A Personal View of Gene Targeting

Mario R. Capecchi

Gene targeting provides the means for creating strains of mice with mutations in virtually any gene.[1] First, the desired mutation is introduced into a cloned copy of the chosen gene by standard recombinant DNA technology. The mutation is then transferred to the genome of a pluripotent mouse embryo-derived stem (ES) cell by means of homologous recombination between the exogenous, mutated DNA sequence and the cognate DNA sequence in the ES cell chromosome. By microinjection of ES cells containing the transferred mutation into blastocysts and by allowing the embryos to come to term in foster mothers, we can generate chimeric mice capable of transmitting the mutation to their offspring (germline chimeras). Finally, interbreeding of heterozygous siblings yields animals homozygous for the desired mutation. Figure 1 outlines the steps, from cultured ES cells to chimeric mouse, used to generate mice with targeted mutations.

The power of gene targeting is that the investigator chooses which gene to modify and has virtually complete control over the way in which that gene's DNA sequence is modified. This technology permits the evaluation of the functions of genes in an intact mammal and the systematic dissection of the most complex of biologic processes such as development and learning. Because nearly all biologic phenomena are mediated or influenced by genes, this technology will have an impact on the analysis of all such phenomena in mammals, including the study of cancer, immunology, neurobiology, and human genetic disease.

Figure 2 shows a schematic representation of the gene-targeting reaction at the DNA level. The uppermost line (*X*) represents the targeting vector carrying sequences of the chosen gene with a precise modification introduced by recombinant DNA technology. One point that I wish to stress is that the modification depicted by the *X* can be almost any change in the cloned DNA sequence—the deletion, addition, or substitution of a single base pair, or the deletion, addition, or substitution of thousands of base pairs. Thus, the investigator has enormous freedom on how to modify the chosen gene. On entering the cell nucleus, the targeting vector associates with the cell's homologous recombination machinery, searches the entire genome (approximately 3×10^9 base pairs), finds the cognate sequence in the chromosome, aligns with this sequence, and, with the aid of the recombination machinery, exchanges infor-

