March 1999 (Volume 40, Number 3)

**Manipulating Mammalian Genome by Gene Targeting**

*Sinisa Volarevicæ, Mario Pende, Nicholas Pullen*

Department of Growth Control, Friedrich Miescher Institute, Basel, Switzerland

The development of strategies which allow the inactivation of specific murine genes by homologous recombination in embryonic cells has revolutionized biological science in the last 10 years. A large number of mice carrying genetic lesions, generated by gene targeting technology, has tremendously increased our knowledge in many areas of biology, culminating in the identification of mouse models for human genetic disorders. These findings have been recently complemented by “conditional” gene targeting technology, allowing gene inactivation in a defined tissue and at a specific time point during development or adulthood, thereby extending the sophistication and potential of this technology.

Key words: base sequence; DNA, recombinant; genetic engineering; genetic library; genetic techniques; mice, knockout; mice, transgenic; recombination, genetic; vectors, genetic

According to last predictions, the entire human genome should be sequenced by the year 2003 (1). This task is being achieved with the combined efforts of public and private research institutions, pharmaceutical companies, and governments around the world. In addition to the human genome project, genome sequences of many other species are becoming increasingly available, thus providing an important additional input in predicting and understanding the function of all human genes.

Recently, remarkable advances have also been made towards the next exciting step in science, understanding the function of all human genes (2). Biotechnology and bioinformatics, in combination with classical biochemistry and genetics, are the tools now being used to dissect the flood of genomic sequence information. With a human sequence in hand, it is possible, by gene targeting, to generate a mutation in the homologous mouse gene and determine the physiological consequences of this mutation. Since many human diseases have a genetic basis, mouse models generated using this strategy should greatly increase our understanding of the pathophysiology of diseases and lead to the design of new therapies.

Gene targeting is a process by which desired changes are introduced into the nucleotide sequence of a chosen gene (3). The mouse is currently the animal model of choice for gene targeting (4). In addition to evolutionary similarity with humans, mouse physiology, embryology, and genetics are well studied and understood. Human and mouse genomes are approximately of the same size, containing around 100,000 genes that are similarly ordered in the chromosomes. Homologous proteins from human and mouse typically share over 90% identity in their amino acid sequence. However, the most important reason for using the mouse as an animal model for gene targeting lies in the possibility of isolating embryonic stem cells, in which any gene can be modified and then transferred through the germ line to generate novel mutant mice. The scope of this text is not to comprehensively review gene targeting technology but rather to discuss its current status and future prospects and, where appropriate, provide specific examples to illustrate its successful application.

**Classical Gene Targeting Strategies**

Mammalian cells integrate foreign DNA by either a process of non-homologous or homologous recombination. The ability of mammalian cells to mediate recombination between homologous DNA sequences is the basis for gene targeting technology (5). The entire genome can be potentially manipulated, such that intron sequences or gene promoters or enhancers are equally effective and useful targets as exon sequences.

![Figure 1](general scheme for classical gene targeting. The exon-intron structure of the germline locus is indicated with empty boxes representing the coding regions. The black box indicates the region of the exon to be disrupted following homologous recombination with the targeting vector. The targeting construct, containing the exon to be disrupted with a neomycin gene (neo), as well as the endogenous gene segments and the selectable thymidine kinase gene (TK) is indicated. A hypothetical Southern blot analysis of a BamH1 digested genomic DNA before and after homologous recombination using an intron specific probe is shown in right panel.)
anterior commissure, resulting in a failure of neurons to link the temporal lobes (7). However, Nuk is a modification. Null Nuk mice displayed aberrant migration of axons forming the posterior member of the Eph receptor tyrosine kinase family, which had been studied by deletion and insights into the

There are many examples where the classical gene knockout approach has yielded extremely useful knowledge of the functions of particular gene products during development and in progression to adulthood.

