Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 9374–9377, September 1996
Colloquium Paper

This paper was presented at a colloquium entitled "Biology of Developmental Transcription Control," organized by Eric H. Davidson, Roy J. Britten, and Gary Felsenfeld, held October 26–28, 1995, at the National Academy of Sciences in Irvine, CA.

DNA sequence insertion and evolutionary variation in gene regulation

(mobile elements/long terminal repeats/*Alu* sequences/factor-binding sites)

ROY J. BRITTEN

Division of Biology, California Institute of Technology, 101 Dahlia Avenue, Corona del Mar, CA 92625

ABSTRACT **Current evidence on the long-term evolutionary effect of insertion of sequence elements into gene regions is reviewed, restricted to cases where a sequence derived from a past insertion participates in the regulation of expression of a useful gene. Ten such examples in eukaryotes demonstrate that segments of repetitive DNA or mobile elements have been inserted in the past in gene regions, have been preserved, sometimes modified by selection, and now affect control of transcription of the adjacent gene. Included are only examples in which transcription control was modified by the insert. Several cases in which merely transcription initiation occurred in the insert were set aside. Two of the examples involved the long terminal repeats of mammalian endogenous retroviruses. Another two examples were control of transcription by repeated sequence inserts in sea urchin genomes. There are now six published examples in which *Alu* sequences were inserted long ago into human gene regions, were modified, and now are central in control/enhancement of transcription. The number of published examples of *Alu* sequences affecting gene control has grown threefold in the last year and is likely to continue growing. Taken together, all of these examples show that the insertion of sequence elements in the genome has been a significant source of regulatory variation in evolution.**

Repeated sequences were considered to be candidates for roles in the 5′ regions of genes (1) and it was argued that they could move and supply evolutionary variation (2). Recent observations have shown that a number of sequences (known mobile elements and repeated sequences) were frequently inserted into the DNA of gene regions and can influence the regulation of a gene's expression. In a number of cases, the inserts contained sequences that bind nuclear proteins or the sequences were modified after insertion so that binding sites were formed that became effective parts of the regulation system for the gene. Where there is evidence that the insertions occurred far in the past, the examples show that insertion of DNA sequence elements was a source of variation leading to a useful positively selected change in gene regulation that survived a long period of evolution. Thus, this mechanism can be considered to be one of the sources of the DNA sequences that form control regions of genes. As a large amount of data has become available, genes can now be examined in which evolutionary sequence divergence has not fully obscured the evidence that an insert contributed new regulatory elements. Little is known of the evolutionary origin of the control regions, but there is growing literature (for reviews, see refs. 3 and 4). Sequence change, obscuring the original structure, has occurred in the long history, and the underlying rate of base substitution that is responsible is known (5).

The requirements for a convincing example are: (*i*) that there be a trace of a known class of elements present in gene region; (*ii*) that there is evidence that it has been there long enough to not just be a transient mutation; (*iii*) that some sequence residue of the mobile element or repeat participates in regulation of expression of the gene; (*iv*) that the gene have a known function. Examples in which the insert is only known to contain the site of transcription initiation have been set aside, as have all cases that might be transient mutations. The following paragraphs describe some known examples that meet the four requirements in animal genomes. Table 1 lists the cases for quick reference. Susan Wessler (unpublished results) described the story for plants at the colloquium and, although there are some good examples among plant genes, they are not included here. Some of the included examples are derived from a previous collection (6). The number of examples has been rapidly growing with a doubling time of less than a year; some of these examples and closely related issues are listed below.

EXAMPLES

Mouse Sex-Linked Protein. The human genome has two tandem complement C4 genes, *C4A* and *C4B*. However, the mouse has two closely related tandem C4-like genes (7) but only one produces a C4 protein while the other is very differently regulated and produces the sex-linked protein only in males. The androgen dependence of this mouse gene is due to the insertion of a retrovirus-related element in the 5′ region. Initiation of transcription occurs in the long terminal repeat (8–12) under control of androgen responsive sites in the insert. Evidence indicates that the sex-linked protein has a significant function in the mouse (13). The mobile element or retrovirus that was inserted is still more or less complete but the two long terminal repeats have been subject to mutation and differ from each other. The coding regions are much damaged so the conclusion is that the insertion occurred quite a long time ago and the functional sex-linked protein gene and certain control region sequences in the long terminal repeat have been preserved by selection. Stavenhagen and Robins (8) write, "The association of this transposable element with *SLP* [sex-linked protein] regulation thus provides a long-sought example of an insertional mutation that has been maintained in evolution." This example meets the four requirements.

