CellPress

Molecular Cell
Review

Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire

Peter J. Thompson,[1,3] Todd S. Macfarlan,[2,*] and Matthew C. Lorincz[1,*]
[1]Department of Medical Genetics, Life Sciences Institute, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[2]Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD 20892, USA
[3]Present address: Diabetes Center, Department of Medicine, University of California, San Francisco, San Francisco, CA 94143, USA
*Correspondence: todd.macfarlan@nih.gov (T.S.M.), mlorincz@mail.ubc.ca (M.C.L.)
http://dx.doi.org/10.1016/j.molcel.2016.03.029

The life cycle of endogenous retroviruses (ERVs), also called long terminal repeat (LTR) retrotransposons, begins with transcription by RNA polymerase II followed by reverse transcription and re-integration into the host genome. While most ERVs are relics of ancient integration events, "young" proviruses competent for retrotransposition—found in many mammals, but not humans—represent an ongoing threat to host fitness. As a consequence, several restriction pathways have evolved to suppress their activity at both transcriptional and post-transcriptional stages of the viral life cycle. Nevertheless, accumulating evidence has revealed that LTR sequences derived from distantly related ERVs have been exapted as regulatory sequences for many host genes in a wide range of cell types throughout mammalian evolution. Here, we focus on emerging themes from recent studies cataloging the diversity of ERV LTRs acting as important transcriptional regulatory elements in mammals and explore the molecular features that likely account for LTR exaptation in developmental and tissue-specific gene regulation.

Introduction

Retrotransposons, which replicate via a transcription and reverse-transcription "copy-and-paste" mechanism, account for greater than 40% of the human and mouse genomes (Venter et al., 2001; Waterston et al., 2002). These parasitic sequences can be classified into two major groups. Those lacking long terminal repeats (LTRs), including long and short interspersed nuclear elements (LINEs and SINEs, respectively) and SINE variable-number tandem-repeat Alu (SVA) elements, comprise ~30%–35% of the genome, while those with LTRs, termed endogenous retroviruses (ERVs) or LTR retrotransposons, comprise ~8% and 10% of the human and mouse genomes, respectively (Cordaux and Batzer, 2009; Friedli and Trono, 2015; Stocking and Kozak, 2008) (Figure 1A). ERVs are the descendants of exogenous retroviruses that integrated into the genome of germ cells. Most subsequently lost the ability to exit the host cell. Thus, those ERVs that may be defective for infection but are still competent for retrotransposition expand in their host genome by vertical transmission (Mager and Stoye, 2015; Magiorkinis et al., 2012). In addition to their 5′ and 3′ LTRs, which are identical in sequence following reverse transcription and integration, autonomous proviral elements typically harbor several open reading frames (ORFs) that encode proteins essential for viral replication, including *gag*, which encodes a group-specific retroviral antigen, and *pol*, which encodes the reverse transcriptase (Figure 1B). A third ORF encodes an envelope protein (*env*), although the vast majority of ERVs have truncated or mutated *env* sequences.

While the general threat of insertional mutagenesis due to unmitigated ERV transcription and subsequent retrotransposition is minimized by epigenetic mechanisms, including DNA methylation, histone lysine methylation, and small noncoding RNAs (Castro-Diaz et al., 2015; Wolf et al., 2015a), recent studies have revealed that ERVs have also played a prominent role in expanding the regulatory landscape of mammalian genomes (Cordaux and Batzer, 2009; Feschotte and Gilbert, 2012; Gifford et al., 2013; Jern and Coffin, 2008; Rebollo et al., 2012a). The "controlling element" theory that transposable elements (TEs) may participate in gene regulation was postulated over 60 years ago by Barbara McClintock (McClintock, 1950) and was later expanded upon by Britten and Davidson's gene battery hypothesis (Britten and Davidson, 1969). Genome-wide studies have indeed confirmed that species-specific ERV LTRs exert regulatory effects on genes in many cell types during development to modulate the transcriptome (Cowley and Oakey, 2013; Gifford et al., 2013; Isbel and Whitelaw, 2012; Robbez-Masson and Rowe, 2015). However, the molecular mechanisms whereby these heterologous sequences are converted into regulatory elements for host genes remain obscure. Here, we highlight recent studies that have advanced our understanding of how LTR sequences are exapted into species-specific *cis*-regulatory elements. We begin by exploring why LTR retrotransposons are particularly suitable for co-option by the host and subsequently review recent experimental evidence supporting a model of reiterative exaptation of LTRs in mammals as tissue-specific promoters or enhancers for protein-coding genes and long noncoding RNAs (lncRNAs). We conclude with a discussion of recent functional studies of the role of specific exapted LTRs in gene regulation and outstanding questions to be addressed in future studies.

Solo LTRs: Autonomous Regulatory Modules

Several studies in mammals indicate that ERVs have been more frequently exapted as *cis*-regulatory elements relative to other

