# Retrotransposons and regulatory suites

James A. Shapiro

**Summary**
**Cellular differentiation and multicellular development require the programmed expression of coregulated suites of genetic loci dispersed throughout the genome. How do functionally diverse loci come to share common regulatory motifs? A new paper finds that retrotransposons (RTEs) may play a role in providing common regulation to a group of functions expressed during the development of oocytes and preimplantation embryos. Examining cDNA libraries, Peaston et al.[1] find that 13% of all processed transcripts in full-grown mouse oocytes contain RTE sequences, mostly from the MT family of retroviral-like elements. Smaller but still significant percentages of RTE sequences are found in cDNA libraries from 2-cell embryos and blastocysts. A quarter of these RTE sequences are at the 5′ ends of chimeric transcripts that also contain exons from endogenous mouse loci. These chimeric transcripts display restricted expression in oocytes and preimplantation embryos and presumably originate from developmentally regulated LTR promoters. Some, but not all, chimeric transcripts encode novel protein products.** ***BioEssays*** **27:122–125, 2005.**


## Why so many mobile elements in genomes?

More than half the human genome is repetitive DNA.[2] The transcriptionally active euchromatic portion of our DNA contains 20–30 times more nucleotides in mobile genetic elements than in protein-coding sequences. The genomes of other multicellular organisms also have large excesses of repetitive DNA. What does all this repeat DNA do? Functional genomics is starting to produce some answers.

When Barbara McClintock first discovered transposable components in the maize genome, she called them ''controlling elements'' because they could alter patterns of developmental expression when inserted near any genetic locus.[3] Although her discovery of genome fluidity has received widespread acceptance, McClintock's ideas about transposable regulatory modules fell out of favor. However, a new paper from the Jackson Laboratory in *Developmental Cell* by Anne Peaston, Alexei Evsikov and their colleagues may help change that view.[1] Examining expression patterns in mouse oocytes and preimplantation embryos, they found a major developmental role for retrotransposons (RTEs) in determining transcription initiation and mRNA structure. This paper is part of a rapidly growing literature that implicates dispersed mobile elements in the control of genome expression.

Department of Biochemistry and Molecular Biology, University of Chicago, 920 E. 58th Street, Chicago, IL 60637.
E-mail: jsha@uchicago.edu
DOI 10.1002/bies.20192
Published online in Wiley InterScience (www.interscience.wiley.com).

## Retrotransposon sequences in oocyte- and preimplantation-embryo-specific transcripts

Peaston et al. found that 13% of the cDNAs in an EST library prepared from full-grown oocytes (FGOs) contained RTE sequences. At least one quarter of these transcripts were chimeric mRNAs, containing RTE sequences at the 5′ end attached to identified cellular coding sequences (i.e. >3% of the total ESTs). Most of the RTEs were related to retroviruses and had long terminal repeats (LTRs). So it appeared that RTE LTRs could serve as promoters for cellular coding sequences (Fig. 1).

The significance of these initial observations was bolstered by differences found between FGO, 2-cell embryo and blastocyst EST libraries. Overall RTE trancript abundance was greatest in FGOs (13%), lower in 2-cell embryos (8%) and lowest in blastocysts (0.5%). Developmental specificities were even sharper when particular RTEs were examined. The two most heavily transcribed examples are the MT (mouse transcript) family of nonautonomous elements and the MuERV-L endogenous retrovirus. MT transcripts constitute 12.7% of FGO, 2% of 2-cell embryo, and 0.01% of blastocyst cDNAs. Because MT transcripts are highest in FGOs and progressively disappear during embryonic development, they seem to be an important part of the maternal contribution to post-fertilization growth and differentiation. In contrast, MuERV-L sequences are absent from FGO and blastocyst EST libraries but are in 3.2% of 2-cell embryo transcripts, apparently encoded by the zygote genome. Clearly, RTE transcription is subject to tight developmental control, and RT-PCR analysis of different oocyte and embryo stages confirms that each RTE displays its own characteristic regulatory pattern.

