Conspiracy of silence among repeated transgenes

Steven Henikoff*

Summary

Transgenic experiments in vertebrates often involve the insertion of tandem multiple-copy arrays at single sites. For many transgenes, expression is unpredictable from site to site, a phenomenon usually attributed to a repressive environment caused by nearby sequences. However, an alternative explanation comes from evidence that transgene repeat arrays in flies condense into heterochromatin, suggesting that low levels of expression in vertebrate transgene arrays might result from interactions between repeats within the array. A recent experiment using transgenic mouse lines demonstrates that reduction in copy number of silenced transgenes within an array leads to a striking increase in expression,[1] demonstrating that silencing is intrinsic to the array, and is not attributable to position effects of nearby sequences. This work calls into question functions that have been attributed to vertebrate locus control regions and boundaries, and draws attention to the notion that repeat-induced gene silencing is a system for protection of eukaryotic genomes against threatening sequence elements. *BioEssays* **20**:532–535, 1998. © 1998 John Wiley & Sons, Inc.

A bold conjecture

"A heterochromatic segment should arise every time that a minute euchromatic region undergoes repeated reduplication in the genotype and the replicas remain adjacent to each other in the chromosome" - Pontecorvo, 1944.[2]

It seems astonishing that the idea that tandem repeats form heterochromatin predates the realization that DNA is the genetic substance. What was the basis for this bold conjecture? Heterochromatin is the chromosomal material that appears condensed throughout the cell cycle, unlike euchromatin, which decondenses at telophase. The cytologic distinction between euchromatin and heterochromatin is especially prominent in the polytene chromosomes of *Drosophila* salivary glands. A fine level of detail is visible in polytene chromosomes, which consist of >1,000 precisely aligned interphase chromatids from each homologue. In polytene chromosomes, the gene-rich euchromatic arms are readily distinguished from the gene-poor heterochromatin, which coalesces to form a distinctive chromocenter. Unlike the banded euchromatin, the polytene chromocenter has a dense grainy appearance. Pontecorvo realized that the forces of somatic pairing, which cause the precise alignment of euchromatin in the arms, may act between repeats present in the heterochromatin of all chromosomes to cause the formation of the amorphous chromocenter. So Pontecorvo's conjecture was motivated by the notion that heterochromatin is repetitive, whereas euchromatin is not.

In subsequent decades, the presence of tandem repeats in abundance in heterochromatin but not euchromatin was amply confirmed. In *Drosophila*, heterochromatin consists primarily of simple sequence repeat arrays, in which units of 5-, 7-, 10-, or 12-mers are repeated in tandem,[3] whereas euchromatin consists of primarily single-copy sequences. This distinction is not absolute: Single-copy genes are found embedded in heterochromatin, and dispersed middle-repetitive sequences and short microsatellite arrays make up a large minority of euchromatin. Furthermore, middle-repetitive sequences derived from mobile elements make up a large minority of heterochromatin, typically in clusters. Adding to the confusion concerning the sequence organization of heterochromatin is the fact that repeat units are not always simple sequences, but can be much longer. Examples of these

Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA.
*Correspondence to: Steven Henikoff, Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109–1024. E-mail: steveh@muller.fhcrc.org

relatively complex arrays are the ~10-megabase-long 359-bp repeat array on the *D. melanogaster* X chromosome and the 171-bp alpha satellite repeat that is found at all human centromeres and comprises ~10% of the African green monkey genome. Despite these striking correlations between large repeat blocks and heterochromatin, the possibility that repeats alone *cause* heterochromatin to condense was almost absent from modern discussions of heterochromatin formation, until recently. Rather, specific sequences were hypothesized to initiate and terminate heterochromatin propagation along a chromosome.[4]

Transgene repeat arrays can have heterochromatic properties

Research on heterochromatin has been dominated by studies in *Drosophila* since the 1920s. Nevertheless, a mouse study provided the first direct evidence that the Pontecorvo conjecture might be correct. When vertebrates are transformed with exogenous DNA, chromosomal insertions almost always consist of long tandem arrays of the transfected transgene-containing DNA.[5] Sometimes transgene arrays can comprise hundreds or even thousands of copies. In a cytologic study of one such extreme example, Manuelidis found that an 11 megabase transgene array in a mouse line was unmistakably heterochromatic in appearance, suggesting that repeats can be sufficient for heterochromatin formation.[6]

