Choose your target

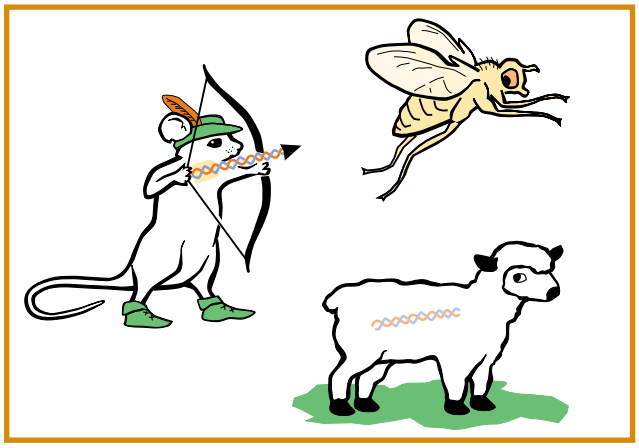
Mario R. Capecchi

The technology of modifying endogenous genes has recently been extended from mice to *Drosophila* and sheep. Concurrently, genomic sequencing is uncovering thousands of previously uncharacterized genes. Armed with today's technologies, what are our best options for delineating the functions of these new genes?

Considering the enormous power of gene targeting for dissecting and analysing biological phenomena[1], it is remarkable that, for more than a decade, application of this technology has been restricted to either relatively simple organisms (the bacteria and fungi) or the mouse. A pair of recent studies open up new vistas in their demonstration of gene targeting in *Drosophila melanogaster*[2] and sheep[3]. The strategies of these different studies are conspicuously different. In *Drosophila*, the gene-targeting reaction—homologous recombination between an exogenous modified DNA sequence and the cognate, endogenous genomic sequence—occurs in the intact animal (Fig. 1). In contrast, transgenic sheep were generated though a targeting reaction in cultured fetal fibroblasts (Fig. 2).

The method developed by Rong and Golic[2] for generating targeted mutations in *Drosophila* is composed of three parts: (i) a transgene that expresses the site-specific yeast recombinase *Flp*; (ii) a transgene that expresses the site-specific endonuclease I-*Sce*I; and (iii) a transgenic donor construct that contains the recognition sites (FRT and I-*Sce*I) for both enzymes, in addition to the modified, exogenous DNA sequences to be used to mutate the endogenous cognate chromosomal sequence (Fig. 1). The Flp recombination target (FRT) sites flank the exogenous DNA sequences, and the I-*Sce*I site is within the region of homology between the exogenous and endogenous DNA sequences. The transgenes encoding the Flp-recombinase and the I-*Sce*I-endonuclease are under the control of a heat-inducible promoter.

Flies containing each transgene were constructed separately and then crossed to bring all three transgenes together in the progeny. Heat shock of the triple-transgenic flies activates Flp recombinase which, in turn, triggers the release of the donor fragment from the chromosome as a circle. The circular fragment is then cleaved by the I-*Sce*I endonuclease. The linearized DNA fragment serves as a substrate for the endogenous homologous recombination machinery of the cell, finds the cognate sequence on the chromosome, and engages in a homologous recombination reaction that transfers the desired modification to the endogenous gene.



Size matters

The homologous-recombination machinery in *Drosophila* seems to be much more mouse-like than yeast-like[4,5]. For example, optimal targeting frequencies in *Drosophila* and mice require extensive DNA sequence homology between the donor DNA fragment and the recipient chromosomal sequence, rather than the short stretches of homology (of approximately 100 bp) that suffice in yeast. The original targeting experiments reported by Rong and Golic[2] used 8 kb of homology between the exogenous and endogenous DNA sequences. Reducing the homology to 2.5 kb reduced the targeting frequency by more than tenfold[6]. Homologous stretches of 5 kb yielded an intermediate targeting frequency. Such sensitivity of the targeting frequency to the extent of homology between the targeting construct and the endogenous chromosomal sequence is very similar to that observed in mammalian cells, including mouse embryonic stem (ES) cells[7,8].