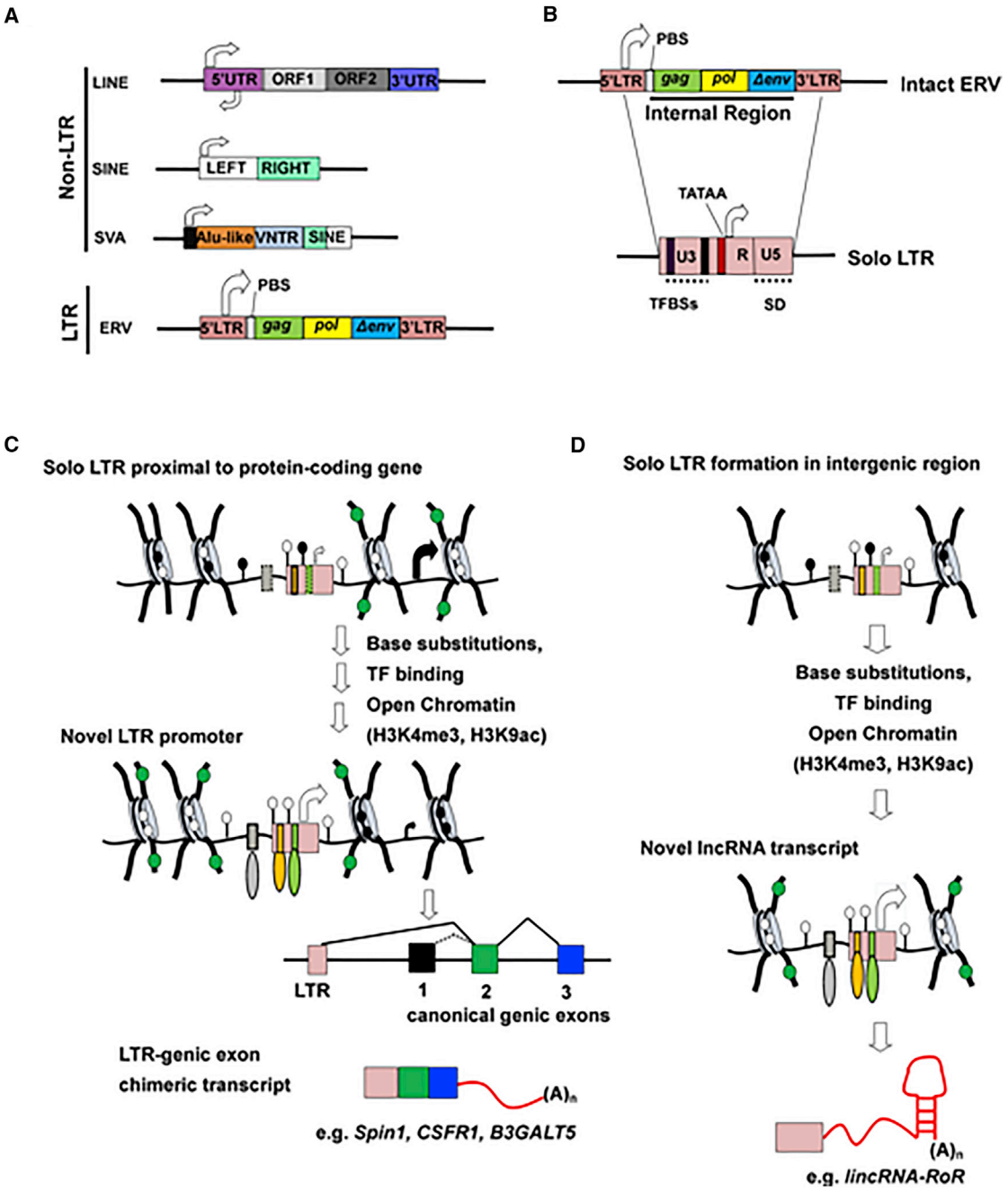

Figure 1. Structure of an Intact ERV and Solo LTR and the Molecular Mechanisms of LTR Exaptation as Protein-Coding or lncRNA Promoters

(A) Schematic of non-LTR retrotransposons, which include SINEs (i.e., Alus), LINEs (i.e., L1Hs), and SVAs (in humans), and LTR retrotransposons, which include many lineage/species-specific subfamilies. Most LINE elements are truncated at the 5′ end, thus lacking the 5′ UTR promoter and TSS.

(B) Full-length ERVs have 5′ and 3′ LTRs, and an "internal" region that includes a primer-binding site (PBS) involved in priming reverse transcription and retroviral ORFs *gag*, *pol*, and a truncated or mutated *env* gene (*Δenv*). Recombination between 5′ and 3′ LTRs deletes the internal region, generating "solo" LTRs (not to scale), which consist of unique 3′ (U3) and 5′ (U5) regions and a regulatory region (R) containing the TSS (white arrow). LTRs often harbor different combinations of TFBSs (green and orange rectangles) in addition to core polymerase II promoter elements (such as TATA box, shown in red) and may also contain a splice donor (SD) site (dashed line) within the U5 region.

(C) LTR exaptation as a protein-coding gene promoter. In a developmental/tissue-specific context, particularly in cell types undergoing epigenetic reprogramming (e.g., early embryo, placenta, or germline), a hypomethylated solo LTR (pink rectangle) 5′ of a protein-coding gene (or in an intragenic region) may become exapted as a novel promoter (black circles represent DNA methylation). The process may involve base substitutions in a neighboring near-consensus TFBS (gray rectangle with dash outline) and a near-consensus site within the LTR (green rectangle with dash outline), which then form a positive genetic interaction with another LTR-derived TFBS (orange rectangle), a mechanism termed "epistatic capture" (Emera and Wagner, 2012a). This leads to synergy in the binding of several TFs (gray, orange, and green ovals); deposition of "active" histone modifications, such as H3K4me3 and H3K9ac (green circles); and robust

(legend continued on next page)

Table 1. Examples of LTRs Exapted as Regulatory Elements in Human and Mouse and Their Classification

Species	ERV Class	ERV Family	Examples of LTR Subtypes
Mouse	I	ERV1	LTR17
	I	ERV3	MER77
	II	ERVK	LTR10C, LTR10B, LTR13D5, BGLII
	III	MaLR	MT-A, MT-B, MT-C, ORR1A0
	III	ERVL	MT2, MT2B, MT2C
Human	I	ERV1	LTR7, MER39, MER41, LTR12C
	I	ERV3	MER21C
	I	ERV9	LTR9
	II	ERVK	LTR5
	II	ERVK3	LTR3B
	II	ERVK14	LTR14B
	III	MaLR	THE1A, THE1B, THE1C, MER39
	III	ERVL	MLT2A1, MLT2B3, LTR16A

TEs (Chuong et al., 2013; Jacques et al., 2013; Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012; Sundaram et al., 2014; Xie et al., 2013). Consistent with these observations, ERVs have been reported to evolve more rapidly than other TEs, as evidenced by orthologous ERVs in humans and chimpanzees exhibiting signatures of directional selection since the human-chimp divergence ~five million years ago (Gemmell et al., 2015). These observations are unlikely to be explained by the integration sites of ERVs, as retrotransposons of all types are most prevalent in intergenic regions, and older LTR and LINE elements are underrepresented within 5 kb of gene promoters, perhaps due to their negative impact on expression of proximal genes and, in turn, host fitness (Medstrand et al., 2002). Rather, the frequent co-option of ERV sequences for gene regulation may be due to the relatively high probability of recombination between the 5′ and 3′ LTRs of intact proviruses, which deletes the internal region, leaving a single or "solo" LTR at the original integration site (Belshaw et al., 2007) (Figure 1B). Recombination between 5′ and 3′ LTRs has generated an estimated 577,000 "solo" LTRs in the human genome, representing the vast majority of annotated ERV sequences (Friedli and Trono, 2015). Notably, both full-length intact ERVs and solo LTRs are underrepresented specifically in the sense orientation within introns, likely reflecting the generally deleterious effects of insertion of polyadenylation signals encoded by LTRs (Medstrand et al., 2002; Smit, 1999). As LTRs harbor the regulatory regions required for proviral transcription, generally including combinations of transcription factor binding sites (TFBSs), they have the intrinsic capacity to autonomously recruit cellular TFs and in turn to maximize transcription of proviral mRNA in specific cell types. Indeed, LTR-derived TFBSs are now known to have contributed up to ~20% of functional binding sites for many TFs in human and mouse (Sundaram et al., 2014), including p53, OCT4, SOX2, and NANOG (Bourque et al., 2008; Kunarso et al., 2010; Wang et al., 2007). In contrast with the regulatory properties of LTRs, the majority of LINE1 elements are truncated at the 5′end, which removes the regulatory region and canonical transcription start site (TSS) (Figure 1A) of these RNA polymerase II-driven elements, rendering them transcriptionally "dead on arrival" (Cordaux and Batzer, 2009).

The presence of a conserved splice donor (SD) site within some classes of LTRs also likely contributes to the propensity for LTRs from specific families to be exapted as alternative promoters (Figure 1B). The consensus sequence of MaLR LTRs, for example, including the mouse transcript (MT) subtypes (see Table 1 for classification of LTRs exapted as regulatory elements discussed here), harbors a conserved SD site that is utilized in many MT-initiated chimeric transcripts in oocytes (Peaston et al., 2004). Similarly, a primate-specific MaLR LTR, THE1B, which harbors an intact SD site, is aberrantly reactivated in Hodgkin's lymphoma and drives expression of *CSF1R* transcripts (Lamprecht et al., 2010). Alternatively, mutations within LTRs may generate novel SD sites, as is the case for the highly expressed oocyte-specific *Spin1* transcript, also driven by an MT LTR (Peaston et al., 2004). Furthermore, at specific loci, cryptic SD sites may be present in the flanking genomic sequence downstream of a transcriptionally active LTR. Regardless, the presence of an SD site within or immediately downstream of the LTR minimizes the length of the 5′ UTR. This decreases the likelihood that the transcript will contain a cryptic start codon upstream of the canonical start codon, thus preserving the native ORF in the resulting chimeric mRNA, and may stabilize the nascent RNA, as SD sites may compete with termination signals (Wu and Sharp, 2013).