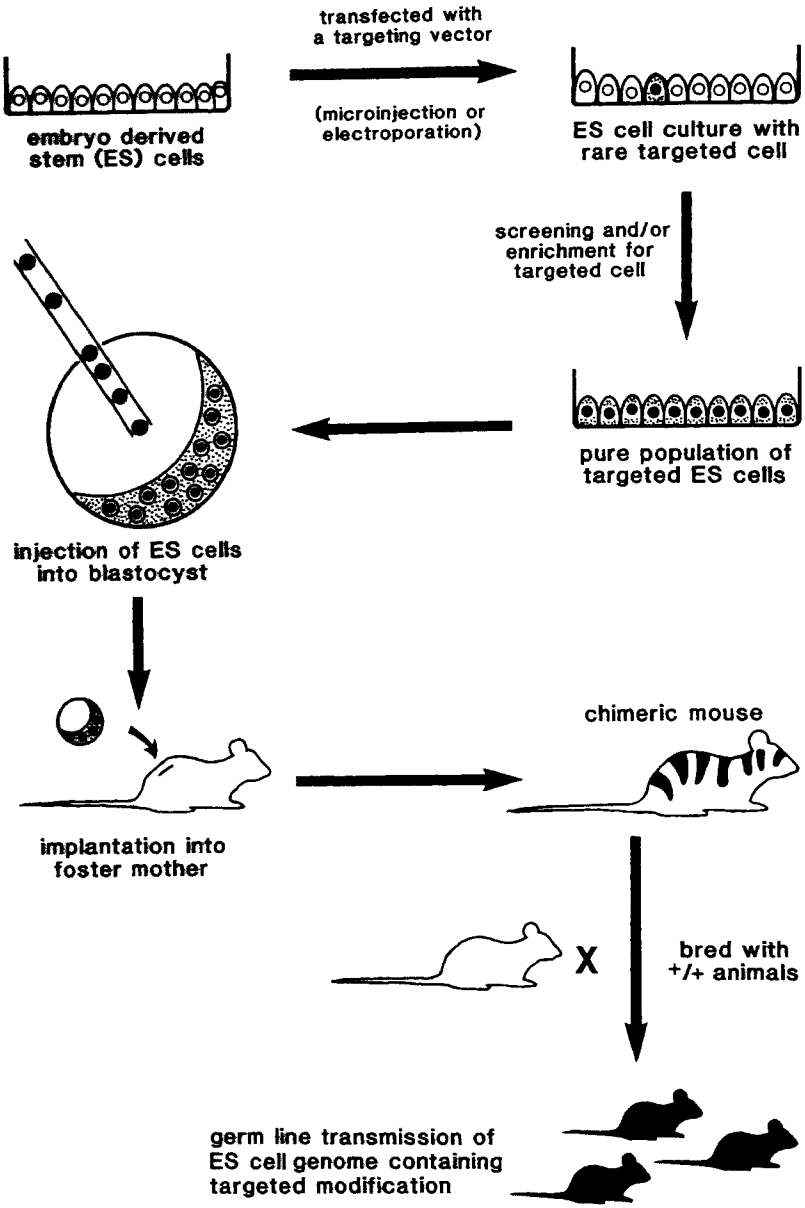

Figure 1. Generation of mouse germline chimeras from embryo-derived stem (ES) cells containing a targeted disruption. The first step involves the isolation of a clonal ES cell line containing the desired targeted disruption. The second step is to use those cells to generate chimeric mice able to transmit the mutant gene to their progeny. To facilitate isolation of the desired progeny, the ES cells and recipient blastocysts are derived from mice with distinguishable coat color alleles. This permits the evaluation of the extent of chimerism by visual assessment of coat color chimerism and the evaluation of ES cell contribution to the formation of the germline by observation of the coat color of chimeric progeny.

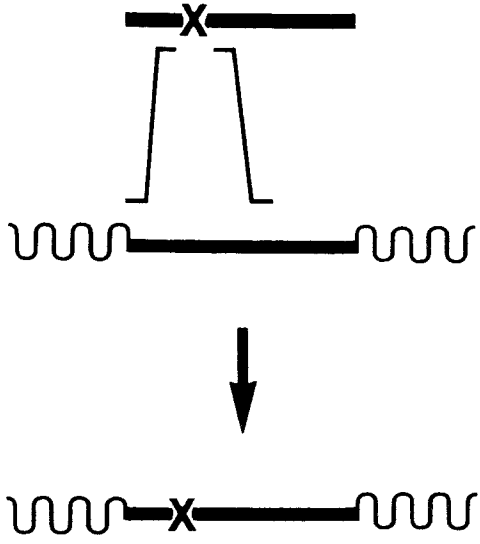

Figure 2. Diagram of the gene-targeting reaction. Homologous recombination between an exogenous mutated copy of the chosen gene and the cognate gene within the host cell chromosome.

mation with the endogenous chromosomal DNA sequence. As a consequence, the specific gene modification created originally in a test tube is transferred to the genome of the living cell.

As a corecipient, with Oliver Smithies, of the 1994 Sloan Prize of the General Motors Cancer Research Foundation, I was asked to document our role in the development of this gene-targeting technology. The chronology that follows should be viewed in that light; it is not a review of the field, but rather a sketch of the factors that influenced our thinking and some of the highlights that maintained the momentum of our research effort.

Our entry into what was going to become the field of gene targeting started in 1977. At this time, I was experimenting with the use of extremely small glass needles to inject DNA directly into nuclei of living mammalian cells. The needles were controlled by hydraulically driven micromanipulators and directed into nuclei with the aid of a microscope. M. Wigler and R. Axel had just reported the successful transfer of a herpes simplex virus–thymidine kinase gene (*HSV-tk*) into tk^- cells by calcium phosphate coprecipitation.[2] Although a very important contribution to somatic cell genetics, their procedure was not very efficient. Approximately 1 in 1 million cells exposed to the calcium phosphate–DNA coprecipitate acquired the exogenous gene in a functional form. However, Wigler and Axel could readily identify such rare cells by growing the treated cells in a medium that killed all cells that did not contain a functional *tk* gene.