In both yeast and mammalian cells, linear DNA molecules are more effective substrates for gene-targeting reactions than are circular molecules[5,9,10]. For this reason, Rong and Golic went to great lengths to generate linear DNA donor molecules for their *in vivo* targeting reaction. They achieved this goal by using the site-specific endonuclease I-*Sce*I. A double-strand break within the region of homology was introduced to favour so-called 'ends in' homologous recombination reactions, which we have called 'insertion' reactions[7] (Fig. 1). Despite the use of this type of DNA donor substrate, Rong and Golic observed (among their products of recombination) equal numbers of molecules that may have been generated by 'ends in' and 'ends out', or what we would call 'insertion' and 'replacement', reactions[7]. This suggests that the location of the cleavage site within the donor construct may not be critical and that, as in mammalian cells, 'insertion' and 'replacement' reactions are equally effective[7,11].

The ability to generate mice with targeted mutations has been dependent on the use of the ES cell as a vehicle to transmit the mutation to the mouse germ line. This is because the frequency

Howard Hughes Medical Institute, Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, USA (e-mail: mario.capecchi@genetics.utah.edu).

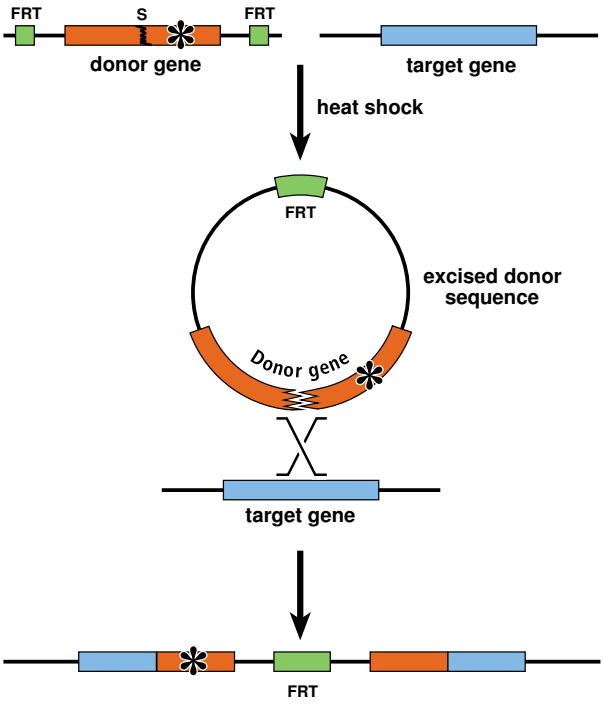


Fig. 1 Gene targeting in *Drosophila*. The gene-targeting protocol described by Rong and Golic[2] occurs within the intact fly. 'FRT' and 'S' denote the recognition sequences for the yeast Flp-recombinase and the I-*Sce*I-site-specific endonuclease, respectively. Following heat shock, the donor fragment is released from the chromosome as a DNA circle, which is then cleaved by the I-*Sce*I endonuclease. The linearized fragment then serves as a substrate for homologous recombination, with the endogenous cognate sequence on the chromosome.

of homologous recombination between the targeting vector and the target locus is low, relative to the frequency of random insertions of the targeting vector into the host cell genome. It is therefore impractical to generate mice with targeted mutations by injecting the targeting vector directly into fertilized eggs. The main obstacle to generating transgenic individuals of other mammalian species is due to a failure to obtain cells capable of contributing to the formation of the germ line. But with the birth of Dolly[12], a cloned sheep generated by nuclear transfer of a somatic cell nucleus into an enucleated egg, an alternative route to gene-targeted mammals was also born. The targeting reaction can be carried out in any karyotypically stable, somatic cell that can be cultured *in vitro*. Particularly attractive for this purpose are fetal fibroblasts, as they are easy to obtain and can be maintained in culture for many passages without compromising their ability to generate cloned animals by nuclear transfer[13]. Once a successful targeting event has been identified in these cultured cells, its nucleus is transferred to an enucleated egg of the same species, creating an embryo. The embryo is then transferred to the uterus of a foster mother and allowed to come to term, generating a live animal with a targeted mutation (Fig. 1).

The study by McCreath *et al.*[3] describes the first successful application of this procedure to generate sheep with targeted mutations. To ensure a high targeting frequency, McCreath *et al.*[3] chose to introduce their targeted modification into a gene that was known to be expressed at high levels in cultured fibroblasts, the ovine I procollagen (*COLIAI*) gene. To enrich for cells containing the targeted mutation (over cells containing random insertions of the targeting vector), they used a selectable gene, *neo*^{r}, that lacked its own promoter. Activation of the selectable gene therefore relied on integration of the vector adjacent to a functional genomic promoter. Most random insertions into the genome would not be predicted to provide a suitable promoter to the promoterless *neo*^{r} gene. McCreath *et al.* also demonstrated that the *COLIAI* locus was a suitable environment to allow a transgene, driven by a promoter specific to mammary epithelium, to be specifically expressed in that tissue, even though *COLIAI* itself is not actively expressed in mammary epithelium. These experiments are of particular interest to drug companies that are exploring the potential of using livestock to produce pharmaceuticals in their milk.