Many intact ERVs are targeted for transcriptional silencing by the rapidly diversifying family of Krüppel-associated box zinc-finger proteins (KRAB-ZFPs), which interact with the co-repressor KAP1 and the histone H3 lysine 9 (H3K9) methyltransferase SETDB1 (Liu et al., 2014; Matsui et al., 2010; Rowe et al., 2010; Turelli et al., 2014; Wolf et al., 2015b). Indeed, chromatin immunoprecipitation sequencing (ChIP-seq) analysis in mouse embryonic stem cells (ESCs) reveals that the solo LTRs of a subset of ERV families, including IAP solo LTRs, are marked by H3K9me3 (Karimi et al., 2011), indicating that for some ERVs, the LTR itself may be bound by specific KRAB-ZFPs. However, while the binding sites of only a few of the >300–400 KRAB-ZFPs in humans and mice have been studied, the majority characterized thus far recognize internal ERV sequences, including the primer binding site, 5′ UTR, *gag*, and 3′ polypurine tract regions (Rowe et al., 2010; Sadic et al., 2015; Wolf and Goff, 2009; Wolf et al., 2015b, Ecco et al., 2016). Since solo LTRs lack these internal sequences, they may escape the

transcription initiation from the LTR-derived promoter. The canonical genic promoter may be DNA methylated as a consequence of such transcription. This process generates LTR-genic exon chimeric transcripts, where exon 1 is derived from the LTR and splicing occurs from the internal LTR SD (or from a cryptic SD site in the intervening genomic sequence downstream of the active LTR) to the first downstream exon with a splice-acceptor site, generally exon 2. Examples of such chimeric transcripts include *Spin1* in mouse and *CSFR1* and *B3GALT5* in human. Arrow sizes indicate relative level of transcription from each promoter. (D) LTR exaptation as a promoter for a novel lncRNA. Through a process as in (C), a newly formed intergenic solo LTR without an SD site could initiate de novo lncRNA transcription, forming a novel lncRNA gene. An example of this is the *lincRNA-RoR* transcript.

KRAB-ZFP/KAP1 silencing machinery directed at full-length elements, facilitating their exaptation as positive regulatory elements by the host.

Once all members of a particular ERV family are effectively silenced by the KRAB-ZFP/KAP1 repression system, the accumulation of inactivating mutations in replication-competent proviruses, i.e., in functional viral protein-coding regions, would over time relieve the positive selective pressure for KRAB-ZFP recognition, allowing mutations to accumulate within the relevant KRAB-ZFP gene, ultimately modifying or ablating the DNA-binding specificity of the encoded protein regardless of whether it binds in the LTR or internal region. The remaining replication-incompetent full-length proviruses and solo LTRs derived from these elements would no longer be recognized by a specific KRAB-ZFP, allowing for selection of LTRs as promoter or enhancer elements of nearby genes (Friedli and Trono, 2015). This does not exclude the possibility that there may be purifying selection of KRAB-ZFP binding sites within otherwise decaying ERV internal regions or LTRs, allowing for ERV exaptation for silencing of nearby genes (Ecco et al., 2016).

Consistent with the presence of TFBSs and their propensity to evade epigenetic silencing, many ERVs and LTRs exhibit tissue-specific expression patterns, especially during embryonic and germline development (Göke et al., 2015; Grow et al., 2015; Jacques et al., 2013; Okahara et al., 2004; Pavlicev et al., 2015; Peaston et al., 2004). Indeed, ERVs have likely been under selection to increase their odds of successful retrotransposition and vertical transmission and therefore exhibit high levels of transcription in the early embryo and reproductive tissues, including primordial germ cells (PGCs) and oocytes (Cohen et al., 2009; Peaston et al., 2004). Thus, it is not surprising that the ERV families present in high copy number are also those competent for expression in the germline. Although H3K9me3 and/or DNA methylation play a role in silencing of ERVs in both undifferentiated and differentiated cell types, specific ERVs likely exploit global reprogramming of epigenetic states, such as during embryonic preimplantation development (Tomizawa et al., 2011; Ziller et al., 2013) or in the placenta (Chuong et al., 2013; Hon et al., 2013; Reiss et al., 2007; Xie et al., 2013), to promote their expression. During these developmental stages, the LTRs that have accumulated mutations that relieve selective pressure for KRAB-ZFP-based silencing in these cell types, or are otherwise not efficiently bound by KRAB-ZFPs due to low level of expression of the relevant KRAB-ZFP or their genomic context, would come under purifying selection for beneficial regulatory effects on neighboring genes. Thus, the combination of autonomous RNA polymerase II promoter/enhancer activity conferred by intact TF binding sites and changes in the repertoire of ERVs bound by KRAB-ZFPs over evolutionary time is likely to provide a unique context for exaptation of solo LTRs for tissue-specific gene regulation.

From LTR to Genic Promoter

In support of this model, a substantial number of LTRs have been reported to function as tissue-specific primary or alternative promoters in a variety of mammalian cell types, including in the early mouse embryo, placenta, human and mouse pluripotent stem cells, mouse erythroid cells, and growing mouse oocytes (Buzdin et al., 2006; Cohen et al., 2009; Faulkner et al., 2009; Fort et al., 2014; Karimi et al., 2011; Macfarlan et al., 2012; Mak et al., 2014; Peaston et al., 2004; Veselovska et al., 2015; Wolf et al., 2015b). Notably, many of the LTRs that have apparently been exapted as genic promoters are not only lineage specific but also show clear differences in transcriptional activity between cell types in the given species. For example, in mouse zygote and two-cell-stage embryos, LTRs from the class III LTR retrotransposon MERVL drive expression of a cohort of stage-specific genes (Evsikov et al., 2004; Macfarlan et al., 2012; Maksakova et al., 2013; Peaston et al., 2004), whereas MaLR and ERVK family LTRs drive expression of many mouse oocyte-specific transcripts (Peaston et al., 2004; Veselovska et al., 2015). Similarly, in human pluripotent stem cells, LTR7, derived from the primate-specific HERV-H, drives transcription of many pluripotency-associated lncRNAs (Durruthy-Durruthy et al., 2016; Lu et al., 2014b; Wang et al., 2014). Furthermore, LTR3B, LTR14B, LTR12C, MLT2A1, THE1A, and LTR5_Hs are expressed at discrete stages during the progression of human preimplantation embryo development from the zygote to the morula stage and serve as promoters for a class of previously unannotated transcripts that may serve important functions at these stages (Göke et al., 2015).

