Epigenetic Regulation of Mammalian Genomes by Transposable Elements

Ahsan Huda and I. King Jordan

School of Biology, Georgia Institute of Technology, Atlanta, Georgia

Transposable element (TE) sequences make up a substantial fraction of mammalian genomes and exert a variety of regulatory influences on mammalian genes. We explore the contributions of TEs to the epigenetic mechanisms that regulate mammalian genomes, emphasizing nucleosome positioning and epigenetic histone modifications. A link between TEs and epigenetics rests on the fact that underlying genetic sequences partially mediate the nature and identity of epigenetic modifications. Here, we review the studies that have uncovered histone modifications that are targeted to mammalian TE sequences and propose a series of hypotheses regarding the potential epigenetic regulatory effects of mammalian TEs. We propose that mammalian TE sequences have specific nucleosome binding properties with regulatory implications for nearby genes, are involved in the phasing of nucleosomes, and recruit epigenetic modifications to function as enhancers; that epigenetic modifications at TE sequences affect the regulation of nearby genes; and that TEs serve as epigenetic boundary elements. It is hoped that these proposed scenarios may help to serve as a roadmap for future investigations into the epigenetic regulatory effects of mammalian TEs.

Key words: **transposable elements; epigenetics; histone modifications; nucleosomes; gene regulation; human genome**

Introduction

Transposable elements (TEs) constitute a vast percentage of mammalian genomes. The ubiquity of TEs has been appreciated for some time; they have been found in the genomes of a wide variety of species from all three domains of life. Accordingly, TEs have played a substantial role in shaping the evolution of these species as evidenced by their profusion and universal distribution. Here, we speculate on the global contribution of TEs to the epigenetic mechanisms that regulate various aspects of mammalian genomes.

The relationship between TEs and epigenetic regulatory mechanisms emanates from the evolutionary tinkering between TEs and their host genomes.[1] Because of the disruptive nature of transposition, it is imperative for host genomes to evolve various tools to suppress element activity and ensure their own survival. This idea forms the core of the "genome defense" model, which proposes that many epigenetic regulatory processes came into existence to defend against the transposition of TEs.[2] Thus, TEs may be the original targets for epigenetic mechanisms that have global impacts on the regulation of gene expression and genome organization.

In eukaryotes, double-stranded DNA wraps around a core of a histone octamer to form a nucleosome. The tails of the histone proteins in the nucleosome core are often covalently modified by the addition of methyl and acetyl groups to various lysine and arginine residues. These epigenetic histone modifications, and combinations thereof, specify various chromatin states

Address for correspondence: I. King Jordan, School of Biology, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, GA 30306, USA. Voice: 404-385-2224. king.jordan@biology.gatech.edu

Natural Genetic Engineering and Natural Genome Editing: Ann. N.Y. Acad. Sci. 1178: 276–284 (2009).
doi: 10.1111/j.1749-6632.2009.05007.x

that carry regulatory "meanings" for the cell. For instance, the tri-methylation of histone H3 at lysine 9 (H3K9me3) signals the formation of heterochromatin and gene silencing, whereas mono-methylation of histone H3 at lysine 4 (H3K4me1) indicates active or open chromatin.

Most of what is currently known regarding the relationship between TEs and epigenetic histone modifications comes from studies on plants and fungi (see reviews Refs. 3–9 and references therein). TEs are abundant in the heterochromatin of a number of plant species. In *Arabidopsis thaliana*, TEs are not only enriched in heterochromatic regions in and around centromeres and knobs, but insertions of TEs into euchromatic regions can also induce the local formation of heterochromatin.[10] This TE-induced repressive chromatin can spread to epigenetically silence nearby genes. The TE rich regions in *Arabidopsis* form heterochromatin by recruiting repressive histone modifications.[11] Similarly, in the yeast *Schizosaccharomyces pombe*, TEs are targeted by repressive H3K9me3 to induce the formation of heterochromatin.[12] Interestingly, the RNA interference (RNAi) pathway is primarily responsible for the targeting of repressive modifications in both *Arabidopsis* and *S. pombe*.[4,12]

Compared to the level of understanding for plants and fungi, there is relatively little known concerning TEs and epigenetics in mammals. In this article, we emphasize the role of TEs in the epigenetic regulation of mammalian genomes via nucleosome positioning and histone modifications. To this end, we (1) review the handful of studies that exist on this subject to date and (2) propose five specific hypotheses regarding mechanisms by which TEs may be involved in the epigenetic regulation of mammalian genomes. For these TE-epigenetic hypotheses, we outline the kinds of analyses that can be done to test them along with the results that could be expected. This approach is taken to help serve as a roadmap for future investigations into the epigenetic regulatory effects of mammalian TEs.

