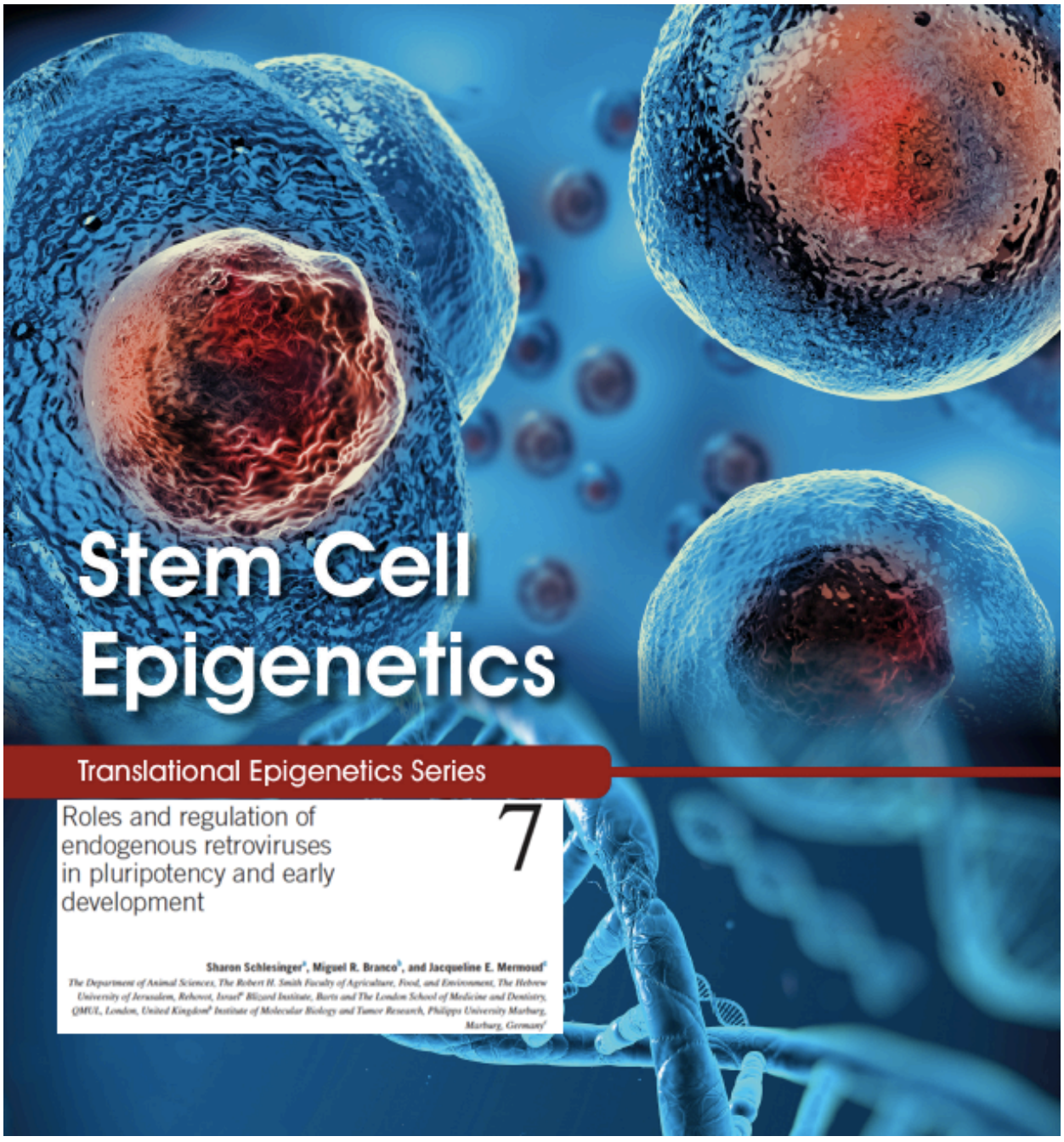
Stem Cell Epigenetics
Translational Epigenetics Series
Roles and regulation of endogenous retroviruses in pluripotency and early development
7
Sharon Schlesinger[a], Miguel R. Branco[b], and Jacqueline E. Mermoud[c]
The Department of Animal Sciences, The Robert H. Smith Faculty of Agriculture, Food, and Environment, The Hebrew University of Jerusalem, Rehovot, Israel[a] Blizard Institute, Barts and The London School of Medicine and Dentistry, QMUL, London, United Kingdom[b] Institute of Molecular Biology and Tumor Research, Philipps University Marburg, Marburg, Germany[c]

CHAPTER

7

Roles and regulation of endogenous retroviruses in pluripotency and early development

Sharon Schlesinger[a], Miguel R. Branco[b], and Jacqueline E. Mermoud[c]

The Department of Animal Sciences, The Robert H. Smith Faculty of Agriculture, Food, and Environment, The Hebrew University of Jerusalem, Rehovot, Israel[a] Blizard Institute, Barts and The London School of Medicine and Dentistry, QMUL, London, United Kingdom[b] Institute of Molecular Biology and Tumor Research, Philipps University Marburg, Marburg, Germany[c]

Introduction

Endogenous retroviruses (ERVs) are the remnants of ancient infections of the germline by retroviruses. Accordingly, ERV sequences have become a rather significant part, about 10%, of the human and mouse genome; they vary greatly in sequence, length, and function and collectively belong to the larger family of transposable elements (TEs).

TEs are defined by (1) their repetitiveness—there is always more than one copy of any given TE in the genome (and usually the numbers are higher, 10–100 copies), and this is due to (2) their potential mobility. TEs are mobile elements that can replicate themselves in a given genome either via a copy-and-paste mechanism through an RNA intermediate (called retrotransposons or class I TEs) or via a cut-and-paste mechanism (called DNA transposons or class II TEs) (Fig. 7.1). Retrotransposons are usually divided into two subclasses, LTR (long terminal repeat) retroelements, populated mostly by ERVs, and non-LTR elements. The second group comprises short and long interspersed nuclear elements (SINEs and LINEs, respectively), which together make up most of the mammalian TE genomic volume [1, 2]. Unlike ERVs, non-LTR retrotransposons do not encode viral genes. While LINE elements replicate autonomously, using their own reverse transcriptase (RT) gene, other members of this subclass utilize LINE enzymes for retrotransposition [3].

TEs were discovered due to their ability to affect phenotypes in a seemingly "unstable" or "mutable" manner [4]. McClintock referred to TEs as "controlling units" or "controlling elements," because in her view the ability of TEs to regulate organismal traits was in many ways more relevant than their mobility [5]. A similarly visionary concept was introduced by Britten and Davidson in the form of a gene regulation model where repetitive elements could be used to establish regulatory networks [6]. As molecular biology caught up, examples started to emerge of TEs that controlled the expression of

Stem Cell Epigenetics. https://doi.org/10.1016/B978-0-12-814085-7.00007-6

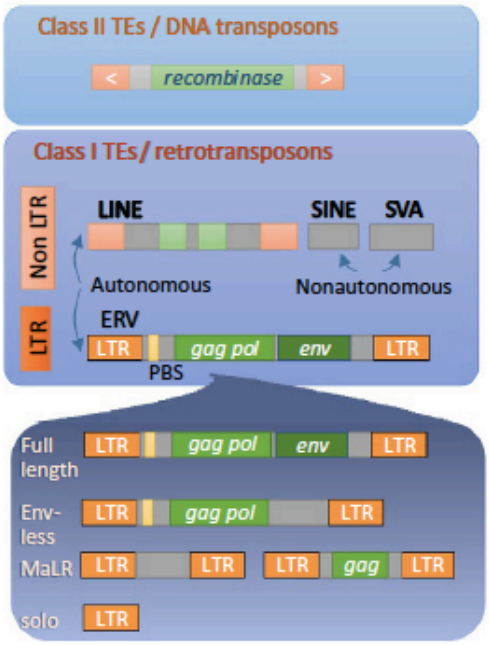

FIG. 7.1

Transposable elements and ERV structure. TE in the mammalian genome include DNA transposons and retrotransposons. Retrotransposons that code for most enzymes necessary for mobility are called autonomous (ERVs and LINEs). Nonautonomous elements, such as SINEs and SVAs, are dependent on LINEs for their mobility. Green boxes demarcate coding regions, and orange boxes mark terminal repeats (inverted or direct). Various ERV structures are depicted in the inset: full length ERVs are essentially identical to the integrated proviruses; they contain two LTRs, a primer binding site (PBS) and a full sequence of the coding genes (*gag*, *pol*, and *env*). "*Env* less" ERVs are elements with a deletion in the *env* gene. MaLR are nonautonomous ERVs since they lack the pol gene. Solo LTRs are single LTRs generated by homologous recombination between the two LTRs of a complete ERV.

nearby genes [7], but an appreciation of the extent by which TEs act as *cis*-regulatory elements has only been possible since the advent of high-throughput sequencing.

In this chapter, we will focus on the unique regulation of ERVs in pluripotency and early development, which is of particular interest for several reasons: (1) The potential damage to genomic integrity of the individual and his offspring that could arise from uncontrolled duplications of those transposition-proficient ERVs. This dictates strict regulation by the host, namely, the germ and stem cells [8]. (2) The "window of opportunity" opening after the wave of demethylation and during epigenetic reprograming that follows, all the way from fertilization to the establishment of the epiblast within the inner cell mass (captured in vitro in the form of self-renewing embryonic stem cells (ESC) [9]. Hence, it is reasonable to assume that some ERVs will be able to exploit the global demethylation for unrestricted transcription. (3) The hypothesis that TEs represent a major evolutionary force contributing to creation of species-specific regulation networks (due to the separate evolutionary phylogenetic line) [10, 11]. This suggests that variances in early development regulation networks (and not the protein-coding genes per se) are a major source of differences between species.

After a brief introduction into the structure and function of ERVs, we discuss the potential roles of ERV activation in human and mouse pluripotent cells and present examples of "domestication" events of ERVs that benefit the host. In the second part, we focus on the mechanisms that minimize the impact of ERVs on host fitness and review the silencing strategies that restrict ERV activity in pluripotency. In addition, we touch on the "arms race" between the "selfish" ERVs and the host, namely, the pluripotent cells. We will finish with conjectures of the influence ERV regulation has on the unique epigenetic features of ESCs, rising some questions yet to be addresses.

The structure and function of ERVs

ERVs are fossils of ancient exogenous retroviral integration events, providing an intriguing connection between the biology of rapidly evolving exogenous retroviruses and mammalian evolution. However, most ERVs lose their ENV gene and the ability to reinfect cells in the germline. Consequently, some ERV subfamilies consist of very few copies [12, 13]. Other ERVs undergo expansions in copy number by acquiring the ability to replicate in the germ cell via copy-and-paste mechanism—effectively becoming LTR retrotransposons [14]. Hence, "ERV" refers to a phylogenetic origin from the family *Retroviridae*, and "LTR retrotransposon" suggests adaptation to an intracellular niche, which could also apply to ERVs. However, since we mostly follow the story of ERV retrotransposition by looking at genomic data, we see only transposition events that are fixed in the germline. Interestingly, we are still in the dark as to questions like the following: What is the frequency of retrotransposition events in a normal somatic tissue and in the germline? Does the environment have an effect on the retrotransposition frequency? And how could that be related to human diseases?

As a rule, all ERVs resemble the products of retroviral integration reactions, with two LTRs and the viral genes, gag, pol, and env (Fig. 7.1) [15]. However, endogenous proviruses are subject to random mutations, constraints, and adaptations, which will render them nonpathogenic and with poor replication capacity overtime. Frequently the coding genes are completely deleted through recombination between 5′ and 3′ LTRs, leaving behind a "solo LTR." Altogether, LTR retroelements represent approximately 8% of the human genome, 1.6% thereof comprise "solo LTRs" (~200,000 elements), 3% consist of more or less complete proviral elements (~30,000 elements), and 3.6% of MaLR elements [16]. In mouse the numbers are mostly similar, but the ratio of the different classes is significantly different—ERV II are highly overrepresented in the mouse genome, while in human and other primates, we find more ERV III repeats (Fig. 7.2, see Box 7.1 for detailed description of the classification to classes and groups). In most cases the full length ERVs carry deletions or point mutations in one or more of their genes, thus left incapable of yielding infectious virus [17]. Moreover, most ERVs are transcriptionally silenced by host defense mechanisms evolved in million years of coevolution [12]. However, in some organisms, some ERVs are known to still be actively replicating and thus considered as not fully "endogenized" or fixed yet. Such "young" ERV is a challenge to the classification system due to their high variability between organisms of the same species. In humans, HERVK-HML2 might answer to this definition, although HERV-H and HERV-W were also found polymorphic in different ethnical groups [13, 18].

Although suggestion for improved nomenclature was recently made [18a], we will use here the accepted names as used in recent publications (Fig. 7.2) or as suggested by the Wicker nomenclature [2]. The basic ERV unit is the ancestral copy or the consensus sequence, which is a representative of the entire subgroup [1, 19]. For more information about ERV classification, see Box 7.1.

