

Endogenous Retroviruses and the Epigenome

Andrew B. Conley and I. King Jordan

Abstract Endogenous retroviruses (ERVs) are the evolutionary remnants of retroviral germline infections, which are no longer capable of intercellular infectivity. Despite being confined within the genomes of their hosts, ERVs are able to replicate and spread via retrotransposition. This replicative process helps to ensure the elements' proliferation and long term evolutionary success, but it also imposes a substantial mutational burden on their host genomes. Accordingly, host organisms have evolved a variety of mechanisms to repress ERV transposition, including epigenetic mechanisms based on the modification of chromatin. In particular, DNA methylation and histone modifications are used to silence ERV transcription thereby mitigating their ability cause mutations via transposition. It has recently become apparent that epigenetic and chromatin based regulation of ERVs can also exert substantial regulatory effects on host genes. In this chapter, we provide a number of examples illustrating how chromatin modifications of ERV insertions relate to host gene regulation including both deleterious cases as well as exapted cases whereby epigenetically activated ERV elements provide functional utility to their host genomes via the provisioning of novel regulatory sequences. For example, we discuss ERV-derived promoter and enhancer sequences in the human genome that are epigenetically modified in a cell-type specific manner to help drive differential expression of host genes. The genomic abundance of ERVs, taken together with their proximity to host genes and their propensity to be epigenetically modified,

A.B. Conley

Georgia Institute of Technology, School of Biology, 310 Ferst Drive, Atlanta, GA 30332, USA

I.K. Jordan (✉)

Georgia Institute of Technology, School of Biology, 310 Ferst Drive, Atlanta, GA 30332, USA

PanAmerican Bioinformatics Institute, Santa Marta, Magdalena, Colombia

e-mail: king.jordan@biology.gatech.edu

suggest that this kind of phenomenon may be far more common than previously imagined. Furthermore, the environmental responsiveness of epigenetic pathways suggests the possibility that ERVs, along with other classes of epigenetically modified TEs, may serve to coordinately modify host gene regulatory programs in response to environmental challenges.

1 Introduction

Endogenous retroviruses (ERVs) are the genomic remnants of retroviruses that integrated into a host genome and subsequently lost the ability to leave the host cell, instead replicating within the host genome (Lower et al. 1996). Evolutionarily, ERVs are members of a broader class of mobile genetic elements known as LTR-containing retroelements; included in this broader set are the LTR retrotransposons. LTR-containing retroelements are named for the Long Terminal Repeats (LTRs) found at their 5' and 3'-ends. These LTRs are direct repeats, identical at the time of insertion, and contain regulatory sequences required for element transcription. The LTRs of ERVs and LTR retrotransposons are highly similar in structure and function (Xiong and Eickbush 1990). The similarity between ERVs and LTR goes beyond the presence of the LTR sequences, however. In fact, LTR retrotransposons have been referred to as being 'retrovirus-like' elements due to their similarity to both ERVs and retroviruses (Lander et al. 2001). Both ERVs and LTR retrotransposons contain coding sequences necessary for their integration into the host genome as well as a region encoding a reverse transcriptase that catalyzes the polymerization of DNA from an RNA template. Comparison of reverse transcriptase sequences from diverse retrotransposons and viruses revealed that retroviruses and ERVs are most closely grouped with LTR retrotransposons (Xiong and Eickbush 1988, 1990; Doolittle et al. 1989). Phylogenetic reconstructions based on reverse transcriptase sequence alignments indicate that retroviruses and ERVs represent a monophyletic subset of overall LTR retroelement diversity and show that the LTR retrotransposons form a basal clade to this group with greater relative diversity. These data were taken to indicate that, at some time in the distant past, retroviruses emerged from within the LTR retrotransposon lineage via the acquisition of an envelope protein coding sequence that conferred intercellular infectivity, *i.e.* the ability to escape the confines of the host cell (Xiong and Eickbush 1990). Thus, ERVs, which are a group of retrovirus-derived sequences that are no longer capable of intercellular infectivity, represent a reversion to the ancestral state of LTR retrotransposons as non-infectious genomic elements.

As with other classes of retrotransposable elements, LTR-containing retroelements, including ERVs, are able to increase their copy number in the genome via retrotransposition. Through retrotransposition, LTR-containing retroelements can achieve high copy number within genomes, *e.g.* ~700,000 insertions in the human genome, comprising 8% of the total genomic sequence (Lander et al. 2001). The

retrotransposition of ERVs and other LTR retroelements can cause deleterious mutations in the host. In mouse, where ERVs are highly active, it has been estimated that 10% of *de novo* mutations result from novel ERV insertions (Maksakova et al. 2006; Waterston et al. 2002). ERV insertions can cause deleterious mutations via a number of mechanisms including the induction of transcriptional aberrations in host genes. For example, integration of the ETn mouse ERV into the second intron of the Fas (tumor necrosis factor receptor superfamily, member 6) gene has been shown to lead to aberrant splicing of Fas transcripts via the donation of splice donor and acceptor sites that cause the inserted ERV to be spliced into the nascent host gene transcript (Wu et al. 1993). This leads to mutant mice with an autoimmune phenotype. More recently, it has been shown that insertion of a mouse ERV into to an intron of the Slc15a2 (solute carrier family 15, member 2) gene can cause premature transcriptional termination at distance via a distinct mechanism that does not involve changes in the splicing of the gene (Li et al. 2012). This same work revealed that similar pre-maturely terminated transcripts occur in ~5% of mouse genes with intronic polymorphisms of ERVs.