Heterochromatin describes a cytologic state, but gene expression can also be used to assay for the formation of heterochromatin, such as in classic position-effect variegation (PEV) in *Drosophila*.[7] In PEV, juxtaposition of a reporter gene to heterochromatin leads to a distinctively variegating phenotype. The phenotype seems to consist of a random mixture of fully mutant and fully wild-type spots and clones, suggesting that the gene is "heterochromatized" and silenced with a frequency that varies between mutant PEV alleles. The gene is not permanently affected in PEV: Expression can be restored by moving the gene away from heterochromatin. The PEV assay has made possible detection of heterochromatin formation at transgene arrays even when the arrays are too small to be visualized as heterochromatin under the microscope. Furthermore, the availability of *trans*-acting suppressors of PEV that are highly specific in their action[8] provides a means of distinguishing heterochromatin-induced silencing from other types of silencing.

The properties of PEV made possible a test of the Pontecorvo conjecture in *Drosophila*. Unlike in mammals, germ-line transformation in *Drosophila* rarely gives rise to tandem arrays; rather the P transposon transformation vector inserts as a single copy. However, P transposase often causes previously inserted copies to duplicate themselves in direct tandem or inverted orientation, so that a single-copy insertion can be expanded into a small cluster. When this was

done using a mini-*white* eye pigment gene reporter, arrays of as few as 3 copies showed PEV-like phenotypes.[9] Furthermore, the frequency of silencing increased with copy number. Copy number-dependent silencing was seen for tandem arrays, although reversals in orientation increased the degree of silencing, supporting the idea underlying Pontecorvo's conjecture, that pairing of repeats would cause heterochromatin to form: Inverted repeats might form hairpin structures that would pair more easily than looped structures. That this variegation was indeed caused by heterochromatin formation was confirmed by showing suppression of silencing by suppressors of PEV, and more recent work has demonstrated additional heterochromatic properties of mini-*white* repeat arrays.[10]

Heterochromatin formation at transgene arrays seems not to be restricted to the animal kingdom. In plants, transgene silencing is common, and in some cases, silencing occurs only when a transgene is locally repeated.[11] Changing the number of copies within a small array at a single site led to the demonstration of repeat-induced gene silencing (RIGS) in *Arabidopsis thaliana*,[12] and this silencing has been shown to alter local chromatin structure.[13] In *Neurospora*, vegetative gene silencing at repeats[14] might also be a manifestation of heterochromatin formation.