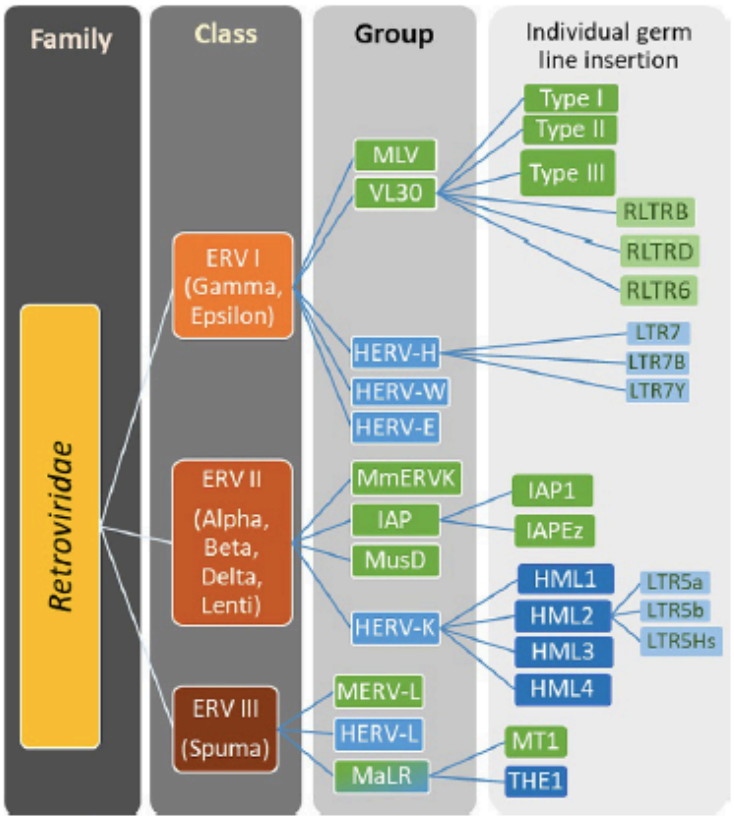

FIG. 7.2

Classification of ERVs. (Family) ERV originated from integration of members of the Retroviridae family into the germline. (Class) Separation into three classes is based on sequences homology to the retroviral "pol" gene. MaLR elements, although lacking the "*pol*" gene, are considered "Class III" due to a slight homology in the "*gag*" sequences. (Groups) Groups are historically named after the organism in which they were first identified (e.g., mouse MuERV, in *green boxes*, or human HERV in *blue boxes*). HERVs are classified on the basis of the tRNA that binds to the viral PBS to prime reverse transcription (e.g., HERV-K, H, and L). Groups can be further divided into subgroups based on differences in their consensus sequence. For example, most IAPs (intracisternal A particle) are lacking the "env" gene, but the IAPE subgroups contain it. Moreover, subgroups can be clustered further to parallel evolutionary trajectories by examining the flanking LTRs and deducing the integration time and the transcriptional pattern of the sequences. In the last level, we get to a single event of germline insertion of a retrovirus that might be presented in several genomic copies with similar consensus sequence due to retrotransposition of the parental ERV.

The "domestication" of ERVs for the benefit of the host is termed cooption or exaptation [20]. The best characterized examples are arguably those where ERVs have become incorporated into coding sequences [21], where evidence for purifying selection can often be found. This includes the RAG genes, which mediate V(D)J recombination in adaptive immunity [22, 23], and Syncytins, which promote cell-cell fusion in the placenta [24], are essential for mouse development, and have been independently derived from different ERVs in different lineages [25]. However, there is also a growing list of cases where ERVs have driven phenotypic diversity and adaptation through their noncoding action, namely,

Box 7.1 ERV classification and age

ERV lineages originate with infection of an ancestor species by an ancient retrovirus, which causes a "germline colonization" event in which a retroviral provirus is integrated into the genome [16a]. Phylogenetic studies looking at the "fossil records" found in the mammalian genome have shown that those ERV sequences are derived from few initial founder events. Due to the accumulation of genomic data, new information of ERV lineages and origin is accumulating, and the original nomenclature used is inadequate. Originally, ERVs were named after either (1) the most closely related exogenous retrovirus (roughly divided to three classes, see Fig. 7.1) or (2) the species in which the ERV was initially identified (3) on the basis of the tRNA that binds to the viral primer binding site (PBS) to prime reverse transcription. Hence, HERV-K is used for all ERVs found in human that use a lysine tRNA, no matter their relationship to one another. For genomic studies, ERVs have been classified by the Repbase database into approximately 80 groups through a systematic cataloguing of all repeats. In this database, groups are represented by a consensus element, which should reflect the sequence of the ancestral active element. However, approximately 10 times more abundant than full length ERVs are the "solo LTRs." Due to high mutation load, they are difficult to identify and are usually found by their sequence similarity to the LTRs of complete ERV or virus. In the Repbase database, those LTRs get unrelated numbers, which increase confusion by naming one subgroup as Class I-ERV9-LTR12 or Class II-HERVK-HML2-LTR5a/b/Hs (Fig. 7.2). Interestingly, since the LTRs are the least conserved part of the ERV, the evolutionary age of the subgroup will be reflected in the identity of the flanking LTRs more than in the relatively conserved internal sequence. The LTRs also dictate the transcriptional regulation of the ERV (by TF binding sites) and the epigenetic landscape of the surrounding DNA region, as will be described in depth later on.

acting as regulators of gene expression in *cis* [26, 27]. For example, TEs have enabled peppered moths to adapt to the industrial revolution [28]; produce the wrinkled-seed pea phenotype described by Mendel [29]; and control anthocyanin production in blood oranges [30], grapes [31], apples [22, 23], and chilies [32]. LTRs also play key functional roles in mammalian gene regulation, namely, driving expression of Dicer in mouse oocytes [33] and of AIM2 in innate immunity [34].

ERV activation in pluripotency

Successful ERVs exploit particular timings in development that favor their incorporation into the germline. Apart from germline development, preimplantation development also offers a reasonable chance for new insertions to be mitotically inherited by the germline. Indeed, de novo TE insertions occur during both of these developmental contexts in mice [35, 36], and there is also evidence in humans that heritable retrotransposition can occur in early development [37]. The expression and genomic expansion of TEs appear therefore to be linked to cellular states of increased developmental potency. One aspect that arguably makes such states amenable to ERV expression is the genome-wide epigenetic reprogramming that occurs during both preimplantation and germline development [38]. Yet, epigenetic alterations are not sufficient to drive strong expression of ERVs. Evolution has favored endogenization of ERVs that harbor within their LTR binding sites for transcription factors expressed in pluripotent cells [39]. Accordingly, differences in the exact timing of expression of different ERV families throughout preimplantation development suggest an association with stage-specific TFs [40, 41] (Fig. 7.3). A relationship between differential ERV activation and the continuum of developmental potency can also be appreciated by manipulating mESC culture conditions [42] (Box 7.2).

Does ERV activation in early development happen solely to serve the selfish needs of ERVs? Or could there be benefits to the host organism? Notably, many ERV families that are no longer mobile

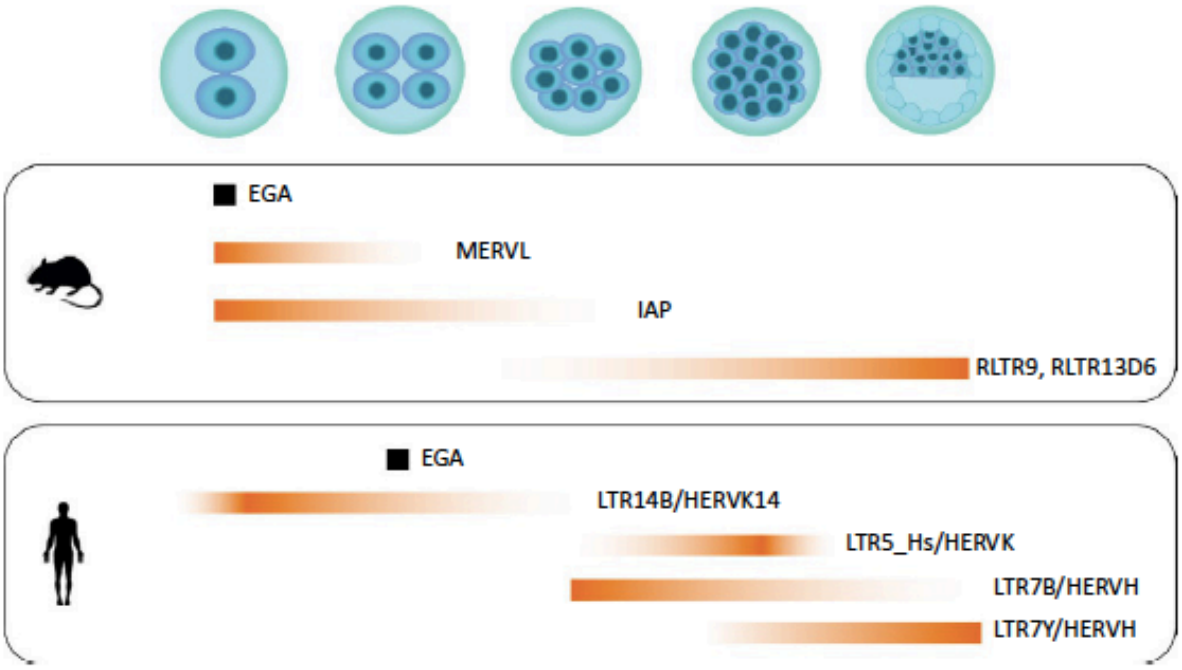

FIG. 7.3

Activation dynamics of ERVs in early embryonic development. Schematic representation of preimplantation development stages shown on the top panel: 2 cells, 4 cells, 8 cells, morula, and blastocyst. ERVs are first transcribed or epigenetically activated during embryonic genome activation (EGA) in mouse two-cell and human four-cell embryos. Later, distinct families of ERVs are activated at each stage of preimplantation development.

continue to become activated in some form, either transcriptionally or epigenetically [43]. It is possible that this activation is inconsequential to the organism, that is, neutral in its effect. Alternatively the host genome could benefit from some ERVs, such that their activation is not so much a failure to silence them but a process that harnesses a subset of elements within a family to increase host fitness [26].

In this section, we discuss the potential impact of ERV activation in pluripotent cells, either during preimplantation development or in ESCs. On one hand, activation of retrotranspositionally competent ERVs is potentially deleterious. On the other hand, there may be benefits to ERV activation, where they could play roles as gene promoters, enhancers, long noncoding RNAs (lncRNAs), and chromatin domain boundaries (Fig. 7.4). The main challenge here is to distinguish those cases where ERVs impact host fitness, from those where their action is neutral despite biochemical evidence suggesting a putative role.

ERV activity during germline maturation

The parental genome undergoes genome-wide loss of DNA methylation, which is reestablished later (before birth in male and after birth in female) as sex-specific germ cell methylation patterns, including methylation marks at imprinted loci [44]. In the oocyte, transcription dictates de novo methylation

Box 7.2 ERVs in ground-state ESC

In recent years, ESC culture conditions presumably mimicking the "ground-state" conditions of the ICM were developed and examined. These cells, usually termed "naïve" ESC in human or "2i" in mESC, possess epigenetic features that are somewhat distinct from the "primed" or the serum-grown ESCs. Therefore it might not come as a surprise that changes in the "transposcriptome" follows and that TEs expression signature could be used as a defining criteria for the developmental stage of the cells in mouse [42] and human [133].

Mouse "2i" ESC—Hackett et al. used different combinations of components in the ESC growth media to simulate distinct phases of pluripotency [42]. Out of 64 pluripotency related differentially expressed TE families, 80% are LTR repeats. For example, while ERV I-LTR45 and ERV II-IAPEy (relative of HERVK elements) are highly expressed in the "2i" ESC, ERV I-LTR9 are upregulated when 2i are depleted. This growth medium-related shift is accompanied by reduction of global DNA methylation [42a], from which many ERVs are protected by TRIM28 binding [107]. Repression of other ERV subsets is maintained by the chromatin modifier SETDB1, which plays a pivotal role in the silencing of IAP elements in naïve ESC [116]. Another elegant study following ERVs expression and DNA methylation of ERVs days after growth media switch to "2i" or "2i + Vitamine C" revealed that not all demethylated ERVs are transcriptionally activated, since alternative silencing mechanism are taking over [108]. Hence, ERVs in naïve ESCs show three typical configurations: Some are transcribed, some are silenced and maintain DNA methylation, and some are silenced by chromatin modifications, without DNA methylation.

Human "naïve" ESC—the culture conditions for naïve hESC are less established, and there is a controversy about/over the identity of HERVs upregulated. Screens designed to identify naïve-specific HERVs came up with conflicting results: either HERVH and its LTR7 [65, 77] or HERVK (HML2) with LTR5Hs were upregulated in naïve ESC [57, 133]. The upregulation was accompanied by DNA demethylation on the LTRs and binding of key pluripotency factors. A third model suggests that both LTR5Hs and LTR7 are upregulated in naïve ESC and repressed in later developmental stages due to epigenetic silencing induced by specific ZFP binding [64]. In all cases HERVs were shown to act as promoters [64, 65, 67], suggesting a functional role for HERVs expression in the regulation of the pluripotency network. HERVK (HML2) and HERVH are among the most recent entrants to the human genome and are primate specific; therefore their function as species-specific regulatory sequences that control such a critical stage is intriguing.

In conclusion the expression signature of ERVs and TEs is dependent of the culture conditions, as it is on the embryonic day and developmental stage. Yet, no major wave of ERVs activation is seen in ground-state ESC, probably due to the complexity, flexibility, and compensation of the cellular regulation network on the endovirome.

by the deposition of H3K36me3 modification [45]. Members of the mouse transcript (MT) family of class III ERVs are particularly active in oocytes, and hundreds of MT LTRs have been coopted as oocyte-specific gene promoters [46]. As a result, active ERVs become signposts and are the first to be fully methylated and silenced following de novo methylation onset, shaping the oocyte-specific methylation pattern. In humans the MT group does not exist, but LTR7 (HERVH), LTR12 (ERV I), and THE1 (ERV III-MaLR group) seem to have similar expression pattern [46], suggesting a possible reason for the distinct methylation patterns of oocyte in mice and man [47].