Using the same experimental paradigm, I asked whether I could introduce a functional *tk* gene into cells by injecting the DNA directly into their nuclei. This procedure turned out to be extremely efficient. One in three cells received the DNA in a functional form and went on to divide and stably pass that DNA on to its daughter cells.[3] The high efficiency of microinjection meant that it was practical for investigators to use this technology to generate transgenic mice by injecting the DNA into one-cell zygotes and allowing the embryos to come to term in foster mothers. Indeed, a few years later, microinjection was used by Gordon et al.,[4] Costantini and Lacy,[5] Brinster et al.,[6] Wagner et al.,[7] and Wagner et al.[8] to generate the first transgenic mice. Generation of transgenic animals in this way involves introduction of exogenous DNA segments at unpredictable locations in the recipient genome, not targeted genetic alterations at defined sites.

To obtain efficient functional transfer of the *tk* gene into cells, I had to add other short viral DNA sequences to the *HSV-tk* gene. The rationale for doing this experiment in the first place was simple. Mammalian viruses have evolved to propagate themselves efficiently within mammalian cells. It thus seemed possible that their genome contained sequences that enhanced their ability to establish themselves within the competitive mammalian host genome. The first viral genome in which I chose to look for such sequences was SV40, a simian, lytic DNA virus. Indeed, I was able to identify a DNA sequence located near the SV40 origin of DNA replication, which when coupled to the *HSV-tk* gene increased its efficiency in conferring a tk^+ phenotype on the tk^- recipient cells by a factor of more than 100. For a number of reasons, I did not believe that this enhancement was the result of *HSV-tk* plasmid replication within the recipient cells. First, the cells I used for these experiments were derived from a mouse, and SV40 is not able to replicate in murine cells. Second, the enhancement of transforming tk^- cells into tk^+ cells was not dependent on including, in the *HSV-tk* plasmid, SV40 sequences that encoded the large T-antigen, which is required for SV40 replication. Southern transfer analysis of the tk^+ cells showed that the newly added *HSV-tk* DNA sequences were integrated into the host genome. I concluded that this efficiency-enhancing sequence was either increasing the frequency with which the exogenous DNA integrated into the host genome, or increasing the probability that the *tk* gene, once integrated into a host genome, would be functional.[3] In collaboration with Luciw, Bishop, and Varmus, we showed that the avian sarcoma virus (ASV) also contained, within its long terminal repeats, sequences that enhanced HSV*tk*-mediated transformation.[3,9] This was of particular interest because the lifestyles of SV40 and the ASV retrovirus were so different, yet both possessed sequences within their genome that functioned similarly to enhance formation of stable tk^+ transformants. These experiments were carried out before the concept of so-called *enhancers* and in fact contributed to their definition.[10]

The observation that I found most fascinating from these early microinjection experiments was that when multiple copies of the *tk* plasmid were injected into a cell, although they were integrated into random locations within the

host chromosomes, they were always present in head-to-tail concatemers. Such highly ordered concatemers could be generated in two ways: (1) by replication (e.g., a rolling circle mechanism), or (2) by homologous recombination. We were able to prove that the concatemers were generated by homologous recombination.[11] The significance of this observation was its demonstration that mammalian cells contained an efficient machinery for mediating homologous recombination.

This was a startling discovery because it was always assumed that the function of homologous recombination in all organisms was to ensure broad dissemination of the parental genetic traits to their offspring by shuffling these traits in the germ cells. Finding evidence for the activity of this machinery in mouse fibroblast cells implied that all cells, somatic and germ cells, were capable of mediating homologous recombination. We suspected that the homologous recombination machinery in somatic cells was very efficient because I could inject more than 100 *tk* plasmid molecules per cell and they were all incorporated into a single, ordered, head-to-tail concatemer. I realized immediately that if we could harness this machinery to carry out homologous recombination between a newly introduced DNA molecule of our choice and the same DNA sequence in the cell's chromosome, we would have the ability to specifically mutate or modify virtually any gene in the cell.

