

REVIEW

Open Access



Co-option of an endogenous retrovirus (LTR7-HERVH) in early human embryogenesis: becoming useful and going unnoticed

Zsuzsanna Izsvák^{1*}, Jin Ma¹, Manvendra Singh^{1,2} and Laurence D. Hurst³

Abstract

While it is straightforward to understand why most mutations affecting functional sequence are harmful, how genomic changes result in new beneficial traits is harder to understand. Domestication of transposable elements (TEs) is an important source of both new genes and new regulatory systems as, for their own propagation, TEs need to have transcription factor binding sites and functional products that predispose to their recruitment. But are such predispositions to gain-of-function sufficient? Here we consider the case of the endogenous retrovirus, HERVH. Knockdown data supports HERVH having roles in pluripotency, self-renewal and defence against transpositionally-active retroelements in the early human embryo. We clarify the pluripotent cell types associated with HERVH expression and, in the process, note a key unresolved issue, framed by the unwanted transcript hypothesis: how can some cell types have 2% of their transcripts being HERVH-derived but survive the multiplicity of cellular devices that suppress foreign transcripts, be this by transcriptional repression or post-transcriptional filtering? We note a common coupling between novelty generation and suppression evasion. For example, pluripotency-associated KLF4 binding is thought to compete with transcriptional suppressor binding. Similarly, HERVH has a strong splice site enabling efficient novel chimeric transcript formation, the resulting exon–intron junctions enabling evasion of the unwanted transcript filters that recognize low or absent intron presence. We conclude that to better understand domestication, a focus on predispositions to avoidance of unwanted transcripts filters, as well as predispositions to gain of functions, is necessary. The same insights will be valuable for transgene design (eg for gene therapy) and instructive of gain-of-function in tumours, as HERVH is known to be involved in onco-exaptation events.

As organisms are well adapted beings, it is straightforward to see why most mutations affecting functional sequence—be they point mutations, deletions or insertions—would be deleterious and in turn why most

selection is purifying. This is evidenced, for example, by the observation that the rate of non-synonymous evolution in coding sequence is typically much lower than at synonymous sites (i.e. $K_a/K_s < 1$) [1], even in mammals, this notwithstanding the fact that purifying selection is expected to be inefficient when the effective population size is low [2] (e.g. in mammals). That selection would favour some novel functions is also clear, but mechanistically how – mutationally speaking – novel functions are generated is not so transparent. This is somewhat akin to asking how one might take a functioning watch and improve it by tinkering with, or adding, components. Why many random changes would degrade, or indeed, break, the watch is transparent (many might also be

*Correspondence:

Zsuzsanna Izsvák
zizsvak@mdc-berlin.de

¹ The Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Robert-Rössle-Straße 10, Berlin 13125, Germany

² Department of Molecular Biology & Genetics, Cornell University, 526 Campus Road, Ithaca, NY 14853, USA

³ The Milner Centre for Evolution, Department of Life Sciences, University of Bath, Bath BA2 7AY, UK



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

effectively neutral). The nature of the changes that add to the watch's functionality are not so clear.

Here we do not intend to review this literature (for recent considerations see [3, 4]). Rather, we note that transposable elements (TEs) are common contributors to gain of functions. Often this is attributed to the fact that to be a successful genomic invader TEs need to have transcription factor binding sites and functional transcriptional and protein products that predispose to their recruitment via domestication [5]. A paradigmatic example is the case of *Syncytins*: retroviral envelope (*env*) genes, originally responsible for membrane fusion, were independently recruited twice in mammals to facilitate membrane fusion in the cells of the trophoblast during placental development [6]. Interestingly, another *env*-derived gene, *Suppressyn*, encoded by HERVH48, retains the receptor-binding domain of the *env* protein but lacks fusogenic function [6]. Instead, it acts as a negative regulator of syncytia formation mediated by syncytin-1 [7], and has been proposed to contribute to antiviral defense mechanisms [8]. We can also point to LTRs of endogenous retroviruses as having multiple transcription factor (TF) binding sites, a prerequisite for being a successful invader but also predisposing to gain, or transfer, of functionality [5, 9, 10].

While such functional predispositions are likely to be important, are they sufficient to understand how functions are gained? Here we consider the case of endogenous retroviruses (ERVs), in particular HERVH this being recruited to pluripotency, self-renewal and anti-mobile element defence in the early human embryo. We argue, in the context of the unwanted transcript hypothesis [11] that predisposition to functionality alone is not sufficient. Rather, there exists a multiplicity of cellular devices to suppress, at all stages of gene expression, the expression of unwanted transcripts, transcripts that serve no utility for the host be these actively parasitic, remnants of parasitic elements, spurious transcripts, mis-spliced forms etc. Indeed, we suggest that the fact that as many as 2% of all transcripts in some early embryo cells are HERVH-derived is the most remarkable of observations: very recently, in evolutionary time, similar HERVH transcripts should have been recognised as foreign and suppressed. How then can they be so abundant, and functional and survive anti-foreign transcript systems? Evasion of the multiple devices to counter unwanted transcripts is, we suggest, key to TE domestication. Just as certain TEs may have predispositions to enable functionality for the host, so any predisposition to escape suppression will make those sequences more likely to be domesticated, although mutations gained later on, possibly as an adaptive response to suppression, may also be important. While, in retrospect, this is possibly an obvious point, more

striking is the sometimes intimate coupling between novelty generation and escape from suppression, as we here discuss.

We start by laying out the biology of ERVs, consider LTR7/HERVH as an example of co-option and outline the evidence for its functionality. We take this opportunity to clarify the HERVH-related activity in different classes of pluripotent cells. We then highlight unknowns, a key one of which is how HERVH transcripts can be so very common in some of these cell types. Examining in detail how suppression is avoided we note that HERVH has some unusual features that simultaneously promote novelty and should act to avoid the filters that evolved to prevent expression. Understanding how some transcripts avoid suppressive filters is, we suggest, not only relevant to the broad question of how over the short evolutionary time novelty evolves. These lessons may well help to understand why some transgenes come to nothing, while others are robustly expressed. This should be of utility for designing transgenes for stable integration, eg for gene therapy. In addition, understanding why HERVH is involved in gain-of-functions on a tumour, rather than evolutionary, timeline (eg onco-exaptations) may well cast light on the same issues.

Endogenous retrovirus as cellular invaders

Endogenous Retroviruses (ERVs) are ancient retroviral integrations that have become part of the host genome. Initially, ERVs retain the ability to exit the host cell, but over time, they lose this capacity and begin to follow the evolutionary life cycle of transposable elements (TEs) [12]. The endogenization process requires integration into the germline, but similar to other TEs, ERVs can also transpose into other lineages [12–14]. This dynamic process allows active TEs to move in and out of the germline, potentially influenced by environmental conditions [15, 16]. In mammals there is a specific connection between the germline and pluripotency as development involves the transition from pluripotency to epiblast after implantation, from which the germline develops so that primordial germ cells (PGCs) regain pluripotency [17]. Indeed, PGCs express transcription factors that are also expressed in pluripotent stem cells (PSCs) and characterise the pluripotent state [18]. This provides important insights on how TEs are co-opted for the fitness of germ cells so that they are transmitted to the next generations and ultimately fixed in populations.

While TEs are active, they can be under selection to have a reduced degree of harm and so self-suppress to some degree [19, 20]. Nonetheless, any activity is likely to incur costs to the hosts and hence there is a continuous "arms race" between them and their host. Throughout evolution, hosts have developed a variety of mechanisms

to suppress active TEs, while TEs have concurrently evolved strategies to evade these defenses (reviewed in refs [15, 21]). Despite these ongoing interactions, the typical evolutionary outcome for TE/ERVs is eventual inactivation. Ectopic recombination is one of the inactivation processes that frequently affect ERVs due to sequence identity in their long terminal repeats (LTRs), leading to deletion of most of the ERV sequence and leaving only ‘solo’ LTRs [22, 23]. After inactivation, accumulation by drift of mutations renders the sequences unrecognisable as TE-derived over time. (H)ERVs make up about 8% of the human genome [24, 25]. The annotated HERVs are categorised into class I (gamma-like), class ii (beta-like)

and class III (spuma-like) (Table 1). Among the oldest HERVs, the HERVL elements are estimated to have integrated into the genome over 100 million years ago (MYA), predating the divergence of placental mammals [26], whereas HERVK are the most recent [27] (Fig. 1), still capable of producing viral particles [28].

However, not all TEs become mere genomic fossils; some undergo a rare process of exaptation (alias co-option) (Table 2). This process can potentially begin while the TE or ERV is still active, providing certain advantages to the host, which may lead to its tolerance. Due to their functional and structural similarities, interactions between exogenous viruses and

Table 1 Copy numbers of different HERVs in the human genome, flanked by LTRs or present as *solo* elements. The table was adapted from ref [29]. Repbase name of the long terminal repeat (LTR) follows in parentheses. Most copy numbers are from ref [30]. Additional information was reported on HERVH [31–34] and HERVK [35]. Note that the ratio of copies flanked by LTRs and *solo* elements are the highest in HERVH (in bold)

Class	HERV family	Rebase	Copies	Solo	Ratio Copies/S olo
Class I	ERV-3	HERV3(LTR4)	100	125	0,8
	ERV-9	HERV9 (LTR12,PTR5)	300	5000	0,06
	ERV-ADP	HERVP-71A I (LTR71A)	40	300	0,13
	ERV-E	ERVE (LTR2,2B,2C)	250	1000	0,25
	ERV-F	HERVFH19I (LTR19)	45	550	0,08
	ERV-Fb	HERVH48I (MER48)	60	100	0,06
	ERV-Fc	HERV46I (LTR46)	2	4	0,5
	ERV-FRD	MER50I (MER50)	50	2000	0,025
	HERV-H	HERVH (LTR7)	1000–1200	1000, 1270	0,8–1
	ERV-HS49C23	MER57I (MER57)	200	1000	0,2
	ERV-I	HERVI (LTR10)	250	1000	0,25
	RRHERV-I	HERV15I (LTR15)	40	250	0,16
	ERV-P	HUERS-P3 (LTR9)	200	1000	0,2
	ERV-Rb	PABL BI (PABL A/B)	8	1000	0,008
	ERV-T	HERVS71 (LTR6A,B)	80	400	0,2
	ERV-W	HERV17 (LTR17)	40	1100	0,36
	ERV-FXA	HERVFH21 (LTR21A)	30	40	0,75
Class II	ERV-K(HML-1)	HERVK14I (LTR14A,B)	68	350	0,19
	ERV-K(HML-2)	HERVK (LTR5)	60, 91	2500, 944	0,024–0,097
	ERV-K(HML-3)	HERVK9I (MER9)	150	700	0,214
	ERV-K(HML-4)	HERVK13I (LTR13)	10	800	0,0125
	ERV-K(HML-5)	HERVK22I (LTR22,-A,-B)	100	600	0,17
	ERV-K(HML-6)	HERVK3I (LTR3, 3B)	50	400	0,125
	ERV-K(HML-7)	HERVK11D1 (MER11D)	20	140	0,14
	ERV-K(HML-8)	HERVK11I (MER11A,B,C)	60	600	0,1
	ERV-K(HML-9)		10	40	0,25
	ERV-K(C4)	HERVKC4 (LTR14)	10	100	0,1
	ERV-K(14C)	HERVK14CI (LTR14C)	15	120	0,125
Class III	ERV16	ERV16 (LTR16, 16A-D)	15	25	0,6
	ERV-L	HERVL (MLT2A1–2B2)	580	6000	0,97
	ERV-S	ERV18 (LTR18, 18B)	50	150	0,33

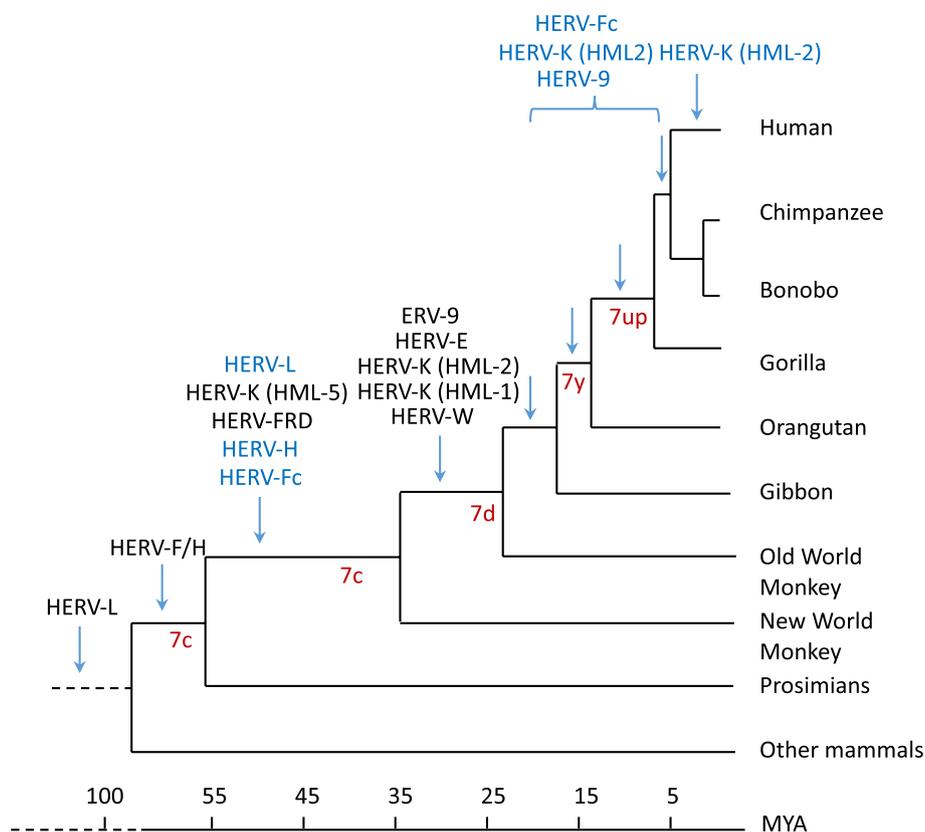


Fig. 1 Invasion of the ancestor of various HERV families during primate evolution. Note that HERVs (in blue) invaded multiple times (adopted from ref [29]). Evolution of HERVH subfamilies (in dark red) during primate evolution adopted from ref [31]

Table 2 The best-characterized co-opted HERVs in the human genome [42]

HERV family	Co-opted function	Role in humans	HERV-derived product
HERV-W (ERVW-1)	Placental trophoblast fusion	Promotes syncytiotrophoblast formation for fetal/maternal exchange	<i>env</i> -derived protein Syncytin-1
HERV-FRD	Immunosuppression in pregnancy	Helps to prevent the maternal immune system from attacking the semi-allogeneic fetus	<i>env</i> -derived protein Syncytin-2
HERVH48	Antiviral defence	Negative regulator of trophoblast differentiation	<i>env</i> -derived protein Suppressyn
HERVL	Marking 8-cells stage ^a	By analogy to MERVL (mice), it might regulate embryonic gene activation (EGA) in early embryogenesis	HERVL-driven gene regulation chimeric transcripts ^a
HERVK (HML-2)	Antiviral defence Regulatory activity in 8/16 cell to early blastocyst stage stage ^a	Protection against de novo retroviral infection Unclear	<i>env</i> -derived protein (<i>Rec</i>) Potential enhancer HERVK-driven transcripts ^a
LTR7-HERVH	Antiviral/TE defence Regulatory activity in the human inner cell mass (ICM- EPI) stages	Self-renewal and pluripotency of stem cells	<i>gag</i> -derived sequences HERVH-derived transcripts lncRNA Chimeric transcripts TADs miRNA sponge

^a needs further clarification

specific ERVs have been observed. For example, HIV-1 or Epstein-Barr virus (EBV) infection can lead to the upregulation of certain ERV elements, which may affect the host's immune response [36–38], but can also interfere with the production of HIV-1 viral particles [39]. Additionally, inactive ERV-derived sequences, such as certain open reading frames (ORFs) (eg *env*, *gag*), transcription regulatory sequences, and transcription factor (TF) binding sites, can be repurposed to support host functioning [23, 40, 41].