Invasion of the 5′ Region of Five Human Amylase Genes. In this complex case, an actin pseudogene fragment is present in all five copies of the amylase gene and thus the insertion occurred before gene duplication. The promoter and transcription start site are derived from "cryptic" sites in the 3′

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Ancient mobile element insertions that affect gene regulation

Insertion	Affected gene	Comment
Alu	Human k18 keratin gene	Carries retinoic acid receptor binding sites Many *Alu* sequences include such sites
Alu	Human CD8 gene	Carries two Lyf-1, one bHLH, and one GATA-3 Binding site likely evolved under positive selection
Alu	Human gamma chain of Fc and T cell receptors	Carries both + and − control elements
Alu	Human parathyroid hormone	Negative calcium response element Occurs in some other *Alu* sequences
Alu	Human breast cancer gene BRCA-1	Estrogen receptor-dependent enhancer A divergent subclass of *Alu* sequences
Alu	Human Wilms tumor gene 1	Silencer in intron Twelve kilobases from promoter
ERV	Mouse sex-linked protein	Androgen response
ERV	Human amylase gene cluster	Activates cryptic promoter
Cassette	Sea urchin metallothionein cassette	Carries six known control sites
RSR	Four sea urchin spec genes	Includes required enhancer sequence Contains four Otx binding sites

Alu, an insertion of a human *Alu* repeated sequence; ERV, an endogenous retrovirus insert; Cassette and RSR, names of sea urchin repeated sequences.

region of the actin gene. At some stage in the evolutionary process, an endogenous retrovirus was inserted within the actin sequence but was not equally present in all five amylase genes (14–16). The evidence shows that these events occurred about 40 million years ago and that the viral insertion activated the promoter in the actin gene 3′ region. Deletion experiments, among others, showed that the functional tissue-specific promoter of human AMY1C is derived from inserted sequences and that parotid expression can be conferred by sequences derived solely from the retrovirus. The evolution of the human amylase multigene family involved a number of consecutive events involving inter- and intrachromosomal crossovers (17). However complex the evolutionary history of the amylase genes, the four requirements are met (18).

Sea Urchin Metallothionein Intron Cassette with Inverted Repeats. This is an example of a sea urchin genome insertion element that the authors identify as a cassette of six regulatory elements (19). It occurs in the first intron of a metallothionein gene (SpMTA) of *Strongylocentrotus purpuratus*. It is present in moderate genomic abundance and six other copies have been cloned that are similar in sequence to the copy in SpMTA but include an additional 3′ 80 nt. They are terminated by 24-nt inverted repeats and are 290- or 370-nt long. The terminal inverted repeat includes a regulatory DNA-binding site P5 (20). The cloned copies are about 90% identical in sequence and thus this family either is highly conserved or there are frequently newly inserted copies from a conserved source. The six regulatory sites are nearly completely identical in all copies. Transcribed copies appear frequently in mRNA in both orientations. The copy in SpMTA is shown to be an active enhancer in comparison with the very similar SpMTB that does not contain the element. Endogenous SpMTA gene expression is localized in aboral ectoderm of normal and metal-induced embryos and in transgenic embryos using a construct including the intron. However, if a 495-nt region of the intron including the cassette is deleted from the construct, the restricted spatial expression is largely eliminated (21). Thus, regulatory-binding sites within the inserted cassette are almost certainly responsible for spatial regulation as the same sites are in other sea urchin genes (22). It is curious that this metallothionein gene is spatially restricted in its expression while others (SpMTB1) are not. It is not clear whether this is an intricate but transient insertion mutation or a variation of long-term value to the sea urchin. Thus, only the third requirement (that the time of insertion is in the far past) is uncertain in this example.

