional innovation, such as in the placenta [38], regulatory sequences exapted from TEs arguably loom larger in our evolutionary history [15]. Indeed, up to 30 % of human and mouse transcription start sites (TSSs) are situated in TEs and display tissue-specific expression patterns [33, 39]. Embryonic human tissues express the greatest diversity of TE-associated TSSs observed to date [33], highlighting the potential of TEs to drive cell type and developmental stage-specific expression, particularly during early embryogenesis when the genome becomes demethylated [40]. In mouse, the LTR promoters of MuERV-L elements regulate a network of genes critical for totipotency and specific to the two-cell stage of embryonic development [41]. TE-derived regulatory sequences likewise contribute to the evolution of regulatory networks in pluripotent stem cells. For example, only about 5 % of Oct4 and Nanog transcription factor (TF) binding sites are shared in mouse and human embryonic stem cells (hESCs). TEs contribute a significant proportion (about 25 %) of the remaining, species-specific, binding sites [42]. Moreover, in vitro knockdown of specific ERVs via RNA interference can lead to a reduction in pluripotency markers [24, 26–28, 43–46]. Thus, TE sequences are broadly and strongly transcribed in the early embryo and can influence pluripotency by being exapted into, or at least adding robustness to, pluripotency networks. These findings underscore the universality and versatility of TEs in driving the evolution of regulatory networks.
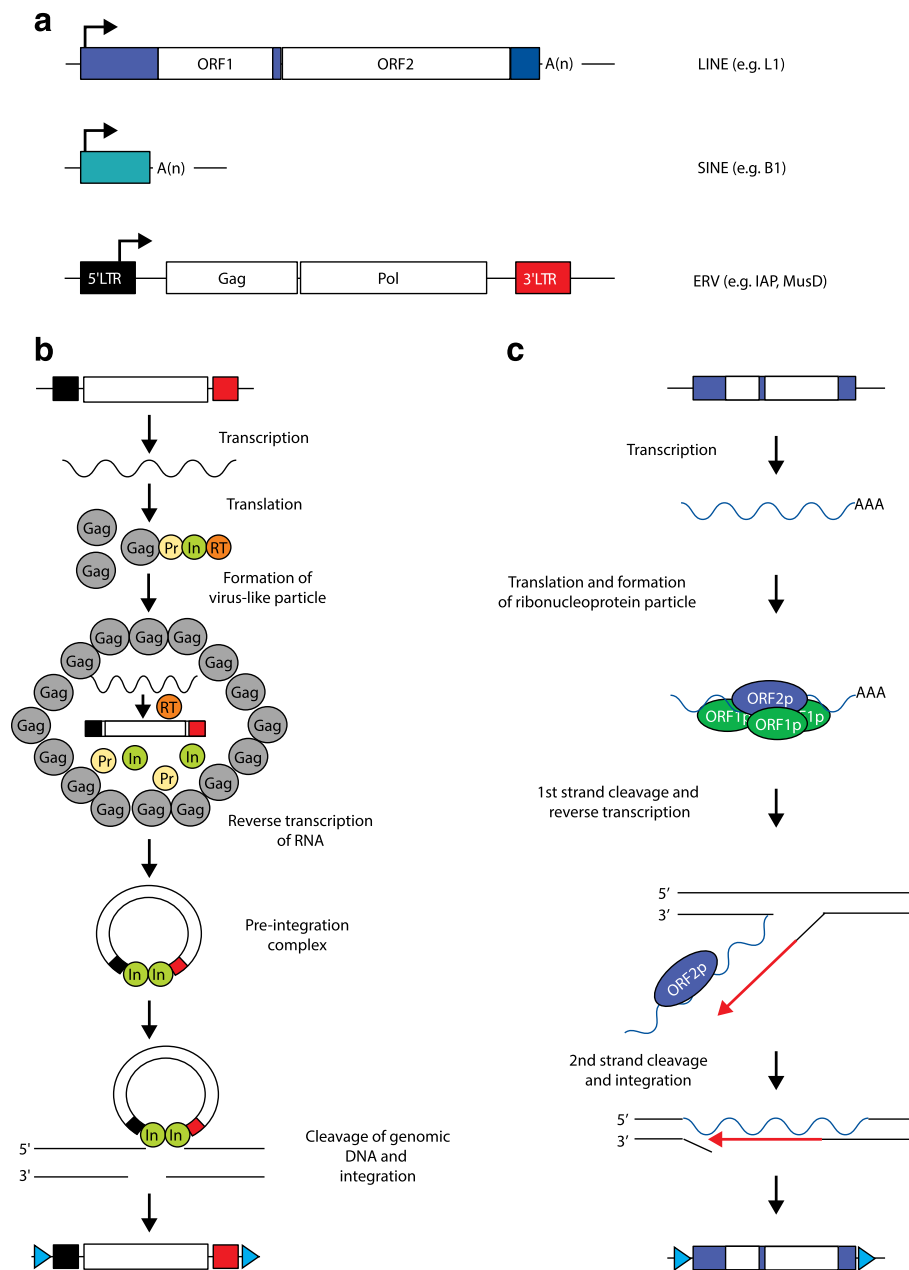
Gerdes *et al. Genome Biology* (2016) 17:100

Page 3 of 17



**Fig. 1** Long terminal repeat (*LTR*) and non-LTR retrotransposition mechanisms. **a** Mammalian retrotransposon structures. A long interspersed element (*LINE*; human L1 shown) typically consists of a 5′ untranslated region (UTR; *blue box*) harboring an internal promoter, two open reading frames (*ORF1*, *ORF2*), a 3′ UTR (*small blue box*), and a poly(A)-tail. A short interspersed element (*SINE*; mouse B1 shown) does not encode proteins and is *trans*-mobilized by LINE proteins. An endogenous retrovirus (*ERV*), such as mouse intracisternal A-type particle (*IAP*) and Mus type-D related retrovirus (*MusD*), lacks an Env protein but encodes functional Gag and Pol proteins flanked by a LTR at the 5′ (*black box*) and 3′ (*red box*) ends. *Arrows* indicate transcription start sites. **b** ERV mobilization starts with mRNA transcription and translation to yield Gag and Gag–Pro–Pol fusion proteins. The fusion proteins consist of a Gag protein (*Gag*), a protease (*Pr*), an integrase (*In*), and a reverse transcriptase (*RT*). Gag proteins build a virus-like particle and encapsulate the fusion proteins, which are processed into separate mature proteins. The ERV mRNA is then reverse transcribed, generating a cDNA. This cDNA and the integrase build a preintegration complex. The integrase then creates a double-strand DNA break, followed by genomic integration of a new ERV copy. Target site duplications (TSDs) are indicated by *blue triangles*. **c** L1 mobilization begins with transcription of an L1 mRNA, which is translated to yield ORF1p and ORF2p. ORF1p, ORF2p, and the L1 mRNA form a ribonucleoprotein particle that re-enters the nucleus. The ORF2p endonuclease cleaves the first genomic DNA strand, while its reverse transcriptase uses a now free 3′ OH group as a primer for reverse transcription of the L1 mRNA. Following second-strand DNA cleavage, a new L1 copy is integrated into the genome and is typically flanked by TSDs

Gerdes et al. Genome Biology (2016) 17:100

Page 4 of 17
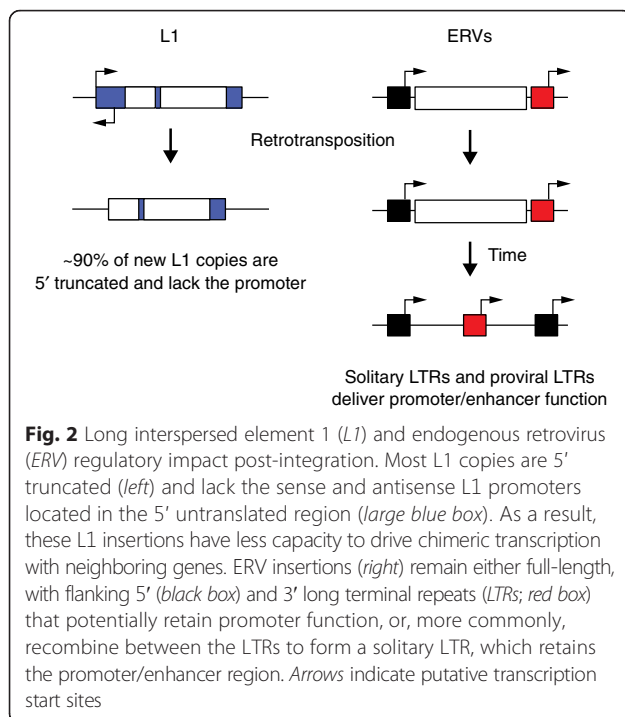
**Box 2. L1 retrotransposons**

The non-long terminal repeat retrotransposon long interspersed element-1 (L1) is the only autonomous, mobile human transposable element [10, 12, 116, 134]. L1 occupies approximately 17 % of the human genome [7]. L1 also mobilizes Alu and SINE–VNTR–Alu (SVA) elements in trans [135, 136]. Mice, by contrast, have three L1 subfamilies ($T_F$, $G_F$, and A) that are autonomous, as well as nonautonomous short interspersed elements (SINEs) retrotransposed by L1 [10]. L1 accounts for 19 % of the mouse genome [8]. A full-length human L1 is approximately 6 kb long and initiates mRNA transcription from a 5' sense promoter active in gametes, stem cells, and various somatic tissues [33, 36, 48, 71, 137–139]. The bicistronic L1 mRNA encodes two proteins, ORF1p and ORF2p, which are flanked by 5' and 3' untranslated regions (Fig. 1a). An L1 antisense peptide (ORF0p) [56] can also be expressed by an adjacent L1 antisense promoter [115]. This antisense promoter is expressed in many spatiotemporal contexts, including in stem cells, and can provide alternative promoters to protein-coding genes [33, 56, 115, 140]. L1 ORF2p presents endonuclease [141] and reverse transcriptase [142] activities and, during retrotransposition, L1 ORF1p, ORF2p, and the canonical L1 mRNA associate in cis to form a cytoplasmic ribonucleoprotein particle (RNP) [143]. The RNP can then enter the nucleus, where the ORF2p endonuclease cleaves genomic DNA, and the ORF2p reverse transcriptase synthesizes a new L1 copy at the cleavage site using the L1 mRNA as a template. This process is called target-site primed reverse transcription (TPRT) [144] (Fig. 1c).

