References

1 Garrod, A.E. (1902) *Lancet* 2, 1616–1620
2 Garrod, A.E. (1908) *Lancet* 2, 73–79
3 Garrod, A.E. (1931) *The Inborn Factors in Disease: An Essay*, Clarendon Press
4 Le Du, B.N., Zannoni, V.G., Laster, L. and Seegmiller, J.E. (1958) *J. Biol. Chem.* 230, 251–260
5 Beadle, G.W. and Tatum, E.L. (1941) *Proc. Nat. Acad. Sci. U. S. A.* 28, 499–506
6 Beadle, G.W. (1946) *Am. Sci.* 34, 31–53
7 Pauling, L., Itano, H.A., Singer, S.J. and Well, I.C. (1949) *Science* 25, 543–548
8 Scott-Moncrieff, R. (1936) *J. Genet.* 32, 117–170
9 Ephrussi, B. (1942) *Quart. Rev. Biol.* 17, 327–338
10 Haldane, J.B.S. (1937) *Perspectives in Biochemistry*, Cambridge University Press
11 Ephrussi, B., Hottinguer, H. and Chimenes, A.M. (1949) *Annal. Inst. Pasteur* 76, 351–367
12 Slonimski, P. and Ephrussi, B. (1949) *Annal. Inst. Pasteur* 77, 47–63
13 Pontecorvo, G. *et al.* (1953) *Adv. Genetics* 5, 141–238
14 Pontecorvo, G. (1962) *Brit. Med. Bull.* 18, 81–84
15 Fernández-Cañón, J.M. and Peñalva, M.A. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 9132–9136
16 Englesberg, D.R. *et al.* (1962) *J. Bacteriol.* 84, 137–146
17 Darlington, A.J. and Scazzocchio, C. (1967) *J. Bacteriol.* 93, 937–940
18 Diallinas, G. *et al.* (1995) *J. Biol. Chem.* 270, 8610–8622
19 Garrod, A.B. (1859) *The Nature and Treatment of Gout and Rheumatic Gout*, Walton and Maberly
20 Fernández-Cañón, J.M. and Peñalva, M.A. (1995) *J. Biol. Chem.* 270, 21199–21205
21 Fernández-Cañón, J.M. *et al.* (1996) *Nat. Genet.* 14, 19–24

Variegated gene expression in mice

KENNETH DOBIE*‡§, MAJID MEHTALI¶, MARGARET McCLENAGHAN‡ AND RICHARD LATHE*
kdobie@aim.salk.edu
mehtali@transgene.fr
maggie.mcclenaghan@bbsrc.ac.uk
rlathe@ed.ac.uk

*CENTRE FOR GENOME RESEARCH, UNIVERSITY OF EDINBURGH, WEST MAINS ROAD, EDINBURGH, UK EH9 3JQ. ‡ROSLIN INSTITUTE (EDINBURGH), ROSLIN, MIDLOTHIAN, UK EH25 9PS. ¶TRANSGENE SA, RUE DE MOLSHEIM, 67000 STRASBOURG, FRANCE. §SALK INSTITUTE, MBVL, 10010 N. TORREY PINES ROAD, LA JOLLA, CA 92027, USA.

Transgenic animals, particularly mice, are most-commonly generated by the microinjection of DNA into fertilized eggs. End-to-end ligation of the construct and/or homologous recombination between circularized molecules is thought to generate a transgene concatemer (often 5–50 copies) that integrates at a single site in the genome. While transgene expression can be stable and heritable, the tandem-repeat nature of the insertion appears to contribute to a phenomenon akin to variegation, as described in insects and plants. This can generate, by stochastic silencing of the transgene, a variable and mosaic pattern of transgene expression. Variegation can complicate the interpretation of some experiments using transgenic animals generated by microinjection. Below, we briefly discuss the evidence for variegation in mice, mechanisms of silencing and routes by which it can be avoided, and the possibility that some endogenous genes might be subject to variegation.

Several different mechanisms can lead to variations in the level of (trans)gene expression within a single line. Variation can be due to strain-specific modifier genes, or to integration or translocation into Y chromosome heterochromatin or to the X chromosome[1]. Loss of function has also been observed when cells expressing the construct have a selective disadvantage, so favouring loss of transgene DNA. For the most part, transgene loss appears uncommon although inverted repeats within a transgene array can favour deletion or rearrangement[2]. Mosaic expression patterns have also been observed with tyrosinase fusion transgenes, although mosaic or striated expression might be an intrinsic property of this locus[3]. As noted above, growing evidence now suggests that some transgenes can be subject to a phenomenon akin to the variegation found in insects and plants.