The RSR Enhancer and Sea Urchin *Spec* Genes. The four known *Spec* genes of the sea urchin *S. purpuratus* have "RSR" repeated sequences in their 5′ regions. The RSR sequences are more than 600 nt in length with ≈200-nt terminal direct repeats symbolized by the two Rs. The RSR element 5′ of the *Spec2a* gene has been examined carefully (23) and includes an enhancer that is defined as the 188-nt RSR enhancer. It is required for the tissue-specific expression of the *Spec2a* gene and contains four Otx binding sites for a bicoid-class nuclear protein of sea urchins, SpOtx (24). A great deal is known about the other control elements in the 5′ region of *Spec2a* from egg injection transformation studies with a variety of constructs. What is important here is that there is no doubt that the RSR region is a central part of the control system for this gene. The RSR elements in the region of the three other *Spec* genes also include Otx-binding sites and although not studied in as great detail they also function in the control of expression of these genes. There are many additional copies of the RSR elements in the genome and they are described as a middle frequency repeat. The RSR elements in the four *Spec* gene 5′ regions were inserted far in the past since they have diverged significantly from each other and occur in different positions and number of copies. There is no simple model by which all of the RSR copies could have derived from an original RSR in an original *Spec* gene and the region later duplicated more than once. The simplest model is that RSR has been repeatedly inserted in the region of the *Spec* genes. This example meets the four requirements if we assume that RSR is an insertion sequence element and was inserted at least once in 5′ control regions of one or more *Spec* genes. It certainly occurred in the distant past and strongly affects gene regulation of genes of known function.

Note on Conservation of Sites in Many Alu Repeats. Nearly complete *Alu* sequences (n = 1500) from human DNA were compared (25) and it was found that certain positions have much reduced mutation rates. The most extreme cases are four specific changes at specific positions that occur at 18% of the average rate of the same changes at other positions. It is also found that the absence of change at many positions is correlated with the absence of change at other positions, as if regions or binding sites were together protected from change in blocks, possibly by sequence-dependent selection. This suggests that many of the *Alu* sequences may have some sort of sequence-dependent role rather than being purely selfish DNA. Another interesting observation is that (at last look) of the 356,000 human expressed sequence tags that were in GenBank in February 1996, 25,000 contain recognizable *Alu* sequences and many include nearly full-length *Alu* sequences located at the 5′ end of the transcripts. These general observations may one day have direct significance, but this review focuses on the *Alu* sequences that have inserted in gene regions

and taken on regulatory roles. At the present time, six examples of *Alu* sequence inserts are known to meet the four requirements of this review and are described below.

Retinoic Acid Receptor-Binding Sites in Many Alu Repeats. Vansant and Reynolds (26) have observed that *Alu* sequences include functional binding sites for retinoic acid receptors. The consensus sequences for evolutionarily recent *Alu* subclasses contain three hexamer half-sites related to the consensus AGGTCA arranged as direct repeats with a spacing of 2 bp, which form apparently effective binding sites. Vansant and Reynolds also studied a particular example of an *Alu* sequence that had been previously implicated in the regulation of the human keratin *K18* gene (27). They showed directly that the keratin K18 *Alu* sequence double half-sites bound bacterially synthesized retinoic acid receptors. These sites were shown to function as a retinoic acid response element in transiently transfected CV-1 cells, increasing transcription of a reporter gene by a factor of about 35. The *Alu* sequence in the keratin gene differs by 10% from the modern consensus and 7% from the older class II consensus (28) and thus has probably been present in this gene for many millions of years. This example meets all of the four requirements listed earlier. However, an additional question is raised about the possible function and dangers of tens of thousands of retinoic acid receptor-binding sites in *Alu* sequences scattered throughout the human genome.

DNA-Binding Sites in an *Alu* Sequence Help Regulate the CD8 Gene Transcription. Hambor *et al.* (29), examining DNase I hypersensitive sites, detected an *Alu* sequence in the last intron of the CD8 gene that is a hypersensitive site and operates as part of an enhancer specific for T lymphocytes. The *Alu* sequence is apparently responsible for the hypersensitivity as well as including nuclear protein-binding sites. Within the *Alu* sequence, four transcription factor-binding sites were found by direct tests to be effective: two LyF-1 sites, one bHLH, and one GATA-3 site. The evidence for positively selected change in the *Alu* sequence, suggesting that this sequence had adapted to function in this location, is as follows. Comparing this *Alu* sequence with its probable source gene, there are seven (non-CpG) differences, four of which are in the 11% of the sequence occupied by the sites just mentioned. Two of the changes are in the GATA-3 site and were shown individually to be necessary for its function. This *Alu* sequence differs by 4% from the modern consensus and 8% from the older class II consensus (28). Thus, it was probably inserted sometime in the relatively recent evolution of the primate lineage after the major class II insertions 30 to 50 million years ago. It would be fascinating to learn what role the *Alu* insert performed in the evolution of the centrally important CD8 gene. This example meets all of the four requirements.