In order to prevent deleterious insertions of ERVs and other LTR-containing retroelements, host genomes have evolved a variety of mechanisms to suppress element transposition (Levin and Moran 2011). Among these mechanisms, epigenetic and chromatin based silencing of insertions by the host limit the ability of the elements to produce mRNA, thereby greatly reducing the likelihood that they will be transposed (Lippman et al. 2004; Leung and Lorincz 2011). A number of recent studies on mammalian chromatin have demonstrated the extent to which ERV element sequences are marked with repressive histone modifications, which presumably limit their transcription. For example, using ChIP-PCR (Chromatin Immuno-Precipitation followed by PCR amplification), Martens et al. demonstrated that Intracisternal A particle (IAP) insertions, a family of ERVs, are subject to the repressive H4K20Me3 (trimethylation of Histone 4 K20) histone modification, while at the same time showing very low levels of the activating mark H3K4Me3 for these same elements (Martens et al. 2005). Similarly, using ChIP-seq (Chromatin Immuno-Precipitation followed by massively parallel sequencing) (Robertson et al. 2007), Mikkelsen et al. found that mouse ERVs are enriched for the epigenetically silencing histone modifications H3K9Me3 and H4K20Me3 (Mikkelsen et al. 2007). Using ChIP-seq data from CD4+ T-cells, Huda et al. also found that human LTR-containing retroelement insertions were enriched for silencing histone modifications (Huda et al. 2010).

While most chromatin studies of ERVs to date have focused on the epigenetic silencing of these elements for the purpose of genome defense, it has become increasingly clear that epigenetic modifications of ERVs and other LTR-containing retroelements can also have profound effects on the regulation of host genes. In other words, epigenetic modifications of ERV sequences are not only used to repress element transcription, but can also be exapted (Brosius and Gould 1992; Gould and Vrba 1982) for the purposes of controlling host gene expression. For example, epigenetic silencing of an ERV insertion near the promoter of a host gene could

possibly reduce the transcriptional activity of that gene. Alternatively, ERV or LTR-containing retroelement insertions could be actively modified and regulated in a way that benefits the host, *e.g.* as an alternative promoter for a host gene or an enhancer that regulates gene expression at distance. Such exapted insertions could help to diversify the host transcriptome as has been seen for an ERV-derived promoter driving the expression of the IL-2 receptor beta gene in human placenta (Cohen et al. 2011). In this chapter, we focus on these kinds of chromatin mediated regulatory exaptations of ERVs and other LTR-containing retroelements. We provide several examples of recent studies showing how epigenetic modifications of these kinds of elements can affect the regulation of host genes in a variety of eukaryotic species. First, we explore host gene regulatory effects exerted by the epigenetic silencing of LTR retroelements (Sects. 2, 3, 4), and then we focus on how activating chromatin modification of these kinds of elements can also effect the regulation of nearby host genes (Sects. 5, 6, 7).

2 Epigenetic Silencing of LTR Retroelement Insertions in *Arabidopsis thaliana*

In an early study on the effect of transposable element (TE) insertions on the local chromatin environment, Lippman et al. characterized the chromatin environment of a genomic region in *Arabidopsis thaliana* which arose from an ancient segmental duplication (Lippman et al. 2004). This duplicated chromosomal region is a so-called ‘knob’, *i.e.* an interstitial heterochromatic region, which was found to contain many LTR retrotransposon and other TE insertions that are not present in its duplicated counterpart. These TE insertions are evolutionarily young indicating that they were inserted into the knob region after the ancient duplication by which it was generated (Fig. 1). The coincidence of heterochromatin and novel TE insertions in the knob region was taken to suggest that these insertions led to the formation of interstitial heterochromatin after duplication, presumably as a result of host chromatin based silencing mechanisms that were targeted to these TEs. Using tiling arrays, Lippman et al. demonstrated that the TE insertions in the knob were in fact marked with DNA methylation and the repressive H3K9Me3 histone modification, with elements of the gypsy family being particularly heavily modified. Knockdown of the DNA methyltransferase *ddm1* resulted in the decrease of the levels these repressive marks in the knob region and an increase in LTR retrotransposon expression therein, mainly from the gypsy family of elements.

This study demonstrated that insertion of LTR-containing retroelements could lead to the in situ formation of heterochromatin in one particular region of a eukaryotic genome in response to host defense mechanisms that silence element expression. These findings suggested that the novel insertions of LTR-containing retroelements could have genome-wide effects via the generation of local heterochromatic regions that can silence nearby host genes.