The L1 5' promoter is the major focus of host genome efforts to prevent L1 mobility, through DNA methylation and transcription factor repression and other pathways [145, 146]. Thus, it appears that L1 in the main persists as a mobile element by avoiding detection of its 5' promoter by host genome surveillance pathways and, where this fails, by harnessing new promoter structures [13]. This could explain the exceptional L1 5' promoter diversity observed even among closely related primates [23]. It should also be noted that the vast majority of L1 copies in the genome are 5' truncated and lack the 5' promoter [13], meaning that the host factors guarding against full-length L1 transcription are not necessarily able to recognize truncated L1s.

**Box 3. Endogenous retroviruses**

Endogenous retroviruses (ERVs) are derived from exogenous retroviruses that, at some point, infected an individual organism's germ cells, integrated into their genome, and were subsequently inherited by their offspring. ERVs are divided into class I, class II, and class III elements, based on the exogenous virus class that they are most similar to [11]. Full-length ERVs are 5–10 kb in length, encode proteins important for mobilization, and are flanked by two identical long terminal repeats (LTRs; 300–1000 bp) that regulate ERV transcription. Loss of the env gene, found in exogenous retroviruses, is a common feature of ERVs as they adopt an intracellular life cycle as a retrotransposon [11, 147, 148]. ERV retrotransposition is initiated by the transcription of the 5' LTR and terminates in the 3' LTR, generating a terminally redundant mRNA that is translated into Gag and Gag–Pro–Pol fusion proteins. Gag proteins encapsulate the mRNA and fusion protein. Pro has protease activity whereas Pol possesses reverse transcriptase, ribonuclease, and integrase domains that generate independent proteins by proteolytic maturation. Together they produce a double-stranded cDNA copy of the ERV and flanking LTRs. This cDNA is then integrated into the genome by the ERV integrase [149] (Fig. 1b).

Human endogenous retroviruses (HERVs) comprise about 8 % of the human genome [7]. All HERVs are considered to be now retrotransposition incompetent [150, 151]. The HERV-K (HML-2) family is exceptional, with several members arising after the divergence of humans and chimpanzees (approximately 6 million years ago) and a handful of polymorphic HERV-K insertions found in human populations [152–155]. Although a mobile HERV-K element has yet to be identified in humans, it is possible that rare, as yet undiscovered polymorphic elements could retain retrotransposition competence [152]. In contrast to humans, ERVs account for approximately 10 % of the mouse genome [8]. Several mouse ERV families are still autonomously active, including intracisternal A-type particle elements [106], Moloney murine leukemia virus [156], and Mus type-D related retrovirus (MusD) [147] elements, as well as the MusD-dependent early retrotransposon family [157]. Together, new mouse ERV insertions are responsible for about 10 % of documented germline mutations in inbred strains [106]. Clade-specific ERVs also occur in other mammals, although genomic ERV content varies significantly between species [11]. Numerous instances of mammalian ERVs contributing regulatory sequences to genes, including examples of convergent evolution [158], are found in pluripotent cells and elsewhere [15, 159, 160].

Gerdes *et al. Genome Biology* (2016) 17:100

Page 5 of 17



**Fig. 2** Long interspersed element 1 (*L1*) and endogenous retrovirus (*ERV*) regulatory impact post-integration. Most L1 copies are 5′ truncated (*left*) and lack the sense and antisense L1 promoters located in the 5′ untranslated region (*large blue box*). As a result, these L1 insertions have less capacity to drive chimeric transcription with neighboring genes. ERV insertions (*right*) remain either full-length, with flanking 5′ (*black box*) and 3′ long terminal repeats (*LTRs*; *red box*) that potentially retain promoter function, or, more commonly, recombine between the LTRs to form a solitary LTR, which retains the promoter/enhancer region. *Arrows* indicate putative transcription start sites
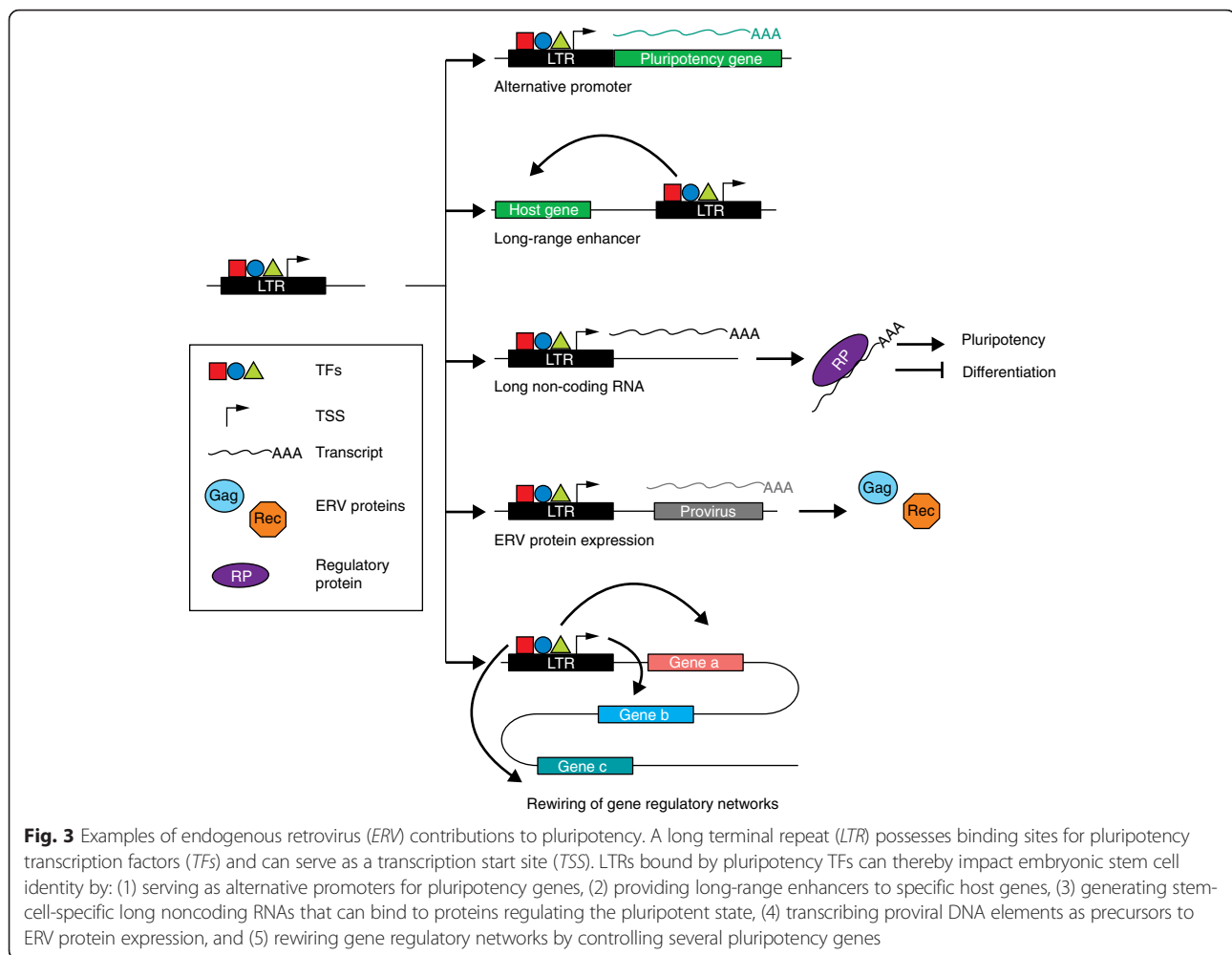
## Independent ERV expression as a hallmark of the pluripotent state

ERV transcription independent of protein-coding genes has also been linked to pluripotency. Despite an apparent lack of retrotransposition activity, specific HERVs are actively transcribed in hESCs and are thought to influence pluripotency maintenance [24, 25, 27–32, 47]. The HERV families HERV-H and HERV-K (HML-2) in particular appear to be connected to early human embryonic development [25, 31]. While stochastic transcriptional derepression of various HERVs [47] as well as non-LTR retrotransposons [48] in pluripotent cells can probably be attributed to a general relaxation of TE silencing [40], specific classes of elements are consistently reactivated across hESC lines, indicating that their expression can serve as a marker for an undifferentiated state [28, 29], further raising the possibility that these elements have a functional link to pluripotency. Distinct HERV families also denote specific embryonic stages, suggesting HERV expression profiles may signify cell identity [25]. It is important to note, however, that, in many cases, only a small fraction of HERVs from a specific family are transcribed [25] and that their genomic context likely plays a pivotal role in their expression. The reasons for HERV families presenting distinct expression patterns during early embryogenesis are currently unclear. To speculate, such patterns could be a reflection of the optimal "ecological niche" of their ancestral exogenous counterparts and may mimic the parallel expression patterns of LTR-binding TFs.

Human oocytes and zygotes (to the cell–cell stage) contain the highest percentages of HERV transcripts observed during development; these are almost certainly deposited maternally prior to embryonic genome activation [25]. Abundant transcription emanating from MaLR and ERVK LTRs has also been documented for mouse oocytes [5, 49]. The provision of ERV transcripts by the maternal genome supports ERV functionality in the early embryo, as these RNAs already seem to be necessary before the embryonic genome is able to generate its own transcripts [31]. However, it is also possible that ERV transcripts do not have a specific function at this early stage but their maternal deposition is permitted because they do not harm the developing embryo. Nevertheless, stage-specific expression from ERV promoters, and of protein-coding genes, LTR-driven chimeric transcripts, and ERV transcripts proper, is a defining feature of early mammalian development.