Negative Calcium Response Element in an *Alu* Sequence. The negative calcium response element in the 5′ flank of the human parathyroid hormone gene, which consists of a palindromic core sequence and several upstream T residues, has been shown to lie within an *Alu* sequence (30). The *Alu* sequence is present as a reverse complement about 3.6 kb upstream of the start of transcription, starting with a long poly(T) sequence. The end of the poly(T) and the next 14 nucleotides form the negative calcium response element site. A GenBank data base search (December 1995) was made with this sequence and among the 111 matching sequences were several in genes that are also affected by calcium concentration suggesting the existence of a number of other *Alu* sequences that affect gene control. The *Alu* sequence in this example is almost complete but is 30% divergent from the modern sequences that recently have been inserted in the human genome. It is a member of the oldest known class I, differing 26% from the consensus of class II. This example meets the four requirements and suggests that there are more examples to come.

Sites in an *Alu* Sequence Control Tissue Specific Expression of IgE Gamma Chain. The transcription control of the gamma chains of the high affinity IgE receptor (FceRI) was examined and two control sites that bind nuclear factors were identified within a 5′ *Alu* sequence in the promoter region (31). The promoter is hematopoietic specific. The gene is constitutively expressed in certain cell types such as basophils but is only expressed in subsets of other cells such as T cells. It consists of five exons, spans four kilobases, and has a principal transcription start site 25 nt 5′ of the ATG. There is an *Alu* sequence from −313 to −582 and two control regions are included within it. Brini and colleagues (31) write, "In conclusion, we have shown that the transcription of the FceRI-gamma chain gene is regulated by positive and negative cis-acting elements. These elements are part of an *Alu* repeat, are only active in hematopoietic cells, and are recognized by DNA-binding proteins in a cell type-specific manner." The *Alu* sequence is most similar to the older class II, differing by 11% from the consensus and differs by 15% from the modern consensus. As in the above example, it is almost certain that this was not a recent insertion and therefore this example meets the four requirements.

***Alu* Sequences that Function as Estrogen Receptor-Dependent Transcriptional Enhancers.** Estrogen responsive promoters/enhancers were picked from a human DNA library using a yeast genetic selection system (32). Sixteen were found, one was examined, and the estrogen responsive region turned out to be within an *Alu* sequence of the older class II family, differing by 22% from the consensus and 27% from the modern consensus (28). The binding sites were identified by sequence comparison to known estrogen receptor binding sites that start with GGTCA and include its palindrome TGACC. The gene *BRCA-1* is also estrogen responsive, so a library of 50 kb of the 5′ region was screened with the same selection method and a fragment was isolated that was effective as an estrogen responsive promoter/enhancer. This clone included a nearly identical *Alu* sequence that contained the same receptor-binding sites as the *Alu* sequence mentioned above. While direct evidence is lacking, this is possibly the region that confers estrogen responsiveness on the *BRCA-1* gene. The *Alu* sequence indicates that it is a member of the older class II. Interestingly, the T of the ggTca is present in the two active sites, whereas the consensus of class II has a C substitution for the T and such *Alu* sequences are inactive as estrogen responsive elements. The consensus for the more recent *Alu* classes also has the C and such *Alu* sequences are also inactive. Norris *et al.* (32) take this T for C substitution as defining a new subclass of *Alu* sequences within class II. Norris and colleagues (32) state: "... this work provides strong evidence that a significant subset [of *Alu* sequences] can confer estrogen responsiveness upon a promoter within which they are located. Clearly, *Alu* sequences must now be considered as important contributors to the regulation of gene transcription in estrogen receptor-containing cells." If this statement is correct, this example meets the four requirements and suggests that there are more examples to come.

An *Alu* Sequence Is a Silencer in the *WT1* Gene. Wilms tumor 1 (*WT1*) is a gene normally expressed in a limited set of tissues, but often highly expressed in a variety of tumors. Twelve kilobases from the promoter, the third intron of the *WT1* gene contains an *Alu* sequence that has been shown to be a silencer capable of strongly reducing the expression of the gene (33). This silencer acts in a position- and orientation-independent manner in many cell lines tested but is inactive in renal cells. The *Alu* sequence is divergent by 17% from class II and 20% from the modern consensus and the authors have identified within it 15 short DNA sequences that potentially could act as binding sites for known transcription factors. While the function of the *WT1* gene has not been identified, it is expressed in specific tissues and stages in man and mouse. Thus, this example appears to meet the four requirements.

DISCUSSION

Since I started on this project a year ago, four of the six examples of ancient *Alu* sequences active in gene control have

been published. Due to their recent publications, this article, to my knowledge, is the first opportunity where all of these examples have been brought together. One common feature of these papers is that none report fingerprint observations that could specifically identify the binding sites that are important to their function. There may be other important sequences in the specific *Alu* repeats in addition to the binding sites identified by sequence similarity to known binding sites.

The example in the CD8 gene is not as ancient and divergent as the other five *Alu* sequences, but they all have been under selection for long periods of time as active elements involved in transcriptional regulation. The types of functions they carry out are very distinct. One is involved as a retinoic acid receptor-binding site and another is involved an estrogen receptor-binding site; both are identified as enhancers. A third sequence is a silencer located far from the promoter region. The Cd8 gene example is also far from the promoter region and forms a nuclease-sensitive chromosomal site as well as including several required binding sites. One example includes both positive and negative control functions. The last includes a negative calcium response element. Three of the six sequences are in the promoter regions of the genes, while the 5′ location is unknown for the *BRCA-1* example. It is likely that these are not unusual examples and many other genes will be found that include in their promoters *Alu* sequences carrying the retinoic acid receptor- and estrogen receptor-binding sites.

There are many significant observations that lie outside the restrictions of this paper and a few are worth brief mention. *Alu* sequences are a major source of mutation and variation (34). They have been identified as negative regulatory elements in at least two examples (35, 36); different families of *Alu* sequences bind different sets of nuclear proteins (37). In a related case, there is a 27-bp sequence important in negative regulation of mouse immunoglobulin kappa light chain that is apparently derived from the mouse B1 repetitive element (38). However, there is no sign of the rest of the B1 element in the gene region and this case differs from the examples discussed in this review because an insertion event cannot be identified. The 27-bp sequence has a barely recognizable homology to the human *Alu* sequence. A nearly identical 27-bp sequence is involved in human and rabbit kappa gene control. This may represent an example where the insertion occurred so far in the past that all evidence except the functional sequence itself has faded away. *Alu* sequences also cause variation by inclusion in coding sequences for proteins and 17 examples have been identified (39); however, all of the examples that have been studied in enough detail cause deleterious effects and no examples are identifiable that are ancient, long lasting, and beneficial.

During their existence for the last 50 to 100 million years, *Alu* sequences have contributed to the function of many useful genes and as a source of mutation and variation they have strongly influenced primate evolution. It must be emphasized that high frequency repeats such as the *Alu* sequence are one of many types of elements that insert into gene regions and are capable of influencing the regulation of transcription. The as yet primitive knowledge of insertion elements in animal genomes shows that five different classes of insertion elements have had lasting effects, and more are likely to identified. The evidence in plants shows a number other elements to be of major importance in evolutionary history (40). It will be fascinating to learn whether or not DNA sequence insertion is the primary source of evolutionary variation in gene expression and thus in biological evolution.

Note Added in Proof. Recently, it has been found that an *Alu* element in the myeloperoxidase gene promoter contains a composite Sp1–thyroid hormone–retinoic acid response element (41). This example, which is similar to the first example in Table 1, appears to meet all four requirements.

This work was supported by grants from the National Institutes of Health.