Position effect variegation in *Drosophila*

Mosaic or 'variegated' expression of endogenous genes[4] and transgenes[5–7] in *Drosophila* is subsumed under the term 'position-effect variegation' (PEV). This can be defined as (a) position-dependent inactivation of gene expression in a fraction of cells that generate a particular tissue, (b) heritable maintenance of this pattern, either 'on' or 'off', through subsequent cell divisions, and (c) incomplete mixing after each cell division such that cells that remain closest in space are those most-closely related by lineage.

The first and best-documented example of PEV concerns a chromosome rearrangement in which an allele of the *white* (w^+) gene was transferred to a site close to the centromere. Expression in the eye is uniform at the original site, but after translocation to a pericentromeric region, the expression of the gene becomes variegated, resulting in patches of pigmented and unpigmented cells in the eye. Further studies have led to a paradigm, still unproved, that pericentromeric localization renders the gene susceptible to heterochromatic condensation spreading out from the centromere[8–10]. PEV has also been documented in yeast for gene sequences integrated close to telomeres[11] or within centromeres[12]. The silencing

PII: S0168-9525(97)01097-4

process in yeast can affect genes up to 20 kb away from the heterochromatin–euchromatin breakpoint[13] while this extends to 2 Mb in *Drosophila*[14].

Variegation of transgene expression in mice

As we have seen, a variety of mechanisms can contribute to variable transgene expression in mice, but do any correspond to variegation as described in *Drosophila*? Despite many suggestive reports, there has been little direct evidence. Two studies in mice now provide firm evidence for sectoring of transgene expression, consistent with stochastic silencing and maintenance of the expression state (on or off) through cell division. *In situ* staining of sections of thymus from mice harbouring a human CD2 transgene revealed some clustering of cells expressing the transgene[15]. Compelling evidence for sectored localization of transgene expression was also obtained in mice harbouring a transgene encoding the ovine milk protein, β-lactoglobulin (BLG). *In situ* hybridization to mammary gland sections revealed clusters of cells strongly expressing the transgene surrounded by cells that fail to express detectable transgene mRNA (Ref. 16; Fig. 1). In two of three lines examined, the level of BLG expression in the milk varied by up to tenfold between individuals of the same line: the level of transgene product in milk appeared to parallel the number and size of cell clusters positive for transgene mRNA (Ref. 16).

While in both these cases the clusters of expressing cells suggest clonal expansion from a committed progenitor, other possibilities have not yet been formally excluded. However, these observations indicate that individual cells are able to suppress the expression of transgene constructs and, subsequently, to propagate the inert state of the transgene through cell division, leading to mosaic or variegated patterns of expression.

FIGURE 1. Variegated gene expression in mammary gland of transgenic mice. *In situ* hybridization was performed to sections of mammary gland from line 7 mice harbouring a β-lactoglobulin (BLG) transgene with a pericentromeric insertion site[16].

How might variegation arise?

The chromosomal localization of the variegating mouse transgenes provides a first clue to the underlying mechanism. *In situ* hybridization studies on chromosome spreads from transgenic animals exhibiting variegated expression of human CD2 revealed that the transgenes were integrated close to a centromere[15]. Similarly, in lines displaying variegated BLG expression the transgene array was also present close to a centromere[16]. These results parallel the situation in *Drosophila* where proximity to heterochromatic regions predisposes to variegation.

Second, there are some indications that transgene copy number can contribute to variegation, also as reported in *Drosophila*. Transgene heterochromatinization in *Drosophila* is not only dependent on proximity to the centromere, but also on the number and orientation of repeats present at the transgene locus; silencing increases in step with copy number[5,7]. Similar copy-number-dependent silencing has been reported in plants[17,18]. The variegating mouse lines discussed above were not designed to address a possible association between copy number and variegation (although the two variegating BLG lines contained more than 17 copies of the transgene while the nonvariegating line harboured less than five copies[16]). However, there is some indirect evidence for linkage between variegation and copy number in mouse. In vertebrates (but not *Drosophila*) DNA methylation is associated with gene inactivity, although whether methylation is a cause or consequence of gene silencing in mammals is a matter of some debate. Nonetheless, we reported previously that the extent of transgene methylation increases in step with copy number[19], consistent with progressive silencing as copy number increases. In addition, higher transgene copy numbers show more-rapid silencing of expression[20]. These observations suggest, but do not prove, that higher copy numbers contribute to transgene silencing in mouse just as they do in *Drosophila*.