ERV activity upon embryonic genome activation (EGA)

In the first stage of embryonic development, the paternal and maternal germ cells are fused to create the zygote. At this stage, they carry distinct methylation patterns on imprinted genes and some ERVs. Once EGA takes place (2 cells in the mouse and 4–8 cells in human), a set of genes (named "2C" genes in mouse) and ERVs (MERVL and HERVL in mouse and human, respectively) become upregulated (Fig. 7.3). Many of the murine EGA genes reside near MERVL LTRs, setting an example for how ERVs

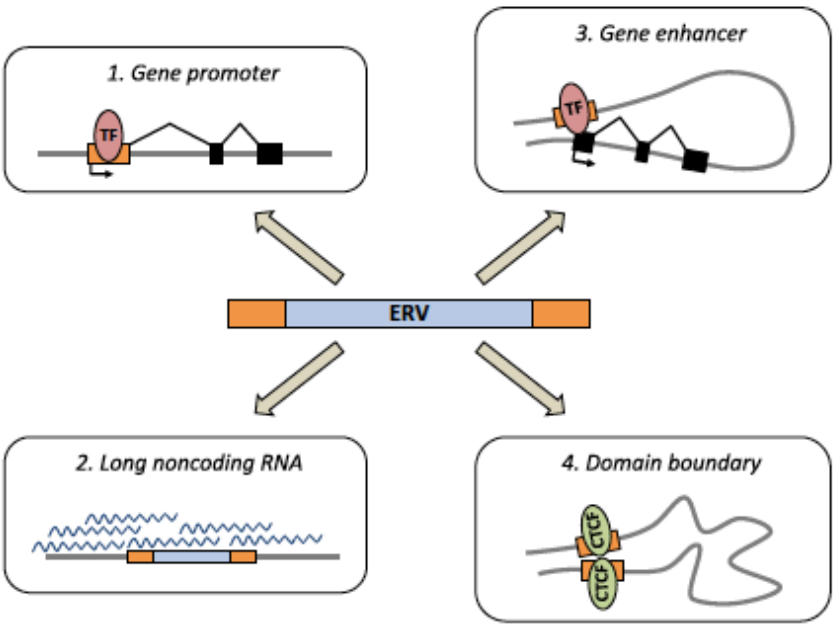

FIG. 7.4

ERVs function as regulatory elements in pluripotent cells. Examples for different types of regulatory activities ERVs exhibit in the host genome. These include effects mediated by *cis*-regulatory LTR elements functioning as promoter to affect gene expression (Panel 1) and *trans* effects mediated by ERV produced regulatory noncoding RNAs (Panel 2). Some LTRs include transcription factor binding sites, which might contribute to their function as enhancers (Panel 3). Last, LTRs might contain CTCF binding sites and thus alter genome topology. ERVs are shown as thick lines, with orange boxes for the LTRs, and boxes with arrow denote the gene promoter (Panel 4).

could have rewired the totipotency network [48, 49]. The transcription factor DUX or DUX4 was shown to play a role in activating MERVL and HERVL, respectively, and their neighboring 2C genes [50–52]. However, DUX depletion does not prevent EGA and normal embryonic development [53], thus putting a question mark on the functional role of DUX-induced MERVL activation in regulating EGA. The transcription of MERVL elements is maintained in a subset of cells (~1%) from the inner cell mass (ICM) of the blastocyst [49], and this "2CLC" state was recently used to map the regulatory landscape of the "2c" stage [54]. The study showed that MERVL activation is the result, and not the cause, of the extensive transcriptional reprogramming that leads to the emergence of the "2c" state. The totipotent nature of those cells and their lineage specification following differentiation onset are still a subject of extensive research.

In humans, several works have shown the differential expression of various HERVs from the 2-cell stage to the blastocyst: HERV-K (flanking with LTR5_Hs) elements are highly expressed at the morula stage, whereas HERV-H elements reach their peak level in embryonic stem cells [41, 55–57] (Fig. 7.3). However, we still lack a systematic analysis of ERV subgroup expression levels, their regulatory effects on transcription of developmental genes *and lineage specification that have not yet been elucidated*, and their contribution to cellular heterogeneity that should be examined.

Functions of ERVs as promoters and enhancers

TEs contribute an astonishing 6%–30% of all transcriptional start sites in human and mouse tissues [58], creating a large potential for the generation of alternative gene promoters. TE mobility has enabled the spreading of TF binding motifs contained within TEs, and as many as 40% of genome-wide TF binding sites are contained within TEs [59]. Elements bound by TFs often bear chromatin signatures of active promoters or enhancers in a tissue-specific fashion [43, 59]. Notably, ERVs (and more specifically, LTRs) make a disproportionate contribution to these putative regulatory elements.

Mouse ERVs

In mouse ESCs, several ERV families harbor motifs for, and are bound by, pluripotency-associated TFs such as OCT4, SOX2, and NANOG [39, 60–62]. For example, many (a subset of) RLTR13D6 and RLTR9 elements (which exist predominantly as solo LTRs) display all of the hallmarks of cell-specific enhancers, including open chromatin, H3K27ac and H3K4me1 deposition, and the ability to enhance gene expression within a reporter plasmid [61, 62]. Yet, out of a total of around 6600 RLTR13D6/RLTR9 elements, only ~10% bear a strong enhancer signature, and it remains unclear what other sequence or genomic environment features define the chromatin status of putative regulatory ERVs [62]. While it is compelling that enhancer-like RLTR13D6 and RLTR9 elements tend to be associated with genes that are preferentially expressed in ESCs, this does not necessarily imply a causal link. CRISPR-based approaches now provide an easy and flexible platform to test such relationships from a functional perspective. CRISPR can be used to genetically excise individual ERVs and evaluate the impact on gene expression. Additionally, several ERV copies can be epigenetically silenced simultaneously using CRISPR interference (CRISPRi), where a KRAB repressive domain is fused to catalytically dead Cas9 (dCas9) and used in conjunction with guide RNAs that bind to multiple elements. When used to silence enhancer-like RLTR13D6 elements, CRISPRi has revealed that only a handful of elements significantly impact on gene expression [62]. Other ERV families may play larger roles (discussed in the succeeding text), and enhancer redundancy may mask some regulatory activities. Nonetheless, experiments such as these demonstrate the need for a careful assessment of enhancer function, especially when dealing with sequences whose regulatory activity was originally aimed at driving the expression of selfish elements.

Human ERVs

Due to the species-specific nature of ERVs, none of the families implicated as enhancers in mouse ESCs are present in humans, and it has been argued that TEs have contributed to species-specific wiring of the pluripotency network [39]. In human ESCs, binding of pluripotency factors is also prominent at ERVs, but at distinct families, including LTR7 and LTR5_Hs [39, 63, 64]. These ERVs become activated during preimplantation development and in naïve ESCs, at the transcriptional and/or chromatin levels [57, 64, 65]. LTR7 elements are hominoid-specific LTRs to HERVH, and these will be discussed in more detail in the next section on lncRNAs, although they also act as *cis*-regulatory elements [65, 66]. LTR5_Hs are human-specific elements that are coupled to HERVK. LTR5_Hs elements bear motifs for, and are bound by, OCT4 and KLF4 in hESCs [57, 64]. They are also enriched for active histone marks (in both preimplantation embryos and hESCs) and binding of the enhancer-associated factor p300 [57, 64]. This enhancer-like signature is consistent with the fact that few elements appear to act as alternative genic promoters [57], as well as the detection of long-range interactions involving LTR5_Hs elements [64]. CRISPRi experiments in naïve hESCs and a teratocarcinoma cell line have shown that

epigenetic editing of LTR5_Hs elements substantially impacts on gene expression, with 87 genes being affected in both cell lines [64, 67]. These observations contrast with those discussed in the preceding text for RLTR13D6 elements in mESCs, where very few genes are deregulated upon CRISPRi. The difference does not appear to be due to indirect effects of LTR5_Hs silencing, suggesting that the contribution of ERVs to pluripotency transcriptional programs depends on the ERV family and its evolution within a species. It is nonetheless surprising that young elements such as LTR5_Hs can play such a large and species-specific role in gene regulation. It will be interesting to test whether such transcriptional roles are important for cell lineage establishment during human embryonic development.

Long noncoding RNAs

LncRNAs control various aspects of development, including X inactivation, gene imprinting, homeotic gene expression, and cellular differentiation. LncRNAs can act as regulators of gene expression through different mechanisms: by inducing transcription-coupled epigenetic changes, recruiting proteins to the vicinity of their transcriptional site (i.e., in *cis*), or acting in *trans* away from their transcriptional locus through various mechanisms [68, 69]. The reduced constraints on lncRNA evolution (when compared with protein-coding genes) make them more tolerant of TE insertions, which may in turn drive species-specific lncRNA evolution. Indeed, around three quarters of human lncRNAs contain an exon derived from TEs, which is in sharp contrast with the low frequency of TE exonization events observed in protein-coding genes [70, 71]. TEs (and ERVs/LTRs in particular) also often act as lineage-specific lncRNA promoters [70, 71].

At least 20 lncRNAs have been implicated in the maintenance of pluripotency in mESCs [72]. In humans, early studies had also identified a number of lncRNAs involved in pluripotency maintenance and reprogramming, including *lncRNA-ES3* and *lncRNA-RoR* [73, 74]. These two lncRNAs are partly derived from LTR7/HERVH elements, which have made a major contribution to the evolution of ESC-expressed lncRNAs [71] and which we will now further describe.

LTR7/HERVH

HERVH is the most abundant ERV in the human genome, whose expression and contribution to pluripotency-associated lncRNAs are conserved from human to gorilla [75, 76]. Expression of LTR7 (and specifically, the LTR7B and LTR7Y subfamilies) is high during preimplantation development, starting at the 8-cell stage and being confined to the epiblast portion of the blastocyst [41]. LTR7/HERVH expression is also elevated in hESCs and hiPSCs when compared with differentiated cells and is particularly associated with a state of naïve pluripotency [41, 65, 71]. Over 200 HERVH insertions are highly expressed in hESCs, around half of which are in a proviral configuration (i.e., are flanked by LTR7 elements) and tend to be evolutionarily younger [66, 77]. Expression of HERVH is driven by binding of NANOG and LBP9 to the LTRs, which also bear motifs for OCT4 and KLF4 binding [41, 65]. Accordingly the LTRs of transcribed HERVH loci are associated with the active H3K4me3 mark, indicative of a role as transcriptional promoters [65, 71].

Depletion of HERVH transcripts leads to differentiation of hESCs and impairment of hiPSC derivation [65, 66], similar to what is seen by depletion of the HERVH-containing *lncRNA-ES3* and *lncRNA-RoR* transcripts [73, 74]. Whereas it is possible that HERVH function is mediated by one or both of these transcripts, HERVH have contributed more widely (and disproportionately) to hESC-expressed lncRNAs [65, 66, 71]. HERVH-derived lncRNAs may also act in *cis* to enhance the expression of

protein-coding genes, presumably through binding of OCT4, P300, and components of the mediator complex to LTR7/HERVH transcripts, which enables recruitment of those activators to the DNA locus [66]. Additionally, some LTR7/HERVH loci serve as promoters for protein-coding genes such as the pluripotency-essential *ESRG* gene [65, 78]. LTR7/HERVH elements may have therefore undergone multiple cooption events that drove the evolution of gene expression programs essential for pluripotency in humans (see also genome topology implications in the succeeding text).

Genome topology

The genome of higher eukaryotes is spatially organized in a manner that establishes compartments of predominantly active and inactive chromatin, prevents spreading of chromatin domains, and allows for the formation of long-range contacts between regulatory elements and the promoters they regulate [79–81]. Many chromatin domain boundaries are defined by the binding of CTCF, and deletion of individual CTCF sites can alter local genome topology [82, 83]. One current model of loop formation involves the extrusion of chromatin by the cohesion complex until it encounters CTCF sites in a convergent orientation [79–81]. Notably, TEs have made a large contribution to the establishment of mammalian CTCF sites (see in the succeeding text) and are thought to have rewired ESC genome topology between mice and humans [39].

Mouse ERVs

As many as 40% of murine CTCF sites overlap repetitive elements [59], 70% of which act as loop anchors [84]. Unlike in the case of the human genome (see in the succeeding text), mouse ERVs have contributed little to this landscape, which is overwhelmingly dominated by B2 SINE elements [59, 60, 84]. This clade of TEs underwent a large expansion in rodents (and particularly in mice), shaping genome organization in a species-specific manner [85].

One particular context where mouse ERVs have been associated with the formation of chromatin domains is the 2-cell stage. In 2-cell-like cells (2CLC), a cell culture model derived from a subpopulation of ESCs [49], chromatin boundaries often occur at MERVL insertion sites [86]. The insulating capacity of MERVL elements is seemingly derived from their LTRs, as both proviral and solo LTR forms can establish domains. The 2CLC-specific nature of these boundaries (which are absent in mESCs) is associated with binding of the DUX TF and transcriptional activation [86]. The genome is weakly organized in the 2-cell stage, and thus activation of MERVL elements may help start defining chromatin domains, as the genome becomes increasingly more structured during preimplantation development [87].

Human ERVs

Similar to the mouse genome, a large proportion of human CTCF sites (23%) overlap repetitive elements, but the array of families involved is dramatically distinct and includes both DNA transposons and ERVs (e.g., MER20, LTR41, LTR50, and LTR13) [39, 59, 84]. Genetic deletion experiments confirm that CTCF loop anchors derived from ERVs are functional from a structural perspective [84]. This suggests a remarkable versatility of different ERV families to dynamically shape the structural organization of the human genome. Indeed, comparison with mouse CTCF sites suggests a model whereby turnover of CTCF sites occurs via TE insertions [84]. While there may be selective pressure to maintain certain structural domains, novel TE insertions may replace existing loop anchors, which can lead to cases where different TEs are independently coopted to orthologous anchor points.