<table>
<thead>
<tr>
<th>Example</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuk</td>
<td>Null Nuk mice displayed aberrant migration of axons forming the posterior tract of anterior commissure, resulting in a failure of neurons to link the temporal lobes (7). However, Nuk is a modified gene after homologous recombination are indicated.</td>
</tr>
</tbody>
</table>

The first step in the targeting strategy begins by generating a targeting vector, containing the desired gene mutation and homologous flanking sequences (Fig. 1). This is introduced into embryonic stem cells by electroporation. In most cells the targeting vector inserts randomly into embryonic stem cell genome. However, in a few cells homologous DNA sequences in the targeting vector pair with homologous sequences in the embryonic stem chromosomal DNA and recombine, introducing the mutation to the genome. If such gene manipulation is designed to prevent expression of the targeted gene it is called a null mutation or complete knockout (Fig. 1). Other modifications, such as point mutations or deletions can be introduced to study different domains and subdomains of the gene product in situ (Fig. 2). Two selectable markers are commonly introduced into the targeting vector to allow identification of embryonic stem cell clones that have undergone successful homologous recombination. A positive selectable marker promotes survival of cells that have integrated the targeting vector either by homologous or non-homologous recombination. A negative selectable marker is used to enrich cells that have integrated the targeting vector by homologous recombination and to eliminate those that have incorporated the targeting vector in a non-homologous manner. The positive selectable marker, typically a neomycin-resistance gene, is flanked by the DNA sequence of the gene to be targeted. The negative selection marker, usually a thymidine kinase gene cassette, is cloned at the end of the targeting vector. When homologous recombination occurs, regions containing the targeting gene sequence, together with the neomycin cassette, replace the corresponding nucleotide sequence in the chromosome. Since the thymidine kinase gene resides outside of the region of homology, it does not integrate into the chromosome and is ultimately degraded. However, if the targeting vector sequence is integrated non-specifically into the chromosome by non-homologous recombination, the whole vector, including the thymidine kinase gene, can be inserted. Treating targeted embryonic stem cells with neomycin kills those that have failed to integrate the targeting vector, and treatment with gancyclovir kills those that have retained the thymidine kinase gene. In this way, it is possible to select cells bearing targeting sequences inserted into the chromosome by homologous recombination. The proper integration of targeting sequences is confirmed by polymerase chain reaction (PCR) or by genomic Southern blotting and, then, selected embryonic stem cells are expanded. Expanded clones are injected into the blastocyst cavity of a pre-implantation mouse embryo and the blastocyst is then transferred into the uterus of a pseudo-pregnant mother. The resulting animal is a chimera, consisting of cells derived from the targeted donor embryonic stem cells and host blastocyst. Embryonic stem cells are usually derived from a mouse homozygous for one coat color allele (white, for example), which is different from the coat color allele of the mouse which provides the recipient blastocyst cells (black, for example). Thus, the coat of the chimeric mouse has patches of both colors. Breeding chimeric mice with wild type mice (black) will yield some heterozygous mutant mice if the targeted embryonic stem cells contributed to germline cells of a chimeric mouse. Genomic screening is then used to identify mice with the correctly targeted allele. Interbreeding of heterozygous mice then yields homozygous animals for the designed genetic lesion (5).

The total number of different mice generated by gene targeting was estimated to be 2,000 in 1998, and this number is rapidly increasing due to a widespread usage of this technology in both research institutions and industry (6). In most cases, the primary goal has been to generate a null allele of the gene of interest and, with the availability of these mice (see also Table 1), has come greater knowledge of the functions of particular gene products during development and in progression to adulthood.

Table 1. Internet addresses of publicly available databases of genetically modified mice

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse Genome Informatics (MGI)</td>
<td><a href="http://www.informatics.jax.org">www.informatics.jax.org</a></td>
</tr>
<tr>
<td>The Jackson Laboratory (JAX)</td>
<td><a href="http://www.jax.org">www.jax.org</a></td>
</tr>
<tr>
<td>The Mouse Genome Database (MGD)</td>
<td><a href="http://www.informatics.jax.org">www.informatics.jax.org</a></td>
</tr>
</tbody>
</table>

Examples of Classical Gene Targeting

There are many examples where the classical gene knockout approach has yielded extremely useful insights into the biological functions of individual genes or gene families. One example is Nuk, a member of the Eph receptor tyrosine kinase family, which had been studied by deletion and modification. Null Nuk mice displayed aberrant migration of axons forming the posterior tract of anterior commissure, resulting in a failure of neurons to link the temporal lobes (7). However, Nuk is a
membrane protein, with both intracellular and extracellular domains. To determine the importance of different domains of Nuk in axon migration, a targeting vector was generated in which a LacZ coding sequence was fused in frame with the Nuk extracellular transmembrane domain and juxtamembrane domains (Fig. 2). Even though this chimeric Nuk receptor lacked the intracellular kinase domain, these mice were not defective in axonal pathfinding. Therefore, these two approaches (Figs. 1 and 2) demonstrated that the extracellular and not the intracellular domain of Nuk is required for the pathfinding of temporal cortical axons.