LTR7-HERVH as an example of ERV co-option

RTVL- retrovirus, the ancestor of HERVH invaded the ancestor of the primate genome around 40 MYA [25, 43]. The invasion occurred multiple times (Fig. 1), and during the endogenization process, which began around 35 MYA [43], the number of HERVH elements expanded through both reinfection and retrotransposition within the genome. Recombination occurred between the different subfamilies [31]. By analysing genomic structures of HERVH elements, it was observed that in contrast to conventional ERVs, the HERVH elements are more frequently flanked by two LTRs than expected from their evolutionary age [43, 44]. The ratio of close-to full length versus *solo* copies is exceptionally (1:1) high, still *solo* LTRs still constitute a significant proportion of HERVH elements in the human genome (Table 1). Despite the fact that HERVH is no longer transposing, approximately 1000–1200 HERVH copies are flanked by LTRs [31–34, 45]. This observation led to the hypothesis that these (or at least some of these) HERVH copies likely have co-opted function(s).

The evolution of the LTR7 sequence (Fig. 1) was shaped by a combination of mutational processes, including point mutations, duplications, and multiple recombination events between subfamilies, which resulted in transcription factor binding motif modules unique to each subfamily [31]. Currently, four subfamilies of HERVH elements are listed in the Dfam [46] and Repbase [47, 48] databases, and they are annotated in the reference human genome based on distinct LTR consensus sequences: LTR7 (formerly Type I), 7b (Type II), 7c, and 7y (Type Ia) [47, 49]. Additional subdivisions of HERVH elements have been proposed based on phylogenetic analyses and structural variations in their internal gene sequences [23, 32, 50] as well as their LTRs [31]. The different subfamilies are expressed in well-defined developmental niches [31, 51]. Curiously, the subfamilies (eg LTR7up1/2) expressed in pluripotent stem cells exhibit a robust increase in transcription [31].

If they are domesticated, what might their functions be? One possible clue comes from their time of activity. Currently, around 300 genomic loci of full-length

LTR7-HERVH are expressed in human pluripotent stem cells. One of the most highly transcribed loci is the HERVH-derived novel gene, ESRG [52].

Over the past decade, research has gradually uncovered several co-opted functions of HERVH (Table 2), demonstrating its multifaceted contributions to genomic regulation and cellular processes (reviewed in ref [53]). The investigation into HERVH's co-opted functions began with the observation that HERVH has binding sites for multiple key pluripotency factors (POU5F1/OCT4, NANOG, TFCP2L1/LBP9, SOX2, and KLF4) [51, 52, 54, 55]. That this binding is functionally relevant is evidenced by the observation that depletion of HERVH RNA in human pluripotent stem cells (hPSCs) leads to a loss of self-renewal and upregulation of differentiation markers [52, 55, 56]. Since then, HERVH expression has been recognized as a marker for pluripotent stem cells, as its activation promotes both the acquisition and maintenance of pluripotent states [57–60]. Notably, LTR7-HERVH-derived transcripts are detected in the human embryo, and embryos are not viable in its absence [61], suggesting their co-option in early human embryogenesis.

The modes of activity of HERVH appears to be highly diverse, and not as simple as old TE creates a *de novo* protein (i.e. not like syncytins). HERVH loci may function as robust alternative promoters and enhancers [9, 51, 62, 63], with several copies overlapping with annotated super-enhancers [64]. Activated HERVH loci frequently generate long non-coding RNAs (lncRNAs) or produce chimeric transcripts with neighbouring protein-coding genes [52, 65]. These transcripts, which contain HERVH-derived and unique chimeric sequences, play specific roles in hPSCs, such as ESRG [52], lncROR [66], and Lnc00458 [67]. There are approximately 50 HERVH genomic loci that form topologically associating domain (TAD) boundaries in hPSCs, especially the highly transcribed ones [68], which actively shape cell type-specific chromatin architecture. Highly transcribed HERVH genomic loci can also establish topologically associating domain (TAD) boundaries in hPSCs [68], actively shaping cell type-specific chromatin architecture. Additionally, HERVH-derived lncRNAs contribute to chromatin remodeling and gene activation during differentiation processes [67, 69]. The relative functional relevance of each of these activities is unknown.

Beyond supporting self-renewal and pluripotency, LTR7-HERVH products are also involved in host defence, such as conserved *gag* sequences that may act as antagonists against viral reinfection [23] or by suppressing the activity of other TEs and ERVs [70]. Consistent with such a role, there is antagonistic expression pattern between potentially mutagenic retrotransposable elements (REs) and LTR7-HERVH around the

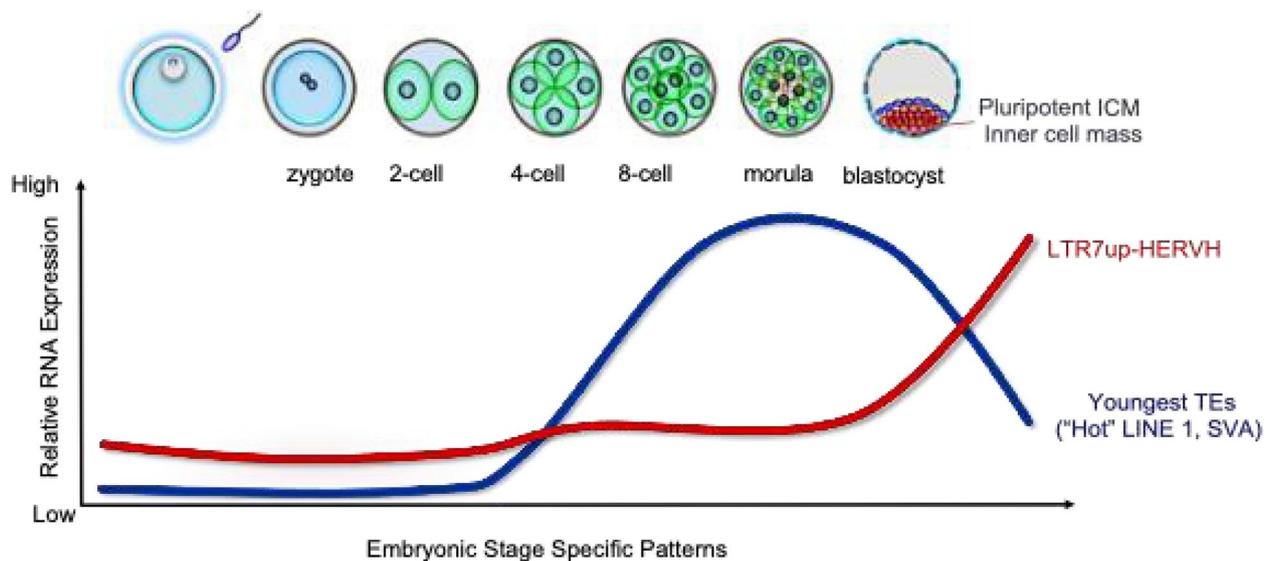


Fig. 2 Antagonistic expression pattern of phylogenetically Young (< 7MY old, eg “Hot” Line 1, SVA) and Old (> 7 MY old, eg HERVH) elements during human early embryonic development (data source: GEO-GSE 36552)

fifth day of embryonic development (after the morula stage) (Fig. 2). HERVH knockdown is indeed associated with depression of younger REs [70]. Mechanistically there may be several modalities to any suppression. HERVH acts as a cis-regulatory element upstream of APOBEC3G as an enhancer for several APOBEC3 genes [63]. The APOBEC3 genes (apolipoprotein-B mRNA-editing catalytic polypeptide-like-3) are an essential component of the innate immune response in humans and play a crucial role in limiting the activity of retroviruses (eg HIV-1) and the autonomous retroelement LINE-1 [71–73]. APOBEC3G particularly targets single-stranded DNA and induces G-to-A hypermutations in retroviral DNA. The HERVH-APOBEC3 locus may also be part of a broader HERVH-based defense mechanism against potentially harmful genomic invaders in early embryogenesis. Analysis of alternative modalities of young RE suppression by HERVH is an active area of research.

There may be other utilities of HERVH. In addition to generating lncRNAs and chimeric transcripts, full-length HERVH copies occasionally encode open reading frames (ORFs), such as *gag* [23]. However, unlike the more evolutionarily recent HERVK, which is shown to produce viral particles [28], HERVH does not seem to do so. Instead, the preservation of the domain containing zinc finger motifs within the HERVH *gag* ORF suggests a co-opted function [23]. It has been proposed that the original function of these zinc finger motifs—enabling retroviral proteins to package RNA for the benefit of an exogenous retrovirus—has been repurposed to bind viral

RNA, potentially providing an immunological benefit to human cells [23]. This repurposing of viral genes to function as host immunity genes is commonly referred to as endogenous viral element-derived immunity (EDI) [74].

Notably, HERVH elements flanked by LTRs were identified, rather than *solo* LTR sequences in co-opted roles, although the *solo* LTRs may still act as enhancers and contribute to tissue-specific or developmental gene expression patterns, which needs to be validated.

HERVH expression in human pluripotent *in vitro* stem cell cultures

While HERVH has been proposed as a precise marker of human pluripotency [57], it is worthwhile expanding on this as there are at least three distinct pluripotent stem cell types that exhibit different but characteristic expression profiles of HERVH and other transposable elements (TEs).

Primed pluripotent stem cells

The primed state of human ESC pluripotency resembles that of mouse epiblast stem cells (EpiSCs) more closely than that of mouse ESCs [75]. In primed pluripotent stem cells, HERVH genomic loci are actively transcribed ($n = 250\text{--}300$) (Table 3), and these loci largely overlap with those expressed in various induced pluripotent stem cell lines (hiPSCs) [52].

Table 3 Features of three human pluripotent cell types. Characteristically expressed TE families are as in [52, 76]

Feature	Primed	Naïve	eFORM
Developmental stage	Post-implantation epiblast	Pre-implantation epiblast/ late morula-early blastocyst	Pre-implantation epiblast/ blastocyst
Pluripotency	Primed for differentiation	Ground state	Ground state
Ability to revert to primed pluripotency	N/A	Low	High
Morphology	Flattened colonies, 2D	Round colonies, 3D	Round colonies, 3D
Epigenetics	Genomic hypermethylation, Xa/Xi inactivation	Genomic hypomethylation, Xa/Xa	Genomic hypomethylation, Xa/Xa
Proliferation rate	24–36 h	18–24 h	42–48 h
Characteristically expressed TE families	L1_Hs L1_PA2	L1_Hs SVA_D LTR5-HERVK LTR7Y_B-HERVH	LTR7-HERVH
Expressed HERVH loci	250–300	50–80	350–400
Genome stability	Stable	Less stable, frequent chromosomal abnormalities	Stable

Naïve pluripotent stem cells

When primed hESCs/hiPSCs are cultured in naïve media [76, 77], the cells transition to a so-called naïve stage. Although mouse and human naïve pluripotent stem cells share a similar morphology (eg, form dome-shaped colonies) and certain well-defined characteristics, the naïve stage is not identical in mice and humans. In mice, naïve cells model the inner cell mass (ICM) of the blastocyst, while in humans, they are thought to represent a developmental stage prior to ICM formation, closely resembling the morula/early blastocyst stage [76]. In terms of TE profiling, human naïve cells exhibit relatively low expression of HERVH (~50–80 genomic loci) [76] (Table 3). Conversely, these cells display high expression of phylogenetically younger TEs (<7MY), such as active L1_Hs and SVA [76]. It is noteworthy that younger L1 elements are partly responsible for determining the pluripotent stage of the mouse [78], whereas the human pluripotent stage has an opposite dynamic, as it has apparently developed through the suppression of younger elements.

eFORM (Naïve-like) human pluripotent stem cells

This cell type, which constitutes approximately 4% of human pluripotent stem cell cultures (hESC/hiPSCs), is morphologically similar to human naïve pluripotent stem cells [52]. However, unlike naïve cells, this "naïve-like" cell type is characterized by high and stable expression of LTR7-HERVH (HERVH^{High}) derived from around 350–400 genomic loci (Table 3) and negligible expression of younger REs (eg. LINE-1, SVA, HERVK) [63]. HERVH expression also provides telomere

protection [52]. Given that LINE-1 expressing cells segregate from the forming embryo while LTR7-HERVH-derived transcripts contribute to human embryogenesis [61], we have named this HERVH^{High} cell type as naïve-like embryo-forming (eFORM).

HERVH^{High} eFORM cells retain their ability to self-renew and can be maintained long-term [59]. However, the elevated levels of HERVH inhibit differentiation [58, 79]. For a successful transition out of the self-renewing pluripotent state, HERVH expression must be reduced. This downregulation is partially mediated by the BTB domain-containing zinc finger protein ZBTB12 [80] or by TUT7 [81].

HERVH transcript abundance and other enigmas in early development

Despite significant progress in understanding the impact of LTR7-HERVH on human physiology, some unresolved issues remain. One concerns the different functionalities – if any – of the different LTRs. Typically activated from the eight-cell stage, HERVH can be driven by variants of LTR7 (eg B/C/Y), each providing transcription at slightly different pre-implantation developmental stages [31, 51]. These different LTR variants are thought to represent an ongoing arms race [82] between host defense mechanisms and HERVH. It is noteworthy that in addition to LTR7, the subfamily LTR7Y also has a 1:1 ratio between LTR-flanked and *solo* elements, whereas the *solo* LTRs in the subfamilies LTR7B and LTR7C exceed the number of nearly proviral copies [31]. Nevertheless, the question remains as to which subfamily, such as LTR7, was used for functional tasks or continues to be captured by host defence mechanisms.

Beyond LTR7B/Y, HERVK, and SVA (SINE-VNTR-Alu) elements are abundantly expressed from the eight-cell stage to the early blastocyst [76] (Fig. 2), suggesting potential functional co-option in early development (Table 2). Indeed, it is tempting to speculate that HERVK derived protein product (eg *Rec*) may serve a protective role against viral infection [28]. However, it remains puzzling that HERVK and SVA are co-expressed with LINE-1, a mutagenic retrotransposon, despite the fact that highly damaged cells are eliminated from the developmental program [63]. This raises the question: if LINE-1 compromises genome integrity, why are HERVK and SVA tolerated during the same developmental window? Given that host defense mechanisms are not yet fully functional at this stage, it remains an open question whether HERVK and SVA elements are indispensable for early development or if their expression is tolerated due to a permissive chromatin state.