Transcriptionally, active human ERVs can also establish chromatin domain boundaries in a cell-specific manner [88]. Many hESC-specific boundaries are associated with HERVH elements, which add to their actions as lncRNAs and enhancers described earlier. Deletion of individual HERVH insertions leads to boundary loss and thus merging of chromatin domains [88]. Similar to MERVL elements in mouse 2CLCs, it is at transcriptionally active HERVHs that hESC boundaries form, with CRISPRi-mediated silencing of these elements leading to boundary loss [88]. HERVH expansion within primates has thus driven structural differences in ESC chromatin organization across species. The curious multitier action of HERVH on the genome could work synergistically in a model whereby the *cis*-regulatory action of HERVH lncRNAs is compartmentalized and insulated by its parallel structural roles, which is supported by the effects of HERVH KOs on local gene expression [88].

Retrotransposition

Mouse ERVs

The potential benefits of ERV activation to the host have to be weighed against the risks associated with their mobility, posing an interesting evolutionary conundrum. In the mouse, several ERVs remain mobile and are responsible for 10%–12% of all known germline mutations [89, 90]. The majority of these germline insertions are caused by elements from the IAP and ETn/MusD families, although others have also been detected (e.g., MLV, MERV-L) [89, 91]. Mutations caused by ERVs have a variety of effects, such as transcript truncation, silencing, exon skipping, and generation of alternative promoters. Arguably the best characterized example is the A^{vy} allele, which involves an IAP promoter that drives DNA methylation-regulated expression of the *agouti* coat color gene and is linked to obesity, diabetes, and increased tumor susceptibility [92]. IAPs and ETn/MusD elements are expressed during preimplantation development [40, 93, 94], raising the possibility that mutations occurring during this period could interfere with developmental progression and in extreme cases lead to embryonic lethality. Notably, there is substantial variation in ERV activity between mouse strains [91], and the commonly used C57BL/6J strain harbors little to no IAP or ETn/MusD activity [36]. Studies aiming to understand the potential deleterious effects of ERV retrotransposition in the pluripotent state should therefore make careful considerations about the choice of model organism.

Human ERVs

In contrast to mice, virtually, all human ERVs have suffered inactivating mutations, and there is no evidence of ERV activity in either the germline or somatic cells. However, the possibility cannot be completely ruled out, as HERV-K elements with intact ORFs have been identified in the human population [95], and recombination could also generate active elements [96]. HERV-K are the most recently active ERVs in the human genome and, as a result, are a source of genetic polymorphisms within the human population [18, 95, 97]. Structural polymorphisms involving HERV-H and HERV-W elements have also been identified [18]. Given the regulatory roles described earlier for HERV-K and HERV-H in pluripotency, it would be interesting to explore whether polymorphic insertions could generate variation in developmental gene expression programs.

Restricting ERV activity

ERVs are exceptionally tightly controlled by multiple host mechanisms that prevent inappropriate ERV activation and restrict mobility to minimize deleterious effects on the host genome and transcriptome. In pluripotent cells the expression of the majority of ERVs is constrained, and inducing their expression experimentally has been associated with elevated transcription of nearby genes, highlighting the need to keep ERVs in check [98–105]. Besides posttranscriptional silencing mechanisms, which take over during epigenetic reprograming, the predominant defense strategy in pluripotent cells is repression of ERV transcription through heterochromatinization of ERV elements, which blocks the binding of transcriptional activators (Fig. 7.5). Chromatin changes driven by histone modification enzymes and nucleosome remodeling activities play crucial roles in transcriptional silencing of ERVs in ESCs, while the contribution of DNA methylation is greater in differentiated cells. Here, we review the major pathways ensuring robust ERV repression including the contribution of RNA-based mechanisms.

Transcriptional silencing mechanisms

In a quest to identify ERV-specific chromatin signatures, a survey of 32 chromatin marks in mESCs revealed that different subsets of ERVs are marked by complex patterns of repressive and activating chromatin modifications [106]. These results are compatible with the model that overlapping epigenetic pathways ensure the control of transposons at the familial and individual level [106–108].

One silencing strategy that is commonly associated with TE silencing is DNA methylation [109]. It has long been argued that DNA methylation evolved due to the selective pressure to maintain TE suppression [110]. Mouse DNA methyltransferase (DNMT) knockout models have demonstrated that this is an important pathway for ERV silencing in differentiated cells, although certain families such as IAPs appear to be more dependent on it [111–113]. IAP elements are also derepressed in the mouse germline upon knockout of the essential DNMT cofactor *Dnmt3l* [114]. In contrast, mESCs devoid of all DNA methylation (through deletion of *Dnmt1*, *Dnmt3a*, and *Dnmt3b*) display far milder transcriptional effects on ERVs when compared with those seen in differentiated cells, and few families are affected [100, 101]. This suggests that DNA methylation-independent mechanisms predominate during preimplantation development, ensuring ERV silencing upon epigenetic reprogramming. Accordingly, experiments using conditional knockout models or manipulation of cell culture conditions (i.e., switching cells to 2i medium, see Box 7.2) have shown that there is a transient upregulation of ERVs upon DNA demethylation before they are targeted by other silencing pathways [108, 115–117]. One way resilencing is achieved is through widespread redistribution of the repressive histone modifications, which occurs in a family-dependent manner (Fig. 7.5).

ESCs primarily rely on histone methylation for ERV repression. Trimethylation of H4K20, H3K64, and H3K27 are all involved, yet methylation of H3K9 stands out and is best characterized [100, 108, 118–122]. Interestingly, distinct H3K9 methylation-based silencing mechanisms catalyzed by separate enzymes operate at different ERV families in mESCs (Fig. 7.5 Panels 1 and 2) [26]. Class III ERVs (e.g., MERVL) are repressed specifically by dimethylated H3K9 deposited by the G9a/GLP lysine methyltransferase (KMT) [49, 100, 123]. H3K9 trimethylation on the other hand is pivotal at class I and II ERVs (e.g., IAPs). The main enzyme responsible for this is the KMT SETDB1/ESET [101, 102, 104]. Suv39h KMT enzymes have additional roles in expanding the repressive H3K9me3

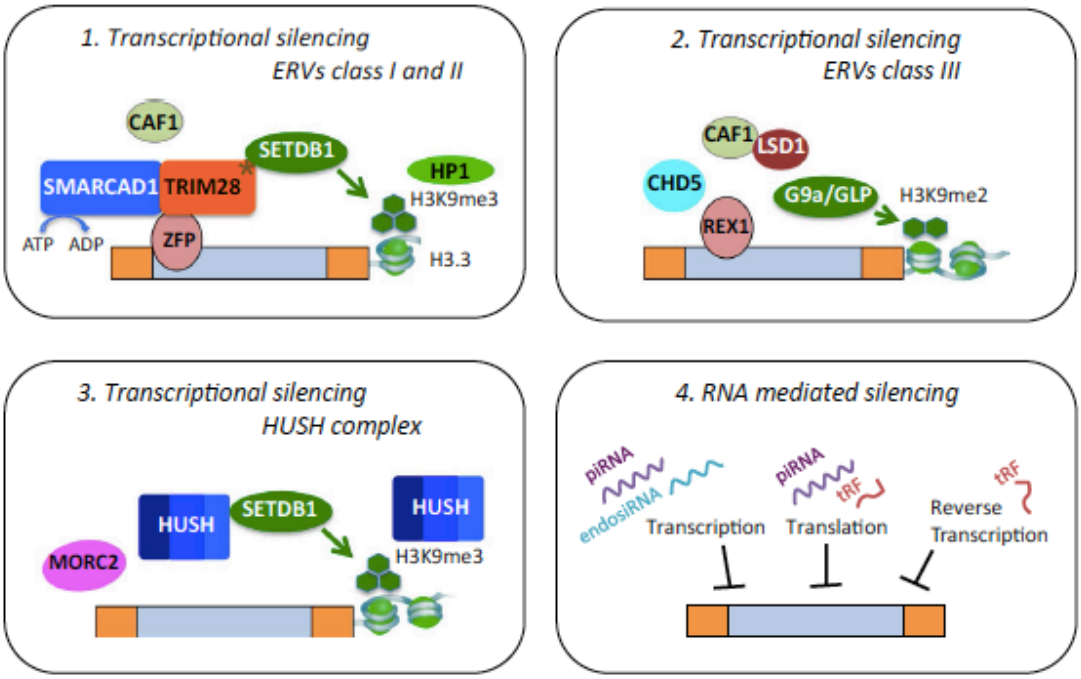

FIG. 7.5

Strategies for the restriction of ERV activity. ERV activity is controlled by covalent histone modifications, for example, repressive H3K9 methylation (green hexagons in Panels 1–3), DNA methylation (not shown), and small RNAs (Panel 4). The prevalence of each pathway varies during different developmental stages. In mouse pluripotent cells, ERV silencing is achieved primarily though histone-modifying enzymes (e.g., H3K9 methytransferases SETDB1 and G9a/GLP and demethylase LSD1/KDM1A), histone mark readers (e.g., HP1 isoforms and HUSH complex), and effectors recruited to ERVs by sequence-specific DNA-binding proteins (e.g., KRAB-ZFPs). A critical corepressor for class I and II ERV families is TRIM28; its SUMOylation (asterix, Panel 1) is required for heterochromatin formation. Other factors associated with silencing of class I and II ERV subsets include the ATP-dependent chromatin remodeler SMARCAD1, the chromatin assembly factor CAF-1 and the histone variant H3.3 (Panel 1) and histone deacetylases (HDACs), hnRNP K, HSP90, RESF1, and ATRX/DAXX (not shown). Linked to silencing class III ERVs are the ATP-dependent chromatin remodeler CHD5, CAF1 (Panel 2) and HDACs, RIF1, and RYBP (not shown). The HUSH complex, which comprises three components and an accessory ATPase (MORC2), mediates recruitment of H3K9me3 writers and the spreading of heterochromatin (Panel 3). Diverse classes of small RNAs (e.g., piRNAs, endosiRNAs and tRNA fragments tRF) regulate ERVs at both the transcriptional and posttranscriptional level and target ERV transcripts to block reverse transcription or translation (Panel 4).

mark at intact ERVs [124]. SETDB1 is selectively targeted to ERVs through direct interaction with the KRAB-associated protein, TRIM28 (KAP1;TIF1beta) [125, 126]. TRIM28 forms a dimer [127–130] that provides a platform for recruitment of not only SETDB1 but also H3K9me3 readers (e.g., heterochromatin protein HP1) and effector proteins required for the formation and expansion of heterochromatin (reviewed in Refs. [119, 131]). A critical role of TRIM28/SETDB1/H3K9me3 in retrotransposon regulation and more broadly in the protection of transcriptome integrity is underscored by several pioneering papers that show that depletion of either SETDB1 or TRIM28 induces the expression of

a broad range of ERVs in mESCs, which is accompanied by activation of proximal genes [100–104]. Although less is known about the extent to which this silencing pathway is conserved in human pluripotency, TRIM28 was found to similarly control endogenous retroelements in human ESCs, especially HERVK and HERVH, whereby subfamily specific repression effects depend on the state of pluripotency [105, 132–134].

It was long believed that the TRIM28/SETDB1 transcriptional silencing pathway is a feature of embryonic development that becomes replaced in adult tissues by DNA methylation. This view has been challenged by the discovery of TRIM28-/SETDB1-mediated repression of distinct ERV families in a number of differentiated cell types, albeit in a context dependent manner [98, 105, 135–139]. However, the TRIM28/SETDB1/H3K9me3 silencing pathway does not operate on evolutionarily old LTRs, which are likely inactivated through the accumulation of genetic mutations nor on the most recent invaders of the genome, which are silenced by DNA methylation [64, 140, 141].

Recruiting silencing components to ERVs

One intriguing question is how silencing pathways access specific ERV families among thousands of host genes and repetitive DNA within the genome. Recognition can principally occur either at the level of RNA or by sequence-specific DNA-binding interactions. RNA-mediated targeting mechanisms are particularly active in the germline, but there is new evidence that they can also occur in pluripotent cells. TET2 enzymes, which function in the repression of MERVL transcripts, are recruited to active MERVLs through association with an RNA-binding protein [142]. Primarily, ESCs use sequence-specific DNA-binding proteins to target ERVs, but most is known about how TRIM28, which itself cannot bind DNA, is recruited. Over 10 years ago, mouse ZFP809, a Kruppel-associated box-zinc finger protein (KRAB-ZFPs), was shown to bring TRIM28 to murine leukemia virus and its endogenous relative via recognition of the tRNA-primer binding site (PBS-pro), a sequence critical for virus replication located at the 5′ end of the retroelement [139, 143, 144]. KRAB-ZFPs present the largest transcription factor family in vertebrate genomes; they exhibit tissue-specific patterns of expression and an exquisite DNA-binding specificity, mediated by multiple zinc finger domains [145]. From the characterization of 101 human KRAB-ZFPs protein interactomes, it emerged that the majority of evolutionary young KRAB-ZFPs associates with TRIM28 [146]. Indeed a growing body of evidence implicates several KRAB-ZFPs in the locus-specific induction of heterochromatin at particular retrotransposon families, both in mouse and in human [98, 99, 133, 146, 147]. One human-specific example is ZNF534, which regulates HERVH in hESCs [133]. ZFP932 and Gm15446 are interesting mouse-specific KRAB-ZFPs, because they dock TRIM28 to the 3′ end of ERVKs at a site important for reverse transcription but distinct from the PBS-prosite [98]. Apart from KRAB-ZFPs, other DNA-binding proteins contribute to the targeting of TRIM28 to distinct ERV families, best illustrated with YY1 and Rex1 in mESCs [106, 148–150]. Overall, TRIM28 can be recruited through a multitude of different site-specific DNA-binding proteins to more than one region within an ERV sequence, which may be important for the stability of silencing.