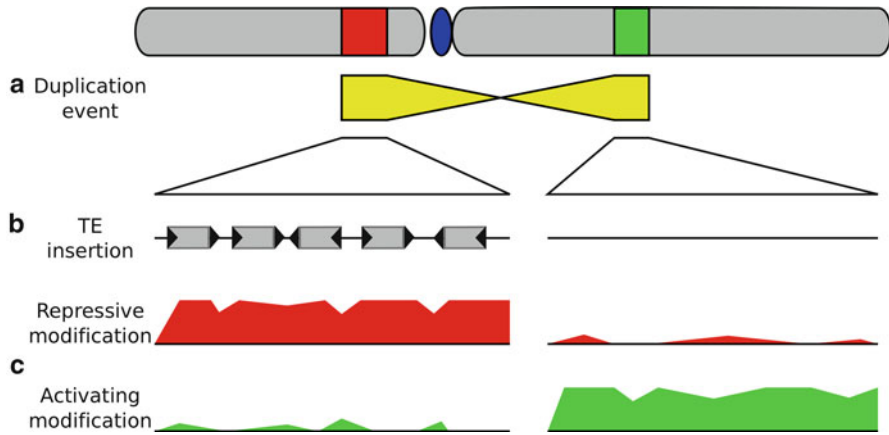


Fig. 1 Generation of an interstitial heterochromatic region driven by transposable element (TE) insertions. **(a)** An ancient segmental duplication in *A. thaliana* led to two paralogous regions. **(b)** One of the duplicated regions is subject to multiple TE insertions (*left*), including numerous LTR retroelements, while the other duplicated region remains largely free of such insertions (*right*). **(c)** The region with TE insertions (*left*) is subject to repressive epigenetic modifications (*red*) and depletion of activating modifications (*green*), while the reverse is seen for the region without the insertions (Figure adopted from Lippman et al. 2004)

3 Epigenetic Silencing of LTR Retroelement Insertions and the Effect on Nearby Genes in *A. thaliana*

The results from Lippman et al. demonstrated that LTR insertions generate novel heterochromatic regions in *A. thaliana*, and they also showed that genes co-located with TEs in the heterochromatic knob-region were expressed at lower levels than their paralogs located in euchromatin. Indeed, if an LTR-containing retroelement insertion near or within a transcribed locus is epigenetically silenced, then it may be possible for the element silencing to affect expression of the gene as well. Based on this line of thinking, Hollister and Gaut sought to characterize the effect of methylated TE insertions, including ERVs and other LTR-containing retroelement insertions, on the expression of nearby genes *A. thaliana* (Hollister and Gaut 2009). Initially, they observed a globally lower expression of genes near TE insertions; however, this did not take into account the epigenetic state of the insertion. Using genome-wide bisulfite sequencing data, they went on to demonstrate a genome-wide depletion of methylated TE insertions near genes, suggesting that such insertions are selected against, perhaps by virtue of their silencing effects on nearby gene expression. In fact, the authors demonstrated that genes proximal to such methylated insertions were expressed at lower levels, indicating that the methylation of TE insertions near genes reduces their expression. In line with the role of selection in removing methylated TEs from the proximity of genes, Hollister and Gaut demonstrated that methylated polymorphic TE insertions near genes were skewed

towards rare variants. Furthermore, this effect was observed only for insertions <1.5 kb from genic loci, pointing to locally confined spreading of methylation from TE insertions into nearby or adjacent genes. Indeed, older methylated TEs were found to be farther from genes, suggesting that selection has not acted on them as it has on younger methylated TEs near genes.

The depletion of LTR-retroelement and other TE insertions within and near genes has been observed for a number of eukaryotic species and itself strongly suggests that such insertions are selected against. The study by Hollister and Gaut provided a specific mechanistic basis for this selection, *i.e.* the fact that methylated insertions within and near genes are deleterious by virtue of their silencing effects on gene expression. Given what these authors observed, it seemed possible that the reduction of neighboring gene expression by the insertion of a TE could also occur in other species that epigenetically silence TE insertions and could perhaps be even more profound in genomes that are denser in repetitive elements.

4 Heterochromatin Spreading from Polymorphic IAP Insertions in the Mouse Genome

The mouse IAP family of ERVs is a highly active, with ~26,000 annotated insertions (Waterston et al. 2002). While Mikkelsen et al. previously showed that IAP insertions in mouse were epigenetically silenced (Mikkelsen et al. 2007), the effect that such silencing would have on nearby genes remained largely unexplored. Recently, Rebollo et al. investigated the possibility that novel IAP insertions in mouse could lead to the formation of local heterochromatin and the spreading of heterochromatin away from the insertion into nearby sequences (Rebollo et al. 2011). To do this, Rebollo et al. characterized IAP insertions which were polymorphic between two mouse cell lines, allowing them to observe the epigenetic state of the IAP insertion site with and without the insertion. It was found that the borders of IAP insertions, both those which were polymorphic between the two cell types and common IAP insertions, were enriched for the repressive H3K9Me3 histone modification. The enrichment of H3K9Me3 was found to spread from the borders of the IAP insertion up to a maximum of 5 kb. Importantly, for polymorphic IAP insertions, Rebollo et al. showed that the pre-insertion site in the cell type without the IAP insertion was not enriched for H3K9Me3, indicating that the novel IAP insertion was the source of the repressive modification.