An intriguing possibility is that HERVH serves as a regulatory hub, counteracting SVA/HERVK/LINE-1 expression to modulate transposon activity. This implies that HERVH may regulate pluripotency while simultaneously suppressing other transposable elements, acting as an early defense mechanism. Its inverse correlation with SVA/HERVK expression (Fig. 2) further supports this hypothesis, although this concept remains to be experimentally validated.

Despite multiple reports highlighting the significance of HERVH in pluripotent stem cells, a major concern remains the conflicting evidence regarding the role of the HERVH-derived ESRG locus in maintaining pluripotency. ESRG (Embryonic Stem Cell Related), is among the most abundantly transcribed LTR7-HERVH locus [52]. It contains at least one putative open reading frame (ORF) (Q1W209), but it remains unclear whether it is functional. Notably, the ORF is predicted to encode a 222-amino-acid protein unique to humans [52], though this prediction is primarily supported by transcription data. Furthermore, Wang et al. demonstrated that knocking down ESRG expression is leading to compromised self-renewal and pluripotency in hESCs [52]. However, this finding was challenged by Takahashi et al., who found ESRG to be dispensable in hPSC lines where it was specifically knocked out (KO) [83]. Li et al. [84] later provided a potential resolution to these conflicting results by suggesting that the temporal knockdown of TP53 during the construction of ESRG knockout cell lines [83] might have contributed to the discrepancies. The cumulative effect of this knockdown, particularly in a TP53-deficient background, could account for the divergent outcomes observed across studies. Additionally, knockout experiments using the CRISPR/Cas9 system may introduce off-target effects. Moreover, while the knockdown affects

multiple HERVH copies, the knockout approach targets only a single genomic locus. If HERVH transcripts function within a network, maintaining a specific threshold of HERVH expression may be crucial for their biological role.

Following on from the high abundance of ESRG transcripts, more generally the relative abundance of HERVH transcripts in self-renewing pluripotent stem cells that will contribute to the embryo is noteworthy and enigmatic. Remarkably, the expression of HERVH-associated lncRNAs is highly efficient, with levels up to eightfold higher compared to non-HERVH-associated lncRNAs in hPSCs [54]. Intriguingly, HERVH products account for approximately 2% of all transcripts in the nuclei of these self-renewing pluripotent stem cells [52, 55, 57]. What remains unclear is first, why are they so very common and second, why HERVH-derived transcripts, unlike younger REs, are not recognized as foreign by the host defense system during early embryogenesis. In the very recent evolutionary past, ERVH would have been recognised as a dangerous endogenous retrovirus. Nonetheless, despite their high expression levels, these transcripts do not trigger an immune response, which suggests that at least certain HERVH copies may have evolved mechanisms to evade detection, or that the early embryonic environment has unique features that allow tolerance of these elements. Understanding this selective recognition—or lack thereof—by the host defense system could provide insights into the regulation of genomic elements during development.

Distinguishing self from non-self, and the “Unwanted Transcript Hypothesis”

The problem of recognition of transcripts derived from HERVH sits within a broader context of control and filters on transcriptions of all types. Current estimates suggest that over 80% of our genome is transcribed [85, 86]. Given that approximately half of the human non-coding genome originates from invasive elements, such as previous viral infections and TE activities, this results in a significant amount of transcriptional junk. The challenge for organisms with large amounts of invasive elements is managing the vast number of transcripts produced by the non-coding genome. According to the ‘unwanted transcript hypothesis’ [11], cells have evolved sophisticated mechanisms to suppress ‘unwanted’ transcripts or, if produced, filter them out and silence ‘non-self’ elements while maintaining the expression of essential host genes. Transcriptional suppression of transcripts recognized as foreign by the HUSH complex, nuclear degradation by the nuclear exosome complex and control of nuclear export are three of the suggested mechanisms [11]. Beyond transcriptional control, transcripts derived

Table 4 Host-encoded suppression mechanisms and their effects on TEs [42] [87–93, 95–106, 109], (r) review article. The symbols indicate references to topics listed in the “Key Players” column

Mechanism	Key Players	Mechanism of Action	Targeted TEs	Examples
Epigenetic Silencing	DNA Methylation* Histone Modification# (H3K9me3, H3K27me3, H4K20me3), HUSH® complex, SALL4 [§] , CAF-1^	Compact chromatin to repress TE transcription	HERVs, LINE-1, SINEs, Satellite DNA	* [42] (r) # [87, 106] (r); [103] @ [88] [109](r) & not yet reported ^ [89]
Post-transcriptional silencing (RNAi)	piRNAs*, siRNAs#, miRNAs®, Argonaute, Dicer	Small RNAs guide silencing complexes to degrade TE transcripts and recruit repressive chromatin marks	ERVs, LINE-1, Alu	[105] (r) # [90] @ [91]
RNA Surveillance	RNA Exosome* NEXT#, PAXT® Nonsense-Mediated Decay (NMD) [§] StaufenMediated Decay (SMD)^	Decay pathways recognize and degrade TE-derived RNAs	TE-derived transcripts, long exon- containing RNA, (> 1 kb) intronretained RNAs, LINE-1, HERVK	* [92] # [93, 94] @ not yet reported & [104] (r) ^ [95]
Nuclear export	TREX/NXT* CRM1! (Exportin-1)#,	Preventing cytoplasmic transfer and translation of TEs	HERVK (<i>Rec</i>), LINE-1,	* [96] # [97]
Sequestration	SAMHD1* NEAT1#	cytoplasmic stress granules nuclear paraspeckles	LINE-1 Alu	* [98] # [99]
Innate Immune Response	RIG-I*, MDA5#, ADAR [§]	Detect TE-derived dsRNA or cDNA, triggering an antiviral-like response	Cytosolic LINE-1 DNA, dsRNA,	* [100] # [101] & [102]

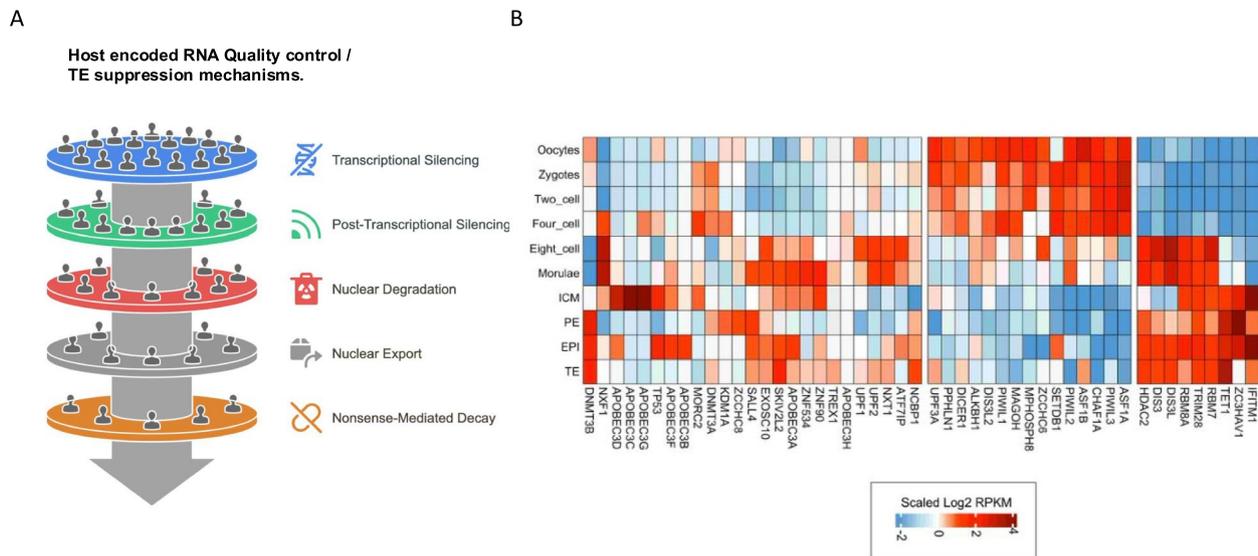


Fig. 3 Host-encoded factors involved in RNA quality control (QC) and TE suppression mechanisms in early human embryonic development. **A** The schematics illustrates the QC-based filtering mechanism of unwanted (TE-derived) transcripts [11]. Transcriptional silencing (i) HUSH: Targets TEs, silencing through H3K9me3 deposition; (ii) SALL4: contributes to H3K27me3 deposition; (iii) CHAF1A: helps to promote the formation of a H3K9me3-marked heterochromatin; TP53: repression of TEs that contain P53 response elements; Post-transcriptional silencing PIWI-piRNA pathway; Nuclear degradation; Exosome targeting (NEXT/PAXT); Nuclear export, TREX/NXT; Nonsense mediated decay (NMD); slow translation; APOBEC3, promotes deaminase induced mutations; Innate immunity, IFITM; **B** Heatmap showing the differential expression of host-encoded factors involved in RNA quality control and TE suppression mechanisms in early human embryonic development (data source from [112])

from TEs are indeed filtered out by several specialized mechanisms [87–103] and reviewed in refs [42, 104–109] (Table 4). These protective mechanisms are specialized,

targeting different types of TEs/ERVs at various stages and levels, often complementing each other (for a review, see [42, 110, 111]).

In humans, transpositionally competent REs, such as LINE-1 and SVA get activated after embryonic gene activation (EGA). The developmental window that enables the activation of mutagenic REs (Fig. 2) suggests that the host's filtering and defence mechanisms are not yet fully functional (Fig. 3). We have provided evidence that these transposition-competent REs can indeed retrotranspose, leading to the elimination of severely damaged cells from the developmental program [63]. This previously unnoticed cell type, termed "REject," lacks commitment markers, expresses DNA damage response genes, is excluded from the developmental process, and eventually undergoes apoptosis [63].

Interestingly, the transcriptional decline of the mutagenic REs around day six of embryonic development is followed by the emergence of transcripts from hundreds of HERVH loci driven by LTR7, peaking at the 6-7th day of blastocyst stage (Fig. 2), such that LTR7(up)-HERVH activation appears to counteract the transcriptional upregulation of REs. In contrast to REject cells, which are eliminated from the forming embryo, HERVH-derived transcripts have been shown to contribute to the formation of the inner cell mass (ICM) of the human embryo [61, 63]. Since certain host defense mechanisms become active in the pluripotent epiblast (Fig. 3), it raises the question of why LTR7-HERVH products are not recognized as foreign and can bypass the host's defense system undetected. For example, interferon-induced transmembrane protein 1 (IFITM1), which has been reported to suppress the expression of certain HERVs (eg HERVK, LTR7Y-HERVH and LTR12D-1), is abundantly expressed in pluripotent stem cells (Fig. 3). However, it may selectively target LTR7Y-HERVH rather than LTR7-HERVH [113]. Even more convincing is the fact that LTR7up-HERVH together with IFITM1 increases the expression of APOBEC3G, APOBEC3C and [32], enzymes known to catalyse the conversion of cytosine to uracil in single-stranded DNA intermediates of various viruses and retroelements [114] (Fig. 3). This suggests that LTR7-HERVH may even play a role in the regulation of other REs.

Becoming useful and going unnoticed

The notion that cells have elaborate systems of transcriptional suppression and of transcript filtering adds to the complexity of the domestication process. It isn't good enough to observe that an HERVHs LTR has transcription factor binding sites, which certainly may predispose to domestication, but we need also to consider how a domesticated transcript escapes all the snares put in place. Here, then we discuss selected examples of HERVH co-option, highlighting how

HERVH has evaded host defense mechanisms, been tolerated, and subsequently co-opted to support pluripotent stem cells in humans. The collected evidence suggests that HERVH has escaped suppression on multiple levels and even that the co-option and avoidance of suppression process may be mechanistically tightly coupled.

The pioneer transcription factor KLF4 may prevent TRIM28/KAP1 binding and recruit additional pluripotency factors

In pluripotent stem cells (PSCs), many ERV/TEs are silenced through tri-methylation of histone H3 at lysine 9 (H3K9me3), a process facilitated by the enzyme SetDB1 (SET domain bifurcated histone lysine methyltransferase 1). SetDB1 is recruited to ERV/TEs by sequence-specific DNA-binding proteins known as Krab zinc-finger proteins (KZFPs). These KZFPs interact with Trim28 (tripartite motif-containing 28, also known as KRAB-associated protein 1 or Kap1), along with other associated factors, to contribute to heterochromatin formation and DNA methylation. The Trim28 suppression machinery is particularly well-characterized in mice, where it plays a major role in regulating TE expression in mouse embryos [115–117].

In humans, TRIM28 is expressed as early as the 8-cell stage (Figs. 3 and 4A). In human pluripotent stem cells (hPSCs), TRIM28-mediated repression has been reported to broadly regulate a wide range of retroelements (REs) [118]. However, only a small subset of HERVH sequences (~0.06%) are directly bound by TRIM28 [76]. Notably, in TRIM28 knockout hPSCs [76, 119], the differentially expressed HERVH loci are predominantly associated with various LTR7 family members (eg LTR7B, LTR7C, LTR7Y) (Table 5 and Table S1). Furthermore, ESRG expression remains unaffected in TRIM28 knockout ESCs, and only three of the differentially expressed loci rank among the top 100 expressed HERVH loci (Table 5 and Table S1).

Beyond regulating TEs, TRIM28 may also play additional roles, such as controlling transcriptional elongation [121], which could explain its crucial function in ESCs. Notably, TRIM28 null mutations impair hPGCLC differentiation, indicating that its silencing in human PSCs compromises germline competence [119]. This underscores the essential role of TRIM28-mediated RE suppression in safeguarding the human germline.

