Mini Review

Gene Targeting
The Precise Manipulation of the Mammalian Genome

Jeffrey Robbins

Genetics is a powerful tool for studying the function of different gene products as well as the physiological consequences of a normal or aberrant polypeptide. Until recently, the most powerful genetic applications have been restricted to relatively simple organisms whose genomes are more easily manipulated. For a number of years, in lower organisms, it has been possible to create defined genetic changes that are targeted to a specific locus or even to a defined codon or transcriptional regulatory sequence. With the development of gene targeting using embryonic stem cells derived from the preimplantation blastocyst of a mouse, it has become possible to extend these experiments to a mammalian system. Via homologous recombination, one can ablate, or “knock out,” a defined genetic locus or mutate a particular set of nucleotides that encodes a peptide domain of interest. These techniques, when applied to genes that underlie normal cardiovascular function, promise to define the exact role(s) different proteins play during the development, growth, and maintenance of the heart. The ability to generate defined animal models of human disease in which the primary genetic defect is known should lead to fundamental advances in the study of the normal and failing heart. (Circulation Research 1993;73:3-9)

KEY WORDS • genetics • gene targeting • mutation • mouse

A genetic approach for studying an organism’s function rests on the ability to easily generate changes in the genetic material and to detect the resultant mutations by developing a rational screening process. The past 50 years provide ample evidence that genetic analyses can be immensely powerful in defining fundamental paradigms of gene and protein function. The majority of detailed genetic analyses have been carried out in the prokaryotes or in the simpler eukaryotes, such as yeast, Drosophila, and Caenorhabditis. However, even these relatively noncomplex systems have provided seminal insights that underlie our current understanding of most of the basic cellular processes including replication, cellular division, and RNA and protein metabolism.

It is necessary, in a general genetic analysis, to screen large numbers of organisms in a relatively short time. Coupled with the difficulties inherent in manipulating the complex mammalian genome in a defined controlled manner, concerted genetic analyses in higher eukaryotes have been limited. An additional difficulty in performing productive genetic experiments in a mammalian system is the potential complex interplay of a number of gene products in producing a detectable phenotype. The success of an undefined random mutational analysis in unraveling a mechanistic basis for a systemic dysfunction remains unproven. Thus, for a complex mammalian developmental system, it seems likely that a lengthy and expensive general screening for undefined mutations would not normally be worth the considerable human and financial resources.

These general caveats, when applied to the cardiovascular system, have severely limited directed mutational analyses of those genes whose products underlie normal and diseased cardiac function. The applicability of information obtained from lower eukaryotic systems, such as Drosophila and Caenorhabditis, concerning the action and function of genes that are expressed in the mammalian heart is doubtful. The organization of the mammalian four-chambered heart and the interplay of the unique isoforms that are found in the myocardium are not accurately reflected in the “hearts” of flies and worms. Although some chronic animal models of cardiovascular disease such as the spontaneous hypertensive rat are available, the underlying genetic defect(s) is not known. The potential value of defining a genetic defect that results in cardiac dysfunction is illustrated by the discovery that the beta-myosin heavy chain locus is responsible for a large number of patients clinically determined to have familial hypertrophic cardiomyopathy. The rapid progress in the diagnosis of asymptomatic patients and in the development of rapid diagnostic procedures based on molecular reagents, as well as the possibility of accurate prognostic capabilities, underscores the rapid progress that can be made if a defined genetic lesion of the cardiovascular system is identified.

The mammalian genome has been most easily manipulated through the use of transgenic technology applied to the mouse, and important information concerning cardiac development and function has been obtained by creating animals whose gene complements have been augmented. Although a detailed treatment of this topic is beyond the scope of this review, overexpression of genes such as the SV40 T antigen or the c-myc protooncogene in the heart has been useful in delineating

Received February 1, 1993; accepted March 12, 1993.
From the Department of Pharmacology and Cell Biophysics, University of Cincinnati (Ohio) College of Medicine.
Correspondence to Department of Pharmacology and Cell Biophysics, University of Cincinnati College of Medicine, Cincinnati, OH 45267-0575 (Dr Robbins).