The spreading of repressive modifications from an IAP insertion raised the question as to whether or not such spreading could lead from the insertion to a nearby promoter (Fig. 2). Indeed, Rebollo et al. were able to find an example of a polymorphic IAP insertion proximal to a mouse gene. There is an IAP insertion upstream of the *B3gatl1* promoter which is present only in the J1 cell type. In the J1 cell type, DNA methylation and the repressive histone modification H3K9Me3 extend from the IAP insertion into the promoter of the *B3gatl1* gene, which is

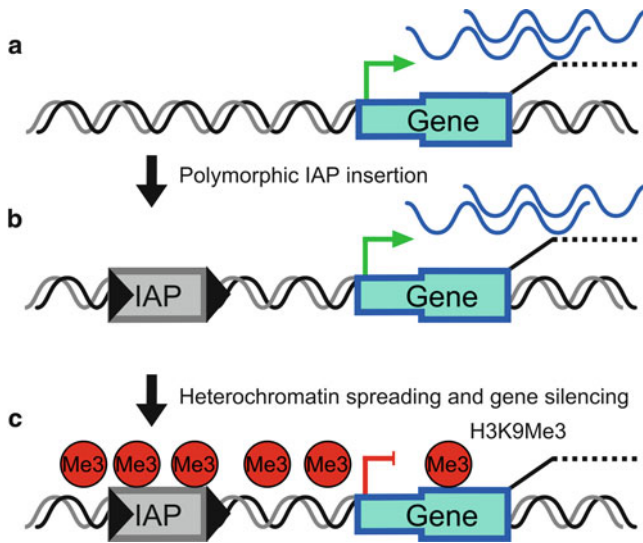


Fig. 2 Spreading of heterochromatin from a novel IAP insertion. (a) An active mouse gene promoter region prior to an IAP insertion. (b) Cell-type specific insertion of an IAP element near the active mouse gene promoter. (c) The IAP insertion is silenced with the repressive histone modification H3K9Me3 (red circles) and this repressive mark spreads to the nearby gene promoter resulting in silencing of the gene (Figure adopted from Rebollo et al. 2011)

accordingly down-regulated in J1 compared to the TT2 cell line that lacks the gene proximal IAP insertion. Such a spreading of heterochromatin from LTR insertions into nearby genes, and the negative regulatory effects caused by such spreading, could explain the apparent negative selection against LTR insertions near promoters previously observed for the mouse and human genomes (Jordan et al. 2003; van de Lagemaat et al. 2003).

It is worth noting that when looking for instances where the insertion of an IAP element led to heterochromatin spreading and alteration of gene expression, Rebollo et al. looked only at those IAP insertions proximal to promoters. In addition to promoters, there are many thousands of enhancers scattered within and between mammalian genes. Visel et al. characterized several thousand enhancers in mouse tissue samples, many of which were active in only one of the cell types analyzed (Visel et al. 2009). Similarly, Ernst et al. characterized many thousands of likely human enhancers based on their profile of active histone modifications (Ernst et al. 2011). Such active histone modifications are likely important in the function of the enhancers, and it stands to reason that an IAP inserted near an enhancer could reduce its function via the spreading of repressive epigenetic histone modifications. Indeed, the insertion of an IAP element near an enhancer could conceivably affect the expression of a gene in a more specific manner than promoter proximal insertions since enhancers tend to be more cell-type specific than promoters.

5 Demethylation of an IAP Insertion Leads to Ectopic Expression of the *agouti* Gene in Mouse

While many ERVs are epigenetically silenced, it is likely, given the large number of insertions present in many genomes, that some will escape such silencing, or even become actively modified. Indeed, Hollister and Gaut showed that not all LTR retroelement insertions are repressed in *A. thaliana*, a large number are demethylated (Hollister and Gaut 2009), and it would not be surprising to find that LTR retroelements in other species could also be demethylated. Given that ERVs contain their own promoters and regulatory sequences, it is conceivable that when demethylated their promoters could potentially transcribe through or away from their inserted sites into nearby genes. Given the genomic abundance of ERVs and other LTR-containing retroelements, it would seem probable that a number of demethylated insertions are likely to transcribe nearby host gene sequences. One such example of this phenomenon occurs at the *agouti* locus in mouse.

The *agouti* gene in mouse controls the pigmentation of mouse coats and hair follicle development. There exist mouse strains which show ectopic expression of the *agouti* gene, predisposing the mice to tumors and obesity (Michaud et al. 1994). Interestingly, the ectopic expression of the *agouti* gene is widely variable: the expression ranges from mice which express it widely, to those which show variegation in expression and those which show no ectopic expression and are otherwise phenotypically normal. It was demonstrated that the ectopic expression was not driven by the canonical promoter of the *agouti* gene, but an IAP insertion upstream of the *agouti* coding exons and that the level of expression driven from this IAP was correlated with the demethylation its LTR (Fig. 3) (Michaud et al. 1994; Morgan et al. 1999).