Only further experimentation will determine whether genes other than *COLIAI* are suitable targets for genomic modification using modified fibroblasts as nuclear donors. In mouse ES cells, any gene can be readily modified if appropriate enrichment techniques are used[14]. Our experience further suggests that if the experiments are carried out appropriately, including the use of isogenic DNA and optimized selection procedures, most mammalian cells in culture have roughly equivalent capacities to mediate homologous recombination between exogenous and endogenous DNA sequences. So most genes should also serve as suitable substrates for targeted modification using the 'fibroblast' system described by MacCreath *et al.*

With the mouse/ES cell system, necessary intermediates in the process of generating mice with targeted mutations are chimaeric mice. These mice are composed of a mixture of cells derived from the targeted ES cells and the recipient blastocyst. As chimaeric mice are not composed entirely of cells containing the targeted mutation, they may not have the full set of properties associated

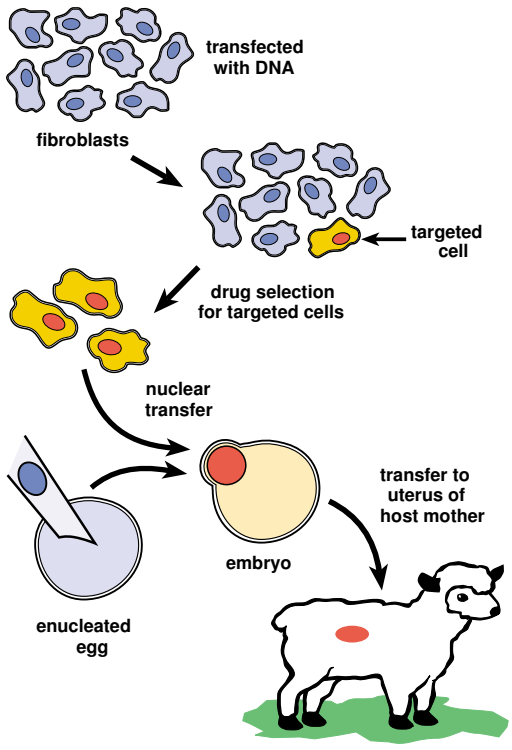


Fig. 2 Gene targeting in sheep. McCreath *et al.*[3] have generated sheep with targeted modifications in the ovine gene encoding procollagen-1. The targeting vector was introduced into cultured primary fetal fibroblasts. Following the identification of cells with the desired targeted modification, its nucleus was transferred to an enucleated egg, creating an embryo. The embryo was then transferred to the uterus of a foster mother and allowed to come to term. This procedure generated live lambs with a targeted mutation.

with mice in which every cell contains the targeted mutation. Such mice are generated by breeding the chimaeric mouse to a suitable mate. An advantage of the fibroblast-nuclear transfer method for generating animals with targeted mutations is that it bypasses the need to generate chimaeric intermediates. This represents a major savings in time and money when it comes to generating livestock with a comparatively lengthy generation time (and hence substantive fees for board and lodging).

Are there sufficient advantages to the fibroblast/nuclear transfer system to use it as the preferred method for generating mice with targeted mutations? It seems not, as the efficiency of generating viable offspring in mice by the fibroblast/nuclear transfer method does not match that of the chimaera method. The cause of extensive pre- and postnatal mortality of cloned animals has yet to be identified, although it has been proposed that a disruption of complex changes in the patterns of genomic demethylation and methylation may be to blame.

Functional genomics

Most of the *Drosophila* genome is now sequenced[15]. Although the analysis is far from complete, current estimates centre around a complement of 14,000 to 15,000 genes. The functions of most of these have yet to be determined. Until very recently, the most reasonable method for generating a comprehensive collection of mutants in *Drosophila* was by insertional mutagenesis using P-elements[16]. But despite extensive efforts, mutant collections cover only one-third of the *Drosophila* genes. The limitation of this approach is largely due to limited sequence specificity of the P-element insertion sites—and not all P-element–mediated insertions generate null mutations.