Early studies of the role of LTR elements as candidate genic promoters relied on single-gene analyses using methods such as 5′ rapid amplification of cDNA ends (RACE) or PCR. Subsequently, higher-throughput approaches were developed, including those based on sequence mining of EST or RefSeq databases (Evsikov et al., 2004; van de Lagemaat et al., 2003; Lipatov et al., 2005; Medstrand et al., 2002; Peaston et al., 2004), or the combination of EST data with high-throughput sequencing by capped analysis of gene expression (CAGE) (Faulkner et al., 2009).

With the widespread use of next-generation sequencing (NGS) technologies and complementary development of bioinformatics tools to exploit such datasets, novel transcripts, including those expressed at relatively low levels, can now be easily identified and enumerated. Indeed, LTR promoter usage in a given cell type can now be readily inferred genome-wide from RNA sequencing (RNA-seq) data. Paired-end RNA-seq data in particular have been used to identify candidate chimeric transcripts (Karimi et al., 2011; Macfarlan et al., 2012). RNA-seq data have also been employed for de novo transcriptome assembly to identify LTR promoter usage in an unbiased manner in the developing oocyte and to identify novel chimeric transcripts initiating in RLTR10B in mouse testis (Isbel et al., 2015; Veselovska et al., 2015). Recent technological advances have led to significant increases in library read depth and standard read lengths, increasing the probability of mapping unique reads within such repetitive elements and in turn the identification of chimeric transcripts showing a broad range of expression levels.

In addition, as active LTR promoters exhibit the same chromatin modification patterns found at active genic promoters, including H3K4me3 and DNase I hypersensitivity, profiling of these features by NGS can also be exploited to identify candidate LTR promoters (Chuong et al., 2013; Jacques et al., 2013; Lynch et al., 2011, 2015; Veselovska et al., 2015) (Figure 1C). For example, using ChIP-seq for H3K4me3 on cyclic AMP and progesterone-treated human decidualized stromal cells, Lynch

et al. (2015) found that ~31% of active promoters mapped in those cells overlap with ancient mammalian TEs, including LTRs (Lynch et al., 2015). Similarly, analysis of DNase I hypersensitivity data from a large panel of human embryonic, adult, and cancer cell lines revealed that up to ~80% of LTRs are located in open chromatin regions in a cell-type-specific manner (Jacques et al., 2013), and intersection with ENCODE H3K4me3 ChIP-seq data revealed that a subset of these LTRs are active promoters.

LTRs as Tissue-Specific Genic Promoters

Several LTRs derived from ancient proviruses that integrated near genes have likely been co-opted as regulatory elements, as indicated by strong purifying selection (Franchini et al., 2012; Lowe et al., 2007). The paucity of additional cases where LTR promoters/enhancers have been clearly shown to evolve under purifying selection may be due to weak selection or the fact that most instances of detectable LTR-derived regulatory elements are of recent origin (i.e., mouse or primate specific), limiting the statistical power to observe signatures of purifying selection by sequence comparisons among different lineages. If LTR-driven transcription is beneficial in a specific cell type, persistence of its promoter activity will be under selective pressure and the expression pattern maintained in that lineage. For example, while the *Dicer1* gene is driven from a CpG island promoter in most tissues where it is expressed, an oocyte-specific isoform in mice is driven by a rodent-specific intragenic MaLR solo LTR of the MT-C subtype (Flemr et al., 2013) (Table 1). Notably, deletion of the MT-C LTR alternative promoter abolishes *Dicer1* expression in the oocyte and causes female sterility, providing strong evidence of the importance of this exapted LTR for host fitness. As this specific LTR is also present in the rat, the ancestral provirus must have integrated prior to the divergence of rats and mice, at least ~25 million years ago (Nei et al., 2001). In contrast with the highly active IAP and ETn/MusD families that are responsible for ~10% of spontaneous mutations in laboratory mouse strains, there is no evidence for recent de novo retrotransposition of MT-C elements (Maksakova et al., 2006; Rebollo et al., 2012b). Nevertheless, LTRs from MT-C, as well as other MT subtypes, are still clearly transcriptionally active specifically in oocytes, reflecting the innate tissue-specific expression profile of these nonautonomous MaLR elements (Evsikov et al., 2004; Peaston et al., 2004; Veselovska et al., 2015). Similarly, the human metabolic gene *B3GALT5* is expressed in many different tissues, but in the colon, a primate-specific MLT2B3 LTR promoter, derived from the ERV-L family, is utilized (Dunn et al., 2003).

A particularly dramatic example of the widespread exaptation of a specific LTR subtype in a specific tissue can be found in early mouse embryogenesis, where MT2 LTRs derived from mouse MERVL elements act as promoters for over 500 two-cell stage-specific gene transcripts (Macfarlan et al., 2012; Maksakova et al., 2013). Although the functions of most of these MT2 LTR chimeric transcripts remain to be determined, a subset may serve important roles in early mouse development, such as *Tcstv1* and *Tcstv3*, which control telomere elongation and genome stability (Zhang et al., 2016). As with MT2 LTRs that serve as genic promoters, intact MERVL elements are also transcriptionally active at the zygote and two-cell stage, but are subsequently inactivated, at least in part as a consequence of a more repressive nuclear architecture instated during differentiation from totipotency to pluripotency, which is regulated by many chromatin modifiers (Hayashi et al., 2016; Hisada et al., 2012; Ishiuchi et al., 2015; Lu et al., 2014a; Macfarlan et al., 2012; Maksakova et al., 2013; Thompson et al., 2015). Thus, these LTR genic promoters likely retain the restricted tissue-specific expression pattern of the full-length ancestral provirus.

Further evidence for strong selective pressure for novel tissue-specific promoters of protein-coding genes can be inferred from the exaptation of different LTRs for orthologous genes in independent lineages. Emera and co-workers (2012) showed that MER39 and MER77 LTRs (Table 1) were independently exapted as novel promoters in primates and rodents, respectively, for the *Prolactin* gene, which is expressed in endometrial cells during pregnancy and essential for normal gestation (Emera et al., 2012). In addition, different LTRs have been independently exapted as promoters for the anti-apoptotic gene *NAIP*. In primates, testis-specific *NAIP* transcripts are driven by the MER21C LTR, while in rodents, *Naip* is expressed in many different tissues from ORR1E or MT-C LTRs (Romanish et al., 2007). Although these are isolated cases, many other instances of exaptation may have occurred earlier in mammalian evolution, with the regulatory elements in question no longer recognizable as LTRs.

In addition to promoting gene expression, the co-option of LTRs as promoters also provides the opportunity for TF-directed repression, as evidenced by a recent study that found that KLF3 enforces transcriptional repression of ORR1A0 LTR-driven transcripts in mouse fetal and adult erythroid cells (Mak et al., 2014). Whether suppression of such ORR1A0 LTR-driven chimeric transcripts serves only to prevent aberrant genic transcription emanating from the LTR remains to be determined. Chromatin modifiers may also direct the silencing of LTR-driven genes, similar to non-TE-derived genic promoters (Isbel et al., 2015; Karimi et al., 2011; Macfarlan et al., 2011; Wolf et al., 2015b). Thus, LTR promoters are apparently as versatile as typical genic promoters, allowing for both positive and negative regulation of their cognate genes. Similarly, purifying selection for KRAB-ZFP-directed gene repression may reflect the persistence of ancient KRAB-ZFP binding sites in degenerate LTRs and/or nonrepetitive regions near genes (Friedli and Trono, 2015).