This *agouti* locus represents a departure from the usual reasoning behind the epigenetic silencing of LTR-containing retroelements and other TE insertions: rather than preventing retrotransposition *per se*, epigenetic silencing of the IAP insertion serves to prevent deleterious transcription from the IAP insertion into the neighboring *agouti* gene. While the *agouti* case was a single example of an ERV altering genomic function when demethylated, the large number of insertions within eukaryotic genomes, ~700,000 and ~850,000 in the human and mouse genomes (Lander et al. 2001; Waterston et al. 2002), virtually guarantees that other such de-repressed LTR retroelement insertions can and do act as promoters. Further, while transcription from the IAP insertion in the *agouti* locus is deleterious, other de-repressed insertions could prove adaptive and become exapted for function in the host genome. Indeed, several hundred promoters derived from LTR-containing retroelement insertions have been characterized in the human genome (Conley et al. 2008), the epigenetic characterization of which we discuss in the next section.

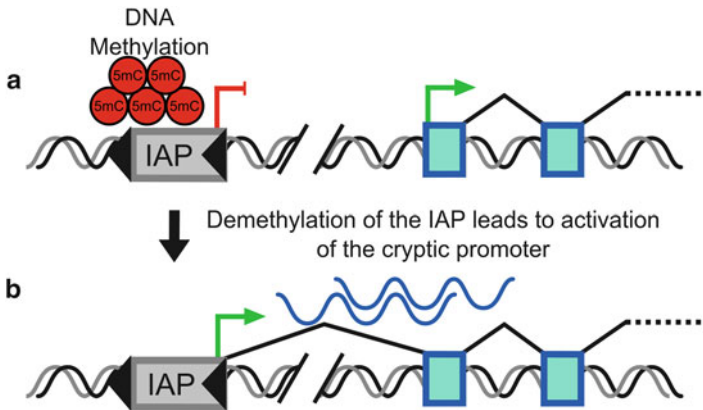


Fig. 3 Demethylation of an IAP leads to ectopic expression of the *agouti* gene. (a) In phenotypically normal mice, the *agouti* proximal IAP insertion is subject to DNA methylation (5mC, red circles) and is inactive. Accordingly, *agouti* gene expression is driven by its canonical promoter in the appropriate tissues. (b) In mice where the IAP insertion is demethylated, it can drive ectopic expression of the nearby *agouti* gene from a cryptic promoter encoded by the IAP insertion (Figure adopted from Morgan et al. 1999)

6 Actively Modified ERVs and Human Gene Promoters

The initial phases of the ENCODE project (Birney et al. 2007; Rosenbloom et al. 2010) have allowed for the unprecedented characterization of the epigenetic state of the large majority of sites in the human genome, including many repetitive elements which could not previously be characterized using array based techniques. Of equal importance, the ENCODE project has allowed for the comparison of the epigenetics state between cell types. Such comparisons allow for the detection of sites with differential modification which could in turn contribute to cell-type specific patterns of gene expression. In Sects. 6 and 7, we review studies of host gene promoters and enhancers respectively, based on ENCODE data from human cell lines, which demonstrate activating epigenetic modifications of ERVs and other LTR-containing retroelements and show how these reactivated insertions may drive cell-type specific patterns of gene expression.

The *agouti* locus in mouse demonstrates that the insertion of an ERV insertion near a gene can lead to the use of the insertion as an alternative promoter for the gene. Indeed, ERV and other LTR-containing retroelement-derived promoters, in both mouse and human, have been characterized in several studies. A 2004 study identified 81 genes expressed in early mouse embryos for which the 5'-end, and thus the promoter, was derived from an LTR retroelement insertion (Peaston et al. 2004). A later study used Paired-End diTag (PET) data (Ng et al. 2005) to characterize 114 distinct ERV-derived promoters in the human genome (Conley et al. 2008), and a

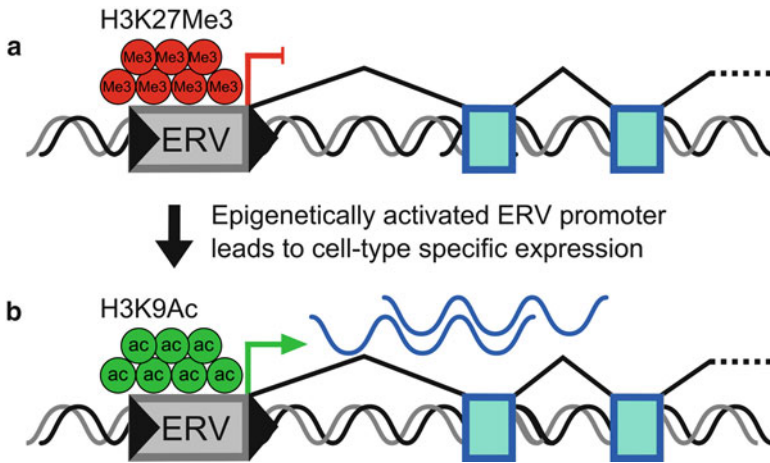


Fig. 4 Cell-type specific epigenetic activation of human ERV-derived promoters. (a) In one cell type, a human ERV insertion is subject to repressive histone modifications and accordingly is not used as a promoter for the adjacent host gene. (b) In a different cell type, the same ERV insertion is marked with activating histone modifications, e.g. H3K9Ac (green circles), leading to active transcription of the adjacent host gene from the ERV promoter (Figure adopted from Huda et al. 2011a)