RIGS in vertebrates

Transgene arrays in vertebrates often show unpredictable levels of expression. Variegation is sometimes observed, which has been interpreted by some authors as evidence for heterochromatin formation.[15,16] Indeed, heterochromatin is capable of silencing mammalian genes, because transgene insertions into mouse heterochromatin cause silencing[17–19] similar to PEV-like silencing of genes into *Drosophila* heterochromatin.[20] However, silencing is generally not proportional to copy number, so that insertions with the same number of copies at different sites result in different levels of expression. This site-to-site variability led to the widespread view that vertebrate transgene arrays are subject to position effects caused by repressive influences of neighboring sequences.[5,21] This finding does not explain why position effects on transgene arrays are *always* repressive. Alternatively, long-range interactions may be at play. For example, moving a *Drosophila* mini-*white* transgene array closer to blocks of natural heterochromatin without changing neighboring sequences strikingly increases the degree of silencing.[10] This and other "heterochromatin distance effects" are thought to result from long-range associations between heterochromatic blocks along a chromosome,[22,23] probably similar to the coalescence of heterochromatin seen at the polytene chromocenter. By this view, a strongly silenced vertebrate transgene array at one site would be closer to a block of natural heterochromatin than is a weakly silenced site. Because comparisons always involved transgene arrays at
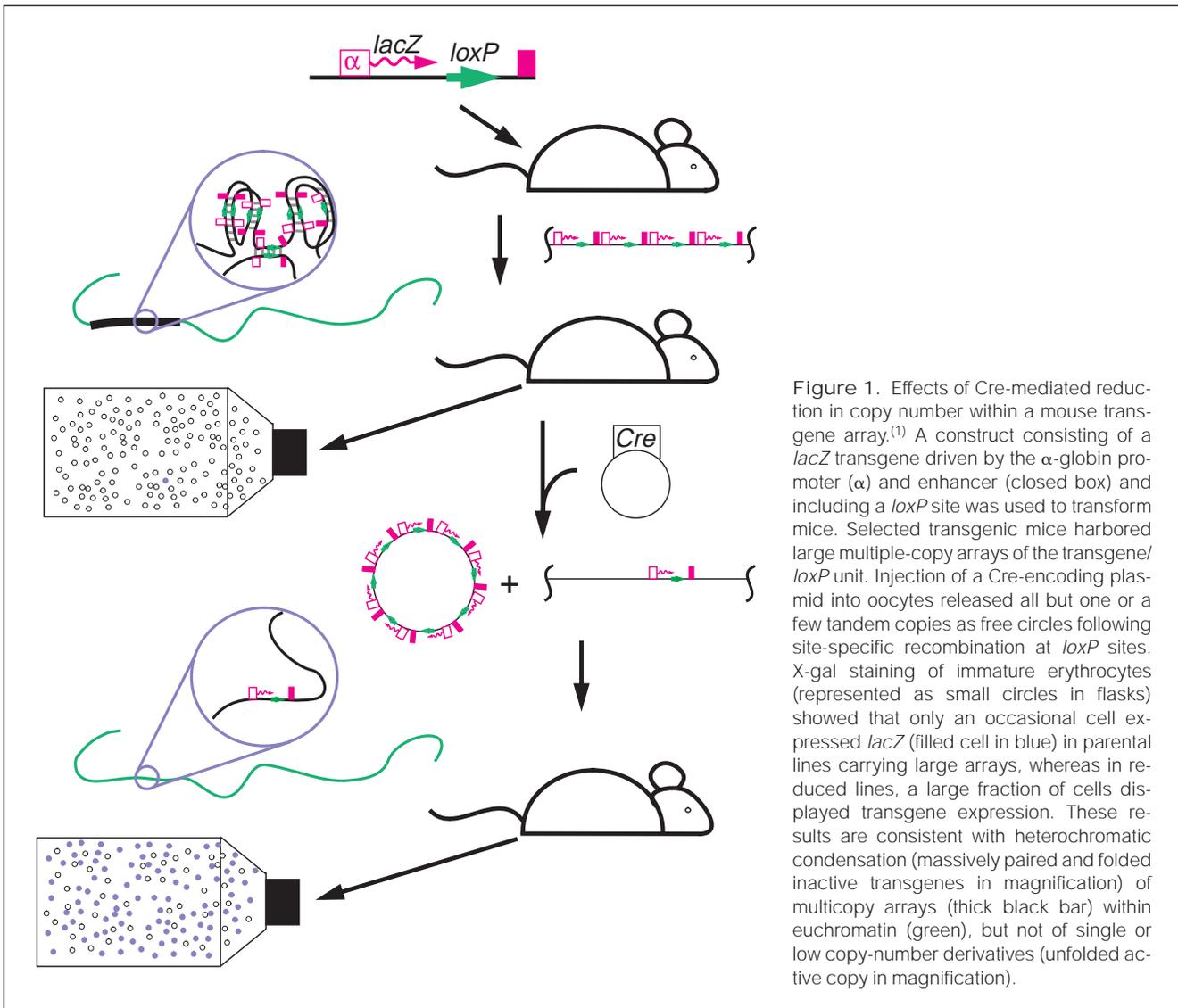
Figure 1. Effects of Cre-mediated reduction in copy number within a mouse transgene array.[1] A construct consisting of a *lacZ* transgene driven by the α-globin promoter (α) and enhancer (closed box) and including a *loxP* site was used to transform mice. Selected transgenic mice harbored large multiple-copy arrays of the transgene/*loxP* unit. Injection of a Cre-encoding plasmid into oocytes released all but one or a few tandem copies as free circles following site-specific recombination at *loxP* sites. X-gal staining of immature erythrocytes (represented as small circles in flasks) showed that only an occasional cell expressed *lacZ* (filled cell in blue) in parental lines carrying large arrays, whereas in reduced lines, a large fraction of cells displayed transgene expression. These results are consistent with heterochromatic condensation (massively paired and folded inactive transgenes in magnification) of multicopy arrays (thick black bar) within euchromatin (green), but not of single or low copy-number derivatives (unfolded active copy in magnification).