Downloaded from http://circres.ahajournals.org/ by guest on April 19, 2017
some of the mechanisms that underlie myocyte proliferation and gene expression. Transgenic experiments, in which a cardiac gene promoter has been used to drive ectopic expression of transcripts not normally expressed in the heart, such as MyoD,\(^7\) have illustrated facets of muscle determination and differentiation.

In some cases, the insertion of the exogenous DNA during creation of the transgenic animal leads to the creation of interesting mutants (reviewed in Reference 8). For example, insertion of a retroviral vector into the \(\alpha 1(I)\) collagen gene reduces gene expression to 1% to 5% of wild-type levels, and the homozygous animals die at midgestation from massive hemorrhage.\(^9\)

Although insertional mutagenesis is very useful in isolating mutations that can be analyzed at the molecular level and can result in "gene knockouts," the process is a random one. In most cases, it is quite laborious to identify the locus in which the insertion has occurred. More often than not, these sequences have not been identified previously, which further complicates the difficulty of characterizing the event. Thus, the major limitations inherent in a genetic analysis of the mammalian genome has been the production of defined mutations capable of being transmitted through the germ line and the ability to detect them. These shortcomings have now been circumvented by the recent development of gene targeting. Unlike traditional transgenic technologies, in which exogenous DNA is inserted randomly in multiple copies into various chromosomal locations, gene targeting depends on homologous recombination between the exogenous DNA sequences and the target locus (Fig 1). Thus, whereas transgenesis normally leaves homologous sequences unaffected (since the input DNA is inserted into a different site), gene targeting places sequences at one site in the genome. Depending on the nature of the construct and how it is prepared, the endogenous sequences can be either replaced and/or disrupted by the insertion. This can have the effect of essentially "knocking out" the target gene, with the subsequent potential of observing the phenotypic consequences of a null mutation.

**Homologous Recombination in Embryonic Stem Cells**

The ability to ablate or modify a particular gene and transfer that mutation to the mammalian germ line depends on isolating a totipotent (stem) recipient cell capable of colonizing the germ line and demonstrating that this cell can undergo homologous recombination with an input DNA. To date, a number of lines of embryonic stem (ES) cells have been derived. However, cell lines capable of colonizing the germ line have all been obtained from the 129Sv mouse.\(^10\) Male cells are used exclusively; their genomic complement appears to be more stable during passage in culture. Additionally, only XY cells will give rise to sperm, and male chimeras can generate many more offspring than females. The ES cells are derived from an inner cell mass of some 60 cells that are present in the 4.5-day preimplantation mouse blastocyst. These cells can be immunoisogenerically separated from the epithelial trophectoderm. When grown on mitotically quiescent fibroblast feeder layers or in the presence of leukemia inhibitory factor, the cells remain undifferentiated and euploid; both appear to be prerequisites for subsequent germ line transmission. Even after many passages, these cells, when injected into a 4.5-day embryo, are capable of colonizing the different cell lineages, including the germ line.\(^11\) Thus, the ge-
netic information in an ES cell is capable of being transmitted in a stable vertical fashion through the resultant offspring and becomes part of the gene pool.

For an effective targeting event to occur at the desired site in the ES cells, homologous recombination must occur between the murine sequences present in the targeting vector and the endogenous locus. Unfortunately, for reasons that are not well understood, homologous recombination in these cells occurs at only a small fraction (on the average, only 1 in 1000 recombination events) of the more frequent nonhomologous events. In most cases, the input DNA is randomly inserted at a site other than the target gene. Thus, it is necessary to be able to detect the desired infrequent homologous recombination from the high background of nonhomologous events.