TEs and Epigenetics in Mammals

The enrichment of repressive H3K9 di-methylation at Alu repeats in human cells was first discovered accidentally by Kondo and Issa in 2003.[13] In this study, they used an H3-Lys methylated antibody in chromatin immunoprecipitation and sequenced the recovered DNA. They found that out of the 47 independent clones sequenced 37 mapped to TEs, and 32 of these were Alu elements. This led to the conclusion that the suppression of Alu element transposition is accomplished by the targeting of H3K9 di-methylation to these sequences, consistent with the "genome defense" model for the epigenetic modifications of TEs.

Another study of the mouse epigenome by Martens *et al.* revealed the association of various repressive histone methylations with different types of repetitive DNA.[14] In this study, lysates of cross-linked chromatin in embryonic stem (ES) cells were sonicated and subjected to chromatin immunoprecipitation using a set of antibodies specific to mono-, di-, and tri-methylations of H3K9, H3K27, and H4K20. The recovered DNA was probed by clusters of primers representing major repeat classes to produce PCR fragments of the element sequences. TEs were found to have variable levels of the histone methylation marks and these seemed to vary greatly between different types of ES cells. On the other hand, tandem repeats displayed a strong affinity for the given set of histone methylations, and these remained relatively constant over different types of ES cells. RT-PCR analysis revealed the presence of double-stranded RNA (dsRNA) produced by tandem repeats indicating the involvement of RNAi mechanism for the recruitment of histone marks.

The Bernstein and Lander groups published a genome-wide map of several histone tail modifications in differentiated and undifferentiated mouse stem cells.[15,16] They used chromatin immunoprecipitation followed by high-throughput sequencing to determine the modification status of various chromatin

regions across the genome. They noted that several classes of repetitive DNA are marked by enrichment of H3K9 and H4K20 tri-methylation repressive histone marks. Among the enriched repetitive DNA classes are tandem repeats at the telomere and satellites as well as LTR-retrotransposon TEs. Intracisternal A particle (IAP) and early transposons (ETn) were the only families of LTR-retrotransponons enriched for these modifications. IAP and ETn are young, recently active families of mouse TEs and their enrichment with repressive histone modifications is consistent with the need for the host to suppress their activity. The authors of this study also predicted the involvement of the RNAi pathway for the recruitment of these modifications to TE sequences, as IAP and ETn both produce double-stranded RNA.

A more recent study characterized various repressive histone modifications in mouse embryonic stem cells using chromatin immunoprecipitation followed by microarray analysis of bound DNA fragments (so called ChIP-Chip) analysis.[17] The experimental analysis was followed by a computational step to obtain a low-resolution "birds-eye" view of the tri-methylation of H3K27 on chromosome 17. This involved dividing chromosome 17 into consecutive windows of 200kb and computing the aggregate H3K27 methylation status as well as the density of genes and different types of TEs. They observed a banding pattern throughout the chromosome with each band spanning several megabases. The enrichment of H3K27me3 was found to be correlated with silent genes and their flanking intergenic regions. H3K27me3 domains were also observed in gene-rich as well SINE-rich regions that carry many active modifications. Alternatively, the gene-poor regions that are also LINE and LTR-rich were found to be depleted in active modifications as well as tri-methylated H3K27. This indicates a global pattern of complimentary LINEs and SINEs clustering in genomic regions that carry specific epigenetic marks. Thus, LINEs and SINEs were found to divide the mouse genome into domains that entail specific epigenetic implications.