## Chimeric RNAs with retrotransposon sequences and endogenous exons

The chimeric RTE-cell sequence transcripts are especially interesting because they encode proteins that may function

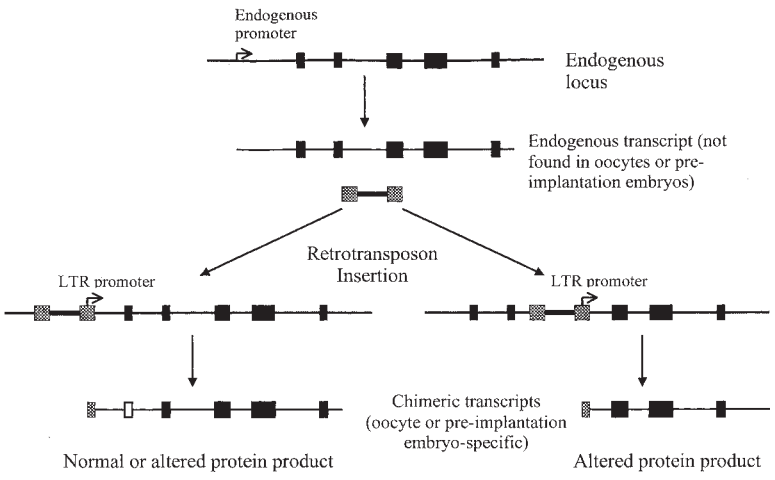


**Figure 1.** Summary of the effects of LTR retrotransposon insertion on expression of endogenous mouse loci described by Peaston et al.[1] The filled boxes show the exons found in cDNA libraries. The hatched boxes are proviral LTRs. The empty boxes show genomic exons downstream of a provirus not found in the chimeric cDNA. Since the mouse genomic sequence includes the provirus, the pre-insertion endogenous locus structure was predicted from genomic analysis.

during oocyte maturation and post-zygotic development. The distribution of specific RTEs in chimeric mRNAs is clearly non-random in the different EST libraries. In FGOs, 55/96 chimeric transcripts contain sequences from the MTA element and its relatives. MTA is interesting because it is the youngest and most-abundant MT subfamily,[4] indicating that over half the maternal chimeric transcripts are of relatively recent origin in mouse evolution. In 2-cell embryos, only 7/41 chimeras contain MT sequences while 23/41 contain sequences from MuERV-L and its relatives. These distinct abundances parallel the contribution of these elements to total transcripts in the two stages. Notably, RTE sequences found in chimeric mRNA from ovaries, oocytes or preimplantation embryos were never detected in cDNA libraries from other developmental stages or tissues.

Comparison of chimeric cDNA and genomic sequences showed that RTEs were located either 5′ to the cellular locus or in one of the introns (Fig. 1). In almost all cases analyzed, at least one of the cellular exons was missing from the chimeric transcript, even when the RTE was located upstream. The exceptions were two loci, each with a single continuous exon, presumably retrogenes.[5] Splicing occurred at a conserved 5′ donor site found in a subset of MT element LTRs and showed a preference for 3′ acceptor sites in AT-rich regions of adjacent host loci. Two-thirds of the chimeric transcripts were predicted to encode truncated or otherwise altered proteins, and only one-third appeared to encode the same protein as the non-chimeric cellular mRNA. Four loci were examined in detail, encoding the following proteins: spindlin, an Hsp40 homolog, voltage-dependent anion channel 2 and IL3-regulated nuclear factor. In all cases, the endogenous and chimeric mRNAs showed different tissue- and stage-specific expression patterns.

## Roles for mobile elements in modulating gene expression and function

The results in Peaston et al.[1] show that RTE-promoted transcription can alter expression of adjacent genetic loci via three different mechanisms:

(i) creating new developmental timing by promoter/enhancer function,
(ii) changing mRNA structure by splicing activity, and
(iii) changing protein structure by inducing alternative splicing and/or initiating transcription from an intron.

The LTR-driven transcripts described in this paper are found only in certain developmental stages, leading Peaston et al.[1] to conclude that RTEs "act as oocyte- and preimplantation-embryo-specific alternative promoters for a wide variety of host genes". This idea is very close to the McClintock notion of controlling elements, and the authors also follow her conceptually in pointing out how different RTE insertions "can result in coregulated gene expression". They further note that the capacity for developing new regulatory suites by retrotransposition events is enhanced in evolutionary potential by the ability of RTE insertions to alter the gene products produced by the coregulated loci.

A virtue of this paper is that it not only discusses the transcriptional and RNA processing contributions of RTEs but also considers the importance of heterochromatin formation in establishing epigenetic patterns of developmental gene expression. The role of repeats and mobile elements in nucleating heterochromatin formation by RNAi-directed silencing provides yet another mechanism by which RTE insertions can influence the expression of nearby coding sequences and act to construct distributed suites of coordinately regulated loci.