**Early Gene-Targeting Experiments**

Initial targeting experiments focused on the gene hprt, which encodes the selectable marker hypoxanthine phosphoribosyl transferase. The gene is present on the X chromosome, so that only one copy is present in a male-derived ES cell line and is subject to both positive and negative selection procedures. Two different types of experiments were done. First, targeting vectors in which the DNA was inserted or part of the endogenous locus was replaced were used. In each case, the locus was disrupted, and the resultant hprt-targeted cells could be selected directly by growth in the base analogue 6-thioguanine, which is lethal for hprt+ cells. The second type of experiment used an ES cell line that carried a mutated inactive hprt. A "correcting" DNA vector was electroporated into these cells. Homologous recombination between the electroporated sequences and the defective gene resulted in hprt+ cells, and the rare corrective event could be selected by growth in HAT medium. Thompson et al. subsequently demonstrated that the corrected gene could be introduced into the mouse germ line. These and other initial experiments demonstrated the feasibility of the general approach; if one could select for the targeting event, it was feasible to repair, ablaze, or mutate a particular site in the murine genome.

**Targeted Ablation of Nonselectable Loci**

Of course, most of the genes in whose function we are interested are not directly selectable, and a second series of gene-targeting experiments addressed this limitation by disrupting nonselectable loci with insertion/ replacement vectors containing a selectable marker. The gene encoding neomycin phosphotransferase, which renders a cell resistant to the antibiotic G-418 (neo+), is often used (Fig 1). If the ES cell does not express the particular gene that is being targeted, neo transcription needs to be driven by a strong promoter present in the targeting vector's sequences, and an initial selection for G-418-resistant colonies should be performed. If the gene being targeted is expressed in the ES cell line, this can be used to enhance the selection for the correct targeting event: the design of the targeting vector is modified such that a promoterless neo is used. Homologous recombination results in neo transcription being driven from the targeted gene's promoter, and a significant enrichment for the desired targeting event in the G-418-resistant colonies is sometimes observed.

Many variations on the basic theme have already been tried, and although a detailed discussion of replacement/insertion vector design is beyond the scope of this review, some characteristics of an efficient targeting vector are worth noting. First, targeting efficiency decreases dramatically if the region of homology falls below a certain length. The numbers differ significantly in both a locus- and construct-dependent manner, but it is clear that the targeting vector should contain 3000 to 10,000 bp of uninterrupted homologous sequence. Isogenic DNA (eg, DNA isolated from a 129sv mouse genomic library) is preferable to minimize any base pair mismatches that might occur as a result of polymorphic variation between different mouse strains.

The actual mechanics of forcing the ES cell to take up the exogenous DNA are fairly straightforward. The targeting vector is electroporated into the ES cells by capacitors; transfection efficiencies usually range between 1 in 104 to 105. Electroporation, in contrast to traditional procedures of eukaryotic DNA transfection, such as precipitation of the nucleic acid with calcium chloride, usually results in a single copy of the DNA being incorporated into the genome at a single site—an important consideration for successful gene-targeting strategies. The transfectants are initially selected by a 10- to 14-day growth period in the presence of G-418. The neo+ cells will form ES colonies on the fibroblast feeder layers. However, most of these will contain neo sequences that have been randomly inserted into the genome, and it is necessary to screen either pools of colonies or colonies derived from individual ES cells for the correct targeting event. Normally, this is done using the polymerase chain reaction. Alternatively, genomic Southern blots analyses carried out on diagnostic restriction endonuclease digestions, which are able to distinguish the targeted locus from the unmodified gene or the randomly inserted targeting sequences, are performed. Detection of a targeted cell can, depending on the targeting frequency (which differs radically between different constructs for the same locus or between different loci), involve the screening of hundreds or thousands of clones; what constitutes a targetable site or the efficiency with which a site can be targeted is still undefined.

The ability to detect the desired targeting event depends on both the efficiency of the screening procedures and the power of the selection methodologies. An enrichment procedure termed positive-negative selection significantly enhances the selection process and the chances of finding the rare targeted cell. This procedure uses a standard selection, such as G-418 resistance, but the targeting construct is designed such that random integration events produce a cell that can be subjected to negative selection. Normally, this is done by placing, in the targeting vector, the herpes simplex virus thymidine kinase gene (HSV-tk) to one side of the sequences that will undergo homologous recombination at the target locus. If the targeting vector undergoes nonhomologous recombination, HSV-tk+ cells are killed. Cells in which a replacement homologous recom-
bination event has occurred between the target vector's sequences and the cognate locus will be HSV-tk'; these will be resistant to the nucleoside analogue treatment. This procedure can result in an enrichment of 10-fold to several thousand-fold for the correct targeting event being present in the G-418-resistant pool.