Many paths ensure suppression of ERV transcription

In addition to TRIM28/SETDB1, a large number of proteins have been implicated in ERV repression (Fig. 7.5), among them, enzymes that remodel chromatin and enzymes that modify chromatin proteins, for instance, by SUMOylation, their cofactors, and the histone variant H3.3 [48, 119, 123, 126, 151–161].

Moreover a distinct H3K9 methylation silencing pathway involving the HUSH complex has been identified. In the succeeding text, we consider recent advances on how the HUSH complex and ATP-dependent chromatin remodelers contribute to ERV repression.

Silencing by HUSH

The human silencing hub, or HUSH complex, participates in the transcriptional shutdown of retroviruses and retroelements and functions by regulating histone methylation of proviral DNA in embryonic and somatic cell types [152, 162–164]. The central functional feature of HUSH silencing is the reading and writing of H3K9me3, which are facilitated by the ability of HUSH to bind H3K9me3 and recruit H3K9me3 methyltransferases like SETDB1 (Fig. 7.5 Panel 3). HUSH consists of three core subunits, the chromodomain protein MPHOSPH8; a H3K9me3 reader, TASOR and PPHLN1; and a biochemically and functionally associated ATPase MORC2 [164, 165]. Each of these proteins was recovered in screens aimed at the identification of provirus silencing factors [152, 161]. HUSH/MORC2 not only bind to ERVs and evolutionary young LINE1 subfamilies in human and mouse ESCs but also are required for their *silencing* [152, 162, 163, 166].

At first sight the HUSH complex employs a similar strategy for the formation of H3K9me3 heterochromatin as exemplified by TRIM28. However, while TRIM28 is principally recruited to retroelements by sequence-specific DNA-binding proteins, HUSH targeting appears less dependent on the underlying DNA sequence than on the chromatin context it is embedded in Refs. [165, 167]. Upon integration of an active transgene into a heterochromatic environment, HUSH facilitates the spreading of preexisting H3K9me3 and subsequent maintenance [165, 167]. A key question is whether HUSH and TRIM28 repress retrotransposons cooperatively or independently of each other. In support for their cooperative function, a range of retrotransposons (ERVs, L1s) are enriched for both proteins, although it has yet to be confirmed that HUSH and TRIM28 bind them simultaneously and it is not known whether their relationship is epistatic. Comparative analysis of transcripts upregulated upon depletion of either TRIM28 or HUSH showed an overlap consistent with TRIM28 and HUSH coregulating primarily young retrotransposons and new genes rewired by retrotransposon noncoding DNA [163].

ERV suppression by chromatin remodeling

ATP-dependent chromatin remodelers shape the chromatin landscape by determining the composition, the position, and the density of nucleosomes, thereby regulating the chromatin access for other proteins (reviewed by Clapier et al. [168], Hota and Bruneau [169], and Narlikar et al. [170]). At the heart of their function lie ATPases related to the SNF2 helicase from yeast. These enzymes use ATP hydrolysis to break histone-DNA contacts. Remodeling outcomes range from the ejection or incorporation of histone variants to sliding or displacing nucleosomes. One could hence envision that ATP-dependent chromatin remodelers participate in the suppression of ERV activity in several ways, for instance, by promoting increased nucleosome density at ERVs or by facilitating histone variant exchange. Historically a link between ATP-dependent chromatin-remodeling enzymes and ERV silencing was proposed based on biochemical observations. TRIM28 was found to interact with CHD3/Mi-2, which comprises one of several mutually exclusive ATPase subunits of the nucleosome and remodeling deacetylase (NuRD) complex [171]. Consequently the NuRD chromatin remodeler is frequently incorporated into models of ERV repression. However, NuRD has not been extensively studied at retrotransposons, and direct

support that NuRD operates at ERVs is lacking. Nevertheless, with the event of successful identification of remodeler binding sites genome wide combined with improved mapping strategies for ERVs, connections between other ATP-dependent chromatin remodelers and ERV restriction have begun to emerge.

SMARCAD1

The SNF2-like remodeler SMARCAD1 was recognized as key component of the TRIM28–SETDB1 ERV silencing machinery [146, 158]. In mESCs, SMARCAD1 is recruited by TRIM28 to class I and II ERV subfamilies, particularly IAPs (Fig. 7.5 Panel 1) [158, 172]. An attractive idea is that remodeling by SMARCAD1 promotes trimethylation of H3K9 and transcriptional repression by facilitating the stable association of TRIM28–SETDB1 complexes at ERVs. Such a model can account for the observation that depletion of SMARCAD1 from mESCs results in reduced TRIM28 occupancy at ERVs, accompanied by reduced histone methylation and inefficient silencing of ERVs and nearby genes [158]. Importantly the catalytic activity of SMARCAD1 is required to maintain and stabilize TRIM28 binding, implying active chromatin remodeling as a central feature in setting up TRIM28-induced heterochromatin formation at ERVs. The challenge for future studies will be to understand the precise mechanisms and to identify the substrate toward which the remodeling activity is directed.

Binding of SMARCAD1 to LTR elements has been observed from budding and fission yeast to pluripotent mouse cells [158, 173–175], pointing to a conserved role for this remodeler family in retrotransposon control. The ortholog of SMARCAD1 in fission yeast represses gypsy class *Tf2* retroelements by stabilizing a critical regulatory nucleosome in the 5' LTR to inhibit productive transcription [176]. The precise mechanism of SMARCAD1 action at ERVs in pluripotent cells is not known. In mammalian cells, SMARCAD1 is a physical and functional partner of TRIM28, but there is no equivalent partner known in yeast [128, 146, 158, 177–179]. It is noteworthy that the depletion of TRIM28 from mESCs negatively impacts the steady-state levels of SMARCAD1, effectively generating a double TRIM28-SMARCAD1 depletion [158, 172]. Consequently, when investigating the influence of prolonged removal of TRIM28 on ERV regulation, it is not possible to unambiguously determine the extent to which observed effects are attributable to TRIM28 alone or to combined TRIM28-SMARCAD1 function [158].

Interestingly, both SMARCAD1 and TRIM28 associate with DNA replication factors, notably PCNA, and have been shown to participate in the heterochromatin propagation through replication [177, 179, 179a]. It is thus tempting to speculate that the partnership between TRIM28 and SMARCAD1 could be critical for preserving ERV heterochromatin integrity through the cell cycle.

ATRX

A screen for mouse proteins that nucleate the formation of repressive chromatin at IAPs identified another SNF2-type ATPase, alpha thalassemia–mental retardation X linked (ATRX) [159]. A number of ChIP studies found ATRX at repeat sequences and heterochromatic loci marked by H3K9me3 and H4K20me3 [159, 180–182]. In pluripotent cells, H3.3 is significantly enriched at ERV class I and II families, particularly at IAPs [151, 183]. Since ATRX can form a heterocomplex with the histone chaperone death domain-associated protein (DAXX), which incorporates the histone variant H3.3 at specific loci [184–186], the major role of ATRX at ERVs was proposed to relate to H3.3 deposition. But emerging evidence for overlapping yet independent functions of DAXX and ATRX at repetitive sequences challenges this hypothesis [119, 151, 159, 180, 182, 187]. The recent identification of an

ATRX-independent DAXX complex involved in ERV repression puts the function of ATRX at ERVs in question [183]. Moreover the impact of ATRX loss on ERV expression is controversial [159, 163, 180, 183]. Nevertheless, IAP derepression is clearly exacerbated when ATRX-depleted cells are challenged by further disruption of heterochromatin organization, for instance, through interfering with TRIM28 function [159, 183] or with DNA methylation [163, 180]. One possible mechanism by which ATRX may help to keep ERVs silenced is by contributing to chromatin compaction [159]. It is conceivable that the presence of ATRX at ERVs provides an additional layer of heterochromatin stabilization, especially when there is a need to compensate for the lack of other pathways of heterochromatin formation.

CHD5

While the SNF2-like remodelers described so far affect primarily the expression of class I and II ERV subfamilies as part of the TRIM28/SETDB1 heterochromatin pathway, a role in silencing of class III ERVs has been demonstrated for the chromodomain helicase DNA-binding protein 5 (CHD5) (Fig. 7.5 Panel 2) [153, 156]. Mouse ESCs deficient in this remodeler considerably upregulate subsets of MERVL elements normally bound by CHD5, accompanied by expression of proximal MERVL LTR-driven 2-cell (2C)-specific genes. In conclusion, ATP-dependent chromatin remodelers clearly act at distinct ERV families, which underscore the specificity of SNF2-like ATPase-mediated control of retrotransposon activity. How chromatin remodeling can translate into repression of ERV activity remains an area for future exploration.

RNA based regulation of ERVs (Fig. 7.5 Panel 4)

RNA interference is yet another layer of control that host organisms use to downregulate TEs, although its contribution to ERV silencing in pluripotent cells is not as great as during other developmental stages. During epigenetic reprogramming, RNA-based retrotransposon silencing pathways deal with ERVs that become active, for example, in the germline and all the way through to implantation [188–191] (reviewed by Schorn and Martienssen [192] and Svoboda [193]). We will mention here the main RNA interference pathways that are involved in ERV silencing.

piRNA (24-31nt) are mostly expressed in the germline, keeping TEs repressed by regulating chromatin modifications and DNA methylation or by posttranscriptional silencing (the ping-pong cycle) [194, 195]. In mice defective for MILI or MIWI2 (the mouse PIWI proteins), IAP (class II ERVs) DNA is demethylated, expression is upregulated, and mice are sterile [196, 197]. In adult human testis the highest piRNA association to ERV was found for the LTR1/HUERS-P2 [46]. However, the mechanisms controlling the generation of piRNAs and their TE-silencing activity in mammals are still enigmatic [198]. Additionally, a large number of studies found high expression of PIWI proteins in the brain and in several types of cancer, suggesting possible roles for this pathway in controlling TEs in somatic cells as well [199]. Lately, piRNA were suggested to function as a first line of defense, or innate genomic protection response, against retroviral sequences in the germline of recently invaded koalas [200].

EndosiRNA (~21nt)-based repression mechanism has also been suggested to be a first line of defense when DNA methylation is erased, as in primordial germ cells or in hypomethylated ESCs [115].

Small tRNA fragments (tRF) are derived from the 3′ end of mature tRNAs and were shown to specifically promote silencing of retrotransposition competent ERVs in ESCs by inhibiting ERV translation and replication [201, 202].

RNA modifications leading to the destabilization of ERV transcripts are another layer of regulation, as exemplified by RNA hydroxymethylation-mediated degradation of MERVL [142].

Summary/open questions

Here, we have highlighted the divergent roles that ERVs play in the plastic and dynamic chromatin landscape of ESC organization [203] and the various restriction strategies ESC apply to control ERV expression and replication.

In the last decade, our knowledge of the functions of ERVs in pluripotency and on ERV control in general has increased exponentially. This is due to advances in ESC culturing technology and the growing understanding of pluripotency on one hand and to the advancement made in the availability and accuracy of high-throughput deep sequencing technologies and mapping of repetitive elements, on the other hand. Our current understanding of how ERVs are kept in check is based mainly on studies performed in mouse ESCs with a focus on the mechanisms and factors responsible for their control but largely ignoring the identities of the specific ERVs subjected to it. Thus a gap in our understanding of ERVs restriction is about how specific it is, both in regard to cells, tissues, and developmental stages (since most studies are done in mESCs) and in regard to the group or subgroup of ERVs targeted (familial and individual level). For example, we do not know why some ERV I/II class LTRs are upregulated in the EGA or in the blastocyst, while others remain silent. Another related issue is the lack of a similar in-depth mechanistic understanding of ERV regulation in humans. Since human ERVome is significantly different than that of the mouse (specifically, less ERV II and more ERV III class elements), this is a question worth considering. Thankfully the relatively new availability of naïve pluripotent human ESC opens the door to this kind of mechanistic studies. When considering the role of ERVs in transcriptional or epigenetic regulation, most of the data come from NGS analysis and are thus correlative by nature. Lately, more studies are accepting the challenge of proving causation, usually by deleting some candidate regulatory ERVs, but are still far from a mechanistic understanding of ERV regulatory functions and from grasping the full magnitude of ERVs that are actually important for preimplantation development and pluripotency.

ERVs, as fossils of past viral infections that remain in the genome for millions of years, were always an intriguing subject for evolutionary biologists. Comparing the different path with homeostasis between the "arms race" and the "exaptation" of ERVs by host cells of different phylogenetic lineages could open a window to a new understanding of pluripotency networks and epigenetic control. Thus more studies targeting the "transposcriptome" of nonmodel organisms during early development are required. Additionally, the habit of masking/discarding repeat data on NGS data is holding the field back and might prevent us from observing the most prominent consequences of our data or analysis. When more scientists become aware of the importance of TEs regulation in the genome, new insights will surely follow.

References

[1] Arkhipova IR. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mob DNA 2017;8:19. https://doi.org/10.1186/s13100-017-0103-2.

[2] Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 2007;8:973–82. https://doi.org/10.1038/nrg2165.