Point mutations can also be introduced by classical gene targeting technology (Fig. 2). For example, Torres et al (8) introduced a point mutation into the α-chain of the immunoglobulin receptor, resulting in a stop codon immediately after the transmembrane region. The resultant mouse mutant exhibited only a small defect in early B lymphocyte development, but a strong impairment in the generation and function of mature peripheral B lymphocytes.

Mouse Models of Human Diseases

Increasingly, examples of animal models for human diseases, such as diabetes (9), cancer (10, 11), and cardiomyopathy (12), are being identified by gene targeting. The deletion of the insulin receptor substrate-2 gene, IRS-2, yielded mice that are diabetic due to increased insulin resistance in peripheral tissues, such as muscle and fat, as the pancreas fails to compensate this defect with increased insulin production (9). Disruption of the p27kip1 gene (10), whose product is an inhibitor of the cyclin-dependent kinases-2, -4, and -6 which regulate the timely progression of the cell cycle, gave rise to mice which displayed increased body size, as a result of organ hyperplasia, and development of pituitary tumors. This result emphasised a role for p27kip1 in tumor suppression, and consistent with this was the finding that p27kip1+/- mice were also pre-sensitized to tumor development, when challenged with irradiation or chemical mutagens (11). Similarly, deletion of MLP, the LIM-only protein that specifically localizes to differentiated striated muscle, generated mice that develop dilated cardiomyopathy and heart failure, producing the first animal model for this condition (12).

Conditional Gene Targeting

Despite the potential of obtaining valuable information on the function of many genes by deleting them from the genome, the underlying genetic program determining the biological processes in mammals is extremely complex. Simple inactivation of genes in embryonic stem cells often only uncovers the essential biological functions of that gene product (13). For some genes, early embryonic lethality prevents researchers from studying their role at later stages of development and into the adult. In other instances, gene products have functions in many different cell types, which makes it very difficult to dissect the cell autonomous role from any humoral role. Another complication is that many mammalian genes have duplicated during evolution, leading to functional redundancy. For example, the existence of insulin receptor substrate-2 (IRS-2) was predicted, and later confirmed, following the phenotype observed in IRS-1 null mice (14). Therefore, a deficient animal may compensate for the loss of one gene product by the over-expression of a homologue, thus limiting pheno-typic characterization. Deletion of a gene in embryonic stem cells has also a disadvantage in exploring the genetic changes that underlie the somatic development of human disease, such as the onset of cancer (6). Additionally, the presence of a positive selection marker in the targeted gene can influence the expression of the gene of interest, resulting in hypomorphic phenotypes of the targeted versus the wild type mice (see below).

However, the development of new tools for tissue specific and timely genetic manipulation, complemented by the imagination of many researchers, is helping to circumvent some of the problems outlined above. A major breakthrough in this field has been the development of "conditional gene targeting", a methodology which allows genome alteration in spatial (tissue or cell specific) and temporal manners (6). Conditional gene targeting uses a combination of gene targeting and site-specific recombination. Site-specific recombinases are enzymes that recognize specific DNA sequences and recombine the intervening DNA sequence between them. The Cre recombinase (causes recombination) of the P1 bacteriophage and Flp recombinase (flipase) of yeast are the most frequently used. Cre and Flp recombinases recognize 34 bp sequences, called loxP and frt, respectively (15). If two recombinase recognition sequences are placed in the same DNA strand in the same orientation a recombination event will result in deletion of the sequence between them (Fig. 3). If the recombinase recognition sequences are placed in the opposite orientation, a recombination event will invert the intervening sequence. The strategy of conditional gene deletion involves the introduction of two recombinase recognition sites within an intron sequence such that they are flanking (or "floxed") an essential part of the gene without affecting its transcription, processing or function. The targeting gene is then integrated into embryonic stem cells by homologous recombination, and normal heterozygous and homozygous mice carrying these modifications are
generated. Gene deletions in vivo can be made conditional for a certain tissue or for a developmental
time point by crossing these mice to transgenic mouse strains expressing Cre or Flp recombinases
under the control of a tissue specific or inducible promoter (Fig. 3).