Whether LTRs functioning as genic promoters generally exhibit substantial sequence differences relative to their ancestral sequence has not been systematically addressed. However, a recent study examining *Prolactin* expression in the placenta, which is driven by an MER39 LTR in various primate lineages, but not in non-ape species (Emera and Wagner, 2012a), sheds some light on the role of "fine-tuning" of LTR promoters. While the ancestral MER39 LTR present in all primates and rodents possessed an intact ETS1 binding site at the time of integration, this LTR was a weak promoter in non-ape species and was replaced by the MER77 LTR as the major *Prolactin* promoter in mice (Emera and Wagner, 2012a). However, over millions of years of ape evolution, MER39 was gradually transformed into a strong promoter by selection for base substitution mutations that synergized with the ancestral ETS1 site in the LTR and

consequently improved the strength of the promoter (Emera and Wagner, 2012a). Thus, although the primordial LTR possessed a functional TFBS, it was likely inefficient to act as a promoter in the placenta and required a series of substitutions to refine its activity. This finding is consistent with previous work showing that species-specific expression of genes near TEs is positively correlated with the number of bound TFBSs in the TE, with a minimum of two bound TFBSs to detect the correlation (Xie et al., 2010). The mechanism termed "epistatic capture" was proposed to describe the process by which a TE-derived TFBS comes under increased purifying selection as a consequence of epistatic interactions with nearby TFBSs refined by mutations over evolutionary time (Emera and Wagner, 2012b) (Figure 1C). Notably, this mechanism also accounts for the tissue specificity of LTR exaptation into promoters/enhancers, since the positive epistatic interactions between the TE-derived ancestral and newly derived TFBSs would be expected to occur only if they enhance recruitment of the TFs relevant to expression in that tissue. After the acquisition and selection for functional TFBSs within LTRs, the accumulation of additional mutations that are nonessential for their transcriptional activity will invariably lead to their progressive divergence from the ancestral sequence (Figure 1C). Indeed, LTRs co-opted as regulatory elements earlier in mammalian evolution may no longer be recognizable as repeat elements using conventional bioinformatics tools, raising the possibility that many more canonical gene promoters are actually derived from ancient LTRs.

While fewer cases have been identified, there is also evidence that recently integrated LTRs can function as *cis*-regulatory elements. Examples include mouse-specific LTR13D5 elements, which act as enhancers in the placenta; the primate-specific LTR9, which enhances *β-globin* gene expression; and the primate MLT2B3, which drives *B3GALT5* expression in the human colon (Chuong et al., 2013; Dunn et al., 2005; Pi et al., 2010). Thus, LTRs can also serve as "ready-made" enhancers or promoters without substantial sequence modification, potentially contributing to rapid evolution of gene regulatory networks (Cohen et al., 2009).

LTRs in lncRNA Expression

The role of TEs in lncRNA expression, function, and evolution is just beginning to emerge (Kapusta and Feschotte, 2014). Recent genome-wide surveys have revealed that 75%–80% of the ~10,000 annotated human lncRNAs contain TE sequences (Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012). Furthermore, LTRs show considerable enrichment in lncRNA transcripts compared with non-LTR elements and other TEs in mouse and human (Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012). While the majority of LTRs transcribed in lncRNAs serve as exons (Kannan et al., 2015), specific families have been co-opted as promoters. For example, many copies of the primate-specific LTR7 derived from human HERV-H are bound by pluripotency factors and function as essential regulatory elements in naive pluripotent stem cells, likely by driving expression of specific lncRNAs (Durruthy-Durruthy et al., 2016; Lu et al., 2014b; Ohnuki et al., 2014; Wang et al., 2014). In addition, genome-wide analysis suggests that many previously unannotated LTR-driven lncRNA transcripts are important for the maintenance of pluripotency in mouse and human (Fort et al., 2014). Deep sequencing of human preimplantation embryos has demonstrated the expression of stage-specific LTR-derived noncoding RNAs from a variety of ERV1, ERVK, and ERVL family ERVs; however, their functions remain to be determined (Göke et al., 2015; Grow et al., 2015).

Due to the versatility of lncRNA biogenesis and function, there are likely to be fewer constraints upon the exaptation of LTRs as lncRNA promoters. In addition to the basic regulatory properties of LTRs relevant to promoters for protein-coding genes, novel lncRNA genes could arise de novo from solo LTRs in intergenic regions (Friedli and Trono, 2015; Kapusta and Feschotte, 2014), which would also not necessarily require an intact SD site (Figure 1D). Since conserved lncRNAs such as *HOTAIR*, *lincRNA-RoR* (a HERV-H-derived lncRNA gene), *lncRNA-p21*, and *TUNAR* have been shown to regulate large cohorts of genes (Froberg et al., 2013; Huarte et al., 2010; Lin et al., 2014; Loewer et al., 2010; Rinn et al., 2007), a single LTR integration into a lncRNA gene has the potential to exert a broad regulatory effect on the transcriptome. Given the relatively rapid origins and turnover of lncRNA genes and their lack of high sequence conservation despite—in some cases—their clear functional conservation (Kapusta and Feschotte, 2014), it will be important to focus future investigations on the genome-wide contribution of LTRs to the genesis of novel lncRNA transcripts during mammalian evolution.

LTRs as Tissue-Specific Enhancers

A number of recent studies have revealed that LTRs have also contributed substantially to the formation of enhancers during mammalian evolution (Emera and Wagner, 2012b; Friedli and Trono, 2015). In fact, the majority of LTRs contributing to placenta-specific gene expression in humans (Pavlicev et al., 2015) and species-specific expression in mouse placenta (Chuong et al., 2013) show signatures of enhancers rather than promoters. In light of the ability of enhancers to act over very long distances, the combinations of TFBSs present in LTRs, and the general selection against ERV integrations near genic promoters (Medstrand et al., 2002), it is not surprising that LTRs have been co-opted as enhancers. Indeed, recent genome-wide surveys reveal that active LTR-derived enhancers exhibit the typical epigenomic signatures of active non-TE-derived enhancer elements, including enrichment of H3K4me1, H3K27ac, DNase I hypersensitivity, DNA hypomethylation, depletion of repressive H3K9me3 and H3K27me3, and TF binding (Chuong et al., 2013; Fort et al., 2014; Jacques et al., 2013; Sundaram et al., 2014; Xie et al., 2013) (Figure 2). In addition to epigenomic signatures, candidate TE-derived enhancers can also be identified from comparative genomic analysis. Using the Marmoset genome, del Rosario et al. (2014) identified noncoding regions constrained in the anthropoid primate lineage that are unconstrained in other distantly related mammals and found 14,546 TE-derived regions covering ~4 Mb of genomic sequence that showed chromatin signatures of anthropoid lineage-specific enhancers, a subset of which were derived from LTRs (del Rosario et al., 2014).