The studies described above addressing the relationship between mammalian TEs and epigenetic histone tail modifications present a complex picture and often contradict each other.[16,18] For instance, Alu elements (SINEs) were shown to be enriched for repressive H3K9me2 marks in human,[13] whereas no mouse TEs (including SINEs) showed enrichment for this particular modification.[14] A later study did show that mouse LTR elements, but not SINEs, were enriched for other repressive histone marks.[15] Furthermore, the nature and identity of histone marks targeted to TEs changed markedly between families of elements and among different cell types.[14,15,17] Given the relative paucity of global epigenetic studies in mammals, along with the contradictory and complicated results these studies yield, a comprehensive picture of the relationship between mammalian TEs and epigenetic histone modifications remains to be established. Clearly, more research is needed to elucidate the connections between mammalian TEs and epigenetics. Below, we propose a series of specific hypotheses regarding the potential contributions of TEs to the epigenetic regulation of the mammalian genome that can be taken as a roadmap for future inquiries into this area of research.

Hypothesis 1: TE Sequences Have Specific Nucleosome Binding Properties with Regulatory Implications for Nearby Genes

In eukaryotes, genomic DNA loops twice around a core of eight histone proteins to form a nucleosome—the fundamental unit of chromatin.[19] In order for this to happen, the DNA has to bend at specific intervals, which is facilitated by the occurrence of certain dinucleotide motifs. In other words, the ability of genomic DNA to bind the histone core and position nucleosomes is determined, to some extent, by

its sequence.[20] This suggests that epigenetic regulatory features, which are mediated by histone modifications, are related to the underlying genetic sequence context, much of which includes TEs in mammals. Thus, we hypothesize that patterns and levels of nucleosome binding at mammalian promoter sequences are mediated in part by the distribution of repetitive sequence elements, TEs in particular, and that the TE profiles of mammalian gene promoters exert tissue-specific regulatory effects by virtue of their binding interactions with nucleosomes.

Consistent with this prediction, there is abundant evidence from other species that points to a connection between repetitive DNA elements, the local chromatin environment and epigenetic gene regulation. For instance, in *Drosophila* and plant species, densely compact heterochromatin is enriched for both TEs and simple sequence repeats (SSRs).[21] The accumulation of TEs in heterochromatin serves to protect the host genome by mitigating deleterious effects associated with transposition and ectopic recombination between dispersed element sequences.[3] In plants it has even been shown that *de novo* heterochromatin formation can be caused by insertions of TEs into euchromatin.[10] When this occurs, the TE induced repressive chromatin environment can spread to nearby genes and silence their expression. The enrichment of TEs in heterochromatin, taken together with the repressive features of this genomic environment, has led to the proposal that heterochromatin originally evolved to serve as a genome defense mechanism by silencing TEs.[22,23] The known relationship between genome repeats, local chromatin environment and gene regulation in *Drosophila* and plants suggests the possibility that TEs may also be involved in regulating expression of genes in even more repeat-rich mammalian genomes by altering the chromatin environment.

To test the relationship between TEs, nucleosome binding and gene regulation in mammals, it is necessary to investigate the extent to which the density of TEs and nucleosome binding covary along proximal promoter sequences. Previous studies strongly suggest that such a connection exists. For instance, it is well known that nucleosome binding affinities are high distal to transcription start sites and the binding affinity decreases closer to transcription start sites where there is a "nucleosome free" region just upstream.[20] Our own work suggests that the relative density of TE insertions in promoter sequences demonstrates a similar trend.[24–26] TEs are relatively abundant far from transcription start sites and decrease steadily along more proximal promoter sequences. Interestingly, however, SSRs show the opposite trend with a marked enrichment just upstream of transcription start sites in the very same region where nucleosome binding reaches its nadir. These data suggest that different classes of genome repeats may be involved in tuning the accessibility of chromatin to transcription factors, either opening or closing depending on the kind of repeat, by virtue of their influence on nucleosome binding.

The ability of transcription factors to access proximal promoter sequences will also have regulatory implications. To explore this possibility, repetitive DNA profiles of proximal promoter regions could be used to group mammalian genes into related clusters. This would amount to a novel way of classifying genes based solely on the density and relative locations of TEs in their proximal promoter regions. Once genes are grouped in this way, the gene expression and functional properties of the resulting clusters could be compared to their characteristic repeat architectures. For example, if TE insertions in proximal promoter regions are deleterious and lead to the repression of gene expression, perhaps by closing the local chromatin, then one would expect that sets of genes with TE-rich promoters would show lower expression than those groups containing TE-depleted gene promoters. It is also tempting to speculate as to how the TE profile of gene promoters, and associated nucleosome binding patterns, may affect tissue-specific patterns of gene expression.