[3] Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science 2016;351:aac7247. https://doi.org/10.1126/science.aac7247.
[4] McClintock B. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A 1950;36:344–55.
[5] McClintock B. Induction of instability at selected loci in maize. Genetics 1953;38:579–99.
[6] Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science 1969;165:349–57.
[7] Britten RJ. DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci 1996;93:9374–7. https://doi.org/10.1073/pnas.93.18.9374.
[8] Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology 2011;411:273–87. https://doi.org/10.1016/j.virol.2010.12.007.
[9] Cantone I, Fisher AG. Epigenetic programming and reprogramming during development. Nat Struct Mol Biol 2013;20:282–9. https://doi.org/10.1038/nsmb.2489.
[10] Robbez-Masson L, Rowe HM. Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology 2015;12:45. https://doi.org/10.1186/s12977-015-0173-5.
[11] Serrato-Capuchina A, Matute DR. The role of transposable elements in speciation. Genes 2018;9. https://doi.org/10.3390/genes9050254.
[12] Mager DL, Stoye JP. Mammalian endogenous retroviruses. Microbiol Spectr Mob DNA III 2015;1079–100. https://doi.org/10.1128/microbiolspec.MDNA3-0009-2014.
[13] Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999;9:657–63. https://doi.org/10.1016/S0959-437X(99)00031-3.
[14] Magiorkinis G, Belshaw R, Katzourakis A. "There and back again": revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. Philos Trans R Soc Lond Ser B Biol Sci 2013;368:20120504. https://doi.org/10.1098/rstb.2012.0504.
[15] Goodier JL, Kazazian HH. Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell 2008;135:23–35. https://doi.org/10.1016/j.cell.2008.09.022.
[16] Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV,

Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J, International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921. https://doi.org/10.1038/35057062.

[16a] Johnson WE. Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol 2019;17:355–70. https://doi.org/10.1038/s41579-019-0189-2.

[17] Stoye JP. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat Rev Microbiol 2012;10:395–406. https://doi.org/10.1038/nrmicro2783.

[18] Thomas J, Perron H, Feschotte C. Variation in proviral content among human genomes mediated by LTR recombination. Mob DNA 2018;9:36. https://doi.org/10.1186/s13100-018-0142-3.

[18a] Gifford RJ, Blomberg J, Coffin JM, et al. Nomenclature for endogenous retrovirus (ERV) loci. Retrovirology 2018;15(1):59. http://dx.doi.org/10.1186/s12977-018-0442-1.

[19] Britten RJ, Kohne DE. Repeated sequences in DNA. Science 1968;161:529–40. https://doi.org/10.1126/science.161.3841.529.

[20] Gould SJ, Vrba ES. Exaptation—a missing term in the science of form. Paleobiology 1982;8:4–15. https://doi.org/10.1017/S0094837300004310.

[21] Joly-Lopez Z, Bureau TE. Exaptation of transposable element coding sequences. Curr Opin Genet Dev 2018;49:34–42. https://doi.org/10.1016/j.gde.2018.02.011.

[22] Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, Wang D, Tian Y, Yang C, Meng M, Yuan G, Kang G, Wu Y, Wang K, Zhang H, Wang D, Cong P. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. Nat Commun 2019;1–13. https://doi.org/10.1038/s41467-019-09518-x.

[23] Zhang Y, Cheng TC, Huang G, Lu Q, Surleac MD, Mandell JD, Pontarotti P, Petrescu AJ, Xu A, Xiong Y, Schatz DG. Transposon molecular domestication and the evolution of the RAG recombinase. Nature 2019;569:79–84. https://doi.org/10.1038/s41586-019-1093-7.

[24] Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta 2012;33:663–71. https://doi.org/10.1016/j.placenta.2012.05.005.

[25] Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T. Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc Natl Acad Sci U S A 2009;106:12127–32. https://doi.org/10.1073/pnas.0902925106.

[26] Cosby RL, Chang N-C, Feschotte C. Host–transposon interactions: conflict, cooperation, and cooption. Genes Dev 2019;33:1098–116. https://doi.org/10.1101/gad.327312.119.

[27] Lisch D. How important are transposons for plant evolution? Nat Rev Genet 2013;14:49–61. https://doi.org/10.1038/nrg3374.

[28] Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. The industrial melanism mutation in British peppered moths is a transposable element. Nature 2016;534:102–5. https://doi.org/10.1038/nature17951.

[29] Bhattacharyya MK, Smith AM, Ellis TH, Hedley C, Martin C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. Cell 1990;60:115–22.

[30] Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell 2012;24:1242–55. https://doi.org/10.1105/tpc.111.095232.

[31] Kobayashi S. Retrotransposon-induced mutations in grape skin color. Science 2004;304:982. https://doi.org/10.1126/science.1095011.

[32] Jung S, Venkatesh J, Kang M-Y, Kwon J-K, Kang B-C. A non-LTR retrotransposon activates anthocyanin biosynthesis by regulating a MYB transcription factor in *Capsicum annuum*. Plant Sci 2019;287:110181. https://doi.org/10.1016/j.plantsci.2019.110181.
[33] Flemr M, Malik R, Franke V, Nejepinska J, Sedlacek R, Vlahovicek K, Svoboda P. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell 2013;155:807–16. https://doi.org/10.1016/j.cell.2013.10.001.
[34] Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science 2016;351:1083–7. https://doi.org/10.1126/science.aad5497.
[35] Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. Genes Dev 2009;23:1303–12. https://doi.org/10.1101/gad.1803909.
[36] Richardson SR, Gerdes P, Gerhardt DJ, Sánchez-Luque FJ, Bodea G-O, Muñoz-Lopez M, Jesuadian JS, Kempen M-JHC, Carreira PE, Jeddeloh JA, Garcia-Perez JL, Kazazian HH, Ewing AD, Faulkner GJ. Heritable L1 retrotransposition in the mouse primordial germline and early embryo. Genome Res 2017;27:1395–405. https://doi.org/10.1101/gr.219022.116.
[37] van den Hurk JAJM, Meij IC, del Seleme MC, Kano H, Nikopoulos K, Hoefsloot LH, Sistermans EA, de Wijs IJ, Mukhopadhyay A, Plomp AS, de Jong PTVM, Kazazian HH, Cremers FPM. L1 retrotransposition can occur early in human embryonic development. Hum Mol Genet 2007;16:1587–92. https://doi.org/10.1093/hmg/ddm108.
[38] Saitou M, Kagiwada S, Kurimoto K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. Development 2012;139:15–31. https://doi.org/10.1242/dev.050849.
[39] Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H, Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet 2010;42:631–4. https://doi.org/10.1038/ng.600.
[40] Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, Carninci P, Torres-Padilla M-E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol 2013;20:332–8. https://doi.org/10.1038/nsmb.2495.
[41] Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, Sachs F, Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell 2015;16:135–41. https://doi.org/10.1016/j.stem.2015.01.005.
[42] Hackett JA, Kobayashi T, Dietmann S, Surani MA. Activation of lineage regulators and transposable elements across a pluripotent spectrum. Stem Cell Rep 2017;8:1645–58. https://doi.org/10.1016/j.stemcr.2017.05.014.
[42a] Leitch HG, McEwen KR, Turp A, et al. Naive pluripotency is associated with global DNA hypomethylation. Nat Struct Mol Biol 2013;20(3):311–6. http://dx.doi.org/10.1038/nsmb.2510.
[43] Jacques P-É, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet 2013;9:e1003504. https://doi.org/10.1371/journal.pgen.1003504.
[44] Ferguson-Smith AC, Surani MA. Imprinting and the epigenetic asymmetry between parental genomes. Science 2001;293:1086–9.
[45] Stewart KR, Veselovska L, Kim J, Huang J, Saadeh H, Tomizawa S, Smallwood SA, Chen T, Kelsey G. Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. Genes Dev 2015;29:2449–62. https://doi.org/10.1101/gad.271353.115.
[46] Brind'Amour J, Kobayashi H, Richard Albert J, Shirane K, Sakashita A, Kamio A, Bogutz A, Koike T, Karimi MM, Lefebvre L, Kono T, Lorincz MC. LTR retrotransposons transcribed in oocytes drive species-specific and heritable changes in DNA methylation. Nat Commun 2018;9:3331. https://doi.org/10.1038/s41467-018-05841-x.
[47] Tang WWC, Dietmann S, Irie N, Leitch HG, Floros VI, Bradshaw CR, Hackett JA, Chinnery PF, Surani MA. A unique gene regulatory network resets the human germline epigenome for development. Cell 2015;161:1453–67. https://doi.org/10.1016/j.cell.2015.04.053.

[48] Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, Pfaff SL. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev 2011;25:594–607. https://doi.org/10.1101/gad.2008511.

[49] Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 2012;487:57–63. https://doi.org/10.1038/nature11244.

[50] De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet 2017;49:941–5. https://doi.org/10.1038/ng.3858.

[51] Eckersley-Maslin M, Alda-Catalinas C, Blotenburg M, Kreibich E, Krueger C, Reik W. Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev 2019; https://doi.org/10.1101/gad.321174.118.

[52] Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, Nix DA, Peterson CM, Tapscott SJ, Carrell DT, Cairns BR. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet 2017;49:925–34. https://doi.org/10.1038/ng.3844.

[53] Chen Z, Zhang Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat Genet 2019;51:947–51. https://doi.org/10.1038/s41588-019-0418-7.

[54] Rodriguez-Terrones D, Gaume X, Ishiuchi T, Weiss A, Kopp A, Kruse K, Penning A, Vaquerizas JM, Brino L, Torres-Padilla M-E. A molecular roadmap for the emergence of early-embryonic-like cells in culture. Nat Genet 2018;50:106–19. https://doi.org/10.1038/s41588-017-0016-5.

[55] Boroviak T, Stirparo GG, Dietmann S, Hernando-Herraez I, Mohammed H, Reik W, Smith A, Sasaki E, Nichols J, Bertone P. Single cell transcriptome analysis of human, marmoset and mouse embryos reveals common and divergent features of preimplantation development. Development 2018;145. https://doi.org/10.1242/dev.167833. dev167833-18.

[56] Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, Yi L, Yu G, Hou Z, Fan D, Tian Y, Liu J, Chen Z-J, Liu J. Chromatin accessibility landscape in human early embryos and its association with evolution. Cell 2018;173:248–59. e15. https://doi.org/10.1016/j.cell.2018.02.028.

[57] Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, Reijo Pera RA, Wysocka J. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature 2015;522(7555):221–5. https://doi.org/10.1038/nature14308.

[58] Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K, Cloonan N, Steptoe AL, Lassmann T, Waki K, Hornig N, Arakawa T, Takahashi H, Kawai J, Forrest ARR, Suzuki H, Hayashizaki Y, Hume DA, Orlando V, Grimmond SM, Carninci P. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet 2009;41:563–71. https://doi.org/10.1038/ng.368.

[59] Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res 2014;24:1963–76. https://doi.org/10.1101/gr.168872.113.

[60] Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew J-L, Ruan Y, Wei C-L, Ng H-H, Liu ET. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 2008;18:1752–62. https://doi.org/10.1101/gr.080663.108.

[61] Sundaram V, Choudhary MNK, Pehrsson E, Xing X, Fiore C, Pandey M, Maricque B, Udawatta M, Ngo D, Chen Y, Paguntalan A, Ray T, Hughes A, Cohen BA, Wang T. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. Nat Commun 2017;8:14550. https://doi.org/10.1038/ncomms14550.

[62] Todd CD, Deniz O, Taylor D, Branco MR. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. eLife 2019;8:1473. https://doi.org/10.7554/eLife.44344.

[63] Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet 2017;13:e1006883. https://doi.org/10.1371/journal.pgen.1006883.

[64] Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. Cell Stem Cell 2019;24:724–35. e5. https://doi.org/10.1016/j.stem.2019.03.012.

[65] Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV, Schumann GG, Chen W, Lorincz MC, Ivics Z, Hurst LD, Izsvák Z. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature 2014;516:405–9. https://doi.org/10.1038/nature13804.

[66] Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, Ng H-H. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat Struct Mol Biol 2014;21:423–5. https://doi.org/10.1038/nsmb.2799.

[67] Fuentes DR, Swigut T, Wysocka J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. eLife 2018;7:861. https://doi.org/10.7554/eLife.35989.

[68] Perry RB-T, Ulitsky I. The functions of long noncoding RNAs in development and stem cells. Development 2016;143:3882–94. https://doi.org/10.1242/dev.140962.

[69] Yao R-W, Wang Y, Chen L-L. Cellular functions of long noncoding RNAs. Nat Cell Biol 2019;21:542–51. https://doi.org/10.1038/s41556-019-0311-8.

[70] Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet 2013;9:e1003470. https://doi.org/10.1371/journal.pgen.1003470.

[71] Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol 2012;13:R107. https://doi.org/10.1186/gb-2012-13-11-r107.

[72] Lin N, Chang K-Y, Li Z, Gates K, Rana ZA, Dang J, Zhang D, Han T, Yang C-S, Cunningham TJ, Head SR, Duester G, Dong PDS, Rana TM. An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol Cell 2014;53:1005–19. https://doi.org/10.1016/j.molcel.2014.01.021.

[73] Loewer S, Cabili MN, Guttman M, Loh Y-H, Thomas K, Park IH, Garber M, Curran M, Onder T, Agarwal S, Manos PD, Datta S, Lander ES, Schlaeger TM, Daley GQ, Rinn JL. Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat Genet 2010;42:1113–7. https://doi.org/10.1038/ng.710.

[74] Ng S-Y, 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;31:522–33. https://doi.org/10.1038/emboj.2011.459.

[75] Ramsay L, Marchetto MC, Caron M, Chen S-H, Busche S, Kwan T, Pastinen T, Gage FH, Bourque G. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. BMC Genomics 2017;18:214. https://doi.org/10.1186/s12864-017-3568-y.

[76] Römer 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.