I first talked publicly about the results from the microinjection experiments at a NATO-sponsored course in Estaril, Portugal in the spring of 1978. Frank Ruddle was also an instructor in that course, and he became a champion and rapid disseminator of our results. A commonly held misconception is that the motivation for our laboratory to pursue gene targeting in mammalian cells was the success of Fink's group in achieving gene targeting in yeast.[12] This is not the case. We were aware a year earlier that mammalian cells possessed the enzymatic machinery for efficient mediation of homologous recombination between newly introduced DNA molecules. It was this awareness that provided us with the incentive to pursue gene targeting in mammalian cells.

We spent the next few years becoming familiar with this machinery by studying recombination between cointroduced DNA substrates. From these experiments, it became evident that this enzymatic machinery could mediate a wide spectrum of reactions, including conservative and nonconservative homologous recombination. Both reciprocal and nonreciprocal recombination events were apparent among the products of recombination, although we observed a distinct bias towards mediating nonreciprocal reactions.[13,14] We found that a cell's ability to carry out homologous recombination depends on a cell's position in the cell cycle, showing a peak of activity in early S phase.[15] Kinetic analysis of homologous recombination between cointroduced DNA molecules indicated that the reaction occurred very rapidly, within 30 minutes after the DNA was introduced into the nucleus.[13] Later, the DNA molecules become refractory to participation in homologous recombination. Exclusion from participation in homologous recombination coincides with the time at which the newly introduced DNA is packaged into chromatin. These studies suggested

that the cellular enzymatic machinery could be exploited to mediate homologous recombination between newly added exogenous DNA sequences and chromosomal sequences.

In 1980, I submitted a National Institutes of Health (NIH) grant proposal outlining experiments intended to test the feasibility of gene targeting in mammalian cells. The proposal elicited an interesting response. In the opinion of the peer reviewers, the probability that the newly introduced DNA would ever find its matching sequence within the host genome was vanishingly small. Aware that the frequency of gene-targeting events could be very low, the experimental paradigm that we proposed allowed for the selection of cells containing the desired recombinant product. One such paradigm (see Figure 3) involved, first, random introduction of a defective copy of a gene, such as *HSV-tk* or *neo*^{r}, into the host chromosome. This was to be followed by introduction into those recipient cells of plasmids containing the gene with a different mutation and then selection for those cells that had generated, by homologous recombination, a functional gene from the two defective parts. This particular paradigm met with skepticism in some circles because it was viewed that the target in the chromosome was artificial (i.e., composed of a randomly introduced DNA sequence). I always viewed this skepticism with amusement because the homologous recombination machinery does not appear to discriminate with respect to DNA sequence content. What appeared to be crucial to this machinery was merely that the sequences between the target and incoming DNA match perfectly.

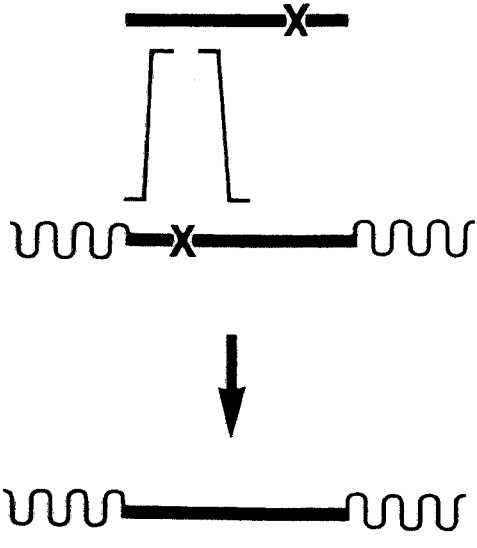

Figure 3. Regenerating a functional gene by homologous recombination from two defective copies of the gene, one located in the host chromosome, and the second newly introduced into the recipient cell.