What might be the mechanism that triggers heterochromatic condensation? In *Drosophila*, it has been suggested that somatic pairing can lead to DNA looping and, hence, sequestration into a heterochromatic compartment[5,7]. A further possibility is suggested to us by the fact that multiple interacting transcription factors bind at many promoter elements. Tandem repetition of the promoter could permit the formation of incorrect (and unproductive) multiprotein complexes between transcription factors bound at adjacent promoter sites, resulting in DNA looping out between them and loss of expression.

These models are consistent with a silencing process comprising three steps (Fig. 2). First, the formation of abnormal (possibly bent?) DNA structures due to pairing or looping, a process favoured by sequence repetition. Second, the specific recognition of these abnormal structures (for instance by the binding of heterochromatin components possibly associated, in mice, with DNA methylation). Third, sequestration of the complexes into heterochromatic regions, and facilitated by proximity to preexisting heterochromatin structures.

It is of note that the contention that silencing extends progressively outwards from early foci of heterochromatinization is consistent with a report that the extinction of expression of a globin transgene increases with age of the animal[20]. However, it is not clear that proximity to preexisting heterochromatic regions is an absolute prerequisite for stochastic silencing to occur.

Is variegation widely encountered?

Tandem repetition and proximity to the centromere (or another heterochromatic region) both appear to

contribute to transgene silencing. Superficially, this supports the argument that many transgenic animals should express transgenes in a stable fashion, because copy numbers are often low and integration close to a centromere is unlikely. This argument might be incorrect. Although no systematic studies have been done, there is evidence that transgene integration into pericentromeric regions may occur frequently. Of 17 lines examined (14 by Festenstein *et al.*[15], three by Dobie *et al.*[16]), more than half (nine) had a pericentromeric localization.

This might indicate that variegation is rather more common than previously recognized, although sectoring of gene expression patterns is, of course, difficult to detect in tissues (such as brain and the immune system) in which individual daughter cells migrate away following cell division. However, there are abundant instances of unexplained variable or mosaic transgene expression[21–24]. To give one example, five out of five transgenic lines harbouring a housekeeping transgene[19] showed extreme variation in the tissue levels of expression in individual animals. Unpredictable swings in the level of transgene expression were observed, ranging over two orders of magnitude [M. Mehtali (1986) PhD Thesis, Strasbourg], particularly at high copy number, and most were consistent with variegation.

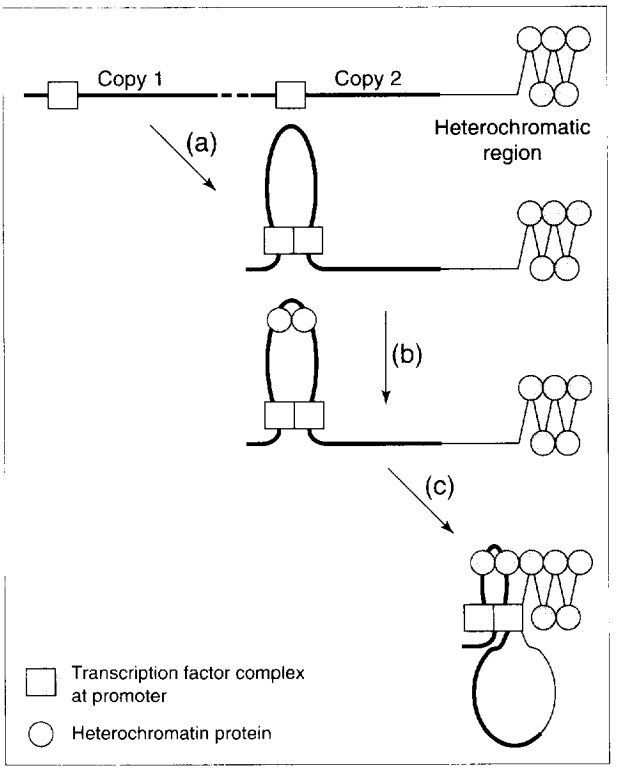

FIGURE 2. A model for heterochromatic condensation of transgene arrays precipitated by (1) repeat DNA sequences and (2) proximity to existing heterochromatin. (a) A multimeric transcription-factor complex is formed between repeat transgene copies. (b) Heterochromatin proteins bind to the abnormal DNA structures. (c) The transgenes are sequestrated into the heterochromatin compartment.

Can variegation in transgenic mice be avoided?