Hypothesis 2: TE Sequences Are Involved in the Phasing of Nucleosomes

The precise positioning of nucleosomes around certain genomic positions (anchors) is referred to as nucleosome phasing. For instance, nucleosomes show characteristic positioning upstream and downstream of transcription start sites in the human genome.[27] Nucleosome phasing is thought to have important regulatory functions by mediating access of transcription factors and RNA polymerase to genomic DNA. We hypothesize that if TEs harbor certain regulatory sequences, such as transcription factor binding sites[28–30] or transcription start sites,[31–34] they may also show characteristic patterns of nucleosome phasing. To evaluate this hypothesis, one could characterize the phasing of nucleosomes in and around various classes of TEs throughout mammalian genomes. This type of genome-scale analysis is becoming more-and-more possible owing to the accumulation of experimentally characterized nucleosome position maps for entire eukaryotic genomes. For instance, Schones *et al.* produced a human genome map of nucleosome positions in $CD4^+$ T cells using chromatin immunoprecipitation followed by high-throughput sequencing.[27] If such nucleosome positioning maps are combined with available TE annotations, typically computed using the RepeatMasker program (http://www.repeatmasker.org), the contribution of TEs to nucleosome phasing could be systematically evaluated. In addition to simply evaluating the ability of TEs to phase nucleosomes, genome-scale TE-nucleosome binding data sets could be scanned for the enrichment of TE-derived transcription factor binding sites and transcription start sites to investigate the presence of nucleosome phasing with respect to these regulatory features in various classes of TEs.

Nucleosome phasing around genomic anchors can be revealed using a recently developed algorithm called the Genomic Signal Aggregation (GSA) measure.[35] GSA evaluates the genomic distribution of short sequence tags that point to the locations of nucleosomes characterized by chromatin immunoprecipitation followed by high-throughput sequencing. Specifically, the GSA algorithm works by centering genome sequence intervals around defined genomic anchors, such as transcription start sites or TE sequences. The counts of sequence tags are aggregated with respect to their relative position around the genomic anchors. This yields position-specific distributions of tag count scores around a genomic anchor. The tag count score distributions are then evaluated for the presence of peaks and valleys. In the case of nucleosome defining sequence tags, this is done by searching for local tag count maxima within a size range approximating the length of nucleosome bound DNA. Once local maxima (i.e., peaks) are calculated in this way, the interpeak distance is calculated as the mean of the distance between adjacent plus strand and adjacent minus strand peaks. Well-positioned nucleosomes have variation of interpeak distances that do not exceed 40bp. In addition, to this algorithmic approach, visual inspection is used to evaluate nucleosome phasing.

The only genomic features known to phase nucleosomes at this time are transcription start sites and binding sites for the transcription factor and insulator protein CTCF.[35] If TE sequences could be shown to phase nucleosomes in a similar way, it would underscore their significance as genomic regulatory elements and further establish their role as mediators of genetic-to-epigenetic interactions.

Hypothesis 3: TE Sequences Recruit Epigenetic Modifications to Function as Enhancers

Enhancers are gene regulatory sequences that exert their effects from a distance, as opposed to proximal promoter sequences that control expression locally. It is of great interest to know the extent to which mammalian

enhancer sequences are derived from TEs. As described above with respect to nucleosome position maps, numerous genome-scale functional data sets have accumulated in recent years. An even more exciting development with respect to epigenetics is the recent publication of a genome-wide map of 38 epigenetic histone tail modifications in human $CD4^+$ T cells.[36] It is now possible to combine a variety of sources of functional genomic data in order to predict and locate TE-derived enhancer sequences. We hypothesize not only that TEs function as enhancers but that they do so by recruiting specific epigenetic histone tail modifications.