Despite rejection of the NIH proposal and skepticism of some, we pursued experiments using the proposed design. Because the experimental approach allowed us to select for rare recombination events in cell culture, we could do many more experiments and thereby establish the parameters that influenced the gene-targeting frequency. These experiments revealed some unexpected features. For example, the frequency of gene targeting was not dependent on the number of targeting vectors introduced into the recipient cell, nor on the number of targets present in the host genome.[14,16] This is in contrast to what is observed in yeast, in which the targeting frequency is proportional to the number of target sequences in the genome.[17,18] In yeast, targeting into the ribosomal RNA genes, which are present at 140 copies per genome, is 100 to 200 times more frequent than targeting into the single copy *leu*-2 gene.

Gene targeting in yeast and mammalian cells differs in another important respect. The frequency depends more on the extent of homology between the exogenous and chromosomal sequences in mammalian cells than it does in yeast. In yeast, this dependence is linear, but in mammalian cells it appears to be exponential.[19,20] Whereas the efficiency of the recombination system in yeast appears to saturate with a few hundred base pairs of perfect homology, in mammalian cells the system does not saturate until approximately 15 kb of sequence homology exists between the added DNA and the target locus. These differences in the properties of the homologous recombination machinery between yeast and mammalian cells may reflect differences in their genome organization. For example, the yeast genome has far fewer repetitive DNA sequences than the mammalian genome. In mammalian somatic cells, although the homologous recombination machinery is an important component of the DNA repair machinery, it is equally important that this machinery not inadvertently mediate recombination between repetitive DNA sequences. Such recombination would result in instability of the mammalian genome and loss of critical genetic information. The dependency of homologous recombination machinery in mammalian cells on long stretches of homologous DNA sequences greatly reduces the probability of homologous recombination between repetitive DNA sequences, without impairing its ability to participate in DNA repair.

By 1984, we were confident that it was feasible to do gene targeting in cultured mammalian cells, and we presented our work at a memorable symposium on homologous recombination held at the Cold Spring Harbor Laboratory.[21] The next question we pursued was whether gene targeting could be extended to a whole animal (i.e., the mouse). Because of the low frequency of targeting events in mammalian cells, it was clear that doing the experiments directly in mouse zygotes would not be practical. Rather, targeting events had to be identified first in cultured cells to allow purification of a clonal cell line containing the desired gene disruption; these cells in turn could be used to generate mice capable of transmitting the mutation in their germline. We were familiar with the frustrations associated with previous attempts to obtain germline chimeras using embryonal carcinoma (EC) cells.

However, in the summer of 1984, I heard at a Gordon Research Confer-

ence a discussion of EK cells, which appeared to be much more promising in their potential for contributing to the germline. EK cells, which were developed in Martin Evans' laboratory in Cambridge, England, differed from EC cells in that they were obtained from the early mouse embryo[22] rather than from a mouse tumor. In the winter of 1985, I arranged to spend a week in Martin Evans' laboratory to learn how to culture and manipulate mouse EK cells. It was just before Christmas, a marvelous time to be in Cambridge, England. While I was there, Martin Evans confided to me that the only other molecular biologist who had shown any interest in EK cells, now called *ES cells,* was Oliver Smithies. I believe that most investigators showed little interest in ES cells because they thought of the mouse as a system for expressing exogenous genes. There were only modest advantages to using ES cells as a vehicle to generate such transgenic mice, and those advantages were not enough to counterbalance the lower efficiency of obtaining germline transmission of the foreign DNA by the ES cell route. However, to pursue gene targeting to generate mutations in the mouse, we had no choice but to take the ES cell route. In short, at that time there were few attractive experiments that could be done with ES cells.

In 1986, our total effort was switched to ES cells. We also decided to use electroporation rather than microinjection as a means of introducing the targeting vector into cells. Although microinjection was orders of magnitude more efficient than electroporation as a means of introducing DNA into cells, injections had to be done one cell at a time, and I was getting tired of doing microinjections. With electroporation, 10^8 cells could be manipulated in a single experiment. I also thought that if electroporation worked, the gene-targeting technology would be more readily transferable to other investigators.

The gene that we chose to disrupt in ES cells was *hprt,* because it provided us the luxury of being able to select directly for cells containing the disrupted gene. Because the *hprt* gene is located on the X chromosome and the ES cells were derived from a male mouse, only a single locus had to be disrupted to yield *hprt*$^-$ cell lines. The strategy that we employed was to use a neomycin-resistance gene (*neo*^{r}) to disrupt the *hprt* genomic sequences and then to select for cells resistant to both G418, a neomycin analogue, and 6-thioguanine (6-TG), a drug toxic to cells with a functional *hprt* gene. All such selected cell lines had lost *hprt* enzymatic activity as a result of targeted disruption of the *hprt* locus. These experiments showed that ES cells were good recipients, able to mediate homologous recombination, and that the selection protocols required to identify cell lines containing the targeted disruption did not alter their pluripotent state in culture.

This system also provided a good experimental model for exploration of the parameters affecting the efficiency of gene targeting.[19] We compared two classes of targeting vectors, sequence replacement vectors, and sequence insertion vectors, and showed that the targeting efficiency of both classes of vectors was strongly dependent on the extent of homology between the exogenous and endogenous DNA sequences. We also outlined a number of enrichment

strategies, including promoter trapping, for extending the technology to modify genes whose disruption does not provide a cell-selectable phenotype. I believe that this study played a pivotal role in the development of the field by encouraging other investigators to begin to use gene targeting as a means for determining the function of genes in mammals.

Although mammalian cells have the machinery to direct the newly introduced DNA to its endogenous target, we found that the targeting vector is more often inserted into a random site within the host genome by nonhomologous recombination. The ratio of homologous to nonhomologous recombination events is approximately 1 to 1,000. Because disruption of most genes is not expected to produce a phenotype selectable at the cell level, an investigator seeking a specific disruption must therefore either screen through many colonies of cells to identify the rare colony that contains the desired targeting event or use selections that enrich for cells containing the targeting event.

Late in 1986 I conceived of a very general strategy to enrich for cells in which the targeting event had occurred. It was based on key observations made during our studies of recombination events involving exogenous DNA in mammalian cells, namely that incorporation of DNA segments at random nonhomologous sites involved insertions of a linearized vector in its entirety through its ends, whereas recombination at a homologous target site, of a replacement type vector, involved crossover events occurring only through homologous sequences in the vector. The strategy based on these observations and known as *positive-negative selection* uses two components (Figure 4). One component is referred to as a *positive selectable gene, neo*^{r}, used as a marker to select *for* cells that have incorporated the targeting vector anywhere in the recipient cell genome (i.e., at the target site by homologous recombination or at random sites by nonhomologous recombination). The second component is called a *negative selectable gene* and is located at one end of the linearized targeting vector and used to select *against* cells containing random insertion of the targeting vector. The net effect is to enrich for cells in which the desired homologous targeting event has occurred.

The strength of this enrichment procedure is that it is independent of the function of the gene and succeeds whether or not the gene is expressed in the recipient ES cells.[23] Positive-negative selection is now the most frequently used procedure to enrich for cells containing targeting events. We evaluated several negative selectable genes, including *xgpt, hprt,* and the *diphtheria-toxin* gene, and finally settled on *HSV-tk*. The origin of the idea for positive-negative selection was not very romantic and involved purely deductive reasoning. Enrichment for cells containing the targeting event could be achieved either by direct selection for cells containing the targeting event or by elimination of cells that contained random insertions of the targeting vector. This truism, coupled with an understanding of how information is transferred between an exogenous and an endogenous DNA sequence by homologous and nonhomologous recombination, naturally led to the concept of positive-negative selection.

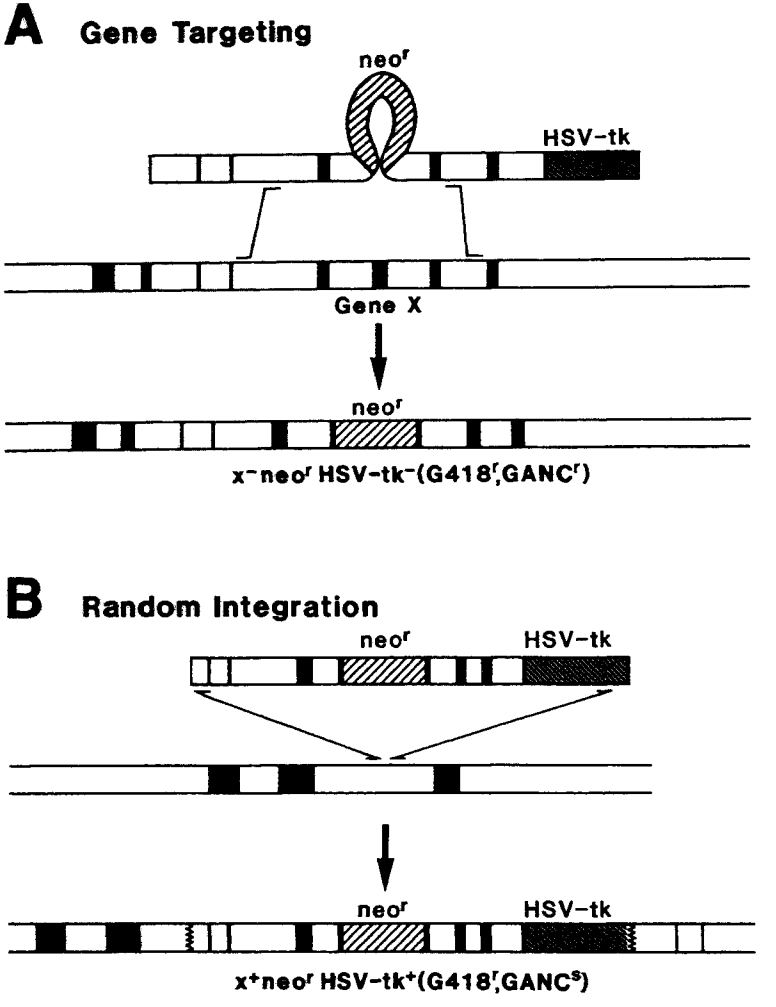

Figure 4. The positive-negative selection procedure used to enrich for ES cells containing a targeted disruption in gene X. (A) A gene X sequence replacement vector, which contains an insertion of the *neo*^{r} gene in an exon of gene X and a linked *HSV-tk* gene, is shown pairing with the chromosomal copy of gene X. Homologous recombination between the targeting vector and genomic X DNA results in the disruption of one copy of genomic gene X and the loss of the vector's *HSV-tk* sequences. Cells resulting from such an event will be X^- neo^{r+} and HSV-*tk*$^-$ and will be resistant to both G418 and ganciclovir. (B) Integration at a random, nonhomologous site. Because nonhomologous insertion of exogenous DNA into the chromosome occurs through the ends of the linearized DNA, the *HSV-tk* gene will remain linked to the *neo*^{r} gene. Cells derived from this type of event will be X^+, neo^{r+} and HSV-*tk*$^+$ and therefore resistant to G418 but sensitive to ganciclovir. Ganciclovir is a nucleoside analogue that specifically kills cells containing a functional *HSV-tk* gene but is not toxic to cells containing the cellular *tk* gene.

The use of gene targeting to evaluate the functions of genes in the living mouse is now a routine procedure. It is very gratifying to be able to pick up almost any major journal in the biologic sciences and find the description of yet another so-called *gene knockout* mouse. In the past 5 years, the *in vivo* functions of well over 250 genes have been determined. It is relatively easy to project where gene-targeting technology will go in the near future. It will continue to serve as the way to determine the roles of individual genes in mammalian biology. This will be accomplished by the generation of null mutations knocking out the genes of interest. Those investigators who desire deeper insights will generate an allelic series of mutations in a chosen gene to evaluate the effects of partial loss-of-function and gain-of-function mutations. To permit the evaluation of potential multiple roles of a gene in multiple tissues, gene targeting will be used to engineer tissue-specific gene disruptions using the cre/loxP system.[24] Further, technology should soon become available that will allow the investigator to turn chosen genes on or off in the adult or during any phase of mouse development. Finally, because most biologic processes are mediated by interactions among a number of genes, such phenomena will be studied by combination of multiple targeted mutations in a single mouse. There is no question that the mouse is a very complex organism. However, the broad range of genetic manipulations now available through gene targeting should provide a means for us to begin deciphering even the most complex of biologic processes such as development and learning.

REFERENCES

1. Capecchi MR. Targeted gene replacement. Sci Am 1994; 270:54–61.
2. Wigler M, Silverstein S, Lee L, Pellicer A, Chen Y, Axel R. Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell 1977; 11:223–232.
3. Capecchi MR. High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell 1980; 22:479–488.
4. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A 1980; 77:7380–7384.
5. Costantini F, Lacy E. Introduction of a rabbit β-globin gene into the mouse germ line. Nature 1981; 294:92–94.
6. Brinster RL, Chen HY, Trumbauer ME, Senear AW, Warren R, Palmiter RD. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. Cell 1981; 27:223–231.
7. Wagner EF, Stewart TA, Mintz B. The human β globin gene and a functional thymidine kinase gene in developing mice. Proc Natl Acad Sci U S A 1981; 78:5016–5020.
8. Wagner TE, Hoppe PC, Jollick JD, Scholl DR, Hodinka RL, Gault JB. Microinjection of a rabbit β globin gene in zygotes and its subsequent expression in adult mice and their offspring. Proc Natl Acad Sci U S A 1981; 78:6376–6380.
9. Luciw PA, Bishop JM, Varmus HE, Capecchi MR. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell 1983; 33:705–716.
10. Levinson B, Khoury G, VandeWoude G, Gruss P. Activation of SV40 genome by 72-base pair tandem repeats of Moloney sarcoma virus. Nature 1982; 295:568–572.

11. Folger KR, Wong EA, Wahl G, Capecchi MR. Patterns of integration of DNA microinjected into cultured mammalian cells: Evidence for homologous recombination between injected plasmid DNA molecules. Mol Cell Biol 1982; 2:1372–1387.
12. Hennen A, Hicks JB, Fink GR. Transformation of yeast. Proc Natl Acad Sci U S A 1978; 75:1929–1933.
13. Folger KR, Thomas KR, Capecchi MR. Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. Mol Cell Biol 1985; 2: 1372–1387.
14. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell 1986; 44:419–428.
15. Wong EA, Capecchi MR. Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. Mol Cell Biol 1987; 7:2294–2295.
16. Zheng H, Wilson JH. Gene targeting in normal and amplified cell lines. Nature 1990; 344:170–173.
17. Petes TD. Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. Cell 1980; 19:765–774.
18. Szostak JW, Wu R. Unequal crossing-over in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature 1980; 284:426–430.
19. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 1987; 51:503–512.
20. Deng C, Capecchi MR. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol Cell Biol 1992; 12:3365–3371.
21. Folger K, Thomas K, Capecchi MR. Analysis of homologous recombination in cultured mammalian cells. Cold Spring Harb Symp Quant Biol 1984; 49:123–138.
22. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature 1981; 292:154–156.
23. Mansour SL, Thomas KR, Capecchi MR. Disruption of the proto-oncogene *int*-2 in mouse embryo-derived stem cells: A general strategy for targeting mutations to nonselectable genes. Nature 1988; 336:348–352.
24. Gu H, Marth JD, Orban PC, Mossmann H, Rajewsky K. Deletion of a DNA polymerase beta gene segment in T cells using cell type specific gene targeting. Science 1994; 265: 103–106.