Additionally, in our analysis of human embryonic stem cells depleted of TRIM28 using shRNA (shTRIM28-hESC_H1) [122], we observed no significant change in the number of expressed LTR7-HERVH genomic loci. While silencing of TRIM28 in human PSCs has no effect on the expression of HERVH, it leads to upregulation of other REs, including LTR5-HERVK (Fig. 4B) and SVA

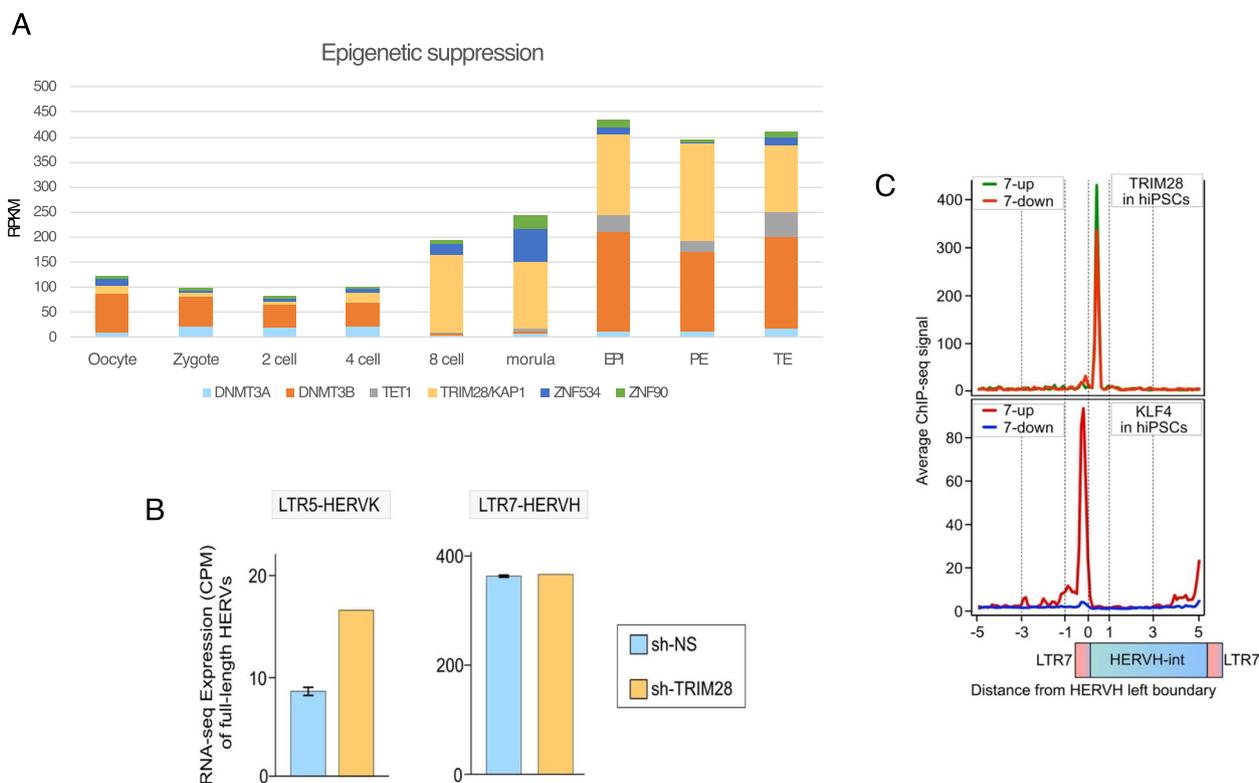


Fig. 4 TRIM28 is not a significant regulator of LTR7/HERVH in human embryonic stem cells. **A** Expression of selected proteins involved in the epigenetic regulation REs, including the TRIM28 suppression system, the LTR7 specific ZNF90 and the LTR7up specific ZNF534 [31, 120] during human early embryonic development (data source from [112]). **B** LTR7-HERVH remains active under normal conditions, whereas LTR5-HERVK is derepressed upon TRIM28/KAP1 depletion (data source from [118]). **C** TRIM28 and KLF4 ChIP-seq counts over LTR7/HERVH genomic loci in hiPSC (data source from [58, 118])

Table 5 Differentially expressed (DE) genomic loci of the LTR7 subfamily members in TRIM28 knockout (T28KO) hESCs [76]. Number of loci expressed in the 100 highest/lowest category. Number of LTR7s expressed in hESCs: 879 [119]

LTR7 subfamily	DE in T28KO	DE in highest 100	DE in lowest 100
LTR7			
Up	8	1	1
Down	1		
LTR7B		1	
Up	27		4
Down	0		
LTR7C			
Up	6		1
Down	0		
LTR7Y			
Up	2		
Down	1	1	
Sum	45	3	6

elements [118]. Together, these findings suggest that, in human PSCs, LTR7-HERVH largely escapes suppression by the TRIM28/KAP1 complex, unlike LTR5-HERVK.

Interestingly, de-repression of Young REs has little immediate effect and is even compatible with short-term self-renewal [80, 119]. In contrast, the biological response in mouse pluripotent cells is more pronounced, as a Trim28 null mutation in mouse cells is incompatible with embryonic stem cell (ESC) self-renewal [123], indicating species-specific differences in TRIM28 function during development. Beyond regulating REs, TRIM28 may also play additional roles, such as controlling transcriptional elongation [121], which could explain its crucial function in ESCs. It is notable that TRIM28 null mutations are detrimental to hPGCLC differentiation, suggesting that silencing TRIM28 in human PSCs impairs germline competence [119]. This finding highlights the critical role of TRIM28-mediated RE suppression in protecting the human germline.

Why, then, does TRIM28 repress many retroelements (REs) but largely not HERVH in pluripotent stem cells?

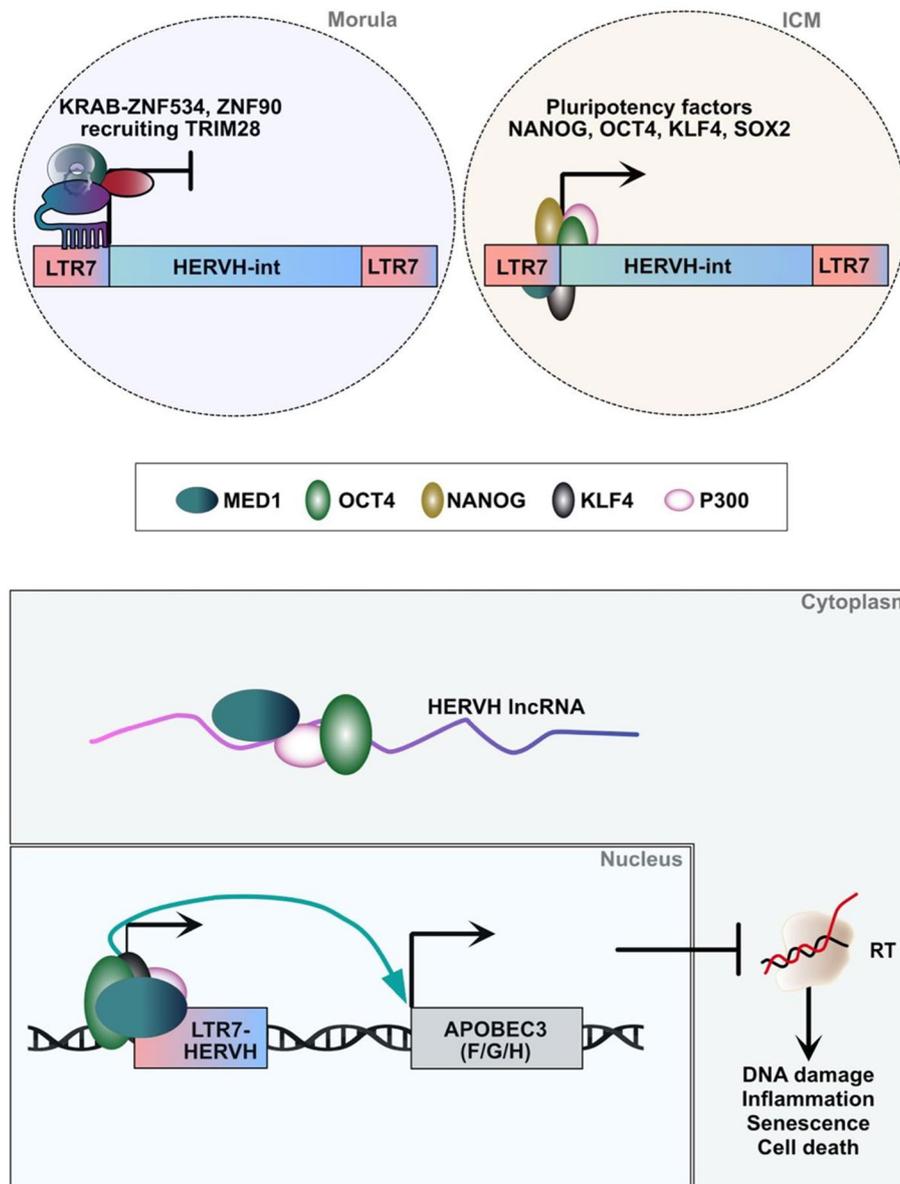


Fig. 5 (Upper panel) The KRAB-ZNF/TRIM28 machinery is a suppressor of the LTR7-HERVH loci in the morula, but not in the ICM, where its expression is controlled by pluripotent transcription factors. (Middle panel) HERVH lncRNA recruits OCT4, P300 and MED1 in the cytoplasm and (Lower panel) enables expression of the LTR7-HERVH/APOBEC3 locus in the nucleus. APOBEC3F/G/H inhibit reverse transcription (RT) activity

One possible explanation is that, although HERVH is specifically targeted by ZNF proteins (eg ZNF90, which recognizes all LTR7 subfamilies, and ZNF534, which preferentially binds LTR7up, these potential repressor proteins are highly expressed during the morula stage but are expressed at low levels in pluripotent stem cells [31] (Figs. 3 and 5). As a result, they suppress LTR7-HERVH only in cellular contexts where these factors are sufficiently expressed. A second explanation, supported by our TRIM28 ChIP-seq data analysis, is that TRIM28

binding does not lead to HERVH silencing per se. Notably, TRIM28 binds within the internal HERVH sequences and not in the regulatory LTR7 regions occupied by pluripotency factors (Fig. 4C). Alternatively, it is possible that TRIM28, when bound to HERVH, undergoes post-translational modifications that alter its function, potentially shifting its role from a repressor to an activator.

Furthermore, LTR7-HERVH provides binding sites for pluripotency transcription factors, including POU5F1/OCT4, NANOG, TFCEP2L1/LBP9 and KLF4 [51, 52, 54,

55]. Given that TRIM28/KAP1 does not seem to be the primary suppressor of HERVH in human ESCs, it has been proposed that pioneer factor KLF4 [124] prevents TRIM28/KAP1 binding and instead recruit additional pluripotency factors, as well as the E1A-binding protein p300 histone acetyltransferase, to HERVH loci [56, 58]. In this model, both the novel pluripotency functions and escape from suppression are possible as KLF4 mediates both (Fig. 5).

In mouse pluripotent stem cells (PSCs), Trim28-induced suppression leads to DNA methylation, resulting in permanent silencing of TEs [125, 126], whereas the relationship between TRIM28 occupancy and DNA methylation is less tight in human PSCs [118]. Either way, the hypothesis that KLF4 outcompeting TRIM28 contributes to the evasion of suppressive methylation is worthy of further scrutiny. While DNA CpG methylation generally suppresses TEs, hypomethylation within specific TE families is associated with tissue-specific enhancer landscapes. Indeed, LTR7-HERVH is significantly hypomethylated in the genomes of PSCs [53, 127].

Underrepresented for canonical TP53 binding sites for transcriptional control

The tumor protein 53 (p53 alias TP53) plays a central role in regulating transposable elements in the germline through conserved mechanisms (reviewed in ref [128]). In germline, p53 enforces genome integrity by inducing cell death via a conserved apoptotic mechanism in cells with DNA damage or abnormal genomes [129, 130]. In human development, the movement of the only autonomous TE, LINE-1 (L1), was long thought to be restricted to the germline. However, we have recently reported that cells exhibiting DNA damage associated with L1 activity are also removed through apoptosis during early human embryogenesis, specifically after the morula stage [63]. Under physiological conditions, L1 activity has also been observed in human neuronal cells [131].

Besides its interplay with piRNA pathways, TP53 can directly suppress TE expression through DNA binding. TP53 binds conserved recognition motifs within LINE-1 elements, specifically at the 5' untranslated region (5'-UTR), where it represses L1 transcription by recruiting repressive chromatin modifiers, leading to heterochromatin formation [132, 133]. Interestingly, approximately 30% of TP53 binding sites overlap with genomic loci of endogenous retroviruses (ERVs) across the genome [134]. Some of these sequences were identified as negative regulators of ERV transcription (eg HERV-1-LTRs) [135]. Curiously, several (1,509/319,000) of the TP53 binding sites identified in ERVs are canonical (RRRCWWGYYY-0-13 bp spacer- RRRCWWGYYY) [136]. Notably, these canonical TP53 sites are significantly underrepresented

in LTR7-HERVH (Table S2). While this could mean that these HERVH loci have escaped TP53-mediated suppression, alternatively, the lack of TP53 binding sites may simply reflect either the absence of evolutionary selection for TP53 regulation or selection against it in pluripotent stem cells, where TP53 suppression of HERVH may not have provided a selective advantage.

Interestingly, TP53 binding sites do not always suppress L1 transcription. In response to double-strand breaks (DSBs), certain L1 loci increase transcription and transposition, creating a positive feedback loop that amplifies genomic stress and triggers TP53-mediated apoptosis [137]. Here, TP53 binds a slightly altered motif (p53 responsive element) within the L1 promoter, promoting L1 mRNA synthesis and retrotransposition. This paradoxical mechanism reinforces genomic stability by amplifying TP53-dependent responses that enforce fidelity through cell death.

During evolution, some TEs have acquired TP53 DNA binding sites, enabling them to regulate neighboring gene expression and form part of a broader regulatory network [132]. Intriguingly, non-canonical TP53 sites (differing in sequence and featuring a 129-bp spacer) have been identified in the *pol* domain of certain HERVH loci [137]. The biological significance of these non-canonical TP53 binding sites in HERVH warrants further investigation.

Releasing of LTR7/HERVH from transcriptional suppression

RNA N6-methyladenosine (m6A) modification is the most abundant epitranscriptomic modification [138]. Post-transcriptional modification of RNA by m6A plays a critical role in regulating the silencing of TEs by influencing histone modifications [139]. m6A-modified RNA are recognized by reader proteins, such as YTHDC1 and YTHDC2. When YTHDC1 binds to m6A-marked TE-derived transcripts, it can lead to either the direct degradation of these transcripts or the recruitment of histone modifiers, which induce the epigenetic silencing of genomic TE loci through histone modifications [140, 141].

HERVH appears to employ a similar suite of players but their effect is the reverse, i.e. activating of HERVH. The RNA m6A reader YTHDC2 specifically interacts with m6A-modified HERVH RNA transcribed from the LTR7-HERVH genomic loci in human PSCs. YTHDC2 not only occupies the genomic loci of LTR7-HERVH but also collaborates with the DNA 5mC 'eraser' TET1 demethylase. The removal of 5mC from LTR7-HERVH by TET1 helps maintain the active state of the LTR7-HERVH loci in hESCs, and influences their neuronal fate commitment [142]. This example illustrates how, in the context of a domestication process, instead of suppressing the LTR7-HERVH locus, the m6A-modified HERVH RNA recruit

TET1, thereby releasing LTR7-HERVH from transcriptional suppression. It is as though HERVH has somehow hijacked the system sent to suppress it to activate instead.

Developmentally programmed HERVH degradation

Uridylation by TUT4/7 inhibits the retrotransposition of LINE-1 elements [143]. TUT7 also uridylylates HERVH RNAs, leading to their degradation [81]. Interestingly, the regulation of HERVH through TUT7-mediated degradation is an integral part of the human developmental program, particularly in neural differentiation. Cells that maintain high levels of HERVH expression (HERVH^{High} cells) are deficient in neural differentiation [58, 79]. However, the programmed degradation of HERVH by TUT7 restores pluripotency and enables these cells to differentiate into the neural lineage [81].

HERVH in chimeric transcripts – splicing

In retroviruses, *pol* and *gag* are not processed by splicing, whereas splicing is essential for *env* expression. In HERVH, while *gag* remains recognizable [23], *pol* does not, and the HERVH loci lack an intact *env* gene due to mutations and deletions. Unlike HERVK, which retains protein-coding potential, HERVH is primarily transcriptionally active.