The possibility that many transgenes are subject to stochastic silencing is of concern, especially because variegation can be difficult to detect without detailed analysis. The outcome of transgenic experiments would be considerably obscured if some animals of a given line express the transgene at high levels in a particular organ while others do not.

Although screening transgenic mice for the chromosomal insertion site and copy number should permit the elimination of variegating lines, there are other routes by which variegation can be avoided. First, single-copy gene insertions do not appear to variegate. This argues for the use of targeted integration of single transgene constructs by homologous recombination in embryonal stem cells[25], or for integration by catalysed recombination in microinjected eggs (J.O. Bishop, pers. commun.). Second, the use of very long constructs (employing yeast or bacterial artificial chromosomes) will increase the distance between repeat elements and can decrease the likelihood of precipitating heterochromatic condensation (A. Colman, pers. commun.). Third, the incorporation of elements permitting efficient site-specific recombination (such as the *loxP* site from bacteriophage P1) could be used (by crossing to an animal expressing the P1 recombinase, Cre, in the germline) to generate progeny in which the number of transgene copies is substantially reduced[26].

The effect of multiple integration and/or a pericentromeric localization can also be attenuated by the inclusion of specific DNA elements in the construct. It was suggested[27] that some transcription enhancers can act to suppress variegation, although it is clear that most enhancers alone do not have this effect. Matrix-attachment regions (MARs; also scaffold attachment regions, SARs) have also been reported to prevent variable expression of some transgene constructs[28,29] although this does not appear to be a systematic feature of such elements[30,31].

Festenstein *et al.*[15] report that the inclusion of the human CD2 locus control region (LCR) in a transgene construct could prevent variegation while, at the same time, rendering the level of expression proportional to copy number. Similar findings were reported for a globin LCR (Ref. 32). While a fuller discussion of the relationship between copy-number-dependence and variegation would be out of place here, this is a potentially important finding with implications across the board from transgenesis in animals to gene therapy in humans. However, it is not known whether the suppression of variegation is a general feature of LCRs or whether particular elements can only suppress variegation in specific cell types. Nevertheless, the incorporation of LCRs into transgene constructs could plausibly provide a general route to ensure reproducible expression levels.

Do any endogenous genes variegate?

If the model of heterochromatic condensation spreading from the centromere (or other heterochromatic regions) to encompass repetitive transgene arrays is even partly correct, this would inevitably raise the question of the fate of genes located close to the transgene or between the transgene insertion and the centromere. Are they stochastically silenced too, and are there genes whose expression naturally variegates?

Chromosome rearrangements located 4–400 kb away from the relevant locus have been associated with some human (e.g. *SOX9*/campomelic

dysplasia; *PAX6*/aniridia) and mouse (*steel*/female sterility) genetic disorders, as reviewed recently[33,34]. Detailed molecular analysis of two aniridia (absence of iris) human pedigrees demonstrated that the disorder is associated with chromosome rearrangements greater than 85 kb downstream of the *PAX6* gene[35]. Although a link between the aberrant phenotypes and the genomic rearrangements has yet to be established, it is tempting to speculate that the rearrangements could in some cases cause relocation of the affected loci into a domain subject to heterochromatic condensation, leading to variegated expression.

'Variegation' could also provide a mechanism for gene silencing during development[36–39] or even for the storing of acquired information: rearrangement of neuronal centromeric satellite DNA was reported during synaptic plasticity[40] while, in *Drosophila*, a protein phosphatase mutation that suppresses variegation also impairs learning and memory[41].

Finally, in plants[17,42] and in insects[43] inactivation of one copy of a gene can, by a phenomenon termed co-suppression or *trans*-inactivation, perturb the expression of another copy situated elsewhere in the genome. Although co-suppression has not so far been documented in mammals, it would come as no surprise if the same were to occur in transgenic mice. More generally, the many examples of variegation, co-suppression and related phenomena discussed here emphasize the conclusion that mechanisms exist by which different copies of a DNA sequence can intercommunicate in some real sense (possibly by direct physical association brought about by pairing and/or looping) such that, jointly, their activity state can differ substantially from that of a single copy acting alone.

Acknowledgements

We thank J.O. Bishop, I. Jackson, D. Kioussis, J. Mullins, M. Steel, J. West, and an anonymous referee for their helpful comments and suggestions.

References

1 Cattanach, B.M. (1974) *Genet. Res. Camb.* 23, 291–306
2 Collick, A. *et al.* (1996) *EMBO J.* 15, 1163–1171
3 Mintz, B. and Bradl, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 9643–9647
4 Karpen, G.H. (1994) *Curr. Opin. Genet. Dev.* 4, 281–291
5 Dorer, D.R. and Henikoff, S. (1994) *Cell* 77, 993–1002
6 Wallrath, L.L. and Elgin, S.C.R. (1995) *Genes Dev.* 9, 1263–1277
7 Sabl, J.F. and Henikoff, S. (1996) *Genetics* 142, 447–458
8 Zuckerkandl, E. (1974) *Biochimie* 56, 937–954
9 Locke, J., Kotarski, M.A. and Tartof, K.D. (1988) *Genetics* 120, 181–198
10 Grigliatti, T. (1991) *Meth. Cell Biol.* 35, 587–627
11 Sandell, L.L. and Zakian, V.A. (1992) *Trends Cell Biol.* 2, 10–14
12 Allshire, R.C., Javerzat, J.P., Redhead, N.J. and Cranston, G. (1994) *Cell* 76, 157–169
13 Gottschling, D.E., Aparico, O.M., Billington, B.L. and Zakian, V.A. (1990) *Cell* 63, 751–762
14 Weiler, K.S. and Wakimoto, B.T. (1995) *Annu. Rev. Genet.* 29, 577–605
15 Festenstein, R. *et al.* (1996) *Science* 271, 1123–1125
16 Dobie, K.W. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 6659–6664
17 Flavell, R.B. (1994) *Proc. Natl. Acad. Sci. U. S. A.* 91, 3490–3496
18 Ye, F. and Signer, E.R. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 10881–10886
19 Mehtali, M., LeMeur, M. and Lathe, R. (1990) *Gene* 91, 179–184
20 Robertson, G. *et al.* (1996) *Nucleic Acids Res.* 24, 1465–1471
21 Jetton, T.L. *et al.* (1994) *J. Biol. Chem.* 269, 3641–3654
22 Morley, S.D. *et al.* (1996) *Mol. Endocrinol.* 10, 585–590
23 Hennighausen, L. *et al.* (1995) *J. Cell. Biochem.* 59, 463–472
24 Robertson, G. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 5371–5375
25 Bronson, S.K. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 9067–9072
26 Jasin, M., Moynahan, M.E. and Richardson, C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 8804–8808
27 Walters, M.C. *et al.* (1996) *Genes Dev.* 10, 185–195
28 McKnight, R.A. *et al.* (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 6943–6947
29 Bonifer, C. *et al.* (1994) *Nucleic Acids Res.* 22, 4202–4210
30 Thompson, E.M., Christians, E., Stinnakre, M.G. and Renard, J.P. (1994) *Mol. Cell. Biol.* 14, 4694–4703
31 De Sepulveda, P. *et al.* (1995) *Biochem. Biophys. Res. Commun.* 211, 735–741
32 Milot, E. *et al.* (1996) *Cell* 87, 105–114
33 Milot, E., Fraser, P. and Grosveld, F. (1996) *Trends Genet.* 12, 123–126
34 Bedell, M.A., Jenkins, N.A. and Copeland, N.G. (1996) *Nat. Genet.* 12, 229–232
35 Fantes, J. *et al.* (1995) *Hum. Mol. Genet.* 4, 415–422
36 Mintz, B. (1974) *Annu. Rev. Genet.* 8, 411–470
37 Eissenberg, J.C., Elgin, S.C.R. and Paro, R. (1995) in *Chromatin Structure and Gene Expression* (Elgin, S.C.R., ed.), pp. 147–171, Oxford University Press
38 Orlando, V. and Paro, R. (1995) *Curr. Opin. Genet. Dev.* 5, 174–179
39 Elgin, S.C.R. (1996) *Curr. Opin. Genet. Dev.* 6, 193–202
40 Billia, F., Baskys, A., Carlen, P.L. and De Boni, C. (1992) *Mol. Brain Res.* 14, 101–108
41 Asztalos, Z. *et al.* (1993) *J. Neurosci.* 13, 924–930
42 Matzke, M.A. and Matzke, A.J.M. (1995) *Trends Genet.* 11, 1–3
43 Dernburg, A.F. *et al.* (1996) *Cell* 85, 745–759

Letters to the Editor

We welcome letters on any topic of interest to geneticists and developmental biologists. Write to:

Mark Patterson
Trends in Genetics,
Elsevier Trends Journals,
68 Hills Road,
Cambridge, UK CB2 1LA.

TIG@elsevier.co.uk
Fax 01223 464430