RNAi operates on RTE-encoded double-stranded RNA.[6] This mechanism is very likely to operate in shutting down RTE-promoted transcripts in preimplantation embryos because Peaston et al. documented anti-sense transcripts coming from MuERV-L proviruses in these stages.

Given (a) the high frequency of RTE transcripts observed in oocytes and preimplantation embryos and (b) the multiple molecular mechanisms documented in this paper that underpin RTE effects on both the timing and nature of products transcribed from a genetic locus, it is not hard to see why Peaston et al. subscribe to the idea that mobile elements are key players in evolutionary genome remodeling and subsequent developmental regulation of gene expression. Genomic analysis indicates that they have played this role often in evolution.[5,7,8] The conventional view is that these roles for RTEs and other mobile elements are the accidental consequences of their lifestyle as genomic parasites. Although usually harmful or selectively neutral, occasionally a random insertion will produce a positively selected beneficial effect. This is the view of the authors.

### A functionalist view of mobile element biology

Conventional wisdom notwithstanding, there is an alternative (and potentially more fruitful) way to think about the regulatory and evolutionary roles of the mobile genetic elements. That is to adopt a ''functionalist'' perspective on repetitive DNA as essential to genome function[9] and fully embrace McClintock's concept of controlling elements. A mobile genetic element is a highly organized composite of signals that moves as an integrated structure. Wherever the element lands, these signals will affect transcription, RNA processing, chromatin remodeling and other aspects of genome function in a complex but reproducible fashion.

In other words, mobile elements are effectively distributed genomic control modules. The data on their regulatory effects continue to grow. In addition to Peaston et al., other recent papers document RTE roles in nucleating heterochromatin formation in fission yeast[6] and maize[10] as well as a role in modulating mammalian transcript elongation.[11] The regulatory contents of mobile elements differ from one element to another. Each is able to play a characteristic role in establishing higher-level genomic regulatory configurations. For example, the glucocorticoid-inducible promoter and chromatin remodeling activity of mouse mammary tumor virus[12] can set up a dispersed network of individual loci responsive to a common hormonal trigger, the chromatin boundary/insulator signal in the *Drosophila gypsy* retrovirus[13] can define the limits of an extended but contiguous chromatin domain, and the binding sites for replication origin complexes in Ty element LTRs[14] can help organize the replication structure of the budding yeast genome.

The argument that mobile elements provide a diversified toolbox for formatting genomic functions predicts that we will find other features enhancing their effectiveness as agents of genome restructuring, such as activation of mobility in response to stress and targeting to preferred locations in the genome.[9,15,16] And that is indeed the case. In conversation, McClintock often called stress activation of controlling elements ''genome shock'', and molecular studies have reproduced her discoveries in bacteria[17–20] and plants.[21] Animals activate mobile elements in their germ lines when normal mating patterns are disrupted.[22–24] Non-randomness (''hotspotting'') of mobile element insertions has long been known in systems as distant as bacteria and *Drosophila*. The insertion specificity of *Drosophila* P elements can be modified by incorporating binding sites for transcription factors so that the engineered transposons ''home'' preferentially to chromosome regions containing loci regulated by those factors.[25–27] Genomic analysis indicates that both Ty elements and murine leukemia virus (MuLV) display a strong preference for insertions 5′ to transcribed loci,[28–30] where they are least likely to disrupt function and most likely to add novel regulatory specificity of the kind described by Peaston et al.[1]

Studies of Ty element targeting in budding yeast have provided mechanistic insight. Insertions of Ty3 can be targeted to the 5′ ends of actively transcribed loci by interactions between integration proteins and components of the RNA polymerase complex,[31] while insertions of Ty5 can be targeted to silenced regions of the genome by interaction with chromatin-silencing proteins.[32] A molecular connection between insertion and the apparatuses for transcription and chromatin remodeling provides a readily comprehensible mechanistic basis for cellular regulation of mobile element specificity. It remains unknown how far regulation of genome restructuring may have played a role in establishing control networks like the oocyte-early embryo regulon. To address this issue, we must devise experimental approaches to explore the origin of functional multi-locus systems. By documenting mobile element potentials for establishing regulatory suites, papers like Peaston et al.[1] have placed hitherto taboo questions about the control of genome remodeling on the research agenda for the 21st century.

### References


1. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, et al. 2004. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell 7:597–606.
2. International Human Genome Consortium. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.
3. McClintock B. 1987. *Discovery And Characterization of Transposable Elements: The Collected Papers of Barbara McClintock.* New York: Garland Press.
4. Smit AF. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res 21:1863–1872.
5. Brosius J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238:115–134.
6. Schramke V, Allshire R. 2003. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science 301:1069–1074.

7. Britten RJ. 1997. Mobile elements inserted in the distant past have taken on important functions. Gene 205:177–182.
8. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. 2003. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68–72.
9. Shapiro JA, Sternberg Rv. 2004. Why repetitive DNA is essential to genome function. Biol Revs 80: In press.
10. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476.
11. Han JS, Szak ST, Boeke JD. 2004. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 429:268–274.
12. Keeton EK, Fletcher TM, Baumann CT, Hager GL, Smith CL. 2002. Glucocorticoid receptor domain requirements for chromatin remodeling and transcriptional activation of the mouse mammary tumor virus promoter in different nucleoprotein contexts. J Biol Chem 277:28247–28255.
13. Gerasimova TI, Byrd K, Corces VG. 2000. A chromatin insulator determines the nuclear localization of DNA. Mol Cell 6:1025–1035.
14. Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, et al. 2001. Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science 14:2357–2360.
15. Shapiro JA. 1999. Genome system architecture and natural genetic engineering in evolution. In: Caporale L, editor. *Molecular Strategies for Biological Evolution*. Ann NY: Acad Sci 870:23–35.
16. Shapiro JA. 2002. Genome Organization and Reorganization in Evolution: Formatting for Computation and Function. In: Speybroeck LV, de Vijver GV, de Waele D, editors. *From Epigenesis to Epigenetics: The Genome in Context*. Ann NY: Acad Sci 981:111–134.
17. Shapiro JA. 1984. Observations on the formation of clones containing *araB-lacZ* cistron fusions. Molec Gen Genet 194:79–90.
18. Shapiro JA. 1997. Genome organization, natural genetic engineering, and adaptive mutation. Trends Genet 13:98–104.
19. Hall BG. 1988. Adaptive evolution that requires multiple spontaneous mutations. I. Mutations involving an insertion sequence. Genetics 120: 887–897.
20. Ilves H, Horak R, Kivisaar M. 2001. Involvement of sigma(S) in starvation-induced transposition of Pseudomonas putida transposon Tn4652. J Bacteriol 183:5445–5448.
21. Wessler SR. 1996. Turned on by stress: Plant retrotransposons. Curr Biol 6:959–961.
22. Bregliano JC, Kidwell M. 1983. Hybrid Dysgenesis. In: Shapiro JA, editor. *Mobile Genetic Elements*. NY: Academic Press; pp. 363–410.
23. O'Neill RJ, O'Neill MJ, Graves JA. 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. Nature 393:68–72.
24. Vrana PB, Fossella JA, Matteson PG, O'Neill MJ, Tilghman SM. 2000. Genetic and Epigenetic Incompatibilities Underlie Hybrid Dysgenesis in Peromyscus. Nature Genet 25:120–124.
25. Hama C, Ali Z, Kornberg TB. 1990. Region-specific recombination and expression are directed by portions of the Drosophila engrailed promoter. Genes Dev 4:1079–1093.
26. Kassis JA, Noll E, Vansickle EP, Odenwald WF, Perrimon N. 1992. Altering the insertional specificity of a Drosophila transposable element. Proc Nat Acad Sci USA 89:1919–1923.
27. Fauvarque MO, Dura JM. 1993. Polyhomeotic regulatory sequences induce developmental regulator-dependent variegation and targeted P-element insertions in Drosophila. Genes Dev 7:1508–1520.
28. Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF. 1998. Transposable Elements and Genome Organization: A Comprehensive Survey of Retrotransposons Revealed by the Complete Saccharomyces cerevisiae Genome Sequence. Genome Res 8:464–478.
29. Wu X, Li Y, Crise B, Burgess SM. 2003. Transcription start regions in the human genome are favored targets for MLV integration. Science 300: 1749–1751.
30. Mitchell RS, Beitzel BF, Schroder ARW, Shinn P, Chen H, et al. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol 2:e234.
31. Kirchner J, Connolly CM, Sandmeyer SB. 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. Science 267:1488–1491.
32. Xie W, Gai X, Zhu Y, Zappulla DC, Sternglanz R, et al. 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. Molec Cell Biol 21:6606–6614.