study by Faulker et al. analyzed a large set of CAGE (Cap Analysis of Gene Expression) (Kodzius et al. 2006) libraries to investigate the potential promoter activity of LTR-containing retroelement insertions in diverse human and mouse tissues (Faulkner et al. 2009). While these studies characterized a breadth of LTR-containing retroelement-derived promoters, the epigenetic status and/or chromatin modifications of these insertions was not investigated.

Huda et al. investigated the epigenetic regulation of TE-derived promoters in the human genome, including those promoters derived from ERV and other LTR-containing retroelement insertions (Huda et al. 2010). The authors identified 1,520 distinct promoters derived from TE insertions, among them over 300 promoters derived from LTR-containing retroelement insertions (Fig. 4). Using ChIP-seq data from the GM12878 and K562 cell lines, Huda et al. characterized the epigenetic environment of the TE-derived promoters, finding an enrichment of activating modifications for active promoters along with a concomitant depletion of the sole repressive mark used, H3K27Me3. Of note, promoters derived from LTR-containing retroelements showed the greatest divergence of histone modification and activity between the GM12878 and K562 cell types. Such a divergence suggests that LTR-containing retroelement insertions have helped to diversify patterns of mammalian gene expression.

This study by Huda et al. demonstrated on a genome wide scale that the epigenetic activation of LTR-containing retroelement insertions can lead to the alteration of host gene expression via the use of the insertions as alternative promoters. This leads to interesting, and still largely open, questions regarding the origin and evolution of such LTR-containing retroelement-derived promoters. In the case of

the *agouti* locus in mouse, ectopic transcription driven by the IAP insertion is deleterious to the mouse (Michaud et al. 1994). Given the intricate control of gene expression, one would expect that such ectopic expression would generally be deleterious. Most would therefore likely be selected against and those that can still be observed represent the few that were adaptive. Indeed, the cell-type specific usage and epigenetic modification of the ERV and other LTR retroelement-derived promoters characterized by Huda et al. is suggestive of their adaptive nature and potential functional utility.

7 Actively Modified ERVs and Human Gene Enhancers

DNaseI hypersensitive sites are regions of the genome that are unusually ‘open’ in terms of their chromatin environment and thus susceptible to degradation by DNaseI. Such sites are often important for gene regulation, *e.g.* active promoters and enhancers. It was previously shown that a large number of DNaseI-hypersensitive sites are derived from ERVs and other LTR-containing retroelement insertions in the human genome (Marino-Ramirez and Jordan 2006), suggesting that these insertions could play roles in gene regulation apart from that of promoters, *e.g.* enhancers. Indeed, functional enhancers derived from other families of TEs are known, such as the AmnSINE1 element derived enhancers that help to drive brain specific expression (Sasaki et al. 2008). Active enhancers are epigenetically modified with activating histone modifications (Heintzman et al. 2007; Ernst et al. 2011), and while LTR-containing retroelement insertions are typically epigenetically silenced (Huda et al. 2010), insertions acting as enhancers would be expected to show the same activating histone modifications (Fig. 5).

In a recent study, Huda et al. used the epigenetic modification patterns of enhancers to predict TE-derived enhancers on a genome-wide scale (Huda et al. 2011b).

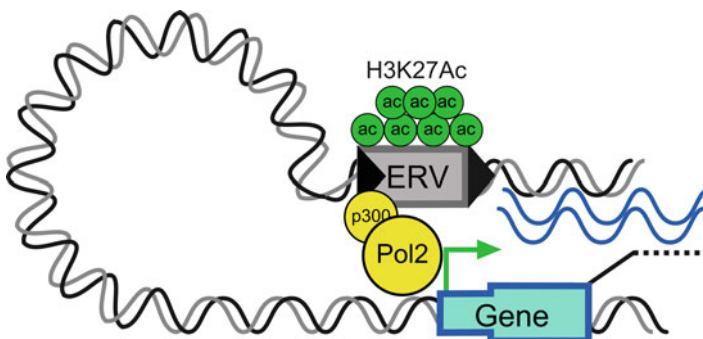


Fig. 5 Epigenetic activation of a human ERV-derived enhancer. An ERV insertion located distal to a host gene is subject to enhancer-characteristic activating histone modifications, *e.g.* H3K27Ac (green circles). When activated, it acts as an enhancer for the distal gene promoter, leading to transcription from the gene promoter (Figure adopted from Huda et al. 2011b)

Using known p300 binding sites as a training set, the authors used ChIP-seq data from the ENCODE project in the GM12878 and K562 cell types to screen DNaseI HS sites for histone modifications similar to those of known enhancers. Nearly 20,000 such sites were identified, several thousand of which were co-located with TE insertions. Of those, over 700 sites were derived from LTR insertions. Importantly, the presence of TE enhancers correlated with the expression of nearby genes, strongly suggesting that the TE-derived enhancers characterized were active and influenced gene expression.