1. Britten, R. J. & Davidson, E. H. (1969) *Science* **165,** 349–358.
2. Britten, R. J. & Davidson, E. H. (1971) *Quart. Rev. Biol.* **46,** 111–138.
3. Britten, R. J. (1996) in *Human Genome Evolution*, eds. Jackson, M., Strachan, T. & Dover, G. (BIOS, Oxford), pp. 211–228.
4. McDonald, J. F. (1993) *Curr. Opin. Genet. Dev.* **3,** 855–864.
5. Britten, R. J. (1986) *Science* **231,** 1393–1398.
6. Britten, R. J. (1996) *Mol. Phylogenet. Evol.* **5,** 13–17.
7. Dodds, A. W. & Law, S. K. (1990) *Biochem. J.* **265,** 495–502.
8. Stavenhagen, J. B. & Robins, D. M. (1988) *Cell* **55,** 247–254.
9. Robins, D. M. & Samuelson, L. C. (1992) *Genetica* **86,** 191–201.
10. Adler, A. J., Danielsen, M. & Robins, D. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11660–11663.
11. Adler, A. J., Scheller, A. & Robins, D. M. (1993) *Mol. Cell. Biol.* **13,** 6326–6335.
12. Robins, D. M., Scheller, A. & Adler, A. J. (1994) *J. Steroid Biochem. Mol. Biol.* **49,** 251–255.
13. van den Berg, C. W., Demant, P., Aerts, P. C. & Van Dijk, H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10711–10715.
14. Samuelson, L. C., Wiebauer, K., Gumucio, D. L. & Meisler, M. H. (1988) *Nucleic Acids Res.* **16,** 8261–8276.
15. Samuelson, L. C., Wiebauer, K., Snow, C. M. & Meisler, M. H. (1990) *Mol. Cell. Biol.* **10,** 2513–2520.
16. Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C. & Meisler, M. H. (1992) *Genes Dev.* **6,** 1457–1465.
17. Groot, P. C., Mager, W. H., Henriquez, N. V., Pronk, J. C., Arwert, F., Planta, R. J., Eriksson, A. W. & Frants, R. R. (1990) *Genomics* **8,** 97–105.
18. Meisler, M. H. & Ting, C. N. (1993) *Crit. Rev. Oral Biol. Med.* **4,** 503–509.
19. Nemer, M., Bai, G. & Stuebing, E. W. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10851–10855.
20. Calzone, F. J., Hoog, C., Teplow, D. B., Cutting, A. E., Zeller, R. W., Britten, R. J. & Davidson, E. H. (1991) *Development (Cambridge, U.K.)* **112,** 335–350.
21. Nemer, M., Stuebing, E. W., Bai, G. & Parker, H. R. (1995) *Mech. Dev.* **50,** 131–137.
22. Thiebaud, P., Goodstein, M., Calzone, F. J., Theze, N., Britten, R. J. & Davidson, E. H. (1990) *Genes Dev.* **4,** 1999–2010.
23. Gan, L., Zhang, W. & Klein, W. H. (1990) *Dev. Biol.* **139,** 186–196.
24. Mao, C.-A., Gan, L. & Klein, W. H. (1994) *Dev. Biol.* **165,** 229–242.
25. Britten, R. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5992–5996.
26. Vansant, G. & Reynolds, W. F. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8229–8233.
27. Thorey, I. S., Cecena, G., Reynolds, W. & Oshima, R. G. (1993) *Mol. Cell. Biol.* **13,** 6742–6751.
28. Britten, R. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6148–6150.
29. Hambor, J. E., Mennone, J., Coon, M. E., Hanke, J. H. & Kavathas, P. (1993) *Mol. Cell. Biol.* **13,** 7056–7070.
30. McHaffie, G. S. & Ralston, S. H. (1995) *Bone* **17,** 11–14.
31. Brini, A. T., Lee, G. M. & Kinet, J.-P. (1993) *J. Biol. Chem.* **268,** 1355–1361.
32. Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., Iglehart, J. D., Deininger, P. L. & McDonnell, D. P. (1995) *J. Biol. Chem.* **270,** 22777–22782.
33. Hewitt, S. M., Fraizer, G. C. & Saunders, G. F. (1995) *J. Biol. Chem.* **270,** 17908–17912.
34. Maraia, R. J. (1995) *The Impact of Short Interspersed Elements (SINEs) on the Host Genome* (Landes, Austin, TX).
35. Saffer, J. D. & Thurston, S. J. (1989) *Mol. Cell. Biol.* **9,** 355–364.
36. Wu, J., Grindlay, G. J., Bushel, P., Mendelsohn, L. & Allan, M. (1990) *Mol. Cell. Biol.* **10,** 1209–1216.
37 Tomilin, N. V., Bozhkov, V. M., Bradbury, E. M. & Schmid, C. W. (1992) *Nucleic Acids Res.* **20,** 2941–2945.
38. Saksela, K. & Baltimore, D. (1993) *Mol. Cell. Biol.* **13,** 3698–3705.
39. Makalowski, W., Mitchell, G. A. & Labuda, D. (1994) *Trends Genet.* **10,** 188–193.
40. Purugganan, M. D. & Wessler, S. R. (1995) *Mol. Ecol.* **4,** 265–269.
41. Piedrafita, F. J., Molander, R. B., Vansant, G., Orlova, E. A., Pfahl, M. & Reynolds, W. F. (1996) *J. Biol. Chem.* **271,** 14412–14420.