To evaluate this hypothesis, enhancers could be operationally characterized as (1) DNaseI hypersensitive regions that (2) contain specific binding sites for transcription factors known to bind enhancers, and (3) specific epigenetic histone modifications that characterize known enhancers, and (4) are located at least several thousand bases away from the nearest transcriptional start site. It should be possible to conduct an integrated genomic-functional analysis to find TEs that conform to this specific set of predictions. For instance, genome-wide data on experimentally characterized DNaseI hypersensitive sites could be co-located with the histone tail modification dataset by associating genes with an enhancer region that lies distal to the start site of transcription. Criteria such as these could be used to identify putative enhancer regions and co-locate transcription factor binding sites with DNaseI hypersensitive regions. Finally, these data could be intersected with TE annotations detailed with RepeatMasker to find TE-derived enhancer sequences.

Furthermore, classes of TEs that are enriched in different transcription factor binding sites can be grouped together for evaluation using the GSA analysis. As discussed in the preceding section, the GSA algorithm designates each transcription factor binding site as an anchor and assigns a score to all genomic positions nearest to it. The score is based on the number of tags that map to each genomic location as well as the distance of the location from the anchor. The aggregate distribution of scores around the transcription factor binding sites can be plotted against the distances from these sites to decipher the pattern and strength of each histone tail modification. Such an analysis would reveal the overall epigenetic environment with respect to each of the 38 histone tail modifications of the different transcription factor binding sites that originate in putative TE-derived enhancer regions.

The approaches described above would serve to identify putative TE-derived enhancer sequences with particular epigenetic modifications but would not yield any information on their ability to actually regulate the expression of host genes. To interrogate the gene regulatory functions of such TE-derived enhancers, one could evaluate whether their most likely target genes are co-regulated. For instance, each enhancer could be uniquely associated to the nearest gene and the gene's $CD4^+$ T cell expression pattern could be evaluated using microarray data. There are numerous genome-scale expression data sets available for mammals that provide expression data for tens-of-thousands of genes over scores of tissues and conditions.[37] This kind of data is ideal for precisely defining genes' expression patterns and uncovering groups of co-regulated genes. Expression can be classified according to the type of transcription factor binding site present in the enhancer as well as the individual histone tail modifications to assess the effect of each modification, and the combinations thereof, on tissue-specific gene expression.

Hypothesis 4: Epigenetic Modifications at TE Sequences Effect the Regulation of Nearby Genes

Mammalian genomes, such as the human genome, are extremely repeat rich. The human genome is at least 50% repetitive DNA, the vast majority of which consists of TEs.[38]

On the other hand, protein coding sequences make up only 1.5% of the human genome. Therefore, any given human gene amounts to a series of relatively tiny exon islands sitting in a sea of TEs, many of which make up regulatory regions of the gene.[25,34] It stands to reason that these abundant TE sequences may have a substantial effect on how nested genes are expressed. Indeed, a number of studies have related the TE environment of mammalian genes to their levels and patterns of expression. For instance, Alu (SINE) elements are enriched in and around highly and broadly expressed genes, whereas L1 (LINE) elements are more abundant in genes with lower levels of expression.[39–42] While demonstrating a connection between TEs and gene expression, such studies do not, for the most part, provide any mechanistic basis for understanding how the TEs help to regulate the genes. We hypothesize that recruiting epigenetic marks is one mechanism by which the TE environment of mammalian genes can exert specific regulatory effects.

To evaluate this hypothesis, one could integrate data on the TE environment of mammalian genes with the presence and distribution of epigenetic marks and gene expression data. Such an integrated study is possible for the human and mouse genomes, both of which have recently published genome wide maps of epigenetic histone modifications[15,36] in addition to abundant data on gene expression and TE sequence distributions. For instance, the status of particular histone modifications, or combinations thereof, as active or repressive is determined by computing the relative enrichment of expression for sets of genes that possess those modifications compared to sets of genes that do not. A similar approach could be taken for TEs. The enrichments of specific families of TEs in and around human or mouse genes could be compared simultaneously to the distribution of specific histone modifications, or combinations, and the expression enrichments of the genes. This kind of approach could point to a role for TEs in recruiting particular sets of histone modifications that entail specific regulatory outcomes. One may expect that for TEs that are targets of repressive modifications, their enrichments in and around human or mouse genes would lead to the epigenetic repression of those genes. This may differ based on the identity of the TE families being examined. Older families of TEs are less prone to be transpositionally competent and are thus not necessarily expected to serve as targets for repressive modifications. Therefore, older TEs may be associated with active histone modifications that help to upregulate co-located genes.