[77] 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. https://doi.org/10.3389/fimmu.2019.01339.

[78] Zhao M, Ren C, Yang H, Feng X, Jiang X, Zhu B, Zhou W, Wang L, Zeng Y, Yao K. Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. Biochem Biophys Res Commun 2007;362:916–22. https://doi.org/10.1016/j.bbrc.2007.08.081.

[79] Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet 2016;17:661–78. https://doi.org/10.1038/nrg.2016.112.

[80] Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nat Rev Genet 2018;19:789–800. https://doi.org/10.1038/s41576-018-0060-8.

[81] Schoenfelder S, Fraser P. Long-range enhancer–promoter contacts in gene expression control. Nat Rev Genet 2019;1–19. https://doi.org/10.1038/s41576-019-0128-0.

[82] Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 2015;162:900–10. https://doi.org/10.1016/j.cell.2015.07.038.

[83] Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A 2015;112:E6456–65. https://doi.org/10.1073/pnas.1518552112.

[84] Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. Co-opted transposons help perpetuate conserved higher-order chromosomal structures (preprint). Genomics 2018; https://doi.org/10.1101/485342.

[85] Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell 2012;148:335–48. https://doi.org/10.1016/j.cell.2011.11.058.

[86] Kruse K, Diaz N, Enriquez-Gasca R, Gaume X, Torres-Padilla M-E, Vaquerizas JM. Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. bioRxiv 2019; https://doi.org/10.1101/523712.

[87] Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, Ma J, Zhang X, Zhang K, Wang Y, Zhang MQ, Gao J, Dixon JR, Wang X, Zeng J, Xie W. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 2017;547:232–5. https://doi.org/10.1038/nature23263.

[88] Zhang Y, Li T, Preissl S, Amaral ML, Grinstein JD, Farah EN, Destici E, Qiu Y, Hu R, Lee AY, Chee S, Ma K, Ye Z, Zhu Q, Huang H, Fang R, Yu L, Izpisua Belmonte JC, Wu J, Evans SM, Chi NC, Ren B. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat Genet 2019;51:1380–8. https://doi.org/10.1038/s41588-019-0479-7.

[89] Maksakova IA, Romanish MT, Gagnier L, Dunn CA, Van De Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet 2006;2:e2. https://doi.org/10.1371/journal.pgen.0020002.

[90] Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigó R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrim J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R,

Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sugnet C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang S-P, Zdobnov EM, Zody MC, Lander ES. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–62. https://doi.org/10.1038/nature01262.

[91] Gagnier L, Belancio VP, Mager DL. Mouse germ line mutations due to retrotransposon insertions. Mob DNA 2019;10:15. https://doi.org/10.1186/s13100-019-0157-4.

[92] Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999;23:314–8. https://doi.org/10.1038/15490.

[93] Brulet P, Condamine H, Jacob F. Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis. Proc Natl Acad Sci 1985;82:2054–8. https://doi.org/10.1073/pnas.82.7.2054.

[94] Peaston AE, Evsikov AV, Graber JH, De Vries WN, Holbrook AE, Solter D, Knowles BB. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 2004;7:597–606. https://doi.org/10.1016/j.devcel.2004.09.004.

[95] Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. Proc Natl Acad Sci 2016;113:E2326–34. https://doi.org/10.1073/pnas.1602336113.

[96] Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, Heidmann T. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Res 2006;16:1548–56. https://doi.org/10.1101/gr.5565706.

[97] Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. J Virol 2014;88:9529–37. https://doi.org/10.1128/JVI.00919-14.

[98] Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Dev Cell 2016;36:611–23. https://doi.org/10.1016/j.devcel.2016.02.024.

[99] Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature 2014;516:242–5. https://doi.org/10.1038/nature13760.

[100] Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M, Lorincz MC. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell 2011;8:676–87. https://doi.org/10.1016/j.stem.2011.04.004.

[101] Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature 2010;464:927–31. https://doi.org/10.1038/nature08858.

[102] Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development 2013;140:519–29. https://doi.org/10.1242/dev.087585.

[103] Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 2010;463:237–40. https://doi.org/10.1038/nature08674.

[104] Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res 2013;23:452–61. https://doi.org/10.1101/gr.147678.112.

[105] Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res 2014; https://doi.org/10.1101/gr.172833.114.

[106] He J, Fu X, Zhang M, He F, Li W, Abdul MM, Zhou J, Sun L, Chang C, Li Y, Liu H, Wu K, Babarinde IA, Zhuang Q, Loh Y-H, Chen J, Esteban MA, Hutchins AP. Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. Nat Commun 2019;10:34. https://doi.org/10.1038/s41467-018-08006-y.

[107] Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. Epigenetics Chromatin 2018;11:7. https://doi.org/10.1186/s13072-018-0177-1.

[108] Walter M, Teissandier A, Pérez-Palacios R, Bourc'his D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife 2016;5. https://doi.org/10.7554/eLife.11418.

[109] Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. Nat Rev Genet 2019;20:417–31. https://doi.org/10.1038/s41576-019-0106-6.

[110] Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet 1997;13:335–40.

[111] Dunican DS, Cruickshanks HA, Suzuki M, Semple CA, Davey T, Arceci RJ, Greally J, Adams IR, Meehan RR. Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. Genome Biol 2013;14:R146. https://doi.org/10.1186/gb-2013-14-12-r146.

[112] Jackson-Grusby L, Beard C, Possemato R, Tudor M, Fambrough D, Csankovszki G, Dausman J, Lee P, Wilson C, Lander E, Jaenisch R. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet 2001;27:31–9. https://doi.org/10.1038/83730.

[113] Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 1998;20:116–7. https://doi.org/10.1038/2413.

[114] Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature 2004;431:96–9. https://doi.org/10.1038/nature02886.

[115] Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, Sharif J, Olova N, Chandra T, Koseki H, von Meyenn F, Reik W. An endosiRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. Cell Stem Cell 2017;21:694–703. e7. https://doi.org/10.1016/j.stem.2017.10.004.

[116] Deniz O, de la Rica L, Cheng KCL, Spensberger D, Branco MR. SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. Genome Biol 2018;19:6. https://doi.org/10.1186/s13059-017-1376-y.

[117] Sharif J, Endo TA, Nakayama M, Karimi MM, Shimada M, Katsuyama K, Goyal P, Brind'Amour J, Sun M, Sun Z, Ishikura T, Mizutani-Koseki Y, Ohara O, Shinkai Y, Nakanishi M, Xie H, Lorincz MC, Koseki H. Activation of endogenous retroviruses in Dnmt1(–/–) ESCs involves disruption of SETDB1-mediated repression by NP95 binding to hemimethylated DNA. Cell Stem Cell 2016;19:81–94. https://doi.org/10.1016/j.stem.2016.03.013.

[118] Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C, Zeissler U, Lappe M, Schübeler D, Torres-Padilla M-E, Schneider R. H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. Nat Struct Mol Biol 2009;16:777–81. https://doi.org/10.1038/nsmb.1629.

[119] Groh S, Schotta G. Silencing of endogenous retroviruses by heterochromatin. Cell Mol Life Sci 2017; https://doi.org/10.1007/s00018-017-2454-8.

[120] Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev 2010;24:265–76. https://doi.org/10.1101/gad.544410.

[121] Leung DC, Lorincz MC. Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem Sci 2012;37:127–33. https://doi.org/10.1016/j.tibs.2011.11.006.

[122] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M,

Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553–60. https://doi.org/10.1038/nature06008.
[123] Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, Lorincz MC. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin 2013;6:15. https://doi.org/10.1186/1756-8935-6-15.
[124] Bulut-Karslioglu A, De La Rosa-Velázquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galán C, Winter GE, Engist B, Gerle B, O'Sullivan RJ, Martens JHA, Walter J, Manke T, Lachner M, Jenuwein T. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell 2014;55:277–90. https://doi.org/10.1016/j.molcel.2014.05.029.
[125] Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ. 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;16:919–32. https://doi.org/10.1101/gad.973302.
[126] Thompson PJ, Dulberg V, Moon KM, Foster LJ, Chen C, Karimi MM, Lorincz MC. hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. PLoS Genet 2015;11:e1004933. https://doi.org/10.1371/journal.pgen.1004933.
[127] Fonti G, Marcaida MJ, Bryan LC, Träger S, Kalantzi AS, Helleboid P-YJ, Demurtas D, Tully MD, Grudinin S, Trono D, Fierz B, Dal Peraro M. KAP1 is an antiparallel dimer with a functional asymmetry. Life Sci Alliance 2019;2. https://doi.org/10.26508/lsa.201900349.
[128] Lim M, Newman JA, Williams HL, Masino L, Aitkenhead H, Gravard AE, Gileadi O, Svejstrup JQ. A ubiquitin-binding domain that binds a structural fold distinct from that of ubiquitin. Structure 2019;27(1316–1325):e6. https://doi.org/10.1016/j.str.2019.05.003.
[129] Stoll GA, Oda SI, Chong ZS, Yu M, McLaughlin SH, Modis Y. Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. Proc Natl Acad Sci U S A 2019;116:15042–51. https://doi.org/10.1073/pnas.1901318116.
[130] Sun Y, Keown JR, Black MM, Raclot C, Demarais N, Trono D, Turelli P, Goldstone DC. A dissection of oligomerization by the TRIM28 tripartite motif and the interaction with members of the Krab-ZFP family. J Mol Biol 2019;431:2511–27. https://doi.org/10.1016/j.jmb.2019.05.002.
[131] Iyengar S, Farnham PJ. KAP1 protein: an enigmatic master regulator of the genome. J Biol Chem 2011;286:26267–76. https://doi.org/10.1074/jbc.R111.252569.
[132] Tao Y, Yen M-R, Chitiashvili T, Nakano H, Kim R, Hosohama L, Tan YC, Nakano A, Chen P-Y, Clark AT. TRIM28-regulated transposon repression is required for human germline competency and not primed or naive human pluripotency. Stem Cell Rep 2018;10:243–56. https://doi.org/10.1016/j.stemcr.2017.11.020.
[133] Theunissen TW, Friedli M, He Y, Planet E, O'Neil RC, Markoulaki S, Pontis J, Wang H, Iouranova A, Imbeault M, Duc J, Cohen MA, Wert KJ, Castanon R, Zhang Z, Huang Y, Nery JR, Drotar J, Lungjangwa T, Trono D, Ecker JR, Jaenisch R. Molecular criteria for defining the naive human pluripotent state. Cell Stem Cell 2016;19:502–15. https://doi.org/10.1016/j.stem.2016.06.011.
[134] Wang J, Rao S, Chu J, Shen X, Levasseur DN, Theunissen TW, Orkin SH. A protein interaction network for pluripotency of embryonic stem cells. Nature 2006;444:364–8.
[135] Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. Proc Natl Acad Sci U S A 2015;112:8367–72. https://doi.org/10.1073/pnas.1422187112.
[136] Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jönsson ME, Männe C, Turelli P, Jern P, Cammas F, Trono D, Jakobsson J. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. Cell Rep 2015;10:20–8. https://doi.org/10.1016/j.celrep.2014.12.004.
[137] Kato M, Takemoto K, Shinkai Y. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. Nat Commun 2018;9:1–13. https://doi.org/10.1038/s41467-018-04132-9.
[138] Tie CH, Fernandes L, Conde L, Robbez-Masson L, Sumner RP, Peacock T, Rodriguez Plata MT, Mickute G, Gifford R, Towers GJ, Herrero J, Rowe HM. KAP1 regulates endogenous retroviruses in adult human cells

and contributes to innate immune control. EMBO Rep 2018; https://doi.org/10.15252/embr.201745000.e45000-13.
[139] Wolf G, Yang P, Füchtbauer AC, Füchtbauer E-M, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev 2015;29:538–54. https://doi.org/10.1101/gad.252767.114.
[140] Ohtani H, Liu M, Zhou W, Liang G, Jones PA. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res 2018;28:1147–57. https://doi.org/10.1101/gr.234229.118.
[141] Trono D, Marzetta F, Simó-Riudalbas L, Duc J, Planet E, Verp S, Turelli P. The KZFP/KAP1 system controls transposable elements-embedded regulatory sequences in adult T cells. bioRxiv 2019; https://doi.org/10.1101/523597.
[142] Guallar D, Bi X, Pardavila JA, Huang X, Saenz C, Shi X, Zhou H, Faiola F, Ding J, Haruehanroengra P, Yang F, Li D, Sanchez-Priego C, Saunders A, Pan F, Valdes VJ, Kelley K, Blanco MG, Chen L, Wang H, Sheng J, Xu M, Fidalgo M, Shen X, Wang J. RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. Nat Genet 2018;30:733. https://doi.org/10.1038/s41588-018-0060-9.
[143] Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature 2009;458:1201–4. https://doi.org/10.1038/nature07844.
[144] Wolf D, Goff SP. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell 2007;131:46–57. https://doi.org/10.1016/j.cell.2007.07.026.
[145] Ecco G, Imbeault M, Trono D. KRAB zinc finger proteins. Development (Camb Engl) 2017;144:2719–29. https://doi.org/10.1242/dev.132605.
[146] Helleboid P-Y, Heusel M, Duc J, Piot C, Thorball CW, Coluccio A, Pontis J, Imbeault M, Turelli P, Aebersold R, Trono D. The interactome of KRAB zinc finger proteins reveals the evolutionary history of their functional diversification. EMBO J 2019;38:e101220. https://doi.org/10.15252/embj.2018101220.
[147] Imbeault M, Helleboid P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 2017;543:550–4. https://doi.org/10.1038/nature21683.
[148] Guallar D, Pérez-Palacios R, Climent M, Martínez-Abadía I, Larraga A, Fernández-Juan M, Vallejo C, Muniesa P, Schoorlemmer J. Expression of endogenous retroviruses is negatively regulated by the pluripotency marker Rex1/Zfp42. Nucleic Acids Res 2012;40:8993–9007. https://doi.org/10.1093/nar/gks686.
[149] Schlesinger S, Lee AH, Wang GZ, Green L, Goff SP. Proviral silencing in embryonic cells is regulated by Yin Yang 1. Cell Rep 2013;4:50–8. https://doi.org/10.1016/j.celrep.2013.06.003.
[150] Schoorlemmer J, Pérez-Palacios R, Climent M, Guallar D, Muniesa P. Regulation of mouse retroelement MuERV-L/MERVL expression by REX1 and epigenetic control of stem cell potency. Front Oncol 2014;4:14. https://doi.org/10.3389/fonc.2014.00014.
[151] Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszynski LA. Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature 2015;522:240–323. https://doi.org/10.1038/nature14345.
[152] Fukuda K, Okuda A, Yusa K, Shinkai Y. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. Genome Res 2018;28:846–58. https://doi.org/10.1101/gr.227280.117.
[153] Hayashi M, Maehara K, Harada A, Semba Y, Kudo K, Takahashi H, Oki S, Meno C, Ichiyanagi K, Akashi K, Ohkawa Y. Chd5 regulates MuERV-L/MERVL expression in mouse embryonic stem cells via H3K27me3 modification and histone H3.1/H3.2. J Cell Biochem 2016;117:780–92. https://doi.org/10.1002/jcb.25368.
[154] Hisada K, Sanchez C, Endo TA, Endoh M, Roman-Trufero M, Sharif J, Koseki H, Vidal M. RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. Mol Cell Biol 2012;32:1139–49. https://doi.org/10.1128/MCB.06441-11.