Figure 3. General scheme for conditional gene targeting. The structure of the germline locus to be
deleted and the targeting construct, containing two loxP sites (filled triangles) flanking the region that
is later to be deleted, as well as the neomycin (neo) and thymidine kinase selectable genes within
non-coding sequences are indicated. By choosing an appropriate probe sequence it is possible to
detect progeny containing the wild type locus from those with the targeted allele, from those with the
deleted allele following Cre-mediated deletion. Controlling Cre expression with appropriate tissue-
specific or ubiquitously expressing promotors allows the possibility of generating progeny with pre-
determined lesions.

Examples of Spatial Gene Alteration
A number of promoters, such as lck, CaMKII, and insulin, which drive specific Cre expression in T
cells, brain, and b-cells, respectively, have been used with some success to bring about tissue
specific gene deletion (16-18). For example, the ubiquitous deletion of DNA polymerase b in mice is
embryonically lethal. Rajewsky and his colleagues deleted the gene in T cells using Cre recombinase
under the control of the T cell-specific lck promoter and tested its role in T cell antigen receptor re-
arrangement (16). The deletion of DNA polymerase b gene in pre-T cells did not prevent the
generation of mature T cells, indicating that DNA polymerase b is not absolutely required throughout
T cell development. Another example of tissue specific gene deletion was used to determine what role
insulin participates in signaling in pancreatic b-cells. The insulin receptor gene was specifically
deleted from b-cells by placing the Cre recombinase under the control of insulin promoter (18). The
resultant mice showed a loss of insulin secretion in response to glucose and a progressive
impairment in glucose tolerance. These observations support the notion that a defect in insulin
signaling in b-cells may be responsible for the changes in insulin secretion seen in type-2 diabetes
(19).

The power of the loxP/Cre system is that, in addition to tissue specific gene deletions, Cre can be
placed under control of a promoter that drives gene transcription in all tissues, such as the viral EIIA
and CMV promotors. Crossing “floxed” mice with transgenic Cre mice (deleter mice) can then be used
to remove the positive selection marker or to generate a complete gene deletion similar to that
obtained by classical gene targeting (20,21).

Temporal Gene Alteration
Transgenic mice which express Cre under the control of inducible promotors, such as those activated
by interferon-a (INF-a) or tetracycline, have additionally allowed gene product function to be
derdifferentiated in a temporal manner (22). For example, to test what role the B-cell receptor plays in the
physiology of mature B cells, Rajewsky’s group deleted a variable region of the B-cell receptor in adult
mice using the Cre recombinase under control of the INF-a-inducible Mx promoter (23). The classical
gene targeting approach could not be used to study the role of the B-cell receptor in mature B cells,
since it is absolutely required for B-cell maturation. INF-a-induced deletion of B-cell receptor led to a
rapid B-cell death, demonstrating its critical role in providing a survival signal to mature B-cells.
Ultimately, systems that induce recombinase expression in both temporally and spatially controlled
manneres would greatly aid understanding of more complex gene functions. One such example is to
infect tissue locally with adenoviruses carrying the specific recombinase gene. Noda and his co-
workers “conditionally” deleted a gene responsible for familiar adenomatous polyposis (APC gene) by
infecting the colorectal region with an adenovirus carrying the Cre recombinase gene (24). These
mice developed adenomas within four weeks after infection, thus providing a useful model for the
study of the initiation phase of colon cancer.

A la carte Manipulation and the Generation of Mutant Allelic Series
Another important breakthrough in the use of gene targeting has been the generation of mutant allelic
series (Fig. 4). As already mentioned, gene function can be inferred from the analysis of mutations
which lead to the complete disruption of the gene. However, it has been demonstrated that in some
cases, a neomycin selection marker, if left in the intron of a targeted gene can, create a hypomorphic
allele (25). If the neomycin resistance gene is flanked with frt sequences, it can be excised by
crossing these mice with transgenic mice expressing the FIP recombinase under control of a
ubiquitous promoter, such as b-actin. This permits the direct comparison of the phenotypes of the
targeted and the deleted mice progeny and is presently the preferred method of gene targeting
strategy in use. However, analysis of mice carrying a hypomorphic alleles can be very useful in cases
of early embryonic lethality, which prevents study of later functions, or for the study of somatic-based
diseases where a phenotype becomes generated as a consequence of a gradual deterioration in gene function. Since the critical exons in the targeted gene can be flanked by loxP sites, a null allele could be generated by crossing these mice with a transgenic strain ubiquitously expressing Cre recombinase (Fig. 4).