As will have escaped the attention of few, over the past several months estimates of the number of mammalian genes have varied over a wide range—from 40,000 to over 100,000 genes. By the time the mouse genome is complete (in about 3 to 4 years' time), there may be as many as 20,000 mouse lines with targeted disruption of different genes. This early milestone will have been accomplished in the absence of any coordinated program. There is an ongoing debate on whether insertional mutagenesis[17] or gene targeting is the better means of generating mouse mutants. I strongly favour the latter. The greatest investment of time and money in new mouse mutants is not in their generation, but their subsequent analysis. The latter exceeds the former by factors of 20 or more. The precision afforded by gene targeting, which takes advantage of established DNA sequence, allows construction of alleles that favour analysis. With a single construct, multiple alleles can be generated at a specified locus: null alleles, conditional alleles and reporter genes that facilitate visualization of gene activity. As with yeast, the advantages of targeted alleles in facilitating the analysis of the mutant phenotype (compared with mutant alleles generated by random insertion) far outweigh their initially higher up-front costs.

On application

The extension of gene targeting to *Drosophila* and sheep will have different impacts on life as we know it. Gene targeting in *Drosophila*, as in mice, will be used primarily as a basic research tool. It will be used to generate mutant alleles that cannot readily be obtained by genetic screens or by P-element–based mutagenic protocols. The impact of gene targeting in sheep, on the other hand, will be primarily commercial. The technology will be used to improve other livestock, the first examples of which will probably be cattle resistant to specific pathogens such as scrapie. The pressing need for organs suitable for human transplantation has led to the suggestion that organs from domestic animals, such as the pig, might be used instead of human organs. Such approaches are not presently feasible because of the rapidity with which pig tissue would be rejected by the human immune system. Gene targeting offers the possibility of modifying pig tissue so that it is better tolerated by our immune system.

Drosophila is one of the most intensively studied model organisms and the addition of gene targeting to its repertoire of tricks will greatly enhance its usefulness. Similarly, the use of gene targeting in sheep and other livestock will find many uses. Plants are conspicuously absent from the 'gene-targeting' list. This is particularly surprising because an efficient means of producing transgenic plants has been available for a number of years. And intact plants can be generated from cells maintained in culture. The latter property provides the opportunity to fashion gene-targeting strategies based on a cell-culture approach, or by using an *in vivo* strategy such as that used by Rong and Golic[2]. Regardless of biological context, gene targeting permits a systematic dissection of the most complex of biological processes, and so the extension of gene targeting to *Drosophila* and sheep is a welcome advance.

1. Capecchi, M.R. Targeted gene replacement. *Sci. Am.* **270**, 34–41 (1994).
2. Rong, Y.S. & Golic, K.G. Gene targeting by homologous recombination in Drosophila. *Science* **288**, 2013–2018 (2000).
3. McCreath, K.J. *et al.* Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* **405**, 1066–1069 (2000).
4. Capecchi, M.R. Tapping the cellular telephone. *Nature* **344**, 105 (1990).
5. Rothstein, R. Targeting, disruption, replacement, and allele rescue: integrative DNA transformations in yeast. *Methods Enzymol.* **194**, 281–296 (1991).
6. Rong, Y.S. & Golic, K.G. Targeted knockout of a gene at a non-telomeric location in Drosophila. *Science (*in press 2000).
7. Thomas, K.R. & Capecchi, M.R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
8. Deng, C. & Capecchi, M.R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12**, 3365–3371 (1992).
9. Thomas, K.R., Folger, K.R. & Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419–428 (1986).
10. Jason, M. & Berg, P. Homologous integration in mammalian cells without target gene selection. *Genes Dev.* **2**, 1353–1363 (1988).
11. Deng, C., Thomas, K.R. & Capecchi, M.R. Location of crossovers during gene targeting with insertion and replacement vectors. *Mol. Cell. Biol.* **13**, 2134–2140 (1993).
12. Wilmut, I. *et al.* Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
13. Kubota, C. *et al.* Six cloned calves produced from adult fibroblast cells after long-term culture. *Proc. Natl Acad. Sci. USA* **97**, 990–995 (2000).
14. Mansour, S.L., Thomas, K.R. & Capecchi, M.R. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
15. Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
16. Spradling, A.C. *et al.* The Berkeley Drosophila Genome Project Gene Disruption Project: Single *P*-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135–177 (1999).
17. Zambrowicz, B.P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–612 (1998).