Splicing likely contributes to the stabilization of HERVH-derived long non-coding RNA (lncRNA) transcripts. Notably, compared to mRNAs, lncRNAs often undergo inefficient splicing, frequently utilizing cryptic splice sites or non-canonical mechanisms. Indeed, certain HERVH-derived lncRNAs, such as LINC-ROR, exhibit features of cryptic splicing or non-canonical mechanisms, such as intron retention and multiple isoforms [144]. Many of these variants are likely to be lost through random evolutionary processes unless maintained by positive selection. However, specific spliced isoforms that confer functional advantages could be retained.

Importantly, a subset of HERVH possesses functional splice donor (SD) and splice acceptor (SA) sites, enabling the formation of chimeric transcripts. This suggests that HERVH's retroviral-derived splicing machinery has been repurposed to incorporate exons from human genes into novel chimeric transcripts. These spliced transcripts typically integrate HERVH as the 5' sequence, bridging to the 5' splice site of exon 2 of a neighboring gene on the same strand. In contrast, SAFB suppresses exonization of other retroelement-derived transcripts [145]. HERVH elements are also characterised by their ability to efficiently donate transcription start sites (TSSs) to chimeric transcripts [53]. These properties, together with their strong LTR, facilitate the formation of chimeric transcripts.

In addition to HERVH, SVA elements are also capable of generating chimeric transcripts. However, while some SVA elements have their own promoter [146], many are dependent on upstream gene promoters or read-through transcription from host genes. It has been shown that SVA elements are incorporated into newly formed transcripts as exonic sequences via their cryptic splice acceptor sites [147].

At least a portion of the HERVH-derived chimeric transcripts with protein-coding genes are expected to be translated. HERVH-derived chimeric lncRNAs [52, 55], actively recruit splicing factors (eg SF1, SF3A1/B3) [84] and SFRS3/9/10/12 [61], suggesting that they undergo typical post-transcriptional processing, including 5'-capping, splicing, and 3'-polyadenylation (Fig. 6). The absence of splicing, as well as the positioning of the first intron near the ATG codon, plays a crucial role in distinguishing native transcripts from aberrant ones. Splicing is anticipated to enhance the stability and functionality of HERVH-mediated chimeric transcripts, in part by passing the first intron and multi-exon tests necessary for nuclear export. The first intron can also facilitate progressive RNA polymerase II elongation [148]. Additionally, a multi-short exon structure enhances transcript recognition [11].

These HERVH-enhanced chimeric products contribute to gene regulation within the pluripotent niche [52, 56].

Discussion

How transposable elements are domesticated is of focal concern to the problem of the evolutionary gain of function problem (how to improve a swiss watch). While the focus has historically been on TE features recruited to host utility, there exists a second component, we suggest, that has attracted less attention, this being how a domesticated TE might avoid the filters and traps set up to prevent or remove unwanted transcripts. Our central hypothesis proposes that the co-option of HERVH has involved evasion from cellular surveillance mechanisms [11].

A subtle aspect of our model is that it requires that the anti-TE filters are leaky: if a TE escapes one suppressor it still needs to pass through the others which it cannot do if all filters are foolproof: escape from one is of no utility if it faces full suppression later. We argue that the evolution of anti-TE systems predisposes to a system with multiple leaky filters. First, any suppressive system needs to balance the need to effectively express native genes with filters to prevent the expression of non-native ones. It is hard to envisage any system that is so effective that it permits expression of all native genes and only these genes. A foolproof filter against intronless transcripts, for example, would capture many essential transcripts (eg histones) and likely

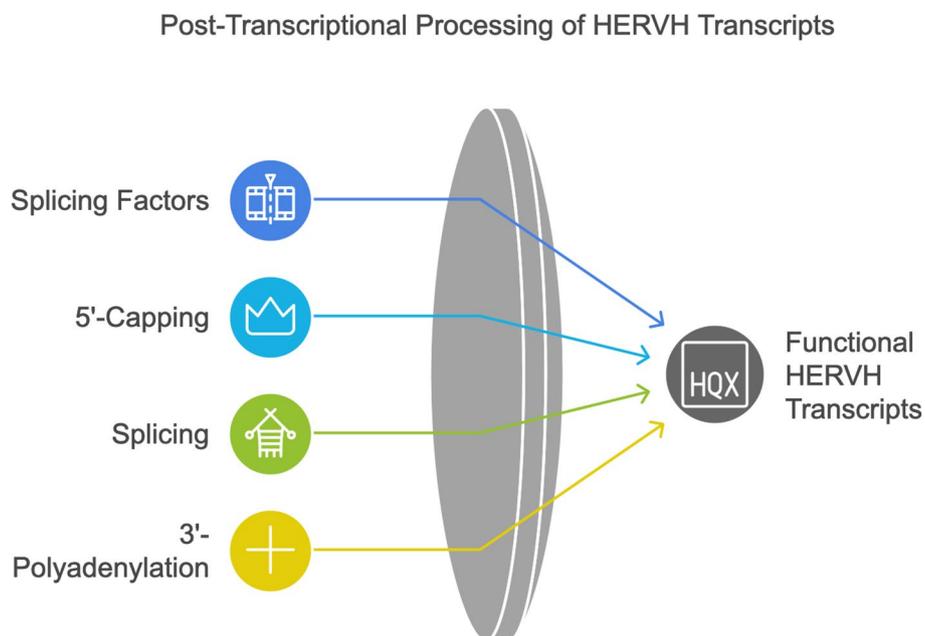


Fig. 6 HERVH-derived transcripts effectively recruit splicing factors, likely undergo typical post-transcriptional processing steps, including 5'-capping, splicing and 3'-polyadenylation

be catastrophic. The filters may thus need to be leaky. Moreover, as with the case of mutation rate reduction, there comes a point where the benefits of reduced TE activity are traded-off against increasing costs of such suppression with equilibrium being a leaky suppressor. That, however, is an equilibrium model and TE suppression systems may not be at equilibrium [149]. Rather they may be regarded as emergency responses in which any limited degree of suppression is better than none (or grossly imperfect suppression). With the possibility of rapid turnover of suppression systems [149], we may expect TE suppressors to be non-optimised. With no filter expected to be foolproof, and TEs under selection to avoid whatever filters are in place, we expect the evolution of multiple leaky pathways. These mechanisms can be broad, such as the HUSH complex (targets potentially mutagenic TEs), or more specific, like the ZNF-KRAB (targets specific TEs) system. Importantly, this leakiness allows TE colonization even in the absence of predispositional escape mechanisms, although naturally we expect that TEs with such predispositions to be more likely to colonize.

We should also highlight that our hypothesis, demonstrating how HERVH has evaded multiple host-encoded filtering mechanisms, aligns with the 'transposon addiction hypothesis' proposed by Cedric Feschotte [150]. Following its invasion, HERVH likely persisted due to genetic drift. However, a subset of genomic copies within the LTR7-HERVH subfamily acquired crucial functions

that the host became dependent on. This dependency prevented the silencing of the LTR7-HERVH subfamily, and may explain why HERVH transcripts are highly expressed during early development.

In the case of HERVH this problem is especially notable as the transcripts are both strikingly abundant in some cell types and would in the recent evolutionary past have been prime candidates for suppression.

The above close consideration suggests that predisposing features of HERVH that avoid filters may well be key. In some cases, filter avoidance and novelty generation are intimately coupled. It is notable that both KLF4 enables novel pluripotent functionality and excludes suppressors. Similarly, HERVH is peculiar in having a strong splice site that enables generation of chimeric multi-exon transcripts. By all accounts these should pass the filters to enable, for example, nuclear export of multi-exon transcripts.

In other regards HERVH appears to have evolved mechanisms to escape classic suppression systems. It is noteworthy that many suppression mechanisms that originally developed in response to TE colonization have now been integrated into general cellular quality control processes and now also fulfil additional tasks [11]. For instance, key regulatory innovations in eukaryotic genomes, including introns, splicing, and nuclear export, are widely believed to have emerged, at least in part, as adaptations to TE invasion [151–153]. Other examples include the TP53 pathway, which initially

functioned to suppress TE-induced genomic instability via the DNA damage response [154] but has since become a master regulator of the cell cycle, orchestrating apoptosis and senescence in response to DNA damage. Similarly, some KRAB-ZFPs, originally dedicated to TE repression, have been repurposed to regulate host genes during embryonic development. The HUSH complex, once solely responsible for silencing newly integrated transposons [155], now also functions in filtering out noisy transcripts [11, 109] and plays a broader role in regulating immune response genes, neuronal function, and maintaining epigenetic stability during early development.

As regards the effects of m6A, it appears to have taken an anti-unwanted transcript filter system and turned it on its head. Similarly, that HERVH is unusual in avoiding canonical TP53 binding sites is consistent with a model that you cannot generate novelty unless you can avoid suppression, although in this instance novelty and suppression are not so obviously coupled. Note that while we regard these features as predispositions, they would only be that if these features were ancestral. Otherwise, they would be regarded as evolved features. For our model we consider that predispositions to evasion of suppression is not necessary but will tilt the balance towards domestication. Other suppression evasion mechanisms may come via mutation after the initial active phase and may indeed be favoured as adaptive means to avoid suppression. We often cannot be sure which are evolved and which ancestral but as some level of gene expression is necessary for initial co-option, it is likely that some are ancestral, similar survival bias potentially explaining why nuclear and cytoplasmic viruses have different nucleotide contents [156]. Note that even if ancestral it is not necessary to evoke the hypothesis that these ancestral states are the product of selection favouring expression in the new host, but rather that, present for whatever reason, they further such expression. It remains an open question as to whether, when TE's move between organisms, selection acting in one host predisposes to successful colonization of a subsequent host.

One should also note that many non-functional HERV-derived sequences in the genome likely evade defense mechanisms. Over time, these sequences undergo stochastic loss and divergence, eventually losing their resemblance to HERVs, posing no threat, and no longer being recognized as 'non-self'. However, some HERVs, despite being fixed for millions of years, remain detectable by host defense machineries under pathological conditions.

This also applies to HERVs with co-opted functions. Notably, tolerance to HERVH-derived transcripts is lost during differentiation, suggesting that even co-opted functions, typically recognized as 'self', may lose host

tolerance under certain conditions. This implies that the balance between immune tolerance and self-defense activation is shaped by factors such as developmental stage, cellular environment, or pathological conditions. Notably, pseudogenes and intronic sequences, despite lacking function, are unequivocally recognized as 'self'.

Understanding these problems is relevant, we suggest, not simply in an evolutionary context but also in the context of transgene design, and indeed especially relevant in the context of gene therapies that seek to incorporate genes stably in to host DNA (less so for RNA vaccines as these need have no transcriptional or nuclear export components). Indeed, non-expression of some transgenes is common, in part because they are thought to be captured by unwanted transcript filters [11]. Are these being treated like a TE with a potential gain of function but falling foul of the transcript filters?

The same insights might also have relevance for cancer pathology. Cancer cells are often thought to resemble in many regards pluripotent cells [70, 157]. Key transcription factors that govern early embryonic fate are also active in certain tumor cells [158]. Deregulated chromatin accessibility may be a common factor linking cancer to early embryogenesis. As HERVH is active in such cells and appears to have anti-unwanted transcript features it may well be a candidate for gain of function not only in evolutionary time but in tumour-time (for want of a better expression). While much TE expression in cancers may reflect nothing more than unwanted transcripts enabled by disturbed epigenetic regulation, the chimeric HERVH-CALB1 transcript detected in cancer [159], has also been observed in human embryonic stem cells and in the epiblast of pre-implantation embryos [52, 160], suggesting that it may represent an exaptation event in normal physiology and also functional in cancer-time. Pathologically activated ERVs can be repurposed as alternative promoters that drive oncogene expression [161–163], sometimes referred to as onco-exaptation. Abnormally activated ERVs may allow cancer cells to exploit and repurpose developmental pathways promoting malignancy [164]. For example, in colorectal cancer, the loss of the BAF chromatin remodeler subunit ARID1A, a tumor suppressor, leads to the derepression of a specific set of HERVH loci, which in turn modulates BRD4-dependent transcription [165]. These onco-exaptation events are often linked to poorer patient outcomes.

In contrast to typical onco-exaptation, a role favourable to the host was observed for the protein product (*Rec*) of the evolutionarily younger HERVK, which counteracts the progression of the cancer to an invasive stage [166]. Additionally, in lung squamous cell carcinoma (LUSC), the production of chimeric transcripts

driven by HERVH and a neighboring gene from the calbindin 1 (CALB1) locus correlates with improved patient survival [159].

Understanding the cancer co-options – both harmful and beneficial – may provide an accessible system to more generally understand how ERV related transcripts both generate novelty and escape unwanted transcript suppression systems. This however comes with an important caveat, namely that, as many cancers are characterised by mutations in the unwanted transcript filters [167] (indicative of the importance of such filters), gain-of-function in cancers need not always be a good model of gain-of-function over evolutionary time when these filters would be operative.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13100-025-00361-0>.

Supplementary Material 1: Table S1. Upregulated HERVH family members upregulated in KO TRIM28 hESCs (data source from [132]). Note that the top expressed ESRG locus is not affected.

Supplementary Material 2: Table S2. Frequency of near-perfect TP53 sites (Two 10-mer (10 nucleotides (nt)) half-sites, linked by a short spacer region of 0–13 nt) in selected ERVs in the human genome (data source from [132]).

Analyses

Involvement of KAP1 in the regulation of LTR7 in hESC_H1
We mined data from [122] (GSE58323). First, we divided LTR7/HERVH genomic loci into active and inactive groups. Then we fetched 5KB upstream and 5KB downstream from the boundaries of LTR7 which makes 10 KB of total length. We divided them into 100 bins of 100 bps each. Then we counted KAP1 ChIP-seq signal over them followed by averaging them and then plotted them with smooth spline parameter 0.2.

Authors' contributions

Z.I. and L.D.H. wrote the review. J.M. performed the TP53 data mining. M.S. edited the review.