Hypothesis 5: TEs Serve as Epigenetic Boundary Elements

An epigenetic boundary element is a DNA sequence that can act as a buffer between active and repressed chromatin by resisting the proliferation of epigenetic changes that are characteristic of each.[43] By blocking the spread of active or repressive chromatin, boundary elements establish genomic domains of gene regulation. Boundary elements can also serve to limit the regulatory effects of enhancers to the domains in which they reside. Only a few examples of boundary elements have been described including the *gypsy* TE in *Drosophila*.[44] We hypothesize that TEs may serve as abundant and dispersed epigenetic boundary elements that help to establish chromatin based regulatory domains along mammalian genomes. This prediction can also be evaluated using the kind of integrated genomic-functional analysis described above for enhancers.

The aforementioned genome-wide map of histone tail modifications[36] includes epigenetic marks that are both active and repressive. One may expect that boundary elements would be characterized by a specific distribution of epigenetic histone marks with active modifications enriched on one side of the boundary element and repressive modifications enriched on the other. Furthermore, adjacent genes in the active region should be expressed, whereas genes in the repressed region are expected to be silent.

To test our hypothesis on TE-derived boundary elements, one could scan genome-wide histone modification maps to look for the kinds of patterns of histone tail modifications around TE sequences that specify boundary elements and integrate these data with the expression patterns of genes that flank the boundary elements. This would entail evaluation of the similarity of histone tail modifications present upstream and downstream of TE sequences. In essence, TE-derived boundary elements would show anti-correlations between the signals for active versus repressive marks upstream and downstream of the element insertion sites.

In addition to searching for TEs that partition chromatin environments, the presence of binding sites for the insulating binding protein CTCF could also be used to search for TE-derived boundary elements. CTCF is known to bind genomic DNA and prevent the spread of heterochromatin,[45] and CTCF binding sites have uncovered discrete chromatin modification domains in the human genome.[46] These global (genome-scale) approaches to evaluating the contributions of TEs to boundary elements could be used to develop algorithms that can uncover specific cases of TE-derived boundary elements.

Acknowledgments

Ahsan Huda and I. King Jordan were supported by the School of Biology, Georgia Institute of Technology. I. King Jordan was supported by an Alfred P. Sloan Research Fellowship in Computational and Evolutionary Molecular Biology.

Conflicts of Interest

The authors declare no conflicts of interest.

References

1. Jordan, I.K. 2006. Evolutionary tinkering with transposable elements. *Proc. Natl. Acad. Sci. USA* **103:** 7941–7942.
2. Yoder, J.A., C.P. Walsh & T.H. Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13:** 335–340.
3. Grewal, S.I. & S.C. Elgin. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447:** 399–406.
4. Grewal, S.I. & S. Jia. 2007. Heterochromatin revisited. *Nat. Rev.* **8:** 35–46.
5. Lippman, Z. & R. Martienssen. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431:** 364–370.
6. Slotkin, R.K. & R. Martienssen. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev.* **8:** 272–285.
7. Suzuki, M.M. & A. Bird. 2008. DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev.* **9:** 465–476.
8. Weil, C. & R. Martienssen. 2008. Epigenetic interactions between transposons and genes: Lessons from plants. *Curr. Opin. Genet. Dev.* **18:** 188–192.
9. Zaratiegui, M., D.V. Irvine & R.A. Martienssen. 2007. Noncoding RNAs and gene silencing. *Cell* **128:** 763–776.
10. Lippman, Z., A.V. Gendrel, M. Black, *et al*. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430:** 471–476.
11. Gendrel, A.V., Z. Lippman, C. Yordan, *et al*. 2002. Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science* **297:** 1871–1873.
12. Volpe, T.A., C. Kidner, I.M. Hall, *et al*. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297:** 1833–1837.
13. Kondo, Y. & J.P. Issa. 2003. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J. Biol. Chem.* **278:** 27658–27662.
14. Martens, J.H., R.J. O'Sullivan, U. Braunschweig, *et al*. 2005. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24:** 800–812.
15. Mikkelsen, T.S., M. Ku, D.B. Jaffe, *et al*. 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560.
16. Bernstein, B.E., A. Meissner & E.S. Lander. 2007. The mammalian epigenome. *Cell* **128:** 669–681.
17. Pauler, F.M., M.A. Sloane, R. Huang, *et al*. 2009. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Research* **19:** 221–233.
18. Tomilin, N.V. 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* **30:** 338–348.