**Introduction of Subtle Mutations Into the Mouse Genome**

All of the above experiments focused on repairing an already defective gene or creating a null allele by insertion into (or replacement of) critical exon sequences. The ability to create null mutations at a particular locus in the mammalian genome should prove immensely powerful in delineating gene function. Nevertheless, it is also important to be able to produce more subtle modifications at a defined site, in which only a few nucleotides may be modified. Techniques to make such changes within translated, untranslated, transcribed, or nontranscribed regions have been developed recently by a number of groups. These strategies, termed "in-out" or "hit and run," involve a two-step targeting design. The entire targeting vector carrying the desired mutation, as well as suitable selective markers (e.g., neo' and HSV-tk), is first inserted into the gene via a single reciprocal recombination event. This produces a partial duplication of the genomic material at that site (similar to the example shown for an insertion vector in Fig 1). In some cases, one of the duplicated copies retains the desired mutation, and these cells are identified as outlined above. The second event involves no exogenous DNA but rather relies on the rare "out" event of intrachromosomal recombination within the duplication. This event can be selected by drug resistance or sensitivity (e.g., neo' and FIAU'), and the presence of the mutation in these cells is confirmed by diagnostic Southern or polymerase chain reaction analyses. Although these techniques have not been widely used to date, their feasibility has been clearly demonstrated, and it should be possible to produce animal models that contain an exact genetic lesion at a chosen site.

**Production of Chimeric Animals**

After the ES cell has been targeted correctly, it must be introduced into a host embryo and be used to generate the germ line. Usually, 10 to 20 heterozygous ES cells are microinjected into blastocysts that have been removed from C57BL/6 mice at approximately 4 days after coitus. The blastocysts are subsequently implanted into the uterus of a pseudopregnant female, and the chimeric offspring are visually identified by coat color (Fig 2). The ES cell-derived coat hairs are easily identified; the 129Sv mouse from which the cells are derived is an agouti (tan). The host blastocyst-derived coat hairs are solid black, and the agouti patches can be clearly detected. To determine if the targeted ES cells have colonized the germ line, the male chimeras are back-crossed to C57BL/6 females. Because agouti is dominant, if any offspring have an agouti coat, germ-line transmission has been obtained. The DNAs of the offspring are then checked to determine which allele (the targeted or nontargeted copy) is present. It is possible to knock out both alleles at the ES cell level and generate the homozygous animal directly. Normally, however, the heterozygote cell is injected, and the mice carrying the desired targeted locus are then bred to homozygosity.

**Knockout of Genes Relevant to the Cardiovascular System**

Null mutations have now been created in a large number of genes. However, only a few are directly relevant to the function of the cardiovascular system. N-myc, which encodes a nuclear phosphoprotein thought to function in transcriptional control, has been ablated; the homozygotes die between 10.5 and 12.5 days of gestation. Heart development is severely affected in these animals, and the 11.5-day heart resembles that of a normal 9-day animal with the characteristic S-shaped morphology. Although the targeting of the muscle creatine kinase and β-cardiac myosin heavy chain genes has been reported in ES cells, the resultant animals remain under development. To date, no mice have been reported in which the genes that encode the major structural proteins of the cardiac contractile apparatus have been ablated or modified. However, knockouts in a number of genes whose function is presumed to be critical to normal skeletal
muscle development have been made, with some unexpected consequences.

The molecular events that underlie development of striated muscle have been the subject of intense investigation. The muscle-specific developmental program and maintenance of the differentiated phenotype is at least partially mediated through the action of a small family of transcriptional activators whose members share a high degree of homology in a basic helix-loop-helix domain with one another. *MyoD* encodes a protein that is able to convert an undifferentiated mesodermal cell to the skeletal muscle lineage, and it and other members of the family, which include *Myogenin*, *Myf-6*, and *Myf-5*, can all activate skeletal muscle–specific gene transcription.²⁴