Funding

J.M. is supported by a scholarship 202308080115 from the Chinese Science Council.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 December 2024 Accepted: 14 May 2025
Published online: 05 July 2025

References

- Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 2002;18(9):486. [https://doi.org/10.1016/s0168-9525\(02\)02722-1](https://doi.org/10.1016/s0168-9525(02)02722-1). (PubMed PMID: 12175810).
- Ohta T. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology, Evolution, and Systematics.* 1992;23:263–86. <https://doi.org/10.1146/annurev.es.23.110192.001403>.
- Carscadden KA, Batstone RT, Hauser FE. Origins and evolution of biological novelty. *Biol Rev Camb Philos Soc.* 2023;98(5):1472–91. Epub 20230413. <https://doi.org/10.1111/brv.12963>. PubMed PMID: 37056155.
- Wu L, Lambert JD. Clade-specific genes and the evolutionary origin of novelty; new tools in the toolkit. *Semin Cell Dev Biol.* 2023;145:52–9. Epub 20220531. <https://doi.org/10.1016/j.semcdb.2022.05.025>. PubMed PMID: 35659164.
- Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 2008;9(5):397–405. <https://doi.org/10.1038/nrg2337>. (PubMed PMID: 18368054; PubMed Central PMCID: PMC2596197).
- Laviolle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Veronchet C, et al. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci.* 2013;368(1626):20120507. Epub 2013/08/14. <https://doi.org/10.1098/rstb.2012.0507>. PubMed PMID: 23938756; PubMed Central PMCID: PMC3758191.
- Sugimoto J, Sugimoto M, Bernstein H, Jinno Y, Schust D. A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci Rep.* 2013;3:1462. <https://doi.org/10.1038/srep01462>. (PubMed PMID: 23492904; PubMed Central PMCID: PMC3598002).
- Frank JA, Singh M, Cullen HB, Kirou RA, Benkaddour-Boumzaouad M, Cortes JL, et al. Evolution and antiviral activity of a human protein of retroviral origin. *Science.* 2022;378(6618):422–8. Epub 20221027. <https://doi.org/10.1126/science.abq7871>. PubMed PMID: 36302021; PubMed Central PMCID: PMC3542854.
- Kunarseo G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet.* 2010;42(7):631–4. Epub 2010/06/08. <https://doi.org/10.1038/ng.600>. PubMed PMID: 20526341.
- Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* 2017;18(2):71–86. Epub 20161121. <https://doi.org/10.1038/nrg.2016.139>. PubMed PMID: 27867194; PubMed Central PMCID: PMC35498291.
- Radrizzani S, Kudla G, Izsvák Z, Hurst LD. Selection on synonymous sites: the unwanted transcript hypothesis. *Nat Rev Genet.* 2024. Epub 20240131. <https://doi.org/10.1038/s41576-023-00686-7>. PubMed PMID: 38297070.
- Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A.* 2004;101(14):4894–9. Epub 20040325. <https://doi.org/10.1073/pnas.0307800101>. PubMed PMID: 15044706; PubMed Central PMCID: PMC387345.
- Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet.* 2008;42:709–32. Epub 2008/08/13. <https://doi.org/10.1146/annurev.genet.42.110807.091501>. PubMed PMID: 18694346.
- Johnson WE. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat Rev Microbiol.* 2019;17(6):355–70. <https://doi.org/10.1038/s41579-019-0189-2>. (PubMed PMID: 30962577).
- Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet.* 2007;41:331–68. <https://doi.org/10.1146/annurev.genet.40.110405.090448>. (PubMed PMID: 18076328; PubMed Central PMCID: PMC3542854).
- Fedoroff NV. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science.* 2012;338(6108):758–67. PubMed PMID: 23145453.

17. Saitou M, Yamaji M. Primordial germ cells in mice. *Cold Spring Harb Perspect Biol.* 2012;4(11). Epub 20121101. <https://doi.org/10.1101/cshperspect.a008375>. PubMed PMID: 23125014; PubMed Central PMCID: PMC3536339.
18. Leitch HG, Smith A. The mammalian germline as a pluripotency cycle. *Development.* 2013;140(12):2495–501. <https://doi.org/10.1242/dev.091603>. (PubMed PMID: 23715543).
19. Lohe AR, Hartl DL. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. *Mol Biol Evol.* 1996;13(4):549–55. <https://doi.org/10.1093/oxfordjournals.molbev.a025615>. (PubMed PMID: 8882498).
20. Saha A, Mitchell JA, Nishida Y, Hildreth JE, Ariberre JA, Gilbert WV, et al. A trans-dominant form of Gag restricts Ty1 retrotransposition and mediates copy number control. *J Virol.* 2015;89(7):3922–38. Epub 20150121. <https://doi.org/10.1128/JVI.03060-14>. PubMed PMID: 25609815; PubMed Central PMCID: PMC34403431.
21. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8(4):272–85. <https://doi.org/10.1038/nrg2072>. (PubMed PMID: 17363976).
22. Belshaw R, Watson J, Katzourakis A, Howe A, Woolven-Allen J, Burt A, et al. Rate of recombinational deletion among human endogenous retroviruses. *J Virol.* 2007;81(17):9437–42. <https://doi.org/10.1128/JVI.02216-06>. (PubMed PMID: 17581995; PubMed Central PMCID: PMC1951428).
23. Gemmell P, Hein J, Katzourakis A. The Exaptation of HERV-H: Evolutionary Analyses Reveal the Genomic Features of Highly Transcribed Elements. *Front Immunol.* 2019;10:1339. Epub 20190709. <https://doi.org/10.3389/fimmu.2019.01339>. PubMed PMID: 31338090; PubMed Central PMCID: PMC6629862.
24. 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(6822):860–921. Epub 2001/03/10. <https://doi.org/10.1038/35057062>. PubMed PMID: 11237011.
25. Li Y, Zhang G, Cui J. Origin and Deep Evolution of Human Endogenous Retroviruses in Pan-Primates. *Viruses.* 2022;14(7). Epub 20220623. <https://doi.org/10.3390/v14071370>. PubMed PMID: 35891351; PubMed Central PMCID: PMC9323773.
26. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol.* 1999;73(4):3301–8. <https://doi.org/10.1128/JVI.73.4.3301-3308.1999>. (PubMed PMID: 10074184; PubMed Central PMCID: PMC104094).
27. Bannert N, Kurth R. Retroelements and the human genome: new perspectives on an old relation. *Proc Natl Acad Sci U S A.* 2004;101 Suppl 2(Suppl 2):14572–9. Epub 20040813. <https://doi.org/10.1073/pnas.0404838101>. PubMed PMID: 15310846; PubMed Central PMCID: PMC521986.
28. 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(7555):221–5. <https://doi.org/10.1038/nature14308>. (PubMed PMID: 25896322; PubMed Central PMCID: PMC4503379).
29. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet.* 2006;7:149–73. <https://doi.org/10.1146/annurev.genom.7.080505.115700>. (PubMed PMID: 16722807).
30. Mager DL, Medstrand P. Retroviral Repeat Sequences. *Nature Encyclopedia of the Human Genome.* 2003; (ed. DN Cooper. London/New York: Nature Pub. Group).
31. Carter TA, Singh M, Dumbovic G, Chobirko JD, Rinn JL, Feschotte C. Mosaic cis-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. *Elife.* 2022;11. Epub 20220218. <https://doi.org/10.7554/eLife.76257>. PubMed PMID: 35179489; PubMed Central PMCID: PMC8912925.
32. Jern P, Sperber GO, Blomberg J. Definition and variation of human endogenous retrovirus H. *Virology.* 2004;327(1):93–110. <https://doi.org/10.1016/j.virol.2004.06.023>. (PubMed PMID: 15327901).
33. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? *BioEssays.* 2016;38(1):109–17. <https://doi.org/10.1002/bies.201500096>. (PubMed PMID: 26735931).
34. Vargiu L, Rodriguez-Tome P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology.* 2016;13:7. Epub 20160122. <https://doi.org/10.1186/s12977-015-0232-y>. PubMed PMID: 26800882; PubMed Central PMCID: PMC4724089.
35. 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. Epub 20111108. <https://doi.org/10.1186/1742-4690-8-90>. PubMed PMID: 22067224; PubMed Central PMCID: PMC3228705.
36. van der Kuyl AC. HIV infection and HERV expression: a review. *Retrovirology.* 2012;9:6. Epub 20120116. <https://doi.org/10.1186/1742-4690-9-6>. PubMed PMID: 22248111; PubMed Central PMCID: PMC3311604.
37. Mantovani F, Kitsou K, Magiorkinis G. HERVs: Expression Control Mechanisms and Interactions in Diseases and Human Immunodeficiency Virus Infection. *Genes (Basel).* 2024;15(2). Epub 20240131. <https://doi.org/10.3390/genes15020192>. PubMed PMID: 38397182; PubMed Central PMCID: PMC10888493.
38. Singh M, Leddy SM, Iniguez LP, Bendall ML, Nixon DF, Feschotte C. Transposable elements may enhance antiviral resistance in HIV-1 elite controllers. *Genome Biol.* 2025;26(1):28. Epub 20250224. <https://doi.org/10.1186/s13059-025-03484-y>. PubMed PMID: 39988678; PubMed Central PMCID: PMC11849351.
39. Terry SN, Manganaro L, Cuesta-Dominguez A, Brinzevich D, Simon V, Mulder LCF. Expression of HERV-K108 envelope interferes with HIV-1 production. *Virology.* 2017;509:52–9. Epub 20170609. <https://doi.org/10.1016/j.virol.2017.06.004>. PubMed PMID: 28605635; PubMed Central PMCID: PMC5538378.
40. Johnson WE, Coffin JM. Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci U S A.* 1999;96(18):10254–60. <https://doi.org/10.1073/pnas.96.18.10254>. (PubMed PMID: 10468595; PubMed Central PMCID: PMC17875).
41. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science.* 2016;351(6277):1083–7. <https://doi.org/10.1126/science.aad5497>. (PubMed PMID: 26941318; PubMed Central PMCID: PMC4887275).
42. Goodier JL. Restricting retrotransposons: a review. *Mob DNA.* 2016;7:16. Epub 20160811. <https://doi.org/10.1186/s13100-016-0070-z>. PubMed PMID: 27525044; PubMed Central PMCID: PMC4982230.
43. 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(2):395–404. <https://doi.org/10.1006/viro.1995.0012>. (PubMed PMID: 7491764).
44. Gemmell P, Hein J, Katzourakis A. Phylogenetic Analysis Reveals That ERVs "Die Young" but HERV-H Is Unusually Conserved. *PLoS Comput Biol.* 2016;12(6): e1004964. <https://doi.org/10.1371/journal.pcbi.1004964>. (PubMed PMID: 27295277; PubMed Central PMCID: PMC4905674).
45. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA.* 2016;7:24. Epub 20161201. <https://doi.org/10.1186/s13100-016-0080-x>. PubMed PMID: 27980689; PubMed Central PMCID: PMC5134097.
46. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA.* 2021;12(1):2. Epub 20210112. <https://doi.org/10.1186/s13100-020-00230-y>. PubMed PMID: 33436076; PubMed Central PMCID: PMC7805219.
47. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6:11. <https://doi.org/10.1186/s13100-015-0041-9>. (PubMed PMID: 26045719; PubMed Central PMCID: PMC4455052).
48. Kojima KK. Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob DNA.* 2018;9:2. Epub 20180104. <https://doi.org/10.1186/s13100-017-0107-y>. PubMed PMID: 29308093; PubMed Central PMCID: PMC5753468.
49. Goodchild NL, Wilkinson DA, Mager DL. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology.* 1993;196(2):778–88. <https://doi.org/10.1006/viro.1993.1535>. (PubMed PMID: 8372448).
50. Jern P, Sperber GO, Ahlsen G, Blomberg J. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. *J Virol.* 2005;79(10):6325–37. <https://doi.org/10.1128/JVI.79.10>.

- 6325-6337.2005. (PubMed PMID:15858016;PubMedCentralPMCID:PMC1091717).
51. Goke J, Lu X, Chan YS, Ng HH, Ly LH, 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(2):135–41. <https://doi.org/10.1016/j.stem.2015.01.005>. (PubMed PMID: 25658370).
 52. Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*. 2014;516(7531):405–9. Epub 2014/10/16. <https://doi.org/10.1038/nature13804>. PubMed PMID: 25317556.
 53. Sexton CE, Tillett RL, Han MV. The essential but enigmatic regulatory role of HERVH in pluripotency. *Trends Genet*. 2022;38(1):12–21. Epub 2021/07/30. <https://doi.org/10.1016/j.tig.2021.07.007>. PubMed PMID: 34340871; PubMed Central PMCID: PMCPCMC8678302.
 54. Kelley DR, Rinn JL. Transposable elements reveal a stem cell specific class of long noncoding RNAs. *Genome Biol*. 2012;13(11):R107. Epub 2012/11/28. <https://doi.org/10.1186/gb-2012-13-11-r107>. PubMed PMID: 23181609.
 55. Lu X, Sachs F, Ramsay L, Jacques PE, Goke 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(4):423–5. Epub 2014/04/01. <https://doi.org/10.1038/nsmb.2799>. PubMed PMID: 24681886.
 56. Romer C, Singh M, Hurst LD, Izsvák Z. How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. *Curr Opin Virol*. 2017;25:49–58. <https://doi.org/10.1016/j.coviro.2017.07.001>. (PubMed PMID: 28750248).
 57. 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(1):111. Epub 2012/12/21. <https://doi.org/10.1186/1742-4690-9-111>. PubMed PMID: 23253934.
 58. 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(34):12426–31. Epub 2014/08/07. <https://doi.org/10.1073/pnas.1413299111>. PubMed PMID: 25097266; PubMed Central PMCID: PMC4151758.
 59. Wang J, Singh M, Sun C, Besser D, Prigione A, Ivics Z, et al. Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH expression. *Nat Protoc*. 2016;11(2):327–46. <https://doi.org/10.1038/nprot.2016.016>. (PubMed PMID: 26797457).
 60. 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(6):558–66. Epub 2014/04/28. <https://doi.org/10.1038/ng.2965>. PubMed PMID: 24777452.
 61. 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(1):44–52. Epub 2015/11/23. <https://doi.org/10.1038/ng.3449>. PubMed PMID: 26595768; PubMed Central PMCID: PMCPCMC4827613.
 62. Barakat TS, Halbritter F, Zhang M, Rendeiro AF, Perenthaler E, Bock C, et al. Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell*. 2018;23(2):276–88 e8. Epub 2018/07/24. <https://doi.org/10.1016/j.stem.2018.06.014>. PubMed PMID: 30033119; PubMed Central PMCID: PMCPCMC6084406.
 63. Singh M, Kondrashkina AM, Widmann TJ, Cortes JL, Bansal V, Wang J, et al. A new human embryonic cell type associated with activity of young transposable elements allows definition of the inner cell mass. *PLoS Biol*. 2023;21(6):e3002162. Epub 2023/06/20. <https://doi.org/10.1371/journal.pbio.3002162>. PubMed PMID: 37339119; PubMed Central PMCID: PMCPCMC10281584.
 64. Jiang Y, Qian F, Bai X, Liu Y, Wang Q, Ai B, et al. SEdb: a comprehensive human super-enhancer database. *Nucleic Acids Res*. 2019;47(D1):D235–43. <https://doi.org/10.1093/nar/gky1025>. (PubMed PMID:30371817;PubMedCentralPMCID:PMCPCMC6323980).
 65. Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J*. 2012;31(3):522–33. Epub 2011/12/24. <https://doi.org/10.1038/emboj.2011.459>. PubMed PMID: 22193719; PubMed Central PMCID: PMC3273385.
 66. 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(1):69–80. Epub 2013/03/28. <https://doi.org/10.1016/j.devcel.2013.03.002>. PubMed PMID: 23541921.
 67. Chen YF, Li YJ, Chou CH, Chiew MY, Huang HD, Ho JH, et al. Control of matrix stiffness promotes endodermal lineage specification by regulating SMAD2/3 via lincRNA LINC00458. *Sci Adv*. 2020;6(6):eaay0264. Epub 2020/02/05. <https://doi.org/10.1126/sciadv.aay0264>. PubMed PMID: 32076643; PubMed Central PMCID: PMCPCMC7002135.
 68. Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet*. 2019;51(9):1380–8. Epub 2019/08/19. <https://doi.org/10.1038/s41588-019-0479-7>. PubMed PMID: 31427791; PubMed Central PMCID: PMCPCMC6722002.
 69. Hsieh FK, Ji F, Damle M, Sadreyev RI, Kingston RE. HERVH-derived lincRNAs negatively regulate chromatin targeting and remodeling mediated by CHD7. *Life Sci Alliance*. 2022;5(1). Epub 2021/10/18. <https://doi.org/10.26508/lsa.202101127>. PubMed PMID: 34663690; PubMed Central PMCID: PMCPCMC8548210.
 70. Singh M, Kondrashkina AM, Hurst LD, Izsvák Z. Staring at the onco-exaptation: the two-faced medley of an ancient retrovirus, HERVH. *J Clin Invest*. 2023;133(14). Epub 2023/07/17. <https://doi.org/10.1172/JCI172278>. PubMed PMID: 37463444; PubMed Central PMCID: PMCPCMC10348764.
 71. Takaori A. Antiviral defense by APOBEC3 family proteins. *Uirusu*. 2005;55(2):267–72. <https://doi.org/10.2222/jsv.55.267>. (PubMed PMID: 16557012).
 72. Harris RS, Hultquist JF, Evans DT. The restriction factors of human immunodeficiency virus. *J Biol Chem*. 2012;287(49):40875–83. Epub 2012/10/05. <https://doi.org/10.1074/jbc.R112.416925>. PubMed PMID: 23043100; PubMed Central PMCID: PMCPCMC3510791.
 73. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, et al. Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A*. 2006;103(23):8780–5. Epub 2006/05/25. <https://doi.org/10.1073/pnas.0603313103>. PubMed PMID: 16728505; PubMed Central PMCID: PMCPCMC1482655.
 74. Aswad A, Katzourakis A. Paleovirology and virally derived immunity. *Trends Ecol Evol*. 2012;27(11):627–36. Epub 2012/08/14. <https://doi.org/10.1016/j.tree.2012.07.007>. PubMed PMID: 22901901.
 75. Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, et al. Mouse cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. 2007;448(7150):196–9. Epub 2007/06/27. <https://doi.org/10.1038/nature05972>. PubMed PMID: 17597760.
 76. Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, et al. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell*. 2016;19(4):502–15. Epub 2016/07/19. <https://doi.org/10.1016/j.stem.2016.06.011>. PubMed PMID: 27424783; PubMed Central PMCID: PMCPCMC5065525.
 77. Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, et al. Derivation of novel human ground state naive pluripotent stem cells. *Nature*. 2013;504(7479):282–6. Epub 2013/11/01. <https://doi.org/10.1038/nature12745>. PubMed PMID: 24172903.
 78. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*. 2018;174(2):391–405 e19. Epub 2018/06/21. <https://doi.org/10.1016/j.cell.2018.05.043>. PubMed PMID: 29937225; PubMed Central PMCID: PMCPCMC6046266.
 79. 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(51):20569–74. Epub 2013/11/22. <https://doi.org/10.1073/pnas.1319061110>. PubMed PMID: 24259714; PubMed Central PMCID: PMC3870695.
 80. Han D, Liu G, Oh Y, Oh S, Yang S, Mandjikian L, et al. ZBTB12 is a molecular barrier to dedifferentiation in human pluripotent stem cells. *Nat Commun*. 2023;14(1):632. Epub 2023/02/09. <https://doi.org/10.1038/>