19. Kornberg, R.D. & Y. Lorch. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98:** 285–294.
20. Segal, E., Y. Fondufe-Mittendorf, L. Chen, *et al*. 2006. A genomic code for nucleosome positioning. *Nature* **442:** 772–778.
21. Dimitri, P. & N. Junakovic. 1999. Revising the selfish DNA hypothesis: New evidence on accumulation of transposable elements in heterochromatin. *Trends Genet.* **15:** 123–124.
22. Henikoff, S. 2000. Heterochromatin function in complex genomes. *Biochim. Biophys. Acta* **1470:** 1–8.
23. Henikoff, S. & M.A. Matzke. 1997. Exploring and explaining epigenetic effects. *Trends Genet.* **13:** 293–295.
24. Huda, A., L. Marino-Ramirez, D. Landsman, *et al*. 2009. Repetitive DNA, nucleosome binding and human gene expression. *Gene*. **436:** 12–22.
25. Jordan, I.K., I.B. Rogozin, G.V. Glazko, *et al*. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19:** 68–72.
26. Marino-Ramirez, L., K.C. Lewis, D. Landsman, *et al*. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* **110:** 333–341.
27. Schones, D.E., K. Cui, S. Cuddapah, *et al*. 2008. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132:** 887–898.
28. Bourque, G., B. Leong, V.B. Vega, *et al*. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18:** 1752–1762.
29. Polavarapu, N., L. Marino-Ramirez, D. Landsman, *et al*. 2008. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics* **9:** 226.
30. Wang, T., J. Zeng, C.B. Lowe, *et al*. 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. USA* **104:** 18613–18618.
31. Conley, A.B., W.J. Miller & I.K. Jordan. 2008. Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet.* **24:** 53–56.
32. Conley, A.B., J. Piriyapongsa & I.K. Jordan. 2008. Retroviral promoters in the human genome. *Bioinformatics* **24:** 1563–1567.
33. Dunn, C.A., M.T. Romanish, L.E. Gutierrez, *et al*. 2006. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* **366:** 335–342.
34. van de Lagemaat, L.N., J.R. Landry, D.L. Mager, *et al*. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **19:** 530–536.
35. Fu, Y., M. Sinha, C.L. Peterson, *et al*. 2008. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4:** e1000138.
36. Wang, Z., C. Zang, J.A. Rosenfeld, *et al*. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Gen.* **40:** 897–903.
37. Su, A.I., T. Wiltshire, S. Batalov, *et al*. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101:** 6062–6067.
38. Lander, E.S., L.M. Linton, B. Birren, *et al*. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
39. Eller, C.D., M. Regelson, B. Merriman, *et al*. 2007. Repetitive sequence environment distinguishes housekeeping genes. *Gene* **390:** 153–165.
40. Ganapathi, M., P. Srivastava, S.K. Das Sutar, *et al*. 2005. Comparative analysis of chromatin landscape in regulatory regions of human housekeeping and tissue specific genes. *BMC Bioinformatics* **6:** 126.
41. Han, J.S., S.T. Szak & J.D. Boeke. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429:** 268–274.
42. Kim, T.M., Y.C. Jung & M.G. Rhyu. 2004. Alu and L1 retroelements are correlated with the tissue extent and peak rate of gene expression, respectively. *J. Korean Med. Sci.* **19:** 783–792.
43. Korenberg, J.R. & M.C. Rykowski. 1988. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* **53:** 391–400.
44. Geyer, P.K. & V.G. Corces. 1992. DNA position-specific repression of transcription by a *Drosophila* zinc finger protein. *Genes Dev.* **6:** 1865–1873.
45. Ohlsson, R., R. Renkawitz & V. Lobanenkov. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17:** 520–527.
46. Barski, A., S. Cuddapah, K. Cui, *et al*. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823–837.