Because of experiments in which transfection with *MyoD* was necessary and sufficient to prompt a nonmuscle cell to undergo the skeletal developmental program, it was assumed that MyoD played a critical even vital role in skeletal muscle development as a “master regulator.” However, when Rudnicki et al.²⁵ generated animals that were homozygous for a null mutation in *MyoD*, muscle development, as examined by histological, immunohistological, and transmission electron microscopy, revealed no obvious defects in fiber organization, number, or type. The *MyoD* null mice were viable and fertile. Examination of the transcriptional patterns of the other myogenic proteins revealed upregulation of *Myf-5*; the steady-state transcript levels were increased 1.8-fold in the *MyoD*– null myotubes and 3.5-fold in the homozygous nulls. The other myogenic gene transcript levels tested, which included myogenin and *myf-6*, were unaffected.

*MyoD* and *Myf-5* encode closely related myogenic proteins, and unlike *Myogenin* and *Myf-6*, both are expressed in undifferentiated myoblasts. Thus, Rudnicki et al.²⁵ speculated that a compensatory upregulation of *Myf-5* might reflect a functional redundancy of the two myogenic proteins. Although *myf-5* may be able to compensate for the absence of MyoD, the two proteins do not function in an identical manner. When a targeted *Myf-5* mouse was produced by the same group, somitic muscle development was slightly slowed, but final development, form, and function appeared to be essentially normal.²⁶ However, the mutation was perinatally lethal because of the absence of the distal ribs. The levels of MyoD, myogenin, and *myf-6* transcripts were unaffected in these animals. Thus, although the exact mechanistic nature of the defect remains undefined at this point, these mice illustrate a new and unexpected role for *myf-5* in the formation of a functional rib cage.

The fact that two proteins that are able to determine the formation of a skeletal muscle lineage (the double *Myf-5*/*MyoD* mutant has not yet been reported) appear to be dispensable for normal skeletal muscle development is surprising and indicates the presence of redundant possibly parallel regulatory circuits for skeletal muscle development. Presently, the regulatory paradigm(s) responsible for cardiac muscle formation and terminal differentiation remains obscure, but these processes may depend on proteins that are related to the skeletal myogenic species. The production of other targeted mice for the myogenic genes and crossbreeding, when possible, between the lines so that double homozygotes or double heterozygotes can be analyzed promise to shed light on the role(s), if any, that the helix-loop-helix proteins play in cardiac development. For example, even though nucleic analyses have been unable to detect the presence of the helix-loop-helix myogenic transcripts in developing cardiac tissue, this does not preclude the possibility that a transient expression of one or more of these genes in a limited progenitor cell population plays a critical role. The production of targeted mice in which early cardiac tube formation is modified or ablated could be immensely valuable in delineating a role for these proteins. Alternatively, cardiac development may depend on a completely different set of factors; normal heart development in the targeted lines would confirm such a hypothesis. Certainly, the above targeting experiments indicate the power of creating a targeted mutation to define the importance and site(s) of action of a gene product.

Dissection of Multifactorial Genetic Predisposition to Cardiovascular Disease

Gene-targeting experiments should significantly enhance our ability to dissect the mechanistic components that underlie those cardiovascular diseases whose genetic bases are complex and multifactorial. For example, the development of atherosclerosis, a major cause of death in the United States and other western developed countries, depends on both environmental and genetic factors. The action of a number of proteins that are involved in lipid metabolism is being explored by a combination of transgenic and gene-targeting approaches. Two groups have now reported a knockout of ApoE,²⁷,²⁸ the gene encoding one of the 10 apolipoproteins. Apolipoprotein E is a structural component of the very low, intermediate, and high density lipoprotein particles. It is expressed in the liver in both mice and humans and plays a critical role in the removal of lipoproteins from the plasma. Although population studies have pointed to the involvement of the gene in a predisposition to the development of atherosclerosis,²⁹ its particular role in the disease's onset is not clearcut. Patients homozygous for an allele (ApoE³/ApoE³) show a >95% reduction in LDL receptor binding, a function critical for apolipoprotein E's action in clearing the intermediate density and very low density lipid particles. However, most people are unaffected, with only 1 in 50 developing type III hyperlipoproteinemia. Interestingly, although 129Av mice are resistant to diet-induced atherosclerosis, the ablation of ApoE resulted in animals that spontaneously developed atherosclerosis when fed either a low fat, normal, or high fat diet. Blood chemistries and histological analyses of the aorta and arteries confirmed the development of atherosclerotic lesions.³⁰ This mouse should be valuable in studying the role of apolipoprotein E in the development of the disease and the exact effects that different dietary and drug regimens have on the disease's progression. Undoubtedly, other knockouts for the related lipoproteins will soon follow, and if the animals are viable (as is the case for the *ApoE* knockout), analyses of the double heterozygotes and homozygotes should prove useful in elucidating the multifactorial basis of the genetic predisposition to this disease.
Future Prospects