While candidate tissue-specific LTR-derived enhancer sequences have been reported by many groups based on the

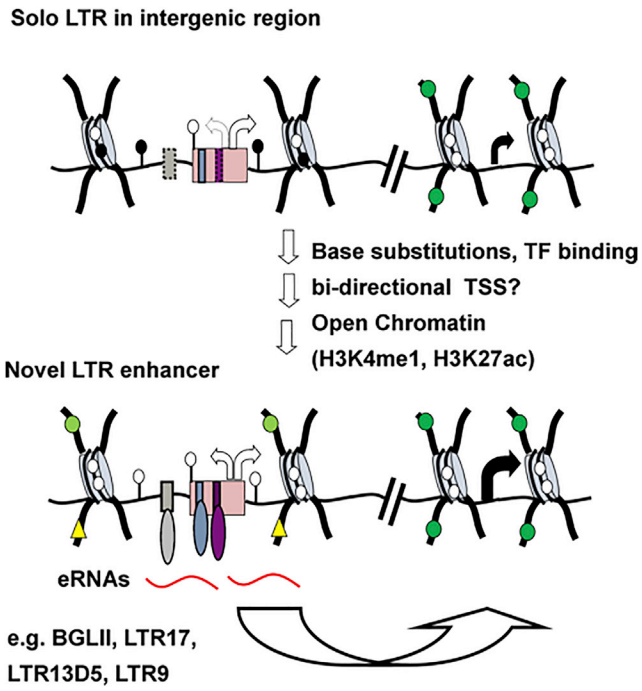

Figure 2. Molecular Mechanisms of Exaptation of LTRs as Enhancers

LTRs located both proximal or distal (i.e., >10 kb) to a genic promoter may be exapted as enhancer elements. Such elements may be solo LTRs (shown) or intact ERVs. While these LTRs may have intrinsic enhancer activity, base substitutions that generate additional TF binding sites (potentially synergizing with pre-existing sites) may over time increase overall enhancer activity and/or refine tissue specificity. Note that in contrast with LTR-derived genic promoters, which are generally in the sense orientation, an LTR integrated in either orientation with respect to the relevant gene could be exapted as an enhancer. Robust enhancer activity also likely requires formation of an open chromatin structure and the generation of enhancer RNA transcripts (eRNAs) in the relevant cell type. The strong association of specific histone marks with enhancers, including H3K4me1 (light green circles) and H3K27ac (yellow triangles), the latter indicative of "active" enhancers, has been widely exploited to identify novel candidate enhancers, including within LTRs. Examples of LTR enhancers include BGLII and LTR17 and perhaps LTR13D5 and LTR9, but whether these latter two LTRs produce eRNAs has not been determined.

presence of specific TFBSs and/or epigenetic marks consistent with enhancer activity, only a few studies have performed confirmatory functional analyses of their activity. Using luciferase-based reporter assays, candidate LTR enhancers have been shown to increase expression from heterologous promoters in cell lines representing the tissue in which they are active, such as rat trophoblast stem cells or human 293T cells, respectively (Chuong et al., 2013; Xie et al., 2013). However, while these assays demonstrate potential for enhancer activity, they do not prove that the LTR-derived sequences function as enhancers in their native genomic context. To address this question, loss- and/or gain-of-function analysis demonstrating putative enhancer function of LTRs have been employed in animal models. For example, deletion of the human LTR9 enhancer located ~100 kb upstream of the *β-globin* gene cluster abolishes *β-globin* gene expression in a transgenic mouse model (Pi et al., 2010). Similarly, transgenic mice were used to demonstrate bona fide enhancer activity of the novel primate-specific TE-derived *ASC192* enhancer (del Rosario et al., 2014). More recently, CRISPR-mediated deletion was used to demonstrate the importance of the MER41 LTR as an enhancer of key innate immunity genes activated by interferons (Chuong et al., 2016). Further studies of candidate LTR-derived enhancers using genome-editing approaches will reveal the extent to which such elements influence target gene expression in vivo.

How are LTRs exapted as novel tissue-specific enhancers? In another example of LTR exaptation in different lineages, Franchini et al. (2011) showed that an ancient SINE and later a MaLR LTR were independently exapted as enhancers to control expression of the vertebrate *Pomc* gene in the pituitary and hypothalamus in independent lineages during mammalian evolution (Franchini et al., 2011). Interestingly, a recent functional analysis of these enhancers showed that while the ancient SINE-derived enhancer nPE2 is only required for ~20% of *Pomc* expression, the MaLR LTR-derived enhancer nPE1 is sufficient to drive ~80% of *Pomc* expression (Lam et al., 2015), suggesting that LTRs may be exapted as enhancers when large increases in gene expression are beneficial and thus selected. In addition to the mechanism of epistatic capture of novel TFBSs described above, which is relevant to both promoters and enhancers, the exaptation of LTRs as enhancers may depend on their capacity to produce bidirectional noncoding transcripts (Figure 2). Indeed, enhancer function may require bidirectional transcription of distinct noncoding RNA species called enhancer RNAs (eRNAs) (Kim et al., 2015; Plank and Dean, 2014).

LTRs that serve as promoters for nuclear lncRNAs have been found in a variety of contexts (Faulkner et al., 2009; Herquel et al., 2013; Lu et al., 2014b), but whether they produce eRNAs had not been addressed. However, a recent comprehensive transcriptome analysis of pluripotent stem cells demonstrated that active BGLII- and LTR17-derived enhancers do indeed express bidirectional eRNAs, and many of these LTR-associated noncoding transcripts are important for the maintenance of ESC pluripotency (Fort et al., 2014). While bidirectional transcription from LTRs has been reported, as in the case of the composite human LTR9/LTR16A that promotes tissue-specific expression of the *DSCR4* and *DSCR8* genes in opposing orientations (Dunn et al., 2006), most LTRs do not produce bidirectional transcripts. Therefore, these findings suggest that LTR exaptation as an enhancer may also depend on the acquisition of substitutions within or adjacent to the LTR that support eRNA transcription (Figure 2). Whether such mutations are distinct from those that support TF binding or acquisition of alternative features essential for enhancer function remains to be determined. Regardless, the number of enhancer-derived LTRs is likely to be high in certain cell types, with their activity restricted in other cell types by the establishment of repressive chromatin, such as the deposition of H3K9me3 by the KRAB-ZFP/KAP1/SETDB1 system (Rowe et al., 2013).

Conclusions

In conclusion, recent transcriptomic and epigenomic studies have revealed that LTRs provide a plethora of novel gene regulatory elements, including tissue-specific promoters and enhancers. Such LTRs are particularly prevalent in early embryonic development, germ cells, and pluripotent stem cells, likely as a consequence of the relaxed epigenetic silencing in these cell

types and regulatory regions optimized for expression in these tissues in the retroviral precursor (Fort et al., 2014; Xie et al., 2013). Future work will likely reveal whether these regions are generally further optimized for tissue-specific expression by the acquisition of TFBSs. In addition, although there are many candidate species-specific enhancers derived from LTRs, few studies have actually addressed their biological significance in vivo with rigorous functional analyses (de Souza et al., 2013). Using CRISPR technology, it is now feasible to inactivate or delete specific LTRs to determine their effects upon the host transcriptome (Yang et al., 2015), providing a powerful tool for systematic analyses of LTR-driven transcripts and candidate enhancers in different cell types and species. While ERVs show striking enrichment in lncRNAs (Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012) and LTRs clearly drive expression of a subset of lncRNA transcripts that appear to play important developmental functions (Durruthy-Durruthy et al., 2016; Fort et al., 2014), it remains to be determined what roles ERV-derived sequences generally play in lncRNA structure, function, and evolution. Furthermore, caution must be exercised when interpreting the results of loss-of-function studies on lncRNAs due to the complex nature of their activities at the transcriptional and post-transcriptional levels (Bassett et al., 2014). Finally, while other retrotransposons have been exapted into both enhancers and insulators in humans (Jjingo et al., 2014; Wang et al., 2015), LTR-derived insulators have not been identified to date. Future investigations into these and related questions will further our understanding of the extent to which mammalian genomes have harnessed the latent regulatory potential of LTRs to control tissue-specific gene expression.

ACKNOWLEDGMENTS

We wish to thank Dixie Mager and Cedric Feschotte for critical reading of the manuscript. This work was supported by Canadian Institutes of Health Research Grant MOP-133417 (M.C.L.) and NIH grant 1ZIAHD008933 (T.S.M.).

REFERENCES

Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W., et al. (2014). Considerations when investigating lncRNA function in vivo. eLife *3*, e03058.

Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., and Tristem, M. (2007). Rate of recombinational deletion among human endogenous retroviruses. J. Virol. *81*, 9437–9442.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. *18*, 1752–1762.

Britten, R.J., and Davidson, E.H. (1969). Gene regulation for higher cells: a theory. Science *165*, 349–357.

Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., and Sverdlov, E. (2006). At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. J. Virol. *80*, 10752–10762.

Castro-Diaz, N., Friedli, M., and Trono, D. (2015). Drawing a fine line on endogenous retroelement activity. Mob. Genet. Elements *5*, 1–6.

Chuong, E.B., Rumi, M.A.K., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat. Genet. *45*, 325–329.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science *351*, 1083–1087.

Cohen, C.J., Lock, W.M., and Mager, D.L. (2009). Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene *448*, 105–114.

Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. *10*, 691–703.

Cowley, M., and Oakey, R.J. (2013). Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet. *9*, e1003234.

de Souza, F.S., Franchini, L.F., and Rubinstein, M. (2013). Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? Mol. Biol. Evol. *30*, 1239–1251.

del Rosario, R.C.H., Rayan, N.A., and Prabhakar, S. (2014). Noncoding origins of anthropoid traits and a new null model of transposon functionalization. Genome Res. *24*, 1469–1484.

Dunn, C.A., Medstrand, P., and Mager, D.L. (2003). An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. Proc. Natl. Acad. Sci. USA *100*, 12841–12846.

Dunn, C.A., van de Lagemaat, L.N., Baillie, G.J., and Mager, D.L. (2005). Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate beta3GAL-T5. Gene *364*, 2–12.

Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. Gene *366*, 335–342.

Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E.J., Davila, J., Mall, M., Wong, W.H., Wysocka, J., et al. (2016). The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. Nat. Genet. *48*, 44–52.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. Dev. Cell *36*, 611–623.

Emera, D., and Wagner, G.P. (2012a). Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc. Natl. Acad. Sci. USA *109*, 11246–11251.

Emera, D., and Wagner, G.P. (2012b). Transposable element recruitments in the mammalian placenta: impacts and mechanisms. Brief. Funct. Genomics *11*, 267–276.

Emera, D., Casola, C., Lynch, V.J., Wildman, D.E., Agnew, D., and Wagner, G.P. (2012). Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. Mol. Biol. Evol. *29*, 239–247.

Evsikov, A.V., de Vries, W.N., Peaston, A.E., Radford, E.E., Fancher, K.S., Chen, F.H., Blake, J.A., Bult, C.J., Latham, K.E., Solter, D., and Knowles, B.B. (2004). Systems biology of the 2-cell mouse embryo. Cytogenet. Genome Res. *105*, 240–250.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. *41*, 563–571.

Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. *13*, 283–296.

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., and Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. Cell *155*, 807–816.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., et al.; FANTOM Consortium (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat. Genet. *46*, 558–566.

Franchini, L.F., López-Leal, R., Nasif, S., Beati, P., Gelman, D.M., Low, M.J., de Souza, F.J.S., and Rubinstein, M. (2011). Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. Proc. Natl. Acad. Sci. USA *108*, 15270–15275.

Franchini, L.F., de Souza, F.S.J., Low, M.J., and Rubinstein, M. (2012). Positive selection of co-opted mobile genetic elements in a mammalian gene: if you can't beat them, join them. Mob. Genet. Elements *2*, 106–109.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu. Rev. Cell Dev. Biol. *31*, 429–451.

Froberg, J.E., Yang, L., and Lee, J.T. (2013). Guided by RNAs: X-inactivation as a model for lncRNA function. J. Mol. Biol. *425*, 3698–3706.

Gemmell, P., Hein, J., and Katzourakis, A. (2015). Orthologous endogenous retroviruses exhibit directional selection since the chimp-human split. Retrovirology *12*, 52.

Gifford, W.D., Pfaff, S.L., and Macfarlan, T.S. (2013). Transposable elements as genetic regulatory substrates in early development. Trends Cell Biol. *23*, 218–226.

Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135–141.

Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature *522*, 221–225.

Hayashi, M., Maehara, K., Harada, A., Semba, Y., Kudo, K., Takahashi, H., Oki, S., Meno, C., Ichiyanagi, K., Akashi, K., and Ohkawa, Y. (2016). Chd5 regulates MuERV-L/MERVL expression in mouse embryonic stem cells via H3K27me3 modification and histone H3.1/H3.2. J. Cell. Biochem. *117*, 780–792.

Herquel, B., Ouararhni, K., Martianov, I., Le Gras, S., Ye, T., Keime, C., Lerouge, T., Jost, B., Cammas, F., Losson, R., and Davidson, I. (2013). Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. Nat. Struct. Mol. Biol. *20*, 339–346.

Hisada, K., Sánchez, C., Endo, T.A., Endoh, M., Román-Trufero, M., Sharif, J., Koseki, H., and Vidal, M. (2012). RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. Mol. Cell. Biol. *32*, 1139–1149.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. *45*, 1198–1206.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409–419.

Isbel, L., and Whitelaw, E. (2012). Endogenous retroviruses in mammals: an emerging picture of how ERVs modify expression of adjacent genes. BioEssays *34*, 734–738.

Isbel, L., Srivastava, R., Oey, H., Spurling, A., Daxinger, L., Puthalakath, H., and Whitelaw, E. (2015). Trim33 binds and silences a class of young endogenous retroviruses in the mouse testis; a novel component of the arms race between retrotransposons and the host genome. PLoS Genet. *11*, e1005693.

Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.-E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol. *22*, 662–671.

Jacques, P.-E., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. *9*, e1003504.

Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. Annu. Rev. Genet. *42*, 709–732.

Jjingo, D., Conley, A.B., Wang, J., Mariño-Ramírez, L., Lunyak, V.V., and Jordan, I.K. (2014). Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. Mob. DNA *5*, 14.

Kannan, S., Chernikova, D., Rogozin, I.B., Poliakov, E., Managadze, D., Koonin, E.V., and Milanesi, L. (2015). Transposable element insertions in long intergenic non-coding RNA genes. Front. Bioeng. Biotechnol. *3*, 71.