- s41467-023-36178-9. PubMed PMID: 36759523; PubMed Central PMCID: PMCPCMC9911396.
81. Takahashi K, Jeong D, Wang S, Narita M, Jin X, Iwasaki M, et al. Critical Roles of Translation Initiation and RNA Uridylation in Endogenous Retroviral Expression and Neural Differentiation in Pluripotent Stem Cells. *Cell Rep.* 2020;31(9): 107715. <https://doi.org/10.1016/j.celrep.2020.107715>. (PubMed PMID:32492424;PubMedCentralPMCID:PMC8195978).
 82. Luo S, Zhang H, Duan Y, Yao X, Clark AG, Lu J. The evolutionary arms race between transposable elements and piRNAs in *Drosophila melanogaster*. *BMC Evol Biol.* 2020;20(1):14. Epub 20200128. <https://doi.org/10.1186/s12862-020-1580-3>. PubMed PMID: 31992188; PubMed Central PMCID: PMCPCMC6988346.
 83. Takahashi K, Nakamura M, Okubo C, Kliesmete Z, Ohnuki M, Narita M, et al. The pluripotent stem cell-specific transcript ESRG is dispensable for human pluripotency. *PLoS Genet.* 2021;17(5):e1009587. Epub 2021/05/26. <https://doi.org/10.1371/journal.pgen.1009587>. PubMed PMID: 34033652; PubMed Central PMCID: PMCPCMC8184003 following competing interests: K.T. is on the scientific advisory board of l Peace, Inc. without salary, S.Y. is a scientific advisor (without salary) of iPS Academia Japan, and other authors have declared that no competing interests exist.
 84. Li S, Liu H, Liu W, Shi N, Zhao M, Wanggou S, et al. ESRG is critical to maintain the cell survival and self-renewal/pluripotency of hPSCs by collaborating with MCM2 to suppress p53 pathway. *Int J Biol Sci.* 2023;19(3):916–35. Epub 20230116. <https://doi.org/10.7150/ijbs.79095>. PubMed PMID: 36778110; PubMed Central PMCID: PMCPCMC9909993.
 85. Kapranov P, Cheng J, Dike S, Nix DA, Dutttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science.* 2007;316(5830):1484–8. Epub 20070517. <https://doi.org/10.1126/science.1138341>. PubMed PMID: 17510325.
 86. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature.* 2012;489(7414):101–8. <https://doi.org/10.1038/nature.11233>. (PubMed PMID:22955620;PubMedCentralPMCID:PMCPCMC3684276).
 87. Bulut-Karslioglu A, De La Rosa-Velazquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell.* 2014;55(2):277–90. Epub 20140626. <https://doi.org/10.1016/j.molcel.2014.05.029>. PubMed PMID: 24981170.
 88. Garland W, Muller I, Wu M, Schmid M, Imamura K, Rib L, et al. Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. *Mol Cell.* 2022;82(9):1691–707 e8. Epub 20220328. <https://doi.org/10.1016/j.molcel.2022.03.004>. PubMed PMID: 35349793; PubMed Central PMCID: PMCPCMC9433625.
 89. Clemot M, Molla-Herman A, Mathieu J, Huynh JR, Dostatni N. The replicative histone chaperone CAF1 is essential for the maintenance of identity and genome integrity in adult stem cells. *Development.* 2018;145(17). Epub 20180828. <https://doi.org/10.1242/dev.161190>. PubMed PMID: 30093554.
 90. Yang N, Kazazian HH, Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol.* 2006;13(9):763–71. Epub 20060827. <https://doi.org/10.1038/nsmb1141>. PubMed PMID: 16936727.
 91. Hamdorf M, Idica A, Zisoulis DG, Gamelin L, Martin C, Sanders KJ, et al. miR-128 represses L1 retrotransposition by binding directly to L1 RNA. *Nat Struct Mol Biol.* 2015;22(10):824–31. Epub 20150914. <https://doi.org/10.1038/nsmb.3090>. PubMed PMID: 26367248.
 92. Belair C, Sim S, Kim KY, Tanaka Y, Park IH, and, et al. The RNA exosome nuclease complex regulates human embryonic stem cell differentiation. *J Cell Biol.* 2019;218(8):2564–82. Epub 20190715. <https://doi.org/10.1083/jcb.201811148>. PubMed PMID: 31308215; PubMed Central PMCID: PMCPCMC6683745.
 93. Wu Y, Liu W, Chen J, Liu S, Wang M, Yang L, et al. Nuclear Exosome Targeting Complex Core Factor Zcchc8 Regulates the Degradation of LINE1 RNA in Early Embryos and Embryonic Stem Cells. *Cell Rep.* 2019;29(8):2461–72 e6. <https://doi.org/10.1016/j.celrep.2019.10.055>. PubMed PMID: 31747613.
 94. Yan R, Qi M, Zhang P, Shen B, Yin J, Chen C, et al. Core factor of NEXT complex, ZCCHC8, governs the silencing of LINE1 during spermatogenesis. *Natl Sci Rev.* 2025;12(1):nwae407. Epub 20241217. <https://doi.org/10.1093/nsr/nwae407>. PubMed PMID: 39758125; PubMed Central PMCID: PMCPCMC11697976.
 95. Lucas BA, Lavi E, Shiue L, Cho H, Katzman S, Miyoshi K, et al. Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay. *Proc Natl Acad Sci U S A.* 2018;115(5):968–73. Epub 20180116. <https://doi.org/10.1073/pnas.1715531115>. PubMed PMID: 29339519; PubMed Central PMCID: PMCPCMC5798355.
 96. Lu YY, Krebber H. Nuclear mRNA Quality Control and Cytoplasmic NMD Are Linked by the Guard Proteins Gbp2 and Hrb1. *Int J Mol Sci.* 2021;22(20). Epub 20211019. <https://doi.org/10.3390/ijms222011275>. PubMed PMID: 34681934; PubMed Central PMCID: PMCPCMC8541090.
 97. O'Carroll IP, Fan L, Kroupa T, McShane EK, Theodore C, Yates EA, et al. Structural Mimicry Drives HIV-1 Rev-Mediated HERV-K Expression. *J Mol Biol.* 2020;432(24):166711. Epub 20201114. <https://doi.org/10.1016/j.jmb.2020.11.010>. PubMed PMID: 33197463; PubMed Central PMCID: PMCPCMC7842262.
 98. Hu S, Li J, Xu F, Mei S, Le Duff Y, Yin L, et al. SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation. *PLoS Genet.* 2015;11(7):e1005367. Epub 20150702. <https://doi.org/10.1371/journal.pgen.1005367>. PubMed PMID: 26134849; PubMed Central PMCID: PMCPCMC4489885.
 99. Chen LL, Carmichael GG. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol Cell.* 2009;35(4):467–78. <https://doi.org/10.1016/j.molcel.2009.06.027>. (PubMed PMID:19716791;PubMedCentral PMCID:PMCPCMC2749223).
 100. Hayakawa S, Shiratori S, Yamato H, Kameyama T, Kitatsuji C, Kashigi F, et al. ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat Immunol.* 2011;12(1):37–44. Epub 20101121. <https://doi.org/10.1038/ni.1963>. PubMed PMID: 21102435.
 101. Yan J, Zhao Y, Du J, Wang Y, Wang S, Wang Q, et al. RNA sensor MDA5 suppresses LINE-1 retrotransposition by regulating the promoter activity of LINE-1 5'-UTR. *Mob DNA.* 2022;13(1):10. Epub 20220412. <https://doi.org/10.1186/s13100-022-00268-0>. PubMed PMID: 35414110; PubMed Central PMCID: PMCPCMC9003951.
 102. Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, et al. Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res.* 2004;14(9):1719–25. <https://doi.org/10.1101/gr.2855504>. (PubMed PMID:15342557;PubMedCentralPMCID:PMC515317).
 103. Groh S, Schotta G. Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci.* 2017;74(11):2055–65. Epub 2017/02/06. <https://doi.org/10.1007/s00018-017-2454-8>. PubMed PMID: 28160052.
 104. Amrani N, Sachs MS, Jacobson A. Early nonsense: mRNA decay solves a translational problem. *Nat Rev Mol Cell Biol.* 2006;7(6):415–25. <https://doi.org/10.1038/nrm1942>. (PubMed PMID: 16723977).
 105. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol.* 2011;12(4):246–58. <https://doi.org/10.1038/nrm3089>. (PubMed PMID: 21427766).
 106. Schlesinger S, Goff SP. Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol Cell Biol.* 2015;35(5):770–7. <https://doi.org/10.1128/MCB.01293-14>. (PubMed PMID:25547290;PubMedCentralPMCID:PMC4323490).
 107. Cosby RL, Chang NC, Feschotte C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* 2019;33(17–18):1098–116. <https://doi.org/10.1101/gad.327312.119>. (PubMed PMID:31481535;PubMedCentralPMCID:PMCPCMC6719617).
 108. Gerdes P, Richardson SR, Mager DL, Faulkner GJ. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol.* 2016;17:100. Epub 20160509. <https://doi.org/10.1186/s13059-016-0965-5>. PubMed PMID: 27161170; PubMed Central PMCID: PMCPCMC4862087.
 109. Muller I, Helin K. Keep quiet: the HUSH complex in transcriptional silencing and disease. *Nat Struct Mol Biol.* 2024;31(1):11–22. Epub 20240112. <https://doi.org/10.1038/s41594-023-01173-7>. PubMed PMID: 38216658.
 110. Goodier JL, Kazazian HH Jr. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell.* 2008;135(1):23–35. <https://doi.org/10.1016/j.cell.2008.09.022>. (PubMed PMID: 18854152).

111. Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet.* 2011;12(9):615–27. Epub 20110818. <https://doi.org/10.1038/nrg3030>. PubMed PMID: 21850042; PubMed Central PMCID: PMC3192332.
112. 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(9):1131–9. Epub 2013/08/13. <https://doi.org/10.1038/nsmb.2660>. PubMed PMID: 23934149.
113. Fu Y, Zhou Z, Wang H, Gong P, Guo R, Wang J, et al. IFITM1 suppresses expression of human endogenous retroviruses in human embryonic stem cells. *FEBS Open Bio.* 2017;7(8):1102–10. Epub 2017/08/07. <https://doi.org/10.1002/2211-5463.12246>. PubMed PMID: 28781951; PubMed Central PMCID: PMC5537067.
114. Harris RS, Dudley JP. APOBECs and virus restriction. *Virology.* 2015;479–480:131–45. Epub 20150326. <https://doi.org/10.1016/j.virol.2015.03.012>. PubMed PMID: 25818029; PubMed Central PMCID: PMC4424171.
115. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* 2002;16(8):919–32. <https://doi.org/10.1101/gad.973302>. (PubMed PMID: 11959841; PubMed Central PMCID: PMC152359).
116. Wolf D, Goff SP. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell.* 2007;131(1):46–57. <https://doi.org/10.1016/j.cell.2007.07.026>. (PubMed PMID: 17923087).
117. Zuo X, Sheng J, Lau HT, McDonald CM, Andrade M, Cullen DE, et al. Zinc finger protein ZFP57 requires its co-factor to recruit DNA methyltransferases and maintains DNA methylation imprint in embryonic stem cells via its transcriptional repression domain. *J Biol Chem.* 2012;287(3):2107–18. Epub 20111205. <https://doi.org/10.1074/jbc.M111.322644>. PubMed PMID: 22144682; PubMed Central PMCID: PMC3265890.
118. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* 2014;24(8):1260–70. Epub 20140530. <https://doi.org/10.1101/gr.172833.114>. PubMed PMID: 24879559; PubMed Central PMCID: PMC4120080.
119. Tao Y, Yen MR, Chitashvili T, Nakano H, Kim R, Hosohama L, et al. TRIM28-Regulated Transposon Repression Is Required for Human Germine Competency and Not Primed or Naïve Human Pluripotency. *Stem Cell Reports.* 2018;10(1):243–56. Epub 20171228. <https://doi.org/10.1016/j.stemcr.2017.11.020>. PubMed PMID: 29290627; PubMed Central PMCID: PMC5768987.
120. Imbeault M, Hellebood PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature.* 2017;543(7646):550–4. <https://doi.org/10.1038/nature21683>. (PubMed PMID: 28273063).
121. Bunch H, Calderwood SK. TRIM28 as a novel transcriptional elongation factor. *BMC Mol Biol.* 2015;16:14. Epub 20150821. <https://doi.org/10.1186/s12867-015-0040-x>. PubMed PMID: 26293668; PubMed Central PMCID: PMC4545989.
122. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, et al. Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 2014;28(13):1397–409. Epub 20140617. <https://doi.org/10.1101/gad.241661.114>. PubMed PMID: 24939876; PubMed Central PMCID: PMC4083085.
123. 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(7278):237–40. <https://doi.org/10.1038/nature08674>. (PubMed PMID: 20075919).
124. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell.* 2015;161(3):555–68. Epub 20150416. <https://doi.org/10.1016/j.cell.2015.03.017>. PubMed PMID: 25892221; PubMed Central PMCID: PMC4409934.
125. Quenneville S, Turelli P, Bojkowska K, Raclot C, Offner S, Kapopoulou A, et al. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep.* 2012;2(4):766–73. Epub 20121004. <https://doi.org/10.1016/j.celrep.2012.08.043>. PubMed PMID: 23041315; PubMed Central PMCID: PMC3677399.
126. 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(3):519–29. <https://doi.org/10.1242/dev.087585>. (PubMed PMID: 23293284; PubMed Central PMCID: PMC34892343).
127. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet.* 2013;45(7):836–41. Epub 20130526. <https://doi.org/10.1038/ng.2649>. PubMed PMID: 23708189; PubMed Central PMCID: PMC3695047.
128. Levine AJ, Ting DT, Greenbaum BD. p53 and the defenses against genome instability caused by transposons and repetitive elements. *Bioessays.* 2016;38(6):508–13. Epub 20160513. <https://doi.org/10.1002/bies.201600031>. PubMed PMID: 27172878; PubMed Central PMCID: PMC45031199.
129. Gartner A, Boag PR, Blackwell TK. Germline survival and apoptosis. *WormBook.* 2008:1–20. Epub 20080904. <https://doi.org/10.1895/wormbook.1.145.1>. PubMed PMID: 18781708; PubMed Central PMCID: PMC4781258.
130. Gartner A, Gartner A. Germ cell apoptosis and DNA damage responses. *Adv Exp Med Biol.* 2013;757:249–76. https://doi.org/10.1007/978-1-4614-4015-4_9. (PubMed PMID: 22872480).
131. 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(7259):1127–31. Epub 20090805. <https://doi.org/10.1038/nature08248>. PubMed PMID: 19657334; PubMed Central PMCID: PMC2909034.
132. Tiwari B, Jones AE, Caillet CJ, Das S, Royer SK, Abrams JM. p53 directly represses human LINE1 transposons. *Genes Dev.* 2020;34(21–22):1439–51. Epub 20201015. <https://doi.org/10.1101/gad.343186.120>. PubMed PMID: 33060137; PubMed Central PMCID: PMC7608743.
133. Wylie A, Jones AE, D'Brot A, Lu WJ, Kurtz P, Moran JV, et al. p53 genes function to restrain mobile elements. *Genes Dev.* 2016;30(1):64–77. Epub 20151223. <https://doi.org/10.1101/gad.266098.115>. PubMed PMID: 26701264; PubMed Central PMCID: PMC4701979.
134. Su D, Wang X, Campbell MR, Song L, Safi A, Crawford GE, et al. Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS Genet.* 2015;11(1):e1004885. Epub 20150108. <https://doi.org/10.1371/journal.pgen.1004885>. PubMed PMID: 25569532; PubMed Central PMCID: PMC4287438.
135. Chang NT, Yang WK, Huang HC, Yeh KW, Wu CW. The transcriptional activity of HERV-L LTR is negatively regulated by its cis-elements and wild type p53 tumor suppressor protein. *J Biomed Sci.* 2007;14(2):211–22. Epub 20061207. <https://doi.org/10.1007/s11373-006-9126-2>. PubMed PMID: 17151828.
136. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A.* 2007;104(47):18613–8. Epub 20071114. <https://doi.org/10.1073/pnas.0703637104>. PubMed PMID: 18003932; PubMed Central PMCID: PMC2141825.
137. Harris CR, Dewan A, Zupnick A, Normart R, Gabriel A, Prives C, et al. p53 responsive elements in human retrotransposons. *Oncogene.* 2009;28(44):3857–65. Epub 2009/09/01. <https://doi.org/10.1038/onc.2009.246>. PubMed PMID: 19718052; PubMed Central PMCID: PMC3193277.
138. Shi H, Wei J, He C. Where, When, and How: Context-Dependent Functions of RNA Methylation Writers, Readers, and Erasers. *Mol Cell.* 2019;74(4):640–50. <https://doi.org/10.1016/j.molcel.2019.04.025>. (PubMed PMID: 31100245; PubMed Central PMCID: PMC6527355).
139. Liu J, Gao M, He J, Wu K, Lin S, Jin L, et al. The RNA m(6)A reader YTHDC1 silences retrotransposons and guards ES cell identity. *Nature.* 2021;591(7849):322–6. Epub 20210303. <https://doi.org/10.1038/s41586-021-03313-9>. PubMed PMID: 33658714.
140. Jambhekar A, Dhall A, Shi Y. Roles and regulation of histone methylation in animal development. *Nat Rev Mol Cell Biol.* 2019;20(10):625–41. Epub 20190702. <https://doi.org/10.1038/s41580-019-0151-1>. PubMed PMID: 31267065; PubMed Central PMCID: PMC6774358.
141. Michalak EM, Burr ML, Bannister AJ, Dawson MA. The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat Rev Mol Cell Biol.*

- 2019;20(10):573–89. Epub 20190703. <https://doi.org/10.1038/s41580-019-0143-1>. PubMed PMID: 31270442.
142. Sun T, Xu Y, Xiang Y, Ou J, Soderblom EJ, Diao Y. Crosstalk between RNA m(6)A and DNA methylation regulates transposable element chromatin activation and cell fate in human pluripotent stem cells. *Nat Genet.* 2023;55(8):1324–35. Epub 20230720. <https://doi.org/10.1038/s41588-023-01452-5>. PubMed PMID: 37474847; PubMed Central PMCID: PMCPCMC10766344.
 143. Warkocki Z, Krawczyk PS, Adamska D, Bijata K, Garcia-Perez JL, Dziembowski A. Uridylation by TUT4/7 Restricts Retrotransposition of Human LINE-1s. *Cell.* 2018;174(6):1537–48 e29. Epub 20180816. <https://doi.org/10.1016/j.cell.2018.07.022>. PubMed PMID: 30122351; PubMed Central PMCID: PMCPCMC6191937.
 144. Mirzadeh Azad F, Taheri Bajgan E, Naeli P, Rudov A, Bagheri Moghadam M, Sadat Akhtar M, et al. Differential Expression Pattern of linc-ROR Spliced Variants in Pluripotent and Non-Pluripotent Cell Lines. *Cell J.* 2022;24(10):569–76. Epub 20221001. <https://doi.org/10.22074/cellj.2022.8205>. PubMed PMID: 36259474; PubMed Central PMCID: PMCPCMC9617025.
 145. Ilik IA, Glazar P, Tse K, Brandl B, Meierhofer D, Muller FJ, et al. Autonomous transposons tune their sequences to ensure somatic suppression. *Nature.* 2024;626(8001):1116–24. Epub 20240214. <https://doi.org/10.1038/s41586-024-07081-0>. PubMed PMID: 38355802; PubMed Central PMCID: PMCPCMC10901741.
 146. Jacobs FM, 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(7530):242–5. Epub 20140928. <https://doi.org/10.1038/nature13760>. PubMed PMID: 25274305; PubMed Central PMCID: PMCPCMC4268317.
 147. Chu C, Lin EW, Tran A, Jin H, Ho NI, Veit A, et al. The landscape of human SVA retrotransposons. *Nucleic Acids Res.* 2023;51(21):11453–65. <https://doi.org/10.1093/nar/gkad821>. (PubMed PMID:37823611; PubMed Central PMCID:PMCPCMC10681720).
 148. Vlaming H, Mimoso CA, Field AR, Martin BJE, Adelman K. Screening thousands of transcribed coding and non-coding regions reveals sequence determinants of RNA polymerase II elongation potential. *Nat Struct Mol Biol.* 2022;29(6):613–20. Epub 20220609. <https://doi.org/10.1038/s41594-022-00785-9>. PubMed PMID: 35681023; PubMed Central PMCID: PMCPCMC9672417.
 149. Koonin EV, Makarova KS, Wolf YI, Krupovic M. Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. *Nat Rev Genet.* 2020;21(2):119–31. Epub 2019/10/16. <https://doi.org/10.1038/s41576-019-0172-9>. PubMed PMID: 31611667.
 150. Arkhipova IR, Burns KH, Chiappinelli KB, Chuong EB, Goubert C, Guarne A, et al. Meeting report: transposable elements at the crossroads of evolution, health and disease 2023. *Mob DNA.* 2023;14(1):19. Epub 20231127. <https://doi.org/10.1186/s13100-023-00307-4>. PubMed PMID: 38012685; PubMed Central PMCID: PMCPCMC10680173.
 151. Roy SW. The origin of recent introns: transposons? *Genome Biol.* 2004;5(12):251. Epub 20041129. <https://doi.org/10.1186/gb-2004-5-12-251>. PubMed PMID: 15575977; PubMed Central PMCID: PMCPCMC545789.
 152. Mikina W, Halakuc P, Milanowski R. Transposon-derived introns as an element shaping the structure of eukaryotic genomes. *Mob DNA.* 2024;15(1):15. Epub 20240727. <https://doi.org/10.1186/s13100-024-00325-w>. PubMed PMID: 39068498; PubMed Central PMCID: PMCPCMC11282704.
 153. Palazzo AF, Lee ES. Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. *Front Genet.* 2018;9:440. Epub 20181017. <https://doi.org/10.3389/fgene.2018.00440>. PubMed PMID: 30386371; PubMed Central PMCID: PMCPCMC6199362.
 154. Wylie A, Jones AE, Abrams JM. p53 in the game of transposons. *Bioessays.* 2016;38(11):1111–6. Epub 20160919. <https://doi.org/10.1002/bies.201600115>. PubMed PMID: 27644006; PubMed Central PMCID: PMCPCMC5142446.
 155. Robbez-Masson L, Tie CHC, Conde L, Tunbak H, Husovsky C, Tchasovnikarova IA, et al. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* 2018;28(6):836–45. Epub 20180504. <https://doi.org/10.1101/gr.228171.117>. PubMed PMID: 29728366; PubMed Central PMCID: PMCPCMC5991525.
 156. Mordstein C, Cano L, Morales AC, Young B, Ho AT, Rice AM, et al. Transcription, mRNA Export, and Immune Evasion Shape the Codon Usage of Viruses. *Genome Biol Evol.* 2021;13(9). <https://doi.org/10.1093/gbe/evab106>. PubMed PMID: 33988683; PubMed Central PMCID: PMCPCMC8410142.
 157. Manzo G. Similarities Between Embryo Development and Cancer Process Suggest New Strategies for Research and Therapy of Tumors: A New Point of View. *Front Cell Dev Biol.* 2019;7:20. Epub 20190307. <https://doi.org/10.3389/fcell.2019.00020>. PubMed PMID: 30899759; PubMed Central PMCID: PMCPCMC6416183.
 158. Borah A, Raveendran S, Rochani A, Maekawa T, Kumar DS. Targeting self-renewal pathways in cancer stem cells: clinical implications for cancer therapy. *Oncogenesis.* 2015;4(11):e177. Epub 20151130. <https://doi.org/10.1038/oncsis.2015.35>. PubMed PMID: 26619402; PubMed Central PMCID: PMCPCMC4670961.
 159. Attig J, Pape J, Doglio L, Kazachenka A, Ottina E, Young GR, et al. Human endogenous retrovirus onco-exaptation counters cancer cell senescence through calbindin. *J Clin Invest.* 2023;133(14). Epub 20230717. <https://doi.org/10.1172/JCI164397>. PubMed PMID: 37192000; PubMed Central PMCID: PMCPCMC10348765.
 160. Singh M, Widmann TJ, Bansal V, Cortes JL, Schumann GG, Wunderlich S, et al. The selection arena in early human blastocysts resolves the pluripotent inner cell mass. *bioRxiv.* 2019. <https://doi.org/10.1101/318329>.
 161. Babaian A, Romanish MT, Gagnier L, Kuo LY, Karimi MM, Steidl C, et al. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene.* 2016;35(19):2542–6. Epub 20150817. <https://doi.org/10.1038/onc.2015.308>. PubMed PMID: 26279299.
 162. Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet.* 2019;51(4):611–7. Epub 20190329. <https://doi.org/10.1038/s41588-019-0373-3>. PubMed PMID: 30926969; PubMed Central PMCID: PMCPCMC6443099.
 163. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med.* 2010;16(5):571–9, 1p following 9. Epub 20100502. <https://doi.org/10.1038/nm.2129>. PubMed PMID: 20436485.
 164. Lynch-Sutherland CF, Chatterjee A, Stockwell PA, Eccles MR, Macaulay EC. Reawakening the Developmental Origins of Cancer Through Transposable Elements. *Front Oncol.* 2020;10:468. Epub 20200505. <https://doi.org/10.3389/fonc.2020.00468>. PubMed PMID: 32432029; PubMed Central PMCID: PMCPCMC7214541.
 165. Yu C, Lei X, Chen F, Mao S, Lv L, Liu H, et al. ARID1A loss derepresses a group of human endogenous retrovirus-H loci to modulate BRD4-dependent transcription. *Nat Commun.* 2022;13(1):3501. Epub 20220617. <https://doi.org/10.1038/s41467-022-31197-4>. PubMed PMID: 35715442; PubMed Central PMCID: PMCPCMC9205910.
 166. Singh M, Cai H, Bunse M, Feschotte C, Izsvák Z. Human Endogenous Retrovirus K Rec forms a Regulatory Loop with MITF that Opposes the Progression of Melanoma to an Invasive Stage. *Viruses.* 2020;12(11). Epub 20201113. <https://doi.org/10.3390/v12111303>. PubMed PMID: 33202765; PubMed Central PMCID: PMCPCMC7696977.
 167. Insko ML, Abraham BJ, Dubbury SJ, Kalthheuner IH, Dust S, Wu C, et al. Oncogenic CDK13 mutations impede nuclear RNA surveillance. *Science.* 2023;380(6642):eabn7625. Epub 20230421. <https://doi.org/10.1126/science.abn7625>. PubMed PMID: 37079685; PubMed Central PMCID: PMCPCMC10184553.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.