Although the number of targeted mice reported is growing exponentially, improvements in the efficiency of gene targeting must occur before the techniques become widely used. What underlies the efficiency with which a particular locus can be targeted? A few of the basic parameters have been established: length of homologous sequence and the use of isogenic DNA. Yet, numerous other parameters need to be optimized to improve the woefully low efficiency with which the process, on the average, occurs. In this respect it might be of some value to carefully dissect, via a systematic mutational analysis, a locus that has been shown to be targetable at high efficiency and to determine the parameters that underlie the "hot spot." Although numerous clever approaches for construct design and targeted cell selection have been proposed, and in some cases used, the generation of chimeric animals via germ-line transmission remains a long process, normally taking at least a year or more. A formidable battery of sophisticated techniques that necessitates a complex support infrastructure must be set in place for these experiments to work.

The molecular genetic aspects are normally straightforward: to begin, it is necessary to isolate the gene of interest from an isogenic gene library. A large number of genes whose products are involved in cardiovascular function and development are now available, although only a few have been isolated from the murine 129Sv gene libraries. However, this can be accomplished quickly, and a suitable construct can be designed. A successful electroporation and isolation of a targeted ES cell capable of colonizing the germ line requires, at this point, considerable skill and a significant investment in both time and money. Taken together with the blastocyst/embryo manipulations that are necessary and the murine genetic infrastructure that is needed, it is not surprising that the amount of time and effort it takes before a single animal can be generated is substantial.

After such an expenditure of energy, it is incumbent on the investigators to devote a significant amount of time to a rigorous analysis of the resultant phenotype, at the molecular, cellular, and systemic levels. The mouse has never been an animal model of choice for physiological studies because of the difficulties inherent in its size. However, our ability to modulate cardiac function in a defined manner and to carry out sophisticated physiological measurements on the murine cardiovascular system is improving rapidly. Many of the methods used to perturb and measure cardiac performance in larger animals can now be applied to the murine heart. These include the development of models of pressure-induced hypertrophy and the ability to study contractile performance in isolated Langendorff and working heart preparations. As physiologists focus on extending whole-animal physiology to the murine system, the perceived limitations of dealing with the mouse should be largely overcome.

However valuable the targeted murine models prove to be, a mouse is not a human nor, probably in a number of cases, the most appropriate animal model. Efforts are under way in a number of laboratories to develop ES cells capable of germ line transmission for the hamster, rat, guinea pig, and rabbit. Although no successes have yet been reported, it may be anticipated that the methodological and technological problems are solvable. This should lead to the ability to create targeted lines in these or other species. The ability to produce a stable animal model carrying a defined genetic defect that underlies a human cardiovascular disease should prove invaluable in understanding the underlying etiology, phenotypic consequences, and developing pathology. Ablating a particular gene and studying the phenotypic consequences at the molecular, structural, biochemical, and physiological levels will help delineate the encoded protein's exact functions and interactions at the cellular and systemic levels. It is hoped that such techniques may eventually be applied to developing innovative gene therapies that will be able to significantly halt or even prevent some of the major cardiovascular diseases.

Acknowledgments

This study was supported by grants HL-46826, HL-22619, and HL-41496 from the National Institutes of Health. I thank Dr. R. Walsh for his suggestions and a critical reading of the manuscript.

References


Gene targeting. The precise manipulation of the mammalian genome.

J Robbins

Circ Res. 1993;73:3-9
doi: 10.1161/01.RES.73.1.3

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/73/1/3

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/