## Regulation of HERV-K and HERV-H by pluripotency factors

As well as gene regulation transacted by ERVs, many studies have revealed how ERVs are in turn regulated by pluripotency genes. For instance, the core pluripotency TFs Oct4 and Nanog (Box 1) bind specific HERV families (Fig. 3) [26, 42]. HERV-K is the most recently active HERV family and many HERV-K copies retain their protein-coding potential [50]. Notably, transcription from the youngest subclass of HERV-K is induced from its LTR, known as LTR5HS (for "human-specific"), at the eight-cell stage, during embryonic genome activation, and continues through to the blastocyst stage (Fig. 4a). LTR5HS contains an Oct4-binding motif that is not present in older LTRs such as LTR5a or LTR5b [31]. DNA hypomethylation and transactivation by Oct4 at LTR5HS synergistically stimulate HERV-K expression and lead to the presence of retroviral and viral-like particles in human preimplantation embryos [31]. HERV-K type 2 proviruses encode the protein Rec, which derives from alternative splicing of the *env* gene and is responsible for nuclear export and translation of viral RNAs [51]. Rec can be found in pluripotent cells and may influence expression of the interferon-induced viral restriction factor IFITM1 in epiblast cells [31, 52]. Consequently, Grow et al. [31] suggested that antiviral responses might be induced by HERV-K proteins, protecting the human embryo against new retroviral infections. Similarly, HERV-K type 1 proviruses encode the protein Np9, which is the product of a new alternative splicing event and coincides with a deletion in the *env* region [53, 54]. Interestingly, Rec and Np9 are not encoded in rodent ERVs, making them a distinguishing feature of primate ERVs and, moreover, hESCs specifically express Rec, Np9, and Gag [29]. It is tempting,

Gerdes *et al. Genome Biology* (2016) 17:100

Page 6 of 17



**Fig. 3** Examples of endogenous retrovirus (*ERV*) contributions to pluripotency. A long terminal repeat (*LTR*) possesses binding sites for pluripotency transcription factors (*TFs*) and can serve as a transcription start site (*TSS*). LTRs bound by pluripotency TFs can thereby impact embryonic stem cell identity by: (1) serving as alternative promoters for pluripotency genes, (2) providing long-range enhancers to specific host genes, (3) generating stem-cell-specific long noncoding RNAs that can bind to proteins regulating the pluripotent state, (4) transcribing proviral DNA elements as precursors to ERV protein expression, and (5) rewiring gene regulatory networks by controlling several pluripotency genes

therefore, to speculate, as per Grow et al. [31], that hESCs allow expression of these HERV-K proteins to fulfill a protective function via, for example, Rec-induced inhibition of viral infection. It is also possible that some HERV-K elements fortuitously escape silencing and manufacture viral proteins as innocuous byproducts of HERV-K transcription in hESCs (Fig. 3).

HERV-H is another primate-specific retrotransposon [55] with a potentially important role in the maintenance of hESC identity and pluripotency (Table 1). HERV-H transcripts are expressed in pluripotent cells at levels much higher than those seen in differentiated cells and, as a result, HERV-H expression is a proposed marker for pluripotency [28]. Interestingly, HERV-H is expressed in some induced pluripotent stem cell (iPSC) lines (Box 1) at higher levels than for other iPSC lines and embryonic stem cells (ESCs) [47]. Developmental HERV-H expression also appears to be cell type and stage specific in vivo (Fig. 4a). For instance, HERV-H and its flanking LTR element LTR7 can only be detected in epiblast cells [25], whereas other related LTR variants that flank

HERV-H (LTR7B and LTR7Y) are detectable at the eight-cell stage and morula [25]. LTR7 incorporates Oct4, Nanog, Klf4, and Lbp9 TF binding sites, which together appear to mediate HERV-H transcriptional activation [28]. Once activated, individual LTR7 copies can generate noncoding RNAs [43] and form chimeric transcripts with protein-coding genes, in some cases supplying multiple promoters to the same gene (Fig. 3) [27, 28, 56]. LTR7 may also be bound by factors central to so-called naïve, or ground state, pluripotency where cells are predisposed to self-renew and lack differentiation markers, showing that ERVs may be involved in fine tuning stem cell phenotype [28, 57]. In sum, HERV-K and HERV-H are clearly activated by pluripotency TFs and their expressed products are, at the very least, markers of pluripotency.

## HERV-derived long noncoding RNAs regulate pluripotency networks

Long noncoding RNAs (lncRNAs) are RNA transcripts greater than 200 nucleotides long that possess no, or
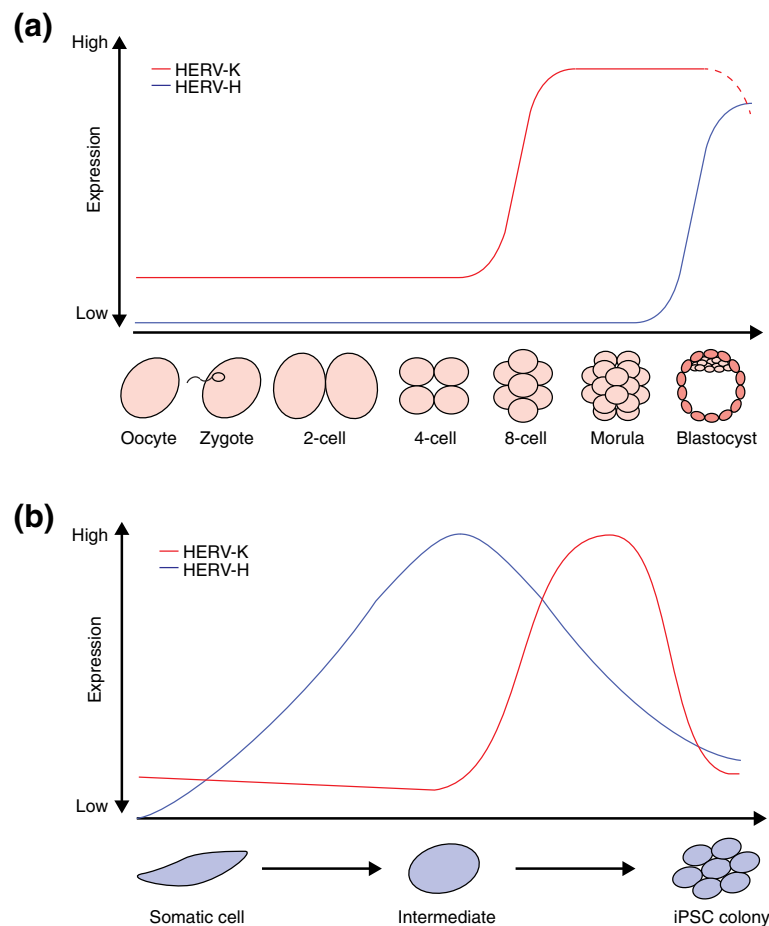
Gerdes *et al. Genome Biology* (2016) 17:100

Page 7 of 17



**Fig. 4** Human endogenous retrovirus (*HERV*) expression patterns in pluripotent cells. **a** HERV-K transcription in human embryogenesis is initiated during embryonic genome activation at the eight-cell stage and remains until the blastocyst stage. *Dashed lines* indicate proposed expression of HERV-K [31]. HERV-H can only be detected in epiblast cells of the late blastocyst [25]. **b** After induction of induced pluripotent stem cell (*iPSC*) reprogramming, HERV-K and HERV-H are derepressed with distinct dynamics. HERV-K transcription reaches its peak shortly before cells are fully reprogrammed. HERV-K expression subsequently decreases in reprogrammed cells and is silenced in iPSCs [32]. HERV-H is highly expressed earlier during reprogramming compared with HERV-K [24]. Note: the time points shown are approximate due to technical differences between studies

very little, protein-coding potential [58–60]. Most lncRNAs are transcribed antisense to protein-coding genes or are intergenic [58, 59]. More than two-thirds of lncRNAs incorporate TE sequences (Fig. 3) and, in cases such as *Xist*, a prototypical lncRNA involved in X chromosome inactivation, TEs are a core component of lncRNA biogenesis [60, 61]. Other than *Xist*, and a few additional examples, lncRNAs have proven difficult to evaluate functionally because, as well as containing TEs, lncRNAs are often expressed at very low levels [30]. However, one of the best established lncRNA functions is to regulate pluripotency, particularly by mediating changes to chromatin [62, 63]. Interestingly, Au et al. [64] reported more than 2000 additional long intergenic noncoding RNA (lincRNA) isoforms, of which 146 were expressed in hESCs. These human pluripotency-associated transcripts (HPATs) typically incorporated ERVs, especially HERV-H [30], and in that regard

were similar to many other hESC-specific lncRNAs [27, 43, 44, 47]. HPATs appear to contribute to formation of the blastocyst ICM, suggesting an essential role for HERV-derived lncRNAs in human embryogenesis [30].

One particularly interesting lincRNA, HPAT5, is hypothesized to be involved in post-transcriptional gene regulation: HPAT5 binds AGO2, a core protein catalyzing microRNA (miRNA) processing [65], and the let-7 miRNA family, which modulates hESC pluripotency [66]. Durruthy-Durruthy et al. [30] have suggested that HPAT5 controls the balance between pluripotency and differentiation by negatively regulating let-7 expression. However, HPAT5 is promoted by the so-called HUERS-P1 ERV, a low copy number TE that has not been investigated very deeply in this context. Interestingly, the HPAT5 promoter is located in the internal Gag sequence of the HUERS-P1 ERV, rather than in an LTR.

Gerdes *et al. Genome Biology* (2016) 17:100

Page 8 of 17

**Table 1** Summary of HERV-H findings to date in human stem cells

| Findings | Reference(s) |
| --- | --- |
| Binding of pluripotency TFs in ESCs and iPSCs within or near LTRs or specific LTR-driven lncRNAs | [24, 28, 42–45, 47] |
| Induction of HERV-H, or LTR7-driven lncRNA/chimeric RNA expression, in ESCs, declining upon differentiation | [26–28, 43–45, 47, 60, 161] |
| Active chromatin marks on specific LTRs in ESCs or iPSCs | [28, 32, 43, 44, 47, 162] |
| DNA hypomethylation at specific LTRs in ESCs | [28, 161] |
| LTR enhancer activity in ESCs | [27] |
| Changes in HERV-H associated with expression changes of genes near LTRs | [27, 28, 32, 42] |
| Induction of HERV-H in iPSCs, declining upon differentiation | [24, 27, 28, 32, 46] |
| Differentiation-defective iPSC clones retain high levels of HERV-H RNA | [46] |
| Knockdown of general HERV-H expression inhibits iPSC formation and causes ESC differentiation | [24, 27, 28] |
| Knockdown of specific LTR-driven RNAs inhibits iPSC formation or causes ESC differentiation | [24, 28, 44, 45] |
| LTR-driven lncRNA acts as miRNA sponge to positively regulate pluripotency TFs | [45] |
| HERV-H RNA associates with coactivators and LTR loci in ESCs | [27] |
| HERV-H expression marks for naïve-like stem cells | [28] |
| HERV-H LTR subtypes expressed sequentially in early development | [25, 31] |

*ESC* embryonic stem cell, *HERV* human endogenous retrovirus, *iPSC* induced pluripotent stem cell, *lncRNA* long noncoding RNA, *LTR* long terminal repeat, *miRNA* microRNA, *TF* transcription factor

Therefore, this promoter likely developed by genetic drift or selection, rather than by harnessing the "ready to use" regulatory motifs found within an LTR. In addition, the let-7 binding site within HPAT5 occurs within an imbedded *Alu* element. HPAT5 is thus an unusual, and yet fascinating, example of retrotransposon-driven regulatory innovation.

More broadly, HERV-driven transcripts contributing to pluripotency networks unique to humans or primates are of particular interest. *lincRNA-RoR*, with its TSS located in a HERV-H element, represents an excellent example of a primate-specific TE found to modulate pluripotency [43]. Notably, *lincRNA-RoR* is expressed more highly in iPSCs than in ESCs and can promote iPSC reprogramming [44], perhaps by serving as an miRNA sponge protecting Sox2 and Nanog from miRNA-mediated degradation [45]. In another example, the gene *ESRG*, which uses a domesticated HERV-H promoter, plays a role unique to human pluripotency [28]. Unusually, *ESRG* encodes an intact open reading frame (ORF) in humans, but possibly not in other primates, and is expressed exclusively in the human ICM and cultured pluripotent cells [67]. *ESRG* knockdown compromises stem cell self-renewal and promotes differentiation, while *ESRG* overexpression aids reprogramming [28]. These case studies demonstrate recurrent incorporation of annotated HERV-derived transcripts into pluripotency networks.

To discover new lncRNAs regulating pluripotency, Fort et al. [26] surveyed in depth the noncoding transcriptomes of mouse and human stem cells. The resulting pluripotency lncRNA catalog included numerous previously unreported antisense, intergenic, and intronic transcripts that initiate in ERVs. Consistent with an earlier report [33], Fort et al. found an exceptional variety of stem cell-specific TSSs that are not directly associated with protein-coding genes. These TSSs often overlap with TEs, especially with ERVK and MaLR LTR subfamilies in mice and ERV1 in humans, and frequently flank enhancer elements. In addition to bidirectional transcription denoting enhancer activity [68, 69], TE-derived enhancer sequences are enriched for bound Nanog, Sox2, Oct4, and the enhancer-related protein p300 [26]. As such, regulation of TE-derived enhancers and lncRNAs by pluripotency TFs can result in the formation of positive-feedback loops, potentially bolstering pluripotency networks [25, 26, 62]. Thus, in agreement with other studies, Fort et al. demonstrated that specific ERVs are major contributors to the stem cell transcriptome and found a plethora of novel stem cell-associated ERV-derived transcripts that await functional characterization, in line with expectation that some of these lncRNAs will be involved in the establishment and maintenance of pluripotency [70].

## ERV expression dynamics during somatic cell reprogramming

Domesticated TEs clearly play important functional roles in stem cell biology. However, TE repression can shift as cells transition through pluripotent states, as encountered during reprogramming. As a result, opportunistic TEs may mobilize, cause insertional mutagenesis and, potentially, compromise the integrity of reprogrammed cells [32, 48, 71]. TE activity in stem cells therefore carries risk as well as benefits for the host genome, along with major incentives for TEs, given potential for early

Gerdes *et al. Genome Biology* (2016) 17:100

Page 9 of 17

embryonic retrotransposition events to be germline transmitted. It follows that, although reprogramming can broadly reactivate TEs, particularly those controlled by TFs expressed dynamically during reprogramming [16, 42], silencing is selectively re-established in the resulting pluripotent cells, potentially ameliorating risk to the host genome. For instance, although HERV-H and HERV-K are both transcriptionally active during reprogramming, HERV-H is expressed in cultured iPSCs, whereas the more recently mobile HERV-K family is silenced [28] (Fig. 4b). This contrast is also found for mouse iPSCs, where Mus type-D related retrovirus (MusD) expression contrasts with intracisternal A-type particle (IAP) silencing [32]. Importantly, more experiments are required to confirm the generality of these observations, as technical considerations in iPSC generation (e.g., reprogramming and culture conditions) can lead to differences in TE expression between iPSC lines [71].

TE repression is dynamic during reprogramming. In a high-resolution analysis of mouse and human iPSC lines, Friedli et al. [32] found that most ERVs peaked in expression shortly before reprogramming was complete and were then repressed in pluripotent cells. Broad TE expression during somatic cell reprogramming may be in itself important for the induction of the pluripotent state. Ohnuki et al. [24] reported, for example, that LTR7 elements (associated with HERV-H) are hyperactivated by Oct4, Sox2, and Klf4 during reprogramming. In the resultant iPSCs, however, LTR7 activity decreased to levels seen in hESCs and, notably, ectopic LTR7 hyperactivity in iPSCs resulted in a differentiation-defective phenotype [24]. Similarly, cumulative HPAT expression rises markedly during reprogramming and is diminished in iPSCs and, as for HPAT5, may influence reprogramming efficiency [30]. Taken together, these data indicate that TE hyperactivity is potentially deleterious to the host genome due to an elevated risk of retrotransposition but may also be a requirement of induced reprogramming.

## ERV silencing in pluripotent states

The machineries responsible for ERV regulation in ESCs are evidence of the complex relationships that can form between TEs and their host genome. Broadly speaking, to reduce the probability of retrotransposon-derived mutagenesis, mammalian genomes target ERVs with DNA methylation, heterochromatin-forming factors, transcriptional repressor complexes, proviral silencing factors, and post-transcriptional arrest or degradation of viral RNAs (Table 2) [19, 20, 72]. Prominently, histone modifications silence ERVs in ESCs [73–75] by making chromatin inaccessible to polymerases and transcription factors [76], although this silencing in itself carries potential for deleterious side effects when nearby genes are also inadvertently repressed [77]. Moreover, some ERVs are marked by H3K9me3 and H4K20me3 for repression in ESCs but not in differentiated cells [6], suggesting that this pathway is used for de novo establishment of

**Table 2** Selected factors silencing ERVs in embryonic stem cells

| Factor | Relevant function in ESCs | Interacting proteins | Reference(s) |
|---|---|---|---|
| Zfp809 | KRAB zinc finger protein, recognizes PBS Pro and recruits Trim28 | Trim28 | [85] |
| YY1 | Yin-Yang 1 (YY1) is a zinc finger protein, initially binds to mouse ERV LTRs (U3 region) and helps assemble Trim28/Eset silencing complex | Trim28 | [92] |
| SHIN | Short heterochromatin inducing sequence initiates heterochromatin and transcriptional repression | KRABZfps (likely) | [93] |
| Trim28 (Kap1, Tif1-β) | Transcriptional co-repressor, acts as a bridge between KRAB-Zfps and other transcriptional repressors, mediates H3K9me3 recruitment | KRABZfps, Eset, NuRD deacetylase complex, HP1 | [74, 86] |
| Eset (Setdb1, Kmt1e) | Histone methyltransferase, trimethylates H3K9 and H4K20, crucial for HP1 binding | Trim28, HP1 | [73] |
| Sumo2 | Sumoylation factor, post-translational sumoylation of Trim28 enhances recruitment of Trim28 to proviral DNA | Trim28 | [72] |
| HP1 | Heterochromatin protein 1, binding to Trim28 might be important for repression of transcription | Trim28, Eset, Atrx | [86, 89] |
| Chaf1a | Histone chaperone, deposits histone H3/H4 which marks proviral DNA for silencing, might execute different silencing mechanisms on different classes of ERVs | Eset, Kdm1a, Hdac1/2, Asf1a/b | [72] |
| Asf1a/b | Histone chaperones, components of the Chaf1a interactome | Chaf1a | [72] |
| Atrx | ATP-dependent helicase, establishes and maintains heterochromatin | Eset, Trim28 (both likely) | [93] |
| Daxx | H3.3-specific chaperone, may facilitate H3.3-deposition on ERV sequences, possible role in SHIN-silencing | Atrx | [93, 95] |

*ERV* endogenous retrovirus, *ESC* embryonic stem cell, *KRAB* Krüppel-associated box, *LTR* long terminal repeat, *SHIN* short heterochromatin inducing sequence, *Zfp* zinc finger protein

Gerdes *et al. Genome Biology* (2016) 17:100

Page 10 of 17

heterochromatin around ERV sequences [75, 78] or, alternatively, is used to maintain repression already established in oocytes [79, 80].

Even ERVs in accessible chromatin can be decisively silenced by DNA methylation. In mice, de novo DNA methylation is regulated by the canonical Zfp/Trim28/Eset machinery [75]. Krüppel-associated box (KRAB) zinc finger proteins (Zfps) play a major role in the initiation of ERV silencing [81, 82]. Indeed, the number of ERVs and Zfp genes in vertebrates are correlated, suggesting coevolution [83]. As an example of the complexity of Zfp-mediated retrovirus silencing, Zfp809 knockout induces the in vivo expression of Moloney murine leukemia virus (MMLV)-like 30 (VL30) provirus [84]. Zfp809 also binds to MMLV and initiates silencing by recruiting Trim28 (also known as Kap1) [74, 85, 86]. Trim28 activity is enhanced by post-translational sumoylation by Sumo2 [72, 87] and binds HP1, which is thought to contribute to the ability of Trim28 to repress transcription in the context of MMLV silencing [86, 88, 89]. Another Zfp, YY1, also binds to MMLV [90, 91] and, together with Zfp809, is thought to recruit Trim28 to ensure a stably DNA-bound silencing complex [92]. In another example, KRAB Zfps have been shown to trigger heterochromatin formation in IAP retrotransposons by binding to a short heterochromatin inducing (SHIN) sequence, dependent on Eset and Trim28 [93], enacting H3K9 and H4K20 trimethylation [73]. Chaf1a facilitates deposition of these H3 and H4 variants and also interacts with Eset [72]. Eset-mediated ERV silencing is also important in mouse primordial germ cells before the onset of de novo DNA methylation [80]. Hence, ERV silencing is enacted by a multilayered and interleaved system that ensures robust and specific repression of ERV families, subsets, and individual loci.

It follows that models explaining ERV silencing are typically complex, which, at times, can lead to differing conclusions. For instance, the SNF2-type chromatin remodeler Atrx is another crucial component for IAP silencing that renders Eset-dependent heterochromatin less accessible [93] and is likely to be recruited to IAPs by Trim28 and Eset [93] (Fig. 5a). Interestingly, Atrx has been reported to interact with the H3.3-specific chaperone Daxx to facilitate H3.3 deposition at telomeric heterochromatin [94]. Yet, it is not clear if H3.3 is required for ERV silencing, despite detection of H3.3 across ERV flanking regions and solo LTRs [95]. In general, Sadic et al. [93] and Elsässer et al. [95] reached opposing conclusions with regards to H3.3 enrichment around ERV sequences (Fig. 5b). One possible explanation here is that Elsässer et al. used chromatin immunoprecipitation sequencing (ChIP-seq) to detect H3.3-enriched regions across the entire mouse genome and found a correlation between H3.3, H3K9me3, and ERV

coordinates. Sadic et al., on the other hand, used an engineered reporter assay to measure ERV silencing which, in H3.3 knockout cells, remained intact. Further study is therefore required to resolve the place of H3.3 in models of ERV silencing. Overall, these and other examples of TE repression in pluripotent cells, such as the silencing of nascent L1 and MMLV insertions in embryonic carcinoma derived cell lines [96, 97], reflect the extraordinary efforts made by the host genome to orchestrate silencing of currently and recently retrotransposition-competent TEs during embryonic development.

## Endogenous L1 mobilization in mammalian somatic cells

The early embryo is a viable niche for the generation of potentially heritable retrotransposon insertions. In particular, L1 mobilization in human and rodent embryos may drive somatic and germline mosaicism [98–101] and, indeed, deleterious human L1 insertions transmitted from mosaic parents to offspring have resulted in sporadic genetic disease [101]. In vitro experiments have likewise provided support for L1 mobilization occurring in pluripotent cells [99–101] and, potentially, the presence of the L1 retrotransposition machinery being required for preimplantation mouse embryo development [102]. Human iPSCs and ESCs allow low-level mobilization of an engineered L1 reporter [22, 48, 99]. Consistently, endogenous L1 promoter hypomethylation and transcriptional activation have been observed in iPSCs [32, 48, 71], as has induction of a primate-specific L1 antisense peptide (ORF0p) that appears to increase L1 mobility in stem cells [56] (Box 2). Endogenous de novo L1 retrotransposition and mobilization of nonautonomous *Alu* and SINE–VNTR–*Alu* (SVA) elements have also been reported by Klawitter et al. [71] in several iPSC lines, as well as an *Alu* insertion in a cultured hESC line. L1 may, therefore, *trans* mobilize *Alu* and other SINEs during development, an important finding due to the high potential of SINEs to impact gene regulation [12, 71, 103, 104]. Klawitter et al. estimated that approximately one de novo L1 insertion occurred per cell in human iPSCs. Strikingly, more than half of the detected de novo L1 insertions were full length and thus potentially able to mobilize further. Klawitter et al. also observed extraordinary induction of L1 mRNA and protein expression after reprogramming. To speculate, numerous L1 ribonucleoprotein particles (RNPs; Box 2) could form as a result and be carried through iPSC culture and differentiation. This could enable L1-mediated insertional mutagenesis in cells descending from those where L1 expression originally occurred, as others have considered for L1 RNPs arising in gametes and carrying over into the zygote [100].
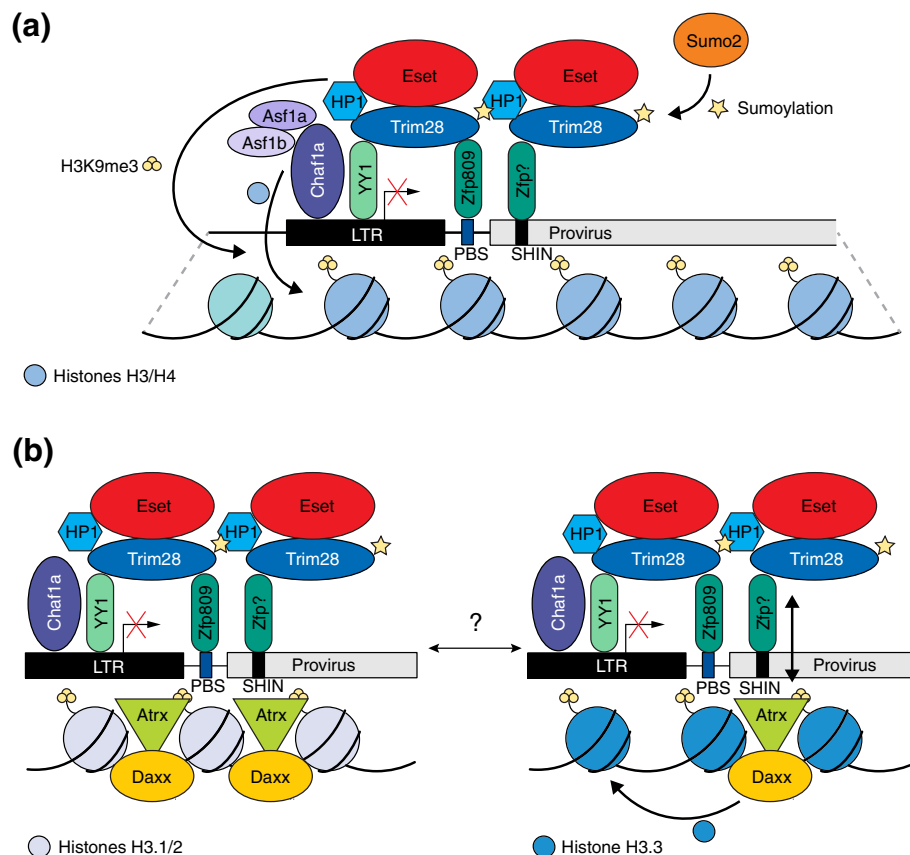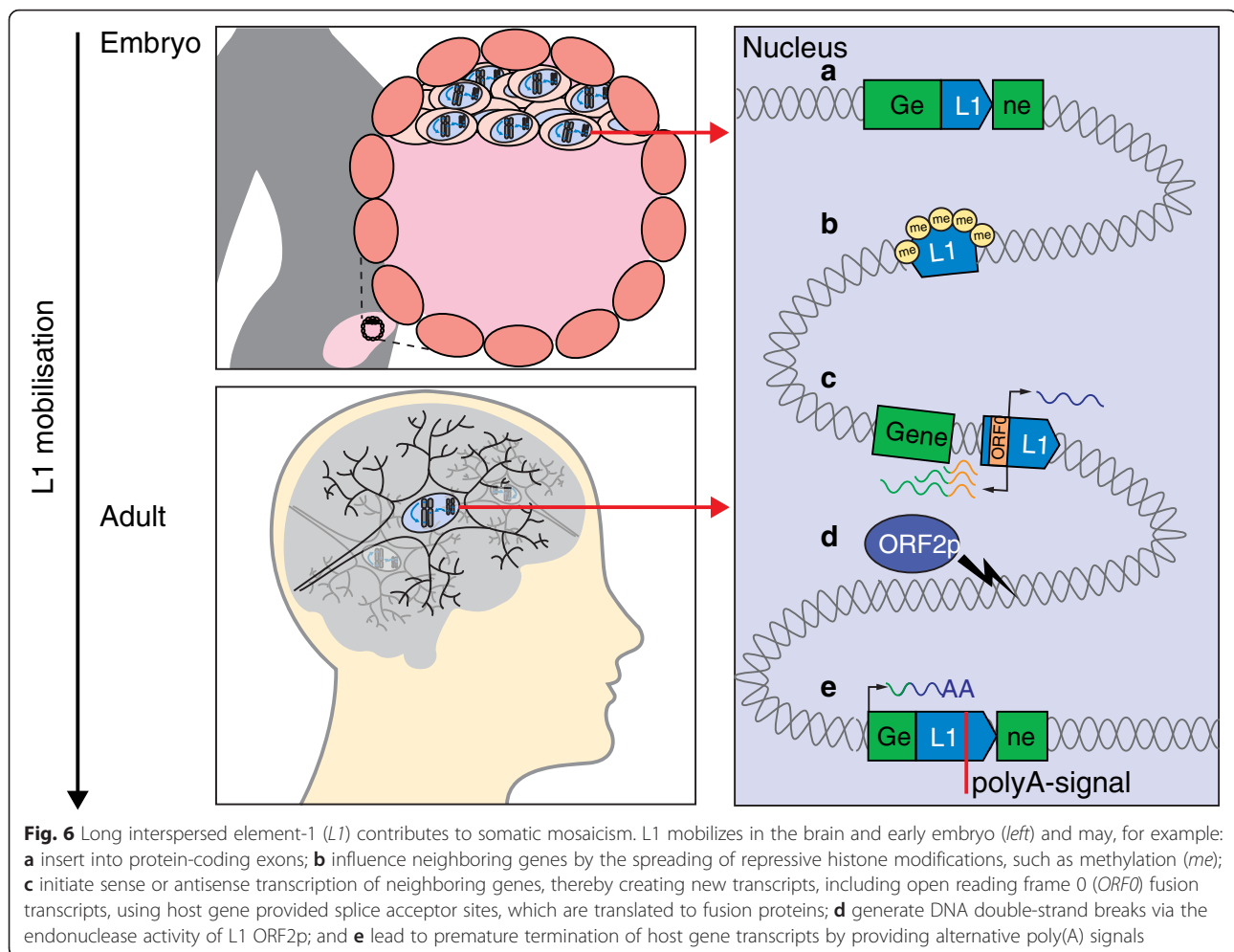
Gerdes *et al. Genome Biology* (2016) 17:100

Page 11 of 17



**Fig. 5** Proposed models of de novo endogenous retrovirus (*ERV*) silencing in embryonic stem cells. **a** To initiate silencing, the Krüppel-associated box (KRAB) zinc finger protein (*Zfp*) Zfp809 interacts with the proline primer binding site (PBS Pro) of some ERV families (e.g., Moloney murine leukemia virus) [85] whereas other KRAB-Zfps bind to a short heterochromatin-inducing (*SHIN*) sequence found in intracisternal A-type particle retrotransposons and other ERV families [93]. Subsequently, Trim28 is recruited by the Zfps [74, 86], assisted by binding of YY1 to the long terminal repeat (*LTR*) and Trim28 [92]. Interaction with HP1 and sumoylation by Sumo2 are thought to contribute to transcriptional repression mediated by Trim28 [72, 86, 89]. Eset also interacts with Trim28 and enables trimethylation of H3K9 and H4K20 [73]. The histone chaperone Chaf1a, aided by Asf1a/b, marks proviral DNA for silencing by depositing histones H3 and H4 and interacts with Eset [72]. **b** Conflicting models of ERV silencing by H3.3 deposition. The Atrx–Daxx complex is suggested to play an important role in SHIN-mediated silencing, which is H3.3-independent. Here, Atrx is thought to promote ERV heterochromatin inaccessibility (*left*) [93]. However, Atrx–Daxx is also proposed to deposit H3.3 and to interact with Trim28, followed by H3.3 being marked with H3K9me3 by Eset (*right*) [95]

Although both L1 and ERV retrotransposons are active in the mouse germline [105, 106], their capacity to mobilize during embryogenesis is less clear than for human L1. Quinlan et al., for instance, concluded de novo retrotransposition in mouse iPSCs did not occur, or was very rare [107], in contrast to results for human iPSCs [22, 48, 71]. However, an earlier study found that engineered L1 reporter genes mobilize efficiently in mouse embryos [100]. Interestingly, the vast majority of engineered L1 insertions in these animals were not heritable, perhaps indicating retrotransposition later in embryogenesis [100]. Targeted and whole-genome sequencing applied to mouse pedigrees has, conversely, revealed that endogenous L1 mobilization in early embryogenesis is relatively common and often leads to heritable L1 insertions (SRR and GJF, unpublished data). Polymorphic

ERV and nonautonomous SINE insertions are also found in different mouse strains [105, 106]. Although the developmental timing of these events is as yet unresolved, we reason that they can occur in spatiotemporal contexts supporting L1 retrotransposition. It follows that both human and mouse L1s, and probably mouse ERVs, can mobilize in embryonic and pluripotent cells (Fig. 6), as well as gametes. The resultant mosaicism can be deleterious to the host organism or their offspring [101], again reinforcing the need for TE restraint during early development.

Somatic L1 retrotransposition can also occur later in development. Over the past decade, it has become accepted that the mammalian brain, particularly cells of the neuronal lineage, accommodate mobilization of engineered and endogenous L1 elements [34–37, 108].

Gerdes *et al. Genome Biology* (2016) 17:100

Page 12 of 17



**Fig. 6** Long interspersed element-1 (*L1*) contributes to somatic mosaicism. L1 mobilizes in the brain and early embryo (*left*) and may, for example: **a** insert into protein-coding exons; **b** influence neighboring genes by the spreading of repressive histone modifications, such as methylation (*me*); **c** initiate sense or antisense transcription of neighboring genes, thereby creating new transcripts, including open reading frame 0 (*ORF0*) fusion transcripts, using host gene provided splice acceptor sites, which are translated to fusion proteins; **d** generate DNA double-strand breaks via the endonuclease activity of L1 ORF2p; and **e** lead to premature termination of host gene transcripts by providing alternative poly(A) signals

Although the frequency of somatic L1 insertions during neurogenesis is disputed [35, 36, 108, 109], this is largely due to differences in the advanced techniques required to discriminate genuine de novo L1 insertions and molecular artifacts arising during whole-genome amplification of individual human neurons. This discrimination can, broadly, be achieved quantitatively, by assuming true-positives will accrue more DNA sequencing reads than artifacts [108], or qualitatively, by analyzing the junction DNA sequences between putative L1 insertions and the flanking genome and excluding examples inconsistent with target-site primed reverse transcription [35]. Despite this debate, there is agreement that L1 mobilization occurs in the brain and can, for the most part, be traced to neuronal precursor cells [35, 36, 109]. Remarkably, neuronal L1 insertions are distributed unevenly genome-wide and are enriched in neurobiological genes and transcribed neuronal enhancers [34, 35]. Somatic L1 insertions oriented in sense to host genes, as the configuration most likely to disrupt transcription [110, 111], are heavily depleted versus random expectation, providing possible evidence of selection against these

events during neurogenesis [35]. Concordantly, somatic L1 insertions in neurobiological genes carry an elevated chance of yielding a molecular phenotype in the brain, especially provided the numerous routes by which L1 insertions can profoundly modify gene structure and expression (Fig. 6) [12, 33, 77, 110, 112–118].

Neuronal L1 insertions impart no obvious evolutionary benefit as they cannot be transmitted to subsequent generations. Thus, it is tempting to speculate that L1 activity is derepressed during neuronal commitment to serve a biological purpose for the host organism, analogous to the potential exaptation of ERV transcription for pluripotency maintenance and following the example of the vertebrate adaptive immune system, where domesticated TEs mediate V(D)J recombination and functional diversification through genomic mosaicism [119]. Similarly, although individual somatic L1 insertions in neurons are not inherited, it is plausible that the cellular mechanisms and factors enabling their production may undergo evolutionary selection [109]. While L1-mediated somatic mosaicism in neurons may eventually be shown to have functional or behavioral

Gerdes *et al. Genome Biology* (2016) 17:100

Page 13 of 17

consequences [109, 118], numerous additional experiments are required to assess this hypothesis. Whether perturbation of L1 regulation and retrotransposition in the brain is connected to neurological disease is not yet clear [35, 120–122]. The available evidence does, however, show conclusively that TE mobilization occurs during embryogenesis and, in a more restricted fashion, later in life.

## Conclusions

The mammalian genome clearly strives to limit TE activity in pluripotent cells. The silencing mechanisms involved are collectively complex and broadly potent and yet are also capable of great specificity and dynamism in targeting individual TE copies [17]. In this regard, ERVs present two contrasting facets: firstly, the control mechanisms that have evolved to restrict ERV activity and, secondly, the domestication of ERV sequences into pluripotency maintenance. Specific ERV families, such as HERV-H and HERV-K, can provide binding sites for pluripotency TFs, produce stem cell-specific protein-coding and noncoding transcripts, and harbor new enhancers. Over time, these contributions have led to the integration of ERVs into gene networks governing embryogenesis and, surprisingly, independent ERV hyperactivity appears to be a harbinger of pluripotent states. Conversely, notwithstanding a need for more experimental data for murine ERVs, L1 appears to be the most successful TE to mobilize in mammalian somatic cells and, at the same time, is arguably less likely to impact their phenotype than ERVs (Fig. 2). During human iPSC reprogramming, for example, L1 and ERVs can both be broadly derepressed, but with divergent repercussions for the host genome and providing different opportunities to each TE family.

Why are TEs active, and apparently essential, in the embryo? The relationship between TEs and the host genome is often referred to as an evolutionary arms race [123, 124]. A review specifically addressing the role of TEs in pluripotency [14] refined this concept to more of a genetic conflict of interest between ERVs and the host genome, where exposure to retrotransposition was a necessary risk of the pluripotent state. The authors, as others have done [28], also considered the possibility that ERVs were active in stem cells by serendipity. Despite their merits, each of these alternatives is contradicted by several considerations. Firstly, L1 mobilization appears to be far more common in the embryo than ERV mobilization, despite ERV domestication being overtly more useful to the host given the many ways ERVs can reinforce pluripotency (Fig. 3). The benefits of unleashing L1 and ERV activity do not seem then, in either case, to be commensurate with the implied risk of doing so. Secondly, ERVs are intrinsic to the pluripotent

state but are now almost, if not fully, immobile in humans. Thirdly, different ERV families are centrally involved in human and mouse pluripotency; convergent evolution driven by the common environmental demands of embryonic development, which are conserved among mammals, is an improbable outcome of chance. Here, time and scale are critical considerations: the vast majority of new ERV insertions will be immediately silenced but, as the retrotranspositional potential of an ERV family is eliminated over time via mutations, pressure to silence the associated LTRs may also diminish, allowing them to regain their regulatory activity. Hence, with sufficient time, distinct ERV families in different species can ultimately come to occupy similar niches, in pluripotency and elsewhere. TEs pervade mammalian genomes and, as such, even the low probability of a de novo ERV insertion immediately escaping silencing presents a reasonable overall chance of such events becoming important to genome-wide regulation. This remains true even if the ERV family is eventually immobilized.

Although not rejecting models based on serendipity or conflict, we highlight that ERVs and other successful TE families commonly arise as low copy number families and then rapidly expand over generations. This scenario could lead to TEs acquiring traits of early pioneers in a potentially hostile genomic landscape. Two not necessarily exclusive strategies can aid TE survival in this environment. One is stealth. For example, adaptation of the L1 5′ promoter (Box 2) enables evasion of host genome surveillance, leading to continuing L1 retrotransposition during development. That most new L1 copies are 5′ truncated, and lack the canonical promoter, also reduces their visibility to surveillance. Although this self-limits the capacity of new L1 insertions to retrotranspose, it also reduces pressure on the host genome to clamp down on L1 activity. The other strategy is gaining acceptance by being useful. ERV promoters are repeatedly found in pluripotency regulatory networks and may, therefore, be intrinsic to the pluripotent state. In this setting, efforts by the host genome to limit ERV activity could be detrimental to pluripotency. As such, ERVs may be able to propagate for longer than would be possible should the host engage in resolute inhibition. Importantly, these strategies are predicated on embryonic retrotransposition having potential for germline transmission, i.e., carrying risk for host genome integrity, as many studies have now found. Even after ERV families are no longer capable of mobilization, their inherent capacity for regulation, especially by solo LTRs, is retained and provides a long-term evolutionary incentive for the host genome to maintain at least one active TE family, as almost all mammals do. As such, rather than an arms race, conflict, or even symbiotic relationship, we would propose that pioneer ERVs adopt peaceful survival

Gerdes *et al. Genome Biology* (2016) 17:100

Page 14 of 17

strategies and that intricate mechanisms for TE repression have evolved to allow the host genome to harness those strategies over time, allowing some ERV families to expand and, as witnessed in the embryo, securely embed themselves by becoming indispensable. In advocating this model, we emphasize that indispensability of ERV-mediated regulatory effects in natural pluripotency and embryogenesis in vivo is still an open question. While difficult to pursue in humans, genetic knockout or deletion of individual mouse ERVs or ERV families implicated in pluripotency is possible [125] and, indeed, ultimately necessary to demonstrate their functional importance to the embryo.

### Abbreviations
ERV: endogenous retrovirus; ESC: embryonic stem cell; HERV: human endogenous retrovirus; hESC: human embryonic stem cell; HPAT: human pluripotency-associated transcript; IAP: intracisternal A-type particle; ICM: inner cell mass; iPSC: induced pluripotent stem cell; KRAB: Krüppel-associated box; L1: long interspersed element-1; lincRNA: long intergenic noncoding RNA; LINE: long interspersed element; lncRNA: long noncoding RNA; LTR: long terminal repeat; miRNA: microRNA; MMLV: Moloney murine leukemia virus; ORF: open reading frame; RNP: ribonucleoprotein particle; SINE: short interspersed element; TE: transposable element; TF: transcription factor; TSS: transcription start site; Zfp: zinc finger protein.

### Author details
[1]Mater Research Institute, University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia. [2]Department of Medical Genetics, Terry Fox Laboratory, British Columbia Cancer Agency, University of British Columbia, Vancouver, BC V5Z 1L3, Canada. [3]School of Biomedical Sciences, University of Queensland, Brisbane QLD 4072, Australia.

Published online: 09 May 2016

### References
1. Tarkowski A. Experiments on the development of isolated blastomers of mouse eggs. Nature. 1959;184:1286–7.
2. Papaioannou VE, Mkandawire J, Biggers JD. Development and phenotypic variability of genetically identical half mouse embryos. Development. 1989; 106:817–27.
3. Latham KE, Schultz RM. Embryonic genome activation. Front Biosci. 2001;6: D748–59.
4. Cantone I, Fisher AG. Epigenetic programming and reprogramming during development. Nat Struct Mol Biol. 2013;20:282–9.
5. Peaston AE, Evsikov AV, Graber JH, Vries WN D, Holbrook AE, Solter D, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004;7:597–606.
6. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448:553–60.
7. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409: 860–921.
8. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420:520–62.
9. de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011;7, e1002384.
10. Ostertag EM, Kazazian HH. Biology of mammalian L1 retrotransposons. Annu Rev Genet. 2001;35:501–38.
11. Mager D, Stoye J. Mammalian endogenous retroviruses. Microbiol Spectr. 2015;3:MDNA3-009-2014.
12. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009;10:691–703.
13. 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.
14. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? Bioessays. 2016;38:109–17.
15. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2011;46:21–42.
16. Bourque G, Leong B, Vega VB, Chen X, Yen LL, Srinivasan KG, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 2008;18:1752–62.
17. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol. 2015;31:429–51.
18. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. 1980;284:601–3.
19. Marchetto MCN, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL, et al. Differential L1 regulation in pluripotent stem cells of humans and apes. Nature. 2013;503:525–9.
20. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology. 2011;411:273–87.
21. Schlesinger S, Goff SP. Retroviral transcriptional regulation and embryonic stem cells: war and peace. Mol Cell Biol. 2015;35:770–7.
22. Wissing S, Montano M, Garcia-Perez JL, Moran JV, Greene WC. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. J Biol Chem. 2011;286:36427–37.
23. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 2014;516:242–5.
24. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, et al. Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proc Natl Acad Sci U S A. 2014;111:12426–31.
25. Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, Sachs F, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell. 2015;16:135–41.
26. Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat Genet. 2014;46:558–66.
27. Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol. 2014;21:423–5.
28. Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014;516:405–9.
29. Fuchs NV, Loewer S, Daley GQ, Izsvák Z, Löwer J, Löwer R. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 2013;10:115.
30. Durruthy-Durruthy J, Sebastiano V, Wossidlo M, Cepeda D, Cui J, Grow EJ, et al. The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. Nat Genet. 2016;48:44–52.

Gerdes *et al. Genome Biology* (2016) 17:100

Page 15 of 17

31. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522:221–5.

32. Friedli M, Turelli P, Kapopoulou A, Rauwel B, Castro-Díaz N, Rowe HM, et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. 2014;24:1251–9.

33. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009;41:563–71.

34. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011;479:534–7.

35. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, et al. Ubiquitous L1 mosaicism in hippocampal neurons. Cell. 2015;161:228–39.

36. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460:1127–31.

37. Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature. 2005;435:903–10.

38. Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta. 2012;33:663–71.

39. Conley AB, Miller WJ, Jordan IK. Human *cis* natural antisense transcripts initiated by transposable elements. Trends Genet. 2008;24:53–6.

40. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. Genes Dev. 2014;28:812–28.

41. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012;487:57–63.

42. Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42:631–4.

43. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13:R107.

44. Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, Park IH, et al. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet. 2010;42:1113–7.

45. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell. 2013;25:69–80.

46. Koyanagi-Aoi M, Ohnuki M, Takahashi K, Okita K, Noma H, Sawamura Y, et al. Differentiation-defective phenotypes revealed by large-scale analyses of human pluripotent stem cells. Proc Natl Acad Sci U S A. 2013;110:20569–74.

47. Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology. 2012;9:111.

48. Wissing S, Muñoz-Lopez M, Macia A, Yang Z, Montano M, Collins W, et al. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. Hum Mol Genet. 2012;21:208–18.

49. Veselovska L, Smallwood SA, Saadeh H, Stewart KR, Krueger F, Maupetit-Méhouas S, et al. Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol. 2015;16:209.

50. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology. 2011;8:90.

51. Löwer R, Tönjes RR, Korbmacher C, Kurth R, Löwer J. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. J Virol. 1995;69:141–9.

52. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013;20:1131–9.

53. Armbruester V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, et al. A novel gene from the human endogenous retrovirus K expressed in transformed cells. Clin Cancer Res. 2002;8:1800–7.

54. Armbruester V, Sauter M, Roemer K, Best B, Hahn S, Nty A, et al. Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X. J Virol. 2004;78:10310–9.

55. Mager DL, Freeman JD. HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. Virology. 1995;213:395–404.

56. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MCN, et al. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell. 2015;163:583–93.

57. Dunn S-J, Martello G, Yordanov B, Emmott S, Smith AG. Defining an essential transcription factor program for naïve pluripotency. Science. 2014;344:1156–60.

58. FANTOM Consortium, RIKEN Genome Exploration Research Group, Genome Science Group, Carninci P. The transcriptional landscape of the mammalian genome. Science. 2005;309:1559–63.

59. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. Science. 2005;309: 1564–6.

60. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 2013;9, e1003470.

61. Duret L, Chureau C, Samain S, Weissenbach J, Avner P. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. Science. 2006;312:1653–5.

62. Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature. 2011;477:295–300.

63. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell. 2007;129:1311–23.

64. Au KF, Sebastiano V, Afshar PT, Durruthy JD, Lee L, Williams BA, et al. Characterization of the human ESC transcriptome by hybrid sequencing. Proc Natl Acad Sci U S A. 2013;110:E4821–30.

65. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature. 2005;436:740–4.

66. Melton C, Judson RL, Blelloch R. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. Nature. 2010;463:621–6.

67. Zhao M, Ren C, Yang H, Feng X, Jiang X, Zhu B, et al. Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. Biochem Biophys Res Commun. 2007;362: 916–22.

68. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. 2015;347:1010–4.

69. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014;507:455–61.

70. Macia A, Blanco-Jimenez E, García-Pérez JL. Retrotransposons in pluripotent cells: impact and new roles in cellular plasticity. Biochim Biophys Acta. 1849; 2015:417–26.

71. Klawitter S, Fuchs NV, Upton KR, Muñoz-Lopez M, Shukla R, Wang J, et al. Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat Commun. 2016;7:10286.

72. Yang BX, El Farran CA, Guo HC, Yu T, Fang HT, Wang HF, et al. Systematic identification of factors for provirus silencing in embryonic stem cells. Cell. 2015;163:230–45.

73. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature. 2010;464:927–31.

74. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010;463:237–40.

75. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, et al. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development. 2013;140:519–29.

76. Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, Crabtree GR. Dynamics and memory of heterochromatin in living cells. Cell. 2012;149:1447–60.

77. Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, et al. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genet. 2011;7, e1002301.

78. Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, et al. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the

Gerdes *et al. Genome Biology* (2016) 17:100

Page 16 of 17

histone methyltransferase Setdb1. Proc Natl Acad Sci U S A. 2014;111: 6690–5.

79. Messerschmidt DM, de Vries W, Ito M, Solter D, Ferguson-Smith A, Knowles BB. Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science. 2012;335:1499–502.

80. Liu S, Brind'Amour J, Karimi MM, Shirane K, Bogutz A, Lefebvre L, et al. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. 2014; 28:2041–55.

81. Wolf G, Greenberg D, Macfarlan TS. Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. Mob DNA. 2015;6:17.

82. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, et al. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nat Biotechnol. 2015;33:555–62.

83. Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. Genome Res. 2011;21:1800–12.

84. Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, Park C, et al. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev. 2015;29:538–54.

85. Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature. 2009;458:1201–4.

86. Wolf D, Goff SP. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell. 2007;131:46–57.

87. Mascle XH, Germain-Desprez D, Huynh P, Estephan P, Aubry M. Sumoylation of the transcriptional intermediary factor 1beta (TIF1beta), the co-repressor of the KRAB multifinger proteins, is required for its transcriptional activity and is modulated by the KRAB domain. J Biol Chem. 2007;282:10190–202.

88. Wolf D, Cammas F, Losson R, Goff SP. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. J Virol. 2008;82:4675–9.

89. Sripathy SP, Stevens J, Schultz DC. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. Mol Cell Biol. 2006;26:8623–38.

90. Flanagan JR, Krieg AM, Max EE, Khan AS. Negative control region at the 5′ end of murine leukemia virus long terminal repeats. Mol Cell Biol. 1989;9: 739–46.

91. Flanagan JR, Becker KG, Ennist DL, Gleason SL, Driggers PH, Levi BZ, et al. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol Cell Biol. 1992;12: 38–44.

92. Schlesinger S, Lee AH, Wang GZ, Green L, Goff SP. Proviral silencing in embryonic cells is regulated by Yin Yang 1. Cell Rep. 2013;4:50–8.

93. Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, et al. Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep. 2015; 16:836–50.

94. Lewis PW, Elsaesser SJ, Noh K-M, Stadler SC, Allis CD. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A. 2010;107:14075–80.

95. Elsässer SJ, Noh K-M, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature. 2015;522:240–4.

96. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC, et al. Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. Nature. 2010;466:769–73.

97. Schlesinger S, Meshorer E, Goff SP. Asynchronous transcriptional silencing of individual retroviral genomes in embryonic cells. Retrovirology. 2014;11:31.

98. Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, et al. Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet. 2002;71:327–36.

99. Garcia-Perez JL, Marchetto MCN, Muotri AR, Coufal NG, Gage FH, O'Shea KS, et al. LINE-1 retrotransposition in human embryonic stem cells. Hum Mol Genet. 2007;16:1569–77.

100. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev. 2009;23:1303–12.

101. van den Hurk JAJM, Meij IC, Seleme MC, Kano H, Nikopoulos K, Hoefsloot LH, et al. L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet. 2007;16:1587–92.

102. Sciamanna I, Vitullo P, Curatolo A, Spadafora C. A reverse transcriptase-dependent mechanism is essential for murine preimplantation development. Genes. 2011;2:360–73.

103. Lunyak VV, Prefontaine GG, Núñez E, Cramer T, Ju B-G, Ohgi KA, et al. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science. 2007;317:248–51.

104. Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature. 2012;491:454–7.

105. Nellåker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, et al. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol. 2012;13:R45.

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

107. Quinlan AR, Boland MJ, Leibowitz ML, Shumilina S, Pehrson SM, Baldwin KK, et al. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. Cell Stem Cell. 2011;9:366–73.

108. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, et al. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. Cell. 2012;151:483–96.

109. Richardson SR, Morell S, Faulkner GJ. L1 retrotransposons and somatic mosaicism in the brain. Annu Rev Genet. 2014;48:1–27.

110. Han JS, Szak ST, Boeke JD. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature. 2004;429:268–74.

111. Ewing AD, Kazazian HH. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res. 2010;20:1262–70.

112. Wheelan SJ, Aizawa Y, Han JS, Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. Genome Res. 2005;15: 1073–8.

113. Belancio VP, Roy-Engel AM, Deininger P. The impact of multiple splice sites in human L1 elements. Gene. 2008;411:38–45.

114. Perepelitsa-Belancio V, Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. Nat Genet. 2003;35:363–6.

115. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol. 2001;21:1973–85.

116. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature. 1988;332:164–6.

117. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol. 2006;357:1383–93.

118. Singer T, McConnell MJ, Marchetto MCN, Coufal NG, Gage FH. LINE-1 retrotransposons: Mediators of somatic variation in neuronal genomes? Trends Neurosci. 2010;33:345–54.

119. Tonegawa S. Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. Proc Natl Acad Sci U S A. 1976;73:203–7.

120. Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K, et al. L1 retrotransposition in neurons is modulated by MeCP2. Nature. 2010;468: 443–6.

121. Coufal NG, Garcia-Perez JL, Peng GE, Marchetto MCN, Muotri AR, Mu Y, et al. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. Proc Natl Acad Sci U S A. 2011;108:20382–7.

122. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, et al. Increased L1 retrotransposition in the neuronal genome in schizophrenia. Neuron. 2014;81:306–13.

123. Dawkins R, Krebs JR. Arms races between and within species. Proc R Soc Lond B Biol Sci. 1979;205:489–511.

124. Van Valen L. A new evolutionary law. Evol Theory. 1973;1:1–30.

125. Yang L, Guell M, Niu D, George H, Lesha E, Grishin D, et al. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). Science. 2015;350: 1101–4.

126. Chambers I, Smith A. Self-renewal of teratocarcinoma and embryonic stem cells. Oncogene. 2004;23:7150–60.

127. Niwa H. How is pluripotency determined and maintained? Development. 2007;134:635–46.

Gerdes *et al. Genome Biology* (2016) 17:100

Page 17 of 17

128. Silva J, Smith A. Capturing pluripotency. Cell. 2008;132:532–6.
129. Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005;122:947–56.
130. Young RA. Control of the embryonic stem cell state. Cell. 2011;144:940–54.
131. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126:663–76.
132. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–72.
133. Hochedlinger K, Jaenisch R. Induced pluripotency and epigenetic reprogramming. Cold Spring Harb Perspect Biol. 2015;7:a019448.
134. Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23:183–91.
135. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH. Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet. 2011;20:3386–400.
136. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 2012;40:1666–83.
137. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. Mol Cell Biol. 1990;10:6718–29.
138. Belancio VP, Roy-Engel AM, Pochampally RR, Deininger P. Somatic expression of LINE-1 elements in human tissues. Nucleic Acids Res. 2010;38:3909–22.
139. Reichmann J, Reddington JP, Best D, Read D, Ollinger R, Meehan RR, et al. The genome-defence gene Tex19.1 suppresses LINE-1 retrotransposons in the placenta and prevents intra-uterine growth retardation in mice. Hum Mol Genet. 2013;22:1791–806.
140. Macia A, Munoz-Lopez M, Cortes JL, Hastings RK, Morell S, Lucena-Aguilar G, et al. Epigenetic control of retrotransposon expression in human embryonic stem cells. Mol Cell Biol. 2011;31:300–16.
141. Feng Q, Moran JV, Kazazian HH, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 1996;87:905–16.
142. Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science. 1991;254:1808–10.
143. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH, et al. Human L1 retrotransposition: *cis* preference versus *trans* complementation. Mol Cell Biol. 2001;21:1429–39.
144. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993;72:595–605.
145. Crowther PJ, Doherty JP, Linsenmeyer M, Williamson MR, Woodcock DM. Revised genomic consensus for the hypermethylated CpG island region of the human L1 transposon and integration sites of full length L1 elements from recombinant clones made using methylation-tolerant host strains. Nucleic Acids Res. 1991;19:2395–401.
146. Kuwabara T, Hsieh J, Muotri A, Yeo G, Warashina M, Lie DC, et al. Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. Nat Neurosci. 2009;12:1097–105.
147. Ribet D, Harper F, Dewannieux M, Pierron G, Heidmann T. Murine MusD retrotransposon: structure and molecular evolution of an "intracellularized" retrovirus. J Virol. 2007;81:1888–98.
148. Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R. Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci U S A. 2012;109:7385–90.
149. Schwartzberg P, Colicelli J, Goff SP. Construction and analysis of deletion mutations in the pol gene of moloney murine leukemia virus: a new viral function required for productive infection. Cell. 1984;37:1043–52.
150. Stoye JP. Endogenous retroviruses: still active after all these years? Curr Biol. 2001;11:R914–6.
151. Boeke JD, Stoye JP. Retrotransposons, endogenous retroviruses and the evolution of retroviruses. In: Coffin JM, Hughes SH, Varmus H, editors. Retroviruses. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997. p. 343–5.
152. Belshaw R, Dawson ALA, Woolven-Allen J, Redding J, Burt A, Tristem M. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. J Virol. 2005;79:12507–14.
153. Hughes JF, Coffin JM. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. Proc Natl Acad Sci U S A. 2004;101:1668–72.
154. Mamedov I, Lebedev Y, Hunsmann G, Khusnutdinova E, Sverdlov E. A rare event of insertion polymorphism of a HERV-K LTR in the human genome. Genomics. 2004;84:596–9.
155. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. Curr Biol. 2001;11:1531–5.
156. Stoye JP, Coffin JM. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. J Virol. 1987;61:2659–69.
157. Sonigo P, Wain-Hobson S, Bougueleret L, Tiollais P, Jacob F, Brûlet P. Nucleotide sequence and evolution of ETn elements. Proc Natl Acad Sci U S A. 1987;84:3768–71.
158. Romanish MT, Lock WM, Van De Lagemaat LN, Dunn CA, Mager DL. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet. 2007;3:0051–62.
159. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene. 2009;448:105–14.
160. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012;13:283–96.
161. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013;153:1134–48.
162. Jacques P-É, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. 2013;9, e1003504.