Kapusta, A., and Feschotte, C. (2014). Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. Trends Genet. *30*, 439–452.

Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. *9*, e1003470.

Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., and Lorincz, M.C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676–687.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. *13*, R107.

Kim, T.-K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. Cold Spring Harb. Perspect. Biol. *7*, a018622.

Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. *42*, 631–634.

Lam, D.D., de Souza, F.S.J., Nasif, S., Yamashita, M., López-Leal, R., Otero-Corchon, V., Meece, K., Sampath, H., Mercer, A.J., Wardlaw, S.L., et al. (2015). Partially redundant enhancers cooperatively maintain mammalian pomc expression above a critical functional threshold. PLoS Genet. *11*, e1004935.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M.A., Richter, J., Soler, E., et al. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat. Med. *16*, 571–579, 1p, 579.

Lin, N., Chang, K.Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., Zhang, D., Han, T., Yang, C.S., Cunningham, T.J., et al. (2014). An evolutionarily conserved long noncoding RNA TUNA controls pluripotency and neural lineage commitment. Mol. Cell *53*, 1005–1019.

Lipatov, M., Lenkov, K., Petrov, D.A., and Bergman, C.M. (2005). Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. BMC Biol. *3*, 24.

Liu, S., Brind'Amour, J., Karimi, M.M., Shirane, K., Bogutz, A., Lefebvre, L., Sasaki, H., Shinkai, Y., and Lorincz, M.C. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. Genes Dev. *28*, 2041–2055.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nat. Genet. *42*, 1113–1117.

Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc. Natl. Acad. Sci. USA *104*, 8005–8010.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014a). Role of Tet proteins in enhancer activity and telomere elongation. Genes Dev. *28*, 2103–2119.

Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G., and Ng, H.-H. (2014b). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. *21*, 423–425.

Lynch, V.J., Leclerc, R.D., May, G., and Wagner, G.P. (2011). Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. *43*, 1154–1159.

Lynch, V.J., Nnamani, M.C., Kapusta, A., Brayer, K., Plaza, S.L., Mazur, E.C., Emera, D., Sheikh, S.Z., Grützner, F., Bauersachs, S., et al. (2015). Ancient transposable elements transformed the uterine regulatory landscape and

transcriptome during the evolution of mammalian pregnancy. Cell Rep. *10*, 551–561.

Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., and Pfaff, S.L. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. *25*, 594–607.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63.

Mager, D.L., and Stoye, J.P. (2015). Mammalian endogenous retroviruses. Microbiol. Spectr. *3*, A3–A0009, 2014.

Magiorkinis, G., Gifford, R.J., Katzourakis, A., De Ranter, J., and Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. Proc. Natl. Acad. Sci. USA *109*, 7385–7390.

Mak, K.S., Burdach, J., Norton, L.J., Pearson, R.C.M., Crossley, M., and Funnell, A.P.W. (2014). Repression of chimeric transcripts emanating from endogenous retrotransposons by a sequence-specific transcription factor. Genome Biol. *15*, R58.

Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet. *2*, e2.

Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin *6*, 15.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927–931.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. USA *36*, 344–355.

Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. *12*, 1483–1495.

Nei, M., Xu, P., and Glazko, G. (2001). Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA *98*, 2497–2502.

Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura, M., Tokunaga, Y., Nakamura, M., Watanabe, A., et al. (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. Proc. Natl. Acad. Sci. USA *111*, 12426–12431.

Okahara, G., Matsubara, S., Oda, T., Sugimoto, J., Jinno, Y., and Kanaya, F. (2004). Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs. Genomics *84*, 982–990.

Pavlicev, M., Hiratsuka, K., Swaggart, K.A., Dunn, C., and Muglia, L. (2015). Detecting endogenous retrovirus-driven tissue-specific gene transcription. Genome Biol. Evol. *7*, 1082–1097.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell *7*, 597–606.

Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J., and Tuan, D. (2010). Long-range function of an intergenic retrotransposon. Proc. Natl. Acad. Sci. USA *107*, 12992–12997.

Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. Mol. Cell *55*, 5–14.

Rebollo, R., Romanish, M.T., and Mager, D.L. (2012a). Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu. Rev. Genet. *46*, 21–42.

Rebollo, R., Zhang, Y., and Mager, D.L. (2012b). Transposable elements: not as quiet as a mouse. Genome Biol. *13*, 159.

Reiss, D., Zhang, Y., and Mager, D.L. (2007). Widely variable endogenous retroviral methylation levels in human placenta. Nucleic Acids Res. *35*, 4743–4754.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Robbez-Masson, L., and Rowe, H.M. (2015). Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology *12*, 45.

Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A., and Mager, D.L. (2007). Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution. PLoS Genet. *3*, e10.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240.

Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res. *23*, 452–461.

Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep. *16*, 836–850.

Smit, A.F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. *9*, 657–663.

Stocking, C., and Kozak, C.A. (2008). Murine endogenous retroviruses. Cell. Mol. Life Sci. *65*, 3383–3398.

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. *24*, 1963–1976.

Thompson, P.J., Dulberg, V., Moon, K.M., Foster, L.J., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. PLoS Genet. *11*, e1004933.

Tomizawa, S., Kobayashi, H., Watanabe, T., Andrews, S., Hata, K., Kelsey, G., and Sasaki, H. (2011). Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. Development *138*, 811–820.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. *24*, 1260–1270.

van de Lagemaat, L.N., Landry, J.-R., Mager, D.L., and Medstrand, P. (2003). Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends Genet. *19*, 530–536.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.

Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Méhouas, S., Arnaud, P., Tomizawa, S., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. Genome Biol. *16*, 209.

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc. Natl. Acad. Sci. USA *104*, 18613–18618.

Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature *516*, 405–409.

Wang, J., Vicente-García, C., Seruggia, D., Moltó, E., Fernandez-Miñán, A., Neto, A., Lee, E., Gómez-Skarmeta, J.L., Montoliu, L., Lunyak, V.V., and

Jordan, I.K. (2015). MIR retrotransposon sequences provide insulators to the human genome. Proc. Natl. Acad. Sci. USA *112*, E4428–E4437.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al.; Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201–1204.

Wolf, G., Greenberg, D., and Macfarlan, T.S. (2015a). Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. Mob. DNA *6*, 17.

Wolf, G., Yang, P., Füchtbauer, A.C., Füchtbauer, E.-M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015b). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev. *29*, 538–554.

Wu, X., and Sharp, P.A. (2013). Divergent transcription: a driving force for new gene origination? Cell *155*, 990–996.

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. (2010). Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Res. *20*, 804–815.

Xie, M., Hong, C., Zhang, B., Lowdon, R.F., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., et al. (2013). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat. Genet. *45*, 836–841.

Yang, L., Güell, M., Niu, D., George, H., Lesha, E., Grishin, D., Aach, J., Shrock, E., Xu, W., Poci, J., et al. (2015). Genome-wide inactivation of porcine endogenous retroviruses (PERVs). Science *350*, 1101–1104.

Zhang, Q., Dan, J., Wang, H., Guo, R., Mao, J., Fu, H., Wei, X., and Liu, L. (2016). Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells. Sci. Rep. *6*, 19852.

Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.-Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. Nature *500*, 477–481.