**Figure 4.** Generation of a mutant allelic series using homologous recombination. The structure of the germline gene to be manipulated and targeting construct, containing a neomycin gene (neo) flanked with loxP sequences (filled triangles) within the intron of the gene, as well as two sequences (empty triangles) flanking two critical exons and a thymidine kinase selectable gene are indicated. Homologous recombination leads to the generation of a targeted, hypomorphic locus, which either re-generates a wild type allele by Cre-specific deletion, or creates a null allele by using flipase.

An additional advantage of this system is possible by generating so-called compound heterozygotes that carry both a hypomorphic and a null allele by appropriate crossing. Compound heterozygotes typically have a more severe phenotype than hypomorphic animals. In an exciting study, Meyers, Lewandoski and Martin demonstrated multiple functions for fibroblast growth factor-8, Fgf8, using this strategy (26). By generating mice with sequences of the Fgf8 gene flanked with loxP sites and crossing these mice with Cre-expressing deleter mice it was possible to generate a complete Fgf8 knockout which died in utero at 10.5 day-stage (E10.5), due to a defect in gastrulation. Homozygous mice harboring a hypomorphic allele died within a day of birth. No major morphological changes in the main body axis were detected, but a lack of olfactory tissue, deletion of posterior midbrain and cerebellar tissue was apparent at E18.5. Furthermore, analysis of heterozygous compound animals at E9.5 revealed an additional role of Fgf8 in heart and craniofacial development. Clearly, the development of transgenic animals or systems that can induce the expression of specific recombinases with greater spatial and temporal selectivity will be paramount in the elucidation of complex systems, of which Fgf8 provides a stark example.

Insertion of DNA Sequences into the Genome by Cre Recombinase

The Cre recombinase has the potential to insert a non-homologous DNA sequence into the genome at a defined site, thus potentially re-inforcing or "knocking-in" new or modified genetic material (27). However, this event is inefficient because the newly inserted sequence is flanked by two loxP sites and can be re-excised by the Cre recombinase. Such problems can be resolved if the targeting and targeted DNA sequences contain both wild type and mutated loxP sequences. In this case mutant loxP sites only recombine with mutant loxP sites and wild type loxP with wild type loxP sites. This insertion event is unable to re-excite because mutated loxP sites do not recombine wild type loxP sites. Using this method it is possible to "knock-in" any mutation of the gene into wild type locus. This exciting possibility could then be used to place one gene under control of another gene promoter. Site specific recombination, in addition to the described examples, has been used to generate point mutations, deletions, and activation of inactive transgenes and "knock-ins" when the systems are carefully designed (28). The Cre recombinase can be also used to make chromosomal alteration, such as deletions, inversions, duplications and inter- as well as intra-chromosomal rearrangements (29,30). These methods could allow generation of mouse models for human diseases that are often associated with specific chromosomal translocations and alterations.

Future Perspectives

The generation of knockout mice is a very demanding and labor-intensive exercise. For these reasons, researches are now trying to optimize genetic screens to obtain mutant mice similar to the ones that have been successfully used over last decades in Drosophila melanogaster, C. elegans, and zebrafish (31). Gene trapping vectors or chemical agents are used to carry out systematic mutagenesis of the embryonic stem cell genome. Recently, a genomics company, Lexicon Genetics, has reported the random disruption of 2,000 genes in mouse embryonic stem cells (32). It is not clear that all the mutations result in null alleles and the level of gene expression of interest needs to be determined before using the embryonic stem cells to generate mutant mice. However, this is clearly an initiative that the scientific community and the pharmaceutical industry will potentially benefit from in the future.

Further improvement of efficiency and specificity of promoters driving Cre and Flp recombinases, as well as characterization of suitable transgenic lines will improve this powerful technology which at this time has some limitations. It is critical that well characterized and efficient transgenic strains expressing recombinases are freely distributed to prevent unnecessary duplication of the effort.

The large amount of information generated from mouse genetics requires the development of updated and comprehensive databases that should be easily accessed by everybody. Table 1 lists some, which are already available over the Internet. Many are maintained by private companies or need to
be updated by whoever generates a mutant strain, but, clