MOLECULAR AND CELLULAR BIOLOGY, July 1992, p. 2919–2923
0270-7306/92/072919-05$02.00/0

Vol. 12, No. 7

High-Fidelity Gene Targeting in Embryonic Stem Cells by Using Sequence Replacement Vectors

KIRK R. THOMAS, CHUXIA DENG, AND MARIO R. CAPECCHI*

HHMI Research Laboratories, Eccles Institute of Human Genetics, Suite 5400, University of Utah Medical Center, Salt Lake City, Utah 84112

Received 10 February 1992/Accepted 26 March 1992

Mutations were targeted to the *Hprt* locus in murine embryonic stem cells by using sequence replacement vectors. When the vector was designed such that the mutated sequences were flanked on both sides by several kilobases of DNA homologous to the target locus, replacement of chromosomal sequences with the exogenous DNA occurred with precision. If, on the other hand, the target-homologous DNA on one arm of the vector was reduced to below 1 kb in length, the fidelity of recombination was diminished.

Gene targeting in mouse embryo-derived stem (ES) cells (22) provides the means to introduce specific mutations into the mouse genome. Analysis of the phenotypes of mice carrying such mutations has provided insight into the biological roles of a number of genes whose functions had previously been only a matter of speculation (3, 4, 8–11, 13, 15–18, 20, 23–25).

The first step in the gene targeting protocol is to mutate, in vitro, a cloned copy of the gene of interest. The mutated gene, carried on a targeting vector, is then introduced into ES cells, some fraction of which incorporates the exogenous sequences via homologous recombination. The most commonly used targeting vector is the replacement vector, which contains a linear array of wild-type genomic DNA on both sides of the mutated sequence (22). The vector is thought to pair via shared homology with its chromosomal counterpart, and subsequent homologous recombination—either a double-reciprocal exchange surrounding the mutation or a gene conversion event—replaces the genomic sequences with sequences from the targeting vector.

The capacity to create predetermined genomic alterations is thus dependent upon the precision of the recombination reaction. Although nonspecific mutagenesis and genomic rearrangements have been shown to accompany homologous recombination (2, 5, 21), such events have generally not been viewed as problematic because of their low frequency of occurrence. Of over 20 animals generated through replacement vector-mediated recombination, only one contained a recombinant locus which deviated from that expected (23). In this case, a dimer of the targeting vector was used as one of the reactants, and a tandem duplication of mutant vector sequences replaced the endogenous locus.

The general assumption that replacement vector-mediated recombination proceeds with high fidelity has recently been challenged (7). This particular examination of gene targeting at the *Hprt* locus in murine ES cells revealed that fewer than 5% of the recombinants were the result of simple legitimate replacement reactions. The vast majority of the events analyzed in this study were accompanied either by vector rearrangements prior to recombination or by genomic rearrangements presumably induced during recombination. One implication from this study is that replacement vector-mediated homologous recombination is inherently inaccurate.

The discrepancies between this latter study and the numerous other published reports of legitimate gene replacement prompted us to scrutinize potential differences in the experimental protocols of the various investigations. One untested variable has been in the design of the replacement vectors, specifically, the lengths of the genome-homologous DNA sequences between the mutation and the ends of the vector. To determine whether the fidelity of homologous recombination could be influenced by the amount of homology flanking the mutation, we examined the products of gene targeting under conditions in which this length was varied.

In this report, we show that replacement vector-mediated recombination is not inherently inaccurate. When the replacement vector is designed such that the desired mutation is flanked on both sides by several kilobases of DNA homologous to the target locus, gene replacement proceeds with fidelity. If, on the other hand, the mutation is flanked on one side by less than 1 kb of target-homologous DNA, abortive recombination can occur.

MATERIALS AND METHODS

Vectors and cell lines. Construction of plasmid pHPRT9.6^{neo} has been described elsewhere (22). The ES cell line, CC1.2 (1), was derived by E. Robertson. Transformation of ES cells by electroporation and selection conditions for gene replacement events by using 6-thioguanine (6-TG) and G418 have been published previously (22).

DNA purification and analysis. ES cell lines resistant to 6-TG and G418 were grown to confluence on a 60-mm plate. The cells were harvested by centrifugation, resuspended in 1.5 ml of lysis buffer (10 mM EDTA, 100 mM NaCl, 1.5% sodium dodecyl sulfate, 20 mM Tris-HCl [pH 7.8]), and digested with protease K (0.2 mg/ml; Beckman) at 37°C for 2 h. DNA was purified from each cell line by the following salting-out protocol, generously provided by Brian Stanton (National Cancer Institute, Frederick, Md.). Five hundred microliters of the cell extract was mixed vigorously with 250 μl of saturated NaCl; after 10 min on ice, the mixture was centrifuged at 3,000 × *g* in a microcentrifuge; 600 μl of the supernatant was removed, and the DNA was precipitated with 2 volumes of ethanol; the DNA was removed with a glass rod and resuspended in 150 μl of a 10 mM Tris-HCl–1 mM EDTA (pH 8.0) solution.

* Corresponding author.

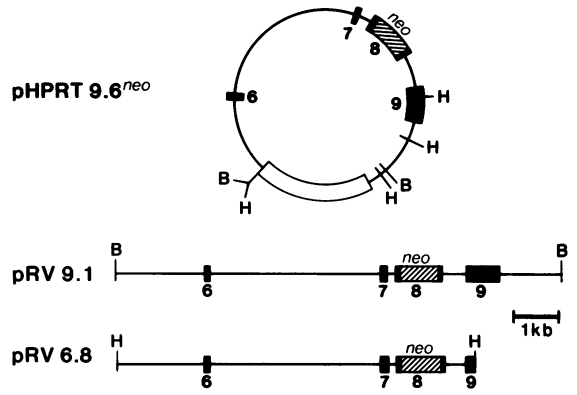

FIG. 1. Gene targeting vectors. Construction of plasmid pHPRT 9.6^{neo} has been described elsewhere (22). To generate the targeting vector, pRV9.1, pHPRT9.6^{neo} was digested with *Bgl*II; to generate pRV6.8, pHPRT9.6^{neo} was digested with *Hin*dIII. B, *Bgl*II; H, *Hin*dIII. Narrow lines represent introns and 3′ noncoding regions of the murine *Hprt* locus; solid boxes represent *Hprt* exons, numbered as defined by Melton et al. (14); striped boxes represent the Neo^r gene from plasmid pMC1Neo; open box represents sequences from plasmid pUC9.

Five micrograms of purified DNA from each cell line was digested with restriction endonuclease and analyzed by Southern transfer analysis as described previously (22). Hybridization probes were as follows. Probe a is a 500-bp *Sph*I-*Bgl*II fragment from intron 5, immediately 5′ to the end of sequences in the targeting vectors; probe b is a 1-kb *Bst*EII-*Sca*I fragment extending from intron 6 to the middle of exon 8, containing sequences present on both targeting vectors; probe c is a 300-bp *Hin*dIII fragment containing sequences outside the *Hprt* locus, beginning 200 bp 3′ from the end of the vector, pRV9.1.

RESULTS

Experimental design. Mutations were targeted to exon 8 of the murine *Hprt* gene. Cells harboring mutations were first identified by virtue of their resistance to G418 and 6-TG. Cells resistant to both G418 and 6-TG were then screened by Southern transfer analysis. Such analysis not only confirms the presence of the targeted mutation but can detect rearrangements in genomic sequences resulting from inappropriate recombination.

The targeting vectors used in these experiments are depicted in Fig. 1. They are derived from plasmid pHPRT9.6^{neo}, which contains 9.6 kb of murine *Hprt* DNA, interrupted in exon 8 by the bacterial Neo^r gene. The Neo^r gene not only disrupts the coding region of the *Hprt* gene but serves as a positive selectable marker to identify cells transfected with the targeting vector. Two replacement vectors were derived from this plasmid by different restriction endonuclease digestions. Digestion with *Bgl*II produces pRV9.1, which contains the mutated exon 8 flanked on the 5′ end by 6.1 kb of *Hprt* sequences and on the 3′ end by 3.0 kb of *Hprt* sequences. The other vector, pRV6.8, is released by digestion with *Hin*dIII and contains the same 6.1 kb of 5′ flanking DNA found in pRV9.1 but only 0.7 kb of DNA 3′ to the Neo^r gene insertion.

Following digestion of the plasmid with the appropriate enzyme, DNA was introduced into ES cells by electroporation. Colonies resistant to both G418 and 6-TG, derived from four separate electroporation experiments, were isolated, and their DNA was analyzed by Southern transfer. DNA was digested with two different restriction endonucleases and probed with three different *Hprt* probes designed to detect rearrangements of both internal and flanking sequences. The targeting frequency, expressed as the ratio of the number of G418^r, 6-TG^r colonies to the number of cells surviving electroporation, was 1×10^{-5} for pRV9.1 and 4×10^{-6} for pRV6.8.

pRV9.1-induced recombinants are of predictable structure. Eighteen Neo^r, 6-TG^r colonies resulting from transfection with pRV9.1 were chosen for analysis by Southern transfer. Figure 2E shows restriction maps of the wild-type genomic *Hprt* locus and the same locus replaced with pRV9.1 sequences. The 9.1-kb *Eco*RI fragment containing the wild-type *Hprt* exon 8 can be detected when DNA from the parental ES line, CC1.2, is digested with *Eco*RI and probed with either sequences flanking the vector (probe a) or an internal fragment (probe b). Insertion of the Neo^r gene in exon 8 introduces an additional *Eco*RI site such that probes a and b are predicted to hybridize to a fragment 8.1 kb in length. All 18 cell lines targeted with pRV9.1 show only the predicted 8.1-kb fragment, implying that no major rearrangements 5′ to the Neo^r gene insertion were generated by the targeting reaction (Fig. 2A and B). (The fainter, 9.1-kb band seen in all lanes is from the wild-type *Hprt* gene present in the feeder cells upon which the ES cells are grown.)

To analyze sequences 3′ to the Neo^r gene insertion, DNA from the same 18 cell lines was digested with *Xba*I. An 8.4-kb *Xba*I fragment is detected in wild-type CC1.2 DNA hybridized to both probe b and probe c, a probe isolated from sequences located 3′ to sequences in the targeting vector (Fig. 2C and D). Targeting of the 1-kb Neo^r gene to exon 8 should increase the length of this *Xba*I fragment by 1 kb. In all cell lines targeted by pRV9.1, only a 9.4-kb *Xba*I fragment is detected by either probe b or probe c. This finding suggests that rearrangements at the 3′ end of the *Hprt* gene did not occur in any of the 18 cell lines examined.

pRV6.8 can induce rearrangements. The precision of recombination obtained with pRV9.1, although consistent with our previous results, is in striking contrast to a recent study that showed rearrangements in over 95% of replacement vector-induced recombinant cells (7). The replacement vector used by Hasty et al. (7) is similar in design to pRV9.1 but contains 2 kb less homology 3′ to the Neo^r gene insertion than does pRV9.1. To determine whether this difference in vector design might be responsible for the difference in results, we reduced the amount of contiguous homology at the terminus of the targeting vector. The vector, pRV6.8, contains the same 6.1 kb of homology 5′ to the Neo^r gene insertion found in pRV9.1 but has only 0.7 kb of homology 3′ to the insertion. Eighteen G418^r, 6-TG^r clones resulting from transfection with pRV6.8 were analyzed by Southern transfer analysis, using the enzymes and probes described for the analysis of pRV9.1-transformed cells. Analysis of sequences 5′ to the Neo^r gene by *Eco*RI digestion and hybridization to probes a and b clearly shows two cell lines containing a fragment deviating from the predicted 8.1-kb length (lines 2m and 2n; Fig. 3A and B). Analysis of the 3′ end with *Xba*I digestion and hybridization to probe b reveals three additional cell lines (2a, 2b, and 2c) with unpredicted structures at the *Hprt* locus (Fig. 3C).

Rearrangements can result from vector concatemerization. The structures of three of the aberrant pRV6.8-induced recombinants can be deduced from the restriction endonuclease digestions shown in Fig. 3. Cell lines 2a, 2b, and 2c all contain a tandem duplication of *Hprt* sequences present on

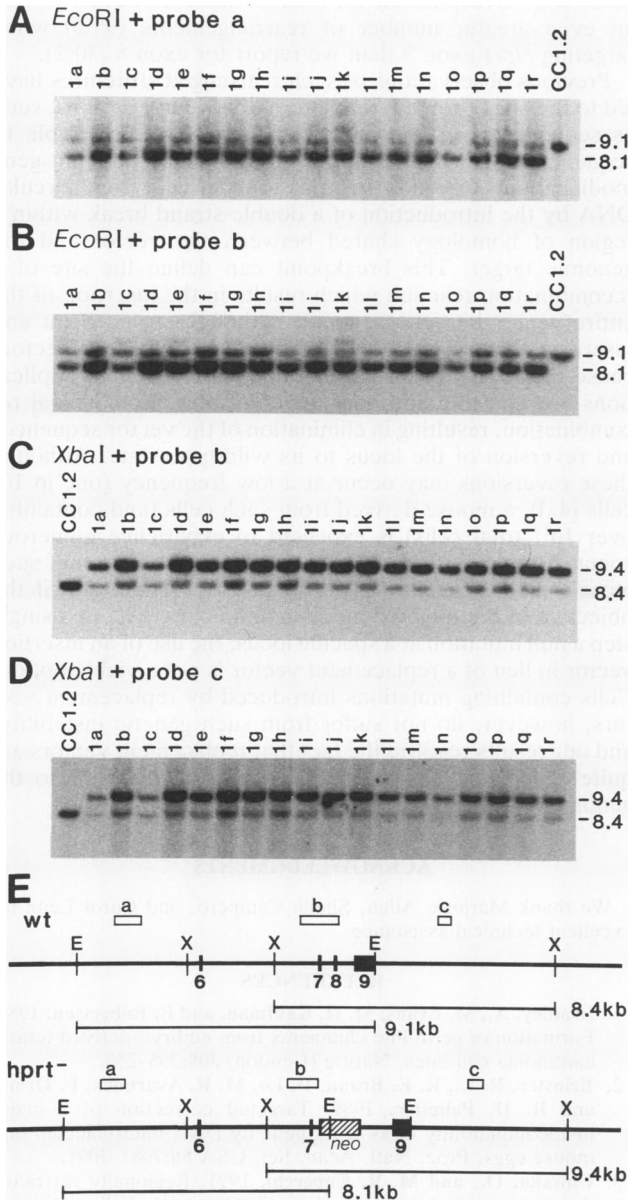

FIG. 2. Gene targeting with pRV9.1. (A to D) Southern transfer analysis. DNA was extracted from 18 G418^r, 6-TG^r cell lines (1a through 1r) transfected with plasmid pRV9.1 and from the parental ES cell line, CC1.2. A 5-μg aliquot from each cell line was digested with either *Eco*RI (A and B) or *Xba*I (C and D), electrophoresed through 0.7% agarose, and transferred to a nitrocellulose membrane. The membranes were then hybridized to random-primed, ^{32}P-labeled probes (Pharmacia) (described in Materials and Methods). Numbers to the right of each panel indicate the sizes of the labeled fragments in kilobases. (E) Restriction endonuclease maps. The upper map represents wild-type *Hprt* DNA from the parental cell line, CC1.2; the lower map depicts the targeted loci from the *hprt*$^-$ cell lines analyzed by Southern transfer. Thin lines represent intron and 3′ flanking sequences; numbered boxes represent exons; striped box represents the Neo^r gene within exon 8; open boxes represent positions of the probes. E, *Eco*RI; X, *Xba*I. Origins of the fragments seen in panels A to D are indicated beneath each map.

the targeting vector. This duplicated structure is distinguished by a 5-kb *Xba*I fragment and an 8.9-kb *Eco*RI fragment that hybridize to probe b. (The 8.9-kb *Eco*RI fragment is masked somewhat by the 9.1-kb fragment from feeder cell DNA.) Only the 5′ copy of the duplication, however, contains the Neo^r gene insertion in exon 8, verified by an 8.4-kb, wild-type *Xba*I fragment detected by probe c, the 3′ flanking probe. The presence of the repeat in the genome is presumably the result of concatemerization of the targeting vector prior to recombination, followed by a gene replacement reaction with crossover points as depicted in Fig. 3E. Although uncommon, similar reactions have been reported previously (6, 19, 23). The structures of the rearranged *Hprt* locus in cell lines 2m and 2n are more complex and have not been fully determined.

DISCUSSION

We have shown that gene targeting via a gene replacement reaction proceeds in a predictable manner, provided that sufficient homology flanks the mutated sequence. Following targeted mutagenesis of the *Hprt* locus with the vector pRV9.1, all 18 analyzed recombinants were shown to contain a single copy of the *Hprt* locus, disrupted in exon 8 by the Neo^r gene. Furthermore, analysis of regions both 5′ and 3′ to the replaced sequences failed to detect rearrangements of adjacent DNA. The fidelity of the replacement reaction could be altered experimentally, however, by reducing the length of end-terminal homology on the targeting vector. When the vector pRV6.8 was used as a reactant, for example, 5 of 18 recombinants showed an unexpected structure of the targeted locus. Of these, at least 3 contained a structure consistent with the replacement of genomic sequences with sequences from a concatemerized targeting vector.

The difference in recombination fidelity between pRV9.1 and pRV6.8 corresponds to a 2.3-kb difference in *Hprt* homology on one arm of the targeting vector. It is likely that the reduction in arm length rather than a reduction in the amount of absolute homology between the targeting vector and the target locus is responsible for the rearrangements induced by pRV6.8. In experiments similar to those presented here, we have examined the recombinant products formed by replacement vectors sharing as little as 5.4 kb of total homology with the target locus. When this homology was distributed equally on the two arms of the replacement vectors, no rearrangements in the recombinant products were detected (22; unpublished results).

The minimum length of homology on both arms of the targeting vector necessary for accurate gene replacement may reflect a requirement for stable pairing between the vector and its target. Stability can also be realized by the pairing of a concatemerized vector, as shown in Fig. 3E. Vector concatemerization is presumably the result of extrachromosomal recombination occurring prior to the gene targeting reaction. Under similar experimental conditions, concatemerization of the two vectors, pRV6.8 and pRV9.1, would be expected to occur at a similar frequency. However, because of minimum homology restrictions, a multimer of the shorter vector, pRV6.8, may show an increased recombinogenic advantage with respect to the pRV6.8 monomer, thus increasing the proportion of duplication events following transfection with this vector. The depression of recombinational fidelity caused by a reduced arm length may be compounded if there is also sequence heterology between the vector arm and its target. Such heterologies can be encountered when vector DNA is derived from a

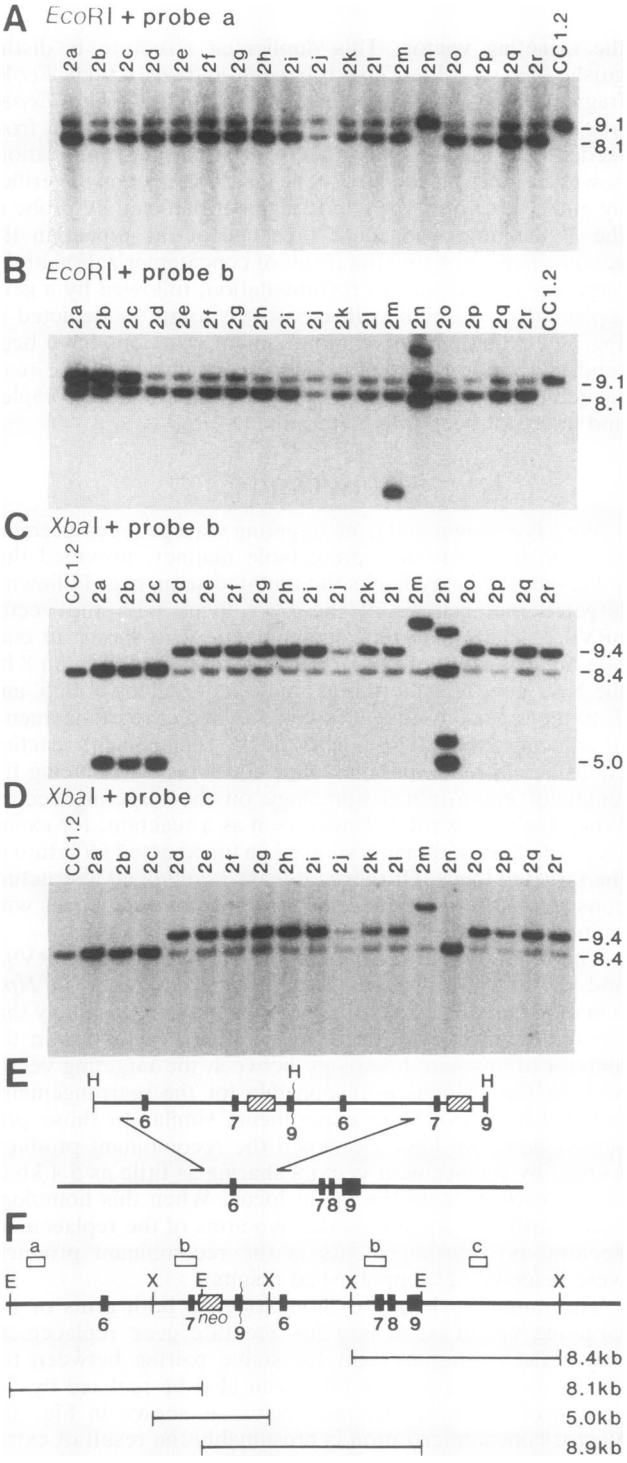

FIG. 3. Gene targeting with pRV6.8. (A to D) Southern transfer analysis. DNA was extracted from 18 G418^r, 6-TG^r cell lines (2a through 2r) transfected with pRV6.8. Analysis was performed as described in the legend to Fig. 2. (E) Duplication of the targeting vector. The upper map represents the head-to-tail dimerization of the *Hin*dIII-linearized vector pRV6.8; the lower map represents the chromosomal *Hprt* locus; diagonal lines highlight potential recombination sites. Thin horizontal lines represent introns and 3′ flanking sequences; numbered boxes represent *Hprt* exons; striped boxes represent the Neo^r gene. H, *Hin*dIII. (F) *Hprt* locus in cell lines 2a, 2b, and 2c, resulting from recombination with a dimerized targeting vector. Positions of the probes used in the Southern transfer analysis are indicated by open boxes, and the origins of the fragments labeled in panels A to D are indicated beneath the map.

mouse strain nonisogenic with that of the target cell line and can vary in degree along the genome. It is possible that such heterology is responsible for the fact that Hasty et al. (7) find an even greater number of rearrangements (95%) when targeting *Hprt* exon 3 than we report for exon 8 (30%).

Previous observations of such rearranged structures have led to the suggestion (7) that alternative vector designs, such as sequence insertion vectors (22), may be preferable to sequence replacement vectors in performing targeted gene modification. Insertion vectors are derived from circular DNA by the introduction of a double-strand break within a region of homology shared between the vector and the genomic target. This breakpoint can define the site of a recombination reaction which results in the insertion of the entire vector into the genome. Although replacement and insertion vectors are equally efficient (22), insertion vectors cause the formation of tandem duplications. Such duplications are unstable and may undergo intrachromosomal recombination, resulting in elimination of the vector sequences and reversion of the locus to its wild-type state. Although these reversions may occur at a low frequency (one in 10^6 cells [4a]), a mouse derived from such cells (and containing over 10^{13} total cells) is expected to experience numerous somatic reversions. The phenotypes of a population of such mosaic animals might vary dramatically. Therefore, if the objective of the targeted modification is to create in a single step a null mutation at a specific locus, the use of an insertion vector in lieu of a replacement vector is not a viable option. Cells containing mutations introduced by replacement vectors, however, do not suffer from such genetic instability, and our results reaffirm the fact that replacement vectors are quite capable of targeting predetermined mutations to the genome of ES cells.

ACKNOWLEDGMENTS

We thank Marjorie Allen, Sheila Campero, and Carol Lenz for excellent technical assistance.

REFERENCES

1. **Bradley, A., M. Evans, M. H. Kaufman, and E. Robertson.** 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature (London) **309:**255–256.
2. **Brinster, R. L., R. E. Braun, D. Lo, M. R. Avarbock, F. Oram, and R. D. Palmiter.** 1989. Targeted correction of a major histocompatibility class II E_α gene by DNA microinjected into mouse eggs. Proc. Natl. Acad. Sci. USA **86:**7087–7091.
3. **Chisaka, O., and M. R. Capecchi.** 1991. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5.* Nature (London) **350:**473–479.
4. **DeChiara, T. M., A. Efstratiadis, and E. J. Robertson.** 1990. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature (London) **345:**78–80.

4a. **Deng, C.** Unpublished data.

5. **Doetschman, T., N. Maeda, and O. Smithies.** 1988. Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA **85:**8583–8587.
6. **Fell, H. P., S. Yarnold, I. Hellström, and K. R. Folger.** 1989. Homologous recombination in hybridoma cells: heavy chain chimeric antibody produced by gene targeting. Proc. Natl. Acad. Sci. USA **86:**8507–8511.
7. **Hasty, P., J. Rivera-Pérez, C. Chang, and A. Bradley.** 1991. Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol. Cell. Biol. **11:**4509–4517.
8. **Joyner, A. L., K. Herrup, B. A. Auerbach, C. A. Davis, and J. Rossant.** 1991. Subtle cerebellar phenotype in mice homozygous

for a targeted deletion of the *En-2* homeobox. Science **251:**1239–1243.

9. **Koller, B. H., P. Marrack, J. W. Kappler, and O. Smithies.** 1990. Normal development of mice deficient in β_2M, MHC class I proteins, and CD8$^+$ T cells. Science **248:**1227–1230.
10. **Kühn, R., K. Rajewsky, and W. Müller.** 1991. Generation and analysis of interleukin-4 deficient mice. Science **254:**707–710.
11. **Lufkin, T., A. Dierich, M. LeMeur, M. Mark, and P. Chambon.** 1991. Disruption of the *Hox-1.6* homeobox gene results in defects in a region corresponding to its rostral domain of expression. Cell **66:**1105–1119.
12. **Mansour, S. L., K. R. Thomas, C. Deng, and M. R. Capecchi.** 1990. Introduction of a *lacZ* reporter gene into the mouse *int-2* locus by homologous recombination. Proc. Natl. Acad. Sci. USA **87:**7688–7692.
13. **McMahon, A. P., and A. Bradley.** 1990. The *Wnt-1* (*int-1*) proto-oncogene is required for development of a large region of the mouse brain. Cell **62:**1073–1085.
14. **Melton, D. W., D. S. Konecki, J. Brennand, and C. T. Caskey.** 1984. Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. Proc. Natl. Acad. Sci. USA **81:**2147–2151.
15. **Mucenski, M. L., K. McLain, A. B. Kier, S. H. Swerdlow, C. M. Schreiner, T. A. Miller, D. W. Pietryga, W. J. Scott, Jr., and S. S. Potter.** 1991. A functional *c-myb* gene is required for normal murine fetal hepatic hematopoiesis. Cell **65:**677–689.
16. **Pevny, L., M. C. Simon, E. Robertson, W. H. Klein, S. Tsai, V. D'Agati, S. H. Orkin, and F. Costantini.** 1991. Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. Nature (London) **349:**257–260.
17. **Rahemtulla, A., W. P. Fung-Leung, M. W. Schilham, T. M. Kündig, S. R. Sambhara, A. Narendran, A. Arabian, A. Wakeham, C. J. Paige, R. M. Zinkernagel, R. G. Miller, and T. W. Mak.** 1991. Normal development and function of CD8$^+$ cells but markedly decreased helper cell activity in mice lacking CD4. Nature (London) **353:**180–184.
18. **Schwartzberg, P. L., S. P. Goff, and E. J. Robertson.** 1989. Germ-line transmission of a *c-abl* mutation produced by targeted gene disruption in ES cells. Science **246:**799–803.
19. **Schwartzberg, P. L., E. J. Robertson, and S. P. Goff.** 1990. Targeted gene disruption of the endogenous *c-abl* locus by homologous recombination with DNA encoding a selectable fusion protein. Proc. Natl. Acad. Sci. USA **87:**3210–3214.
20. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the *c-src* proto-oncogene leads to osteopetrosis in mice. Cell **64:**693–702.
21. **Thomas, K. R., and M. R. Capecchi.** 1986. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. Nature (London) **324:**34–38.
22. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell **51:**503–512.
23. **Thomas, K. R., and M. R. Capecchi.** 1990. Targeted disruption of the murine *int-1* proto-oncogene resulting in severe abnormalities in midbrain and cerebellar development. Nature (London) **346:**847–850.
24. **Tybulewicz, V. L. J., C. E. Crawford, P. K. Jackson, R. T. Bronson, and R. C. Mulligan.** 1991. Neonatal lethality and lymphopenia in mice with a homozygous disruption of the *c-abl* proto-oncogene. Cell **65:**1153–1163.
25. **Zijlstra, M., E. Li, F. Sajjadi, S. Subramani, and R. Jaenisch.** 1989. Germ-line transmission of a disrupted β_2-microglobulin gene produced by homologous recombination in embryonic stem cells. Nature (London) **342:**435–438.