different sites, it was not possible to determine whether neighboring sequences or RIGS was responsible for unpredictable transgene expression.

In a recent paper, Garrick et al. have distinguished models for transgene silencing by comparing the expression of arrays of different size at a single site in the mouse genome[1] (Fig. 1). Their initial transgene construct consisted of α-globin regulatory elements fused to a *lacZ* reporter transgene and a single *loxP* site. They chose two transgenic mouse lines carrying single arrays of >100 copies, in which only a small fraction of the cells showed erythroid-specific *lacZ* expression. A plasmid encoding Cre recombinase was injected into fertilized oocytes, which were then transferred to foster mothers. In both lines, transient expression of recombinase reduced the array, down to five copies in one case and one

copy in the other. These reductions caused a two to three order-of-magnitude increase in the percentage of *lacZ*-expressing cells. Reduction in transgene copy number was accompanied by increased transcription, increased chromatin accessibility and decreased cytosine methylation. These results indicate that the silencing of the large arrays is due to RIGS, and is not repression by neighboring sequences, consistent with the possibility that local heterochromatin formation may be responsible.

The findings of Garrick et al. call for a rethinking of concepts based on the behavior of transgenes in arrays. The concept of a locus control region (LCR) originated with the demonstration that the 3′ regulatory region of the beta globin locus provides copy number-dependent expression in transgene arrays.[24] The concept of a vertebrate boundary ele-

ment likewise originated with the demonstration that sequences flanking the chicken lysozyme gene could confer copy number dependent expression in transgene arrays.[25] These concepts were based on the interpretation that LCRs and boundaries were acting to protect transgene reporters against repressive local chromatin environments. This may be; however, the assays used to define these elements more likely measured their ability to resist RIGS, and thus may not be relevant to properties of these elements in the single-copy context.

Perspectives

Transgenic methodology has been of enormous value for biological research in general, and promises to have a comparable impact in agriculture and medicine. Researchers usually assume that an integrated transgene's behavior reflects normal influences at the site of insertion. However, genomes have mechanisms for recognizing potentially threatening sequence elements, like transposons, and silencing them. Cytosine methylation is believed to serve this purpose in most eukaryotic taxa.[26] RIGS seems to be another genome defense mechanism, one that operates both in organisms that have cytosine methylation such as vertebrates and plants and those that do not such as insects.

If RIGS operates by heterochromatizing repeat arrays and clusters when they appear in a natural context, then this system may be responsible for heterochromatin formation in general. For example, the evolving neo-Y chromosome of *D. miranda* is host for clusters of retrotransposons. These regions have lost genetic functions and have become heterochromatic.[27] In pericentric regions with few genes, a RIGS system acting upon repeat arrays and clusters may be necessary to provide a suitable heterochromatic environment for centromere function.[28] In euchromatin, where most genes reside, a RIGS system that recognizes local repetitiveness may help restrict the expansion of microsatellite repeats. By uncovering RIGS systems in a wide variety of organisms, transgenic methodology may have inadvertently provided us with a probe of important, but poorly understood, processes of genome defense and heterochromatin formation.

References

1 Garrick D, Fiering S, Martin DIK, Whitelaw E (1998) Repeat-induced gene silencing in mammals. *Nat Genet* **18**:56–59.