As in the study of TE-derived promoters by Huda et al. (Huda et al. 2011a), the work on enhancers demonstrated the active epigenetic modification of human LTR-containing retroelement insertions (Huda et al. 2011b), which is in contrast with general the genome-wide enrichment of repressive modifications on such insertions (Huda et al. 2010). Also as in the TE-promoter study, the authors used only two cell types for the analysis of TE-derived enhancers. The large majority of enhancers characterized, however, both those derived from TE insertions and other, were detected in only one of the two cell types. This is in line with what others have observed regarding the cell type specificity of enhancers. For instance, in the large scale analysis of ENCODE ChIP-seq data, Ernst et al. found that while many promoters are active across a number of cell types, the large majority of putative enhancers were active in only one of the cell types investigated (Ernst et al. 2011). This opens the possibility that there are thousands of human enhancers derived from ERVs and other LTR-containing retroelement insertions, many of which would remain unidentified in a study of only two cell-types, and underscores the potential impacting on cell-type expression of thousands of human genes that these ERV-enhancers may exert.

8 Conclusions and Prospects

In this chapter, we reviewed some of the ways in which ERV effects on host gene regulation are mediated by epigenetic and chromatin modifications. ERVs are of course just one class of TEs, and TEs were originally discovered by Barbara McClintock by virtue of the regulatory effects they exert on maize host genes (McClintock 1948). In light of these effects, McClintock referred to TEs as controlling elements, and she ultimately came to believe that TEs could actually re-organize genomes in response to environmental challenges (McClintock 1984). For McClintock, this genome reorganize process was related to the genome dynamics of TEs per se, *i.e.* their ability to transpose and cause genomic rearrangements. Here, we would like to pose the idea that the TE-mediated environmental responsiveness of eukaryotic genomes may also be attributed the epigenetic and chromatin based regulatory effects that they exert on host genes. This notion is based in part on observations that epigenetic changes can in fact occur in response to environmental stimuli (Feil and Fraga 2011). In the case of ERVs, environmentally programmed ERV-mediated chromatin based regulatory changes have been observed for the agouti

locus where environmental exposure to methyl donors leads to increased repression of the upstream IAP thereby mitigating the mutation ectopic expression phenotype (Cropley et al. 2006). Given the abundance of ERVs, their widespread genomic distribution and proximity to genes along with their propensity to be epigenetically modified, these elements may provide a means for host genomes to mount dynamic epigenetically programmed responses to environmental challenges.

References

- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z et al (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799–816. doi:[10.1038/nature05874](https://doi.org/10.1038/nature05874)
- Brosius J, Gould SJ (1992) On “genomenclature”: a comprehensive (and respectful) taxonomy for pseudogenes and other “junk DNA”. *Proc Natl Acad Sci USA* 89(22):10706–10710
- Cohen CJ, Rebollo R, Babovic S, Dai EL, Robinson WP, Mager DL (2011) Placenta-specific expression of the interleukin-2 (IL-2) receptor beta subunit from an endogenous retroviral promoter. *J Biol Chem* 286(41):35543–35552. doi:[10.1074/jbc.M111.227637](https://doi.org/10.1074/jbc.M111.227637)
- Conley AB, Piriyaopangsa J, Jordan IK (2008) Retroviral promoters in the human genome. *Bioinformatics* 24(14):1563–1567. doi:[btn243 \[pii\] 10.1093/bioinformatics/btn243](https://doi.org/10.1093/bioinformatics/btn243)
- Cropley JE, Suter CM, Beckman KB, Martin DI (2006) Germ-line epigenetic modification of the murine A v γ allele by nutritional supplementation. *Proc Natl Acad Sci USA* 103(46):17308–17312. doi:[10.1073/pnas.0607090103](https://doi.org/10.1073/pnas.0607090103)
- Doolittle RF, Feng DF, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. *Q Rev Biol* 64(1):1–30
- Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345):43–49. doi:[10.1038/nature09906](https://doi.org/10.1038/nature09906)
- Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, Schroder K et al (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 41(5):563–571. doi:[10.1038/ng.368](https://doi.org/10.1038/ng.368)
- Feil R, Fraga MF (2011) Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13(2):97–109. doi:[10.1038/nrg3142](https://doi.org/10.1038/nrg3142)
- Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. *Paleobiol* 8:4–15
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39(3):311–318. doi:[10.1038/ng1966](https://doi.org/10.1038/ng1966)
- Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 19(8):1419–1428. doi:[10.1101/gr.091678.109](https://doi.org/10.1101/gr.091678.109)
- Huda A, Marino-Ramirez L, Jordan IK (2010) Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mobile DNA* 1(1):2. doi:[10.1186/1759-8753-1-2](https://doi.org/10.1186/1759-8753-1-2)
- Huda A, Bowen NJ, Conley AB, Jordan IK (2011a) Epigenetic regulation of transposable element derived human gene promoters. *Gene* 475(1):39–48. doi:[S0378-1119\(10\)00476-2 \[pii\] 10.1016/j.gene.2010.12.010](https://doi.org/10.1016/j.gene.2010.12.010)
- Huda A, Tyagi E, Marino-Ramirez L, Bowen NJ, Jjingo D, Jordan IK (2011b) Prediction of transposable element derived enhancers using chromatin modification profiles. *PLoS One* 6(11):e27513. doi:[10.1371/journal.pone.0027513](https://doi.org/10.1371/journal.pone.0027513)
- Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet* 19(2):68–72

- Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D et al (2006) CAGE: cap analysis of gene expression. *Nat Methods* 3(3):211–222. doi:[10.1038/nmeth0306-211](https://doi.org/10.1038/nmeth0306-211)
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921. doi:[10.1038/35057062](https://doi.org/10.1038/35057062)
- Leung DC, Lorincz MC (2011) Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem Sci*. doi:[10.1016/j.tibs.2011.11.006](https://doi.org/10.1016/j.tibs.2011.11.006)
- Levin HL, Moran JV (2011) Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 12(9):615–627. doi:[10.1038/nrg3030](https://doi.org/10.1038/nrg3030)
- Li J, Akagi K, Hu Y, Trivett AL, Hlyniak CJ, Swing DA, Volfvsky N et al (2012) Mouse endogenous retroviruses can trigger premature transcriptional termination at a distance. *Genome Res*. doi:[10.1101/gr.130740.111](https://doi.org/10.1101/gr.130740.111)
- Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K et al (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430(6998):471–476. doi:[10.1038/nature02651](https://doi.org/10.1038/nature02651)
- Lower R, Lower J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci USA* 93(11):5177–5184
- Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet* 2(1):e2. doi:[10.1371/journal.pgen.0020002](https://doi.org/10.1371/journal.pgen.0020002)
- Marino-Ramirez L, Jordan IK (2006) Transposable element derived DNaseI-hypersensitive sites in the human genome. *Biol Direct* 1:20. doi:[10.1186/1745-6150-1-20](https://doi.org/10.1186/1745-6150-1-20)
- Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24(4):800–812. doi:[10.1038/sj.emboj.7600545](https://doi.org/10.1038/sj.emboj.7600545)
- McClintock B (1948) Mutable loci in maize. *Carnegie Inst Wash Year B* 47:155–169
- McClintock B (1984) The significance of responses of the genome to challenge. *Science* 226(4676):792–801
- Michaud EJ, van Vugt MJ, Bultman SJ, Sweet HO, Davisson MT, Woychik RP (1994) Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev* 8(12):1463–1472
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448(7153):553–560. doi:[10.1038/nature06008](https://doi.org/10.1038/nature06008)
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23(3):314–318. doi:[10.1038/15490](https://doi.org/10.1038/15490)
- Ng P, Wei CL, Sung WK, Chiu KP, Lipovich L, Ang CC, Gupta S et al (2005) Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat Methods* 2(2):105–111. doi:[10.1038/nmeth733](https://doi.org/10.1038/nmeth733)
- Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7(4):597–606. doi:[10.1016/j.devcel.2004.09.004](https://doi.org/10.1016/j.devcel.2004.09.004)
- Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, Goyal P et al (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* 7(9):e1002301. doi:[10.1371/journal.pgen.1002301](https://doi.org/10.1371/journal.pgen.1002301)
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G et al (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 4(8):651–657. doi:[10.1038/nmeth1068](https://doi.org/10.1038/nmeth1068)
- Rosenbloom KR, Dreszer TR, Pheasant M, Barber GP, Barber LR, Pohl A, Raney BJ et al (2010) ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res* 38 (Database issue):D620–625. doi:[10.1093/nar/gkp961](https://doi.org/10.1093/nar/gkp961)
- Sasaki T, Nishihara H, Hirakawa M, Fujimura K, Tanaka M, Kokubo N, Kimura-Yoshida C et al (2008) Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci USA* 105(11):4220–4225. doi:[10.1073/pnas.0709398105](https://doi.org/10.1073/pnas.0709398105)

- van de Lagemaat LN, Landry JR, Mager DL, Medstrand P (2003) Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet* 19(10):530–536
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231):854–858. doi:[10.1038/nature07730](https://doi.org/10.1038/nature07730)
- Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915):520–562. doi:[10.1038/nature01262](https://doi.org/10.1038/nature01262)
- Wu J, Zhou T, He J, Mountz JD (1993) Autoimmune disease in mice due to integration of an endogenous retrovirus in an apoptosis gene. *J Exp Med* 178(2):461–468
- Xiong Y, Eickbush TH (1988) Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol Biol Evol* 5(6):675–690
- Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9(10):3353–3362