[155] Hummel B, Hansen EC, Yoveva A, Aprile-Garcia F, Hussong R, Sawarkar R. The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. Nat Struct Mol Biol 2017;24:234–42. https://doi.org/10.1038/nsmb.3368.
[156] Li P, Wang L, Bennett BD, Wang J, Li J, Qin Y, Takaku M, Wade PA, Wong J, Hu G. Rif1 promotes a repressive chromatin state to safeguard against endogenous retrovirus activation. Nucleic Acids Res 2017;45:12723–38. https://doi.org/10.1093/nar/gkx884.
[157] Maksakova IA, Goyal P, Bullwinkel J, Brown JP, Bilenky M, Mager DL, Singh PB, Lorincz MC. H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics Chromatin 2011;4:12. https://doi.org/10.1186/1756-8935-4-12.
[158] Sachs P, Ding D, Bergmaier P, Lamp B, Schlagheck C, Finkernagel F, Nist A, Stiewe T, Mermoud JE. SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. Nat Commun 2019;10:1–16. https://doi.org/10.1038/s41467-019-09078-0.
[159] Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, Theis FJ, Schotta G. Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep 2015;16:836–50. https://doi.org/10.15252/embr.201439937.
[160] Wolf D, Cammas F, Losson R, Goff SP. Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. J Virol 2008;82:4675–9.
[161] Yang BX, El Farran CA, Guo HC, Yu T, Fang HT, Wang HF, Schlesinger S, Seah YFS, Goh GYL, Neo SP, Li Y, Lorincz MC, Tergaonkar V, Lim T-M, Chen L, Gunaratne J, Collins JJ, Goff SP, Daley GQ, Li H, Bard FA, Loh Y-H. Systematic identification of factors for provirus silencing in embryonic stem cells. Cell 2015;163:230–45. https://doi.org/10.1016/j.cell.2015.08.037.
[162] Liu M, Thomas SL, DeWitt AK, Zhou W, Madaj ZB, Ohtani H, Baylin SB, Liang G, Jones PA. Dual inhibition of DNA and histone methyltransferases increases viral mimicry in ovarian cancer cells. Cancer Res 2018;78:5754–66. https://doi.org/10.1158/0008-5472.CAN-17-3953.
[163] Robbez-Masson L, Tie CHC, Conde L, Tunbak H, Husovsky C, Tchasovnikarova IA, Timms RT, Herrero J, Lehner PJ, Rowe HM. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res 2018;28:836–45. https://doi.org/10.1101/gr.228171.117.
[164] Tchasovnikarova IA, Timms RT, Douse CH, Roberts RC, Dougan G, Kingston RE, Modis Y, Lehner PJ. Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. Nat Genet 2017;49:1035–44. https://doi.org/10.1038/ng.3878.
[165] Tchasovnikarova IA, Timms RT, Matheson NJ, Wals K, Antrobus R, Gottgens B, Dougan G, Dawson MA, Lehner PJ. Gene silencing. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. Science 2015;348:1481–5. https://doi.org/10.1126/science.aaa7227.
[166] Cruz-Tapias P, Robin P, Pontis J, Maestro LD, Ait-Si-Ali S. The H3K9 methylation writer SETDB1 and its reader MPP8 cooperate to silence satellite DNA repeats in mouse embryonic stem cells. Genes (Basel) 2019;10:https://doi.org/10.3390/genes10100750.
[167] Timms RT, Tchasovnikarova IA, Lehner PJ. Position-effect variegation revisited: HUSHing up heterochromatin in human cells. Bioessays 2016;38:333–43. https://doi.org/10.1002/bies.201500184.
[168] Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol 2017;18:407–22. https://doi.org/10.1038/nrm.2017.26.
[169] Hota SK, Bruneau BG. ATP-dependent chromatin remodeling during mammalian development. Development 2016;143:2882–97. https://doi.org/10.1242/dev.128892.
[170] Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. Cell 2013;154:490–503. https://doi.org/10.1016/j.cell.2013.07.011.
[171] Schultz DC, Friedman JR, Rauscher FJ. Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. Genes Dev 2001;15:428–43. https://doi.org/10.1101/gad.869501.
[172] Ding XZ, Liang CN, Guo X, Xing CF, Bao PJ, Chu M, Pei J, Zhu XS, Yan P. A novel single nucleotide polymorphism in exon 7 of LPL gene and its association with carcass traits and visceral fat deposition in yak (*Bos grunniens*) steers. Mol Biol Rep 2011; https://doi.org/10.1007/s11033-011-0784-4.

[173] Durand-Dubief M, Will WR, Petrini E, Theodorou D, Harris RR, Crawford MR, Paszkiewicz K, Krueger F, Correra RM, Vetter AT, Miller JR, Kent NA, Varga-Weisz P. SWI/SNF-like chromatin remodeling factor Fun30 supports point centromere function in *S. cerevisiae*. PLoS Genet 2012;8:e1002974. https://doi.org/10.1371/journal.pgen.1002974.

[174] Steglich B, Stralfors A, Khorosjutina O, Persson J, Smialowska A, Javerzat JP, Ekwall K. The Fun30 chromatin remodeler Fft3 controls nuclear organization and chromatin structure of insulators and subtelomeres in fission yeast. PLoS Genet 2015;11. https://doi.org/10.1371/journal.pgen.1005101.

[175] Xiao S, Lu J, Sridhar B, Cao X, Yu P, Zhao T, Chen CC, McDee D, Sloofman L, Wang Y, Rivas-Astroza M, Telugu BP, Levasseur D, Zhang K, Liang H, Zhao JC, Tanaka TS, Stormo G, Zhong S. SMARCAD1 contributes to the regulation of naive pluripotency by interacting with histone citrullination. Cell Rep 2017;18:3117–28. https://doi.org/10.1016/j.celrep.2017.02.070.

[176] Persson J, Steglich B, Smialowska A, Boyd M, Bornholdt J, Andersson R, Schurra C, Arcangioli B, Sandelin A, Nielsen O, Ekwall K. Regulating retrotransposon activity through the use of alternative transcription start sites. EMBO Rep 2016;17:753–68.

[177] Jang SM, Kauzlaric A, Quivy JP, Pontis J, Rauwel B, Coluccio A, Offner S, Duc J, Turelli P, Almouzni G, Trono D. KAP1 facilitates reinstatement of heterochromatin after DNA replication. Nucleic Acids Res 2018;46:8788–802. https://doi.org/10.1093/nar/gky580.

[178] Okazaki N, Ikeda S, Ohara R, Shimada K, Yanagawa T, Nagase T, Ohara O, Koga H. The novel protein complex with SMARCAD1/KIAA1122 binds to the vicinity of TSS. J Mol Biol 2008;382:257–65. https://doi.org/10.1016/j.jmb.2008.07.031.

[179] Rowbotham SP, Barki L, Neves-Costa A, Santos F, Dean W, Hawkes N, Choudhary P, Will WR, Webster J, Oxley D, Green CM, Varga-Weisz P, Mermoud JE. Maintenance of silent chromatin through replication requires SWI/SNF-like chromatin remodeler SMARCAD1. Mol Cell 2011;42:285–96. https://doi.org/10.1016/j.molcel.2011.02.036.

[179a] Taneja N, Zofall M, Balachandran V, et al. SNF2 family protein Fft3 suppresses nucleosome turnover to promote epigenetic inheritance and proper replication. Mol Cell 2017;66(1):50–62.e6. http://dx.doi.org/10.1016/j.molcel.2017.02.006.

[180] He Q, Kim H, Huang R, Lu W, Tang M, Shi F, Yang D, Zhang X, Huang J, Liu D, Songyang Z. The Daxx/Atrx complex protects tandem repetitive elements during DNA hypomethylation by promoting H3K9 trimethylation. Cell Stem Cell 2015;17:273–86. https://doi.org/10.1016/j.stem.2015.07.022.

[181] Voon HP, Hughes JR, Rode C, De La Rosa-Velazquez IA, Jenuwein T, Feil R, Higgs DR, Gibbons RJ. ATRX plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. Cell Rep 2015;11:405–18. https://doi.org/10.1016/j.celrep.2015.03.036.

[182] Voon HP, Wong LH. New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone. Nucleic Acids Res 2016;44:1496–501. https://doi.org/10.1093/nar/gkw012.

[183] Hoelper D, Huang H, Jain AY, Patel DJ, Lewis PW. Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. Nat Commun 2017;8:1193. https://doi.org/10.1038/s41467-017-01206-y.

[184] Drane P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev 2010;24:1253–65. https://doi.org/10.1101/gad.566910.

[185] Goldberg AD, Banaszynski LA, Noh K-MM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chapgier A, DeKelver RC, Miller JC, Lee Y-LL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 2010;140:678–91. https://doi.org/10.1016/j.cell.2010.01.003.

[186] Lewis PW, Elsaesser SJ, Noh KM, Stadler SC, Allis CD. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A 2010;107:14075–80. https://doi.org/10.1073/pnas.1008850107.

[187] Wolf G, Rebollo R, Karimi MM, Ewing AD, Kamada R, Wu W, Wu B, Bachu M, Ozato K, Faulkner GJ, Mager DL, Lorincz MC, Macfarlan TS. On the role of H3.3 in retroviral silencing. Nature 2017;548:E1–3. https://doi.org/10.1038/nature23277.
[188] Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol 2011;12:246–58. https://doi.org/10.1038/nrm3089.
[189] Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev Biol 2004;269:276–85. https://doi.org/10.1016/j.ydbio.2004.01.028.
[190] Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 2008;453:534–8. https://doi.org/10.1038/nature06904.
[191] Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, Surani MA, Sakaki Y, Sasaki H. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 2008;453:539–43. https://doi.org/10.1038/nature06908.
[192] Schorn AJ, Martienssen R. Tie-break: host and retrotransposons play tRNA. Trends Cell Biol 2018;28:793–806. https://doi.org/10.1016/j.tcb.2018.05.006.
[193] Svoboda P. Renaissance of mammalian endogenous RNAi. FEBS Lett 2014;588:2550–6. https://doi.org/10.1016/j.febslet.2014.05.030.
[194] Castaneda J, Genzor P, Bortvin A. piRNAs, transposon silencing, and germline genome integrity. Mutat Res 2011;714:95–104. https://doi.org/10.1016/j.mrfmmm.2011.05.002.
[195] Grivna ST, Beyret E, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev 2006;20:1709–14. https://doi.org/10.1101/gad.1434406.
[196] Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 2007;316:744–7. https://doi.org/10.1126/science.1142612.
[197] Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev 2008;22:908–17. https://doi.org/10.1101/gad.1640708.
[198] Iwasaki YW, Siomi MC, Siomi H. PIWI-interacting RNA: its biogenesis and functions. Annu Rev Biochem 2015;84:405–33. https://doi.org/10.1146/annurev-biochem-060614-034258.
[199] Hyun S. Small RNA pathways that protect the somatic genome. Int J Mol Sci 2017;18. https://doi.org/10.3390/ijms18050912.
[200] Yu T, Koppetsch BS, Pagliarani S, Johnston S, Silverstein NJ, Luban J, Chappell K, Weng Z, Theurkauf WE. The piRNA response to retroviral invasion of the koala genome. Cell 2019;179:632–43. e12 https://doi.org/10.1016/j.cell.2019.09.002.
[201] Jin D, Musier-Forsyth K. Role of host tRNAs and aminoacyl-tRNA synthetases in retroviral replication. J Biol Chem 2019;294:5352–64. https://doi.org/10.1074/jbc.REV118.002957.
[202] Schorn AJ, Gutbrod MJ, LeBlanc C, Martienssen R. LTR-retrotransposon control by tRNA-derived small RNAs. Cell 2017;170:61–71. e11. https://doi.org/10.1016/j.cell.2017.06.013.
[203] Schlesinger S, Meshorer E. Open chromatin, epigenetic plasticity, and nuclear organization in pluripotency. Dev Cell 2019;48:135–50. https://doi.org/10.1016/j.devcel.2019.01.003.