2 Pontecorvo G (1944) Structure of heterochromatin. *Nature* **153**:365–367.
3 Lohe AR, Hilliker AJ (1995) Return of the H-word (heterochromatin). *Curr Opin Genet Dev* **5**:746–755.
4 Tartof KD, Bishop C, Jones M, Hobbs CA, Locke J (1989) Towards an understanding of position effect variegation. *Dev Genet* **10**:162–176.
5 Palmiter RD, Brinster RL (1986) Germ-line transformation of mice. *Annu Rev Genet* **20**:465–499.
6 Manuelidis L (1991) Heterochromatic features of an 11-megabase transgene in brain cells. *Proc Natl Acad Sci USA* **88**:1049–1053.
7 Weiler KS, Wakimoto BT (1995) Heterochromatin and gene expression in *Drosophila. Annu Rev Genet* **29**:577–605.
8 Henikoff S (1996) Dosage-dependent modification of position-effect variegation in *Drosophila. Bioessays* **18**:401–409.
9 Dorer DR, Henikoff S (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila. Cell* **77**:993–1002.
10 Dorer DR, Henikoff S (1997) Transgene repeat arrays interact with distant heterochromatin and cause silencing in *cis* and *trans. Genetics* **147**:1181–1190.
11 Matzke MA, Matzke AJM, Mittelsten-Scheid O (1994) Inactivation of repeated genes: DNA-DNA interaction? In J. Paszkowski (ed): *Homologous Recombination in Plants.* Amsterdam: Kluwer, pp. 271–307.
12 Assaad FF, Tucker KL, Signer ER (1993) Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis. Plant Mol Biol* **22**:1067–1085.
13 Ye F, Signer ER (1996) RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc Natl Acad Sci USA* **93**:10881–10886.
14 Selker EU (1997) Epigenetic phenomena in filamentous fungi: Useful paradigms or repeat-induced confusion? *Trends Genet* **13**:296–301.
15 Martin DIK, Whitelaw E (1996) The vagaries of variegating transgenes. *Bioessays* **18**:918–923.
16 Dorer DR (1997) Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res* **6**:3–10.
17 Festenstein R, Tolaini M, Corbella P, Mamalaki C, Parrington J, Fox M, Miliou A, Jones M, Kioussis D (1996) Locus control region function and heterochromatin-induced position effect variegation in transgenic mice. *Science* **271**:1123–1125.
18 Dobie KW, Lee M, Fantes JA, Graham E, Clark AJ, Springbett A, Lathe R, McClenaghan M (1996) Variegated transgene expression in mouse mammary gland is determined by the transgene integration locus. *Proc Natl Acad Sci USA* **93**:6659–6664.
19 Milot E, Stouboulis J, Trimborn T, Wijgerde M, de Boer E, Langeveld A, Tan-Un K, Vergeer W, Yannoutsos N, Grosveld F, Fraser P (1996) Heterochromatic effects on the frequency of duration of LCR-mediated gene transcription. *Cell* **87**:105–114.
20 Wallrath LL, Elgin SCR (1995) Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes Dev* **9**:1263–1277.
21 Bonifer C, Hecht A, Saueressig H, Winter DM, Sippel AE (1991) Dynamic chromatin: The regulatory domain organization of eukaryotic gene loci. *J Cell Biochem* **47**:99–108.
22 Csink AK, Henikoff S (1996) Genetic modification of heterochromatic associations and nuclear organization in *Drosophila. Nature* **381**:529–531.
23 Dernburg AF, Borman KW, Fung JC, Marshall WF, Philips J, Agard D, Sedat JW (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* **85**:745–759.
24 Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* **51**:975–985.
25 Stief A, Winter DM, Stratling WH, Sippel AE (1989) A nuclear DNA attachment element mediates elevated and position-independent gene activity. *Nature* **341**:343–345.
26 Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**:335–340.
27 Steinemann M, Steinemann S (1997) The enigma of Y chromosome degeneration: TRAM, a novel retrotransposon is preferentially located on the Neo-Y chromosome of *Drosophila miranda. Genetics* **145**:261–266.
28 Wines DR, Henikoff S (1992) Somatic instability of a *Drosophila* chromosome. *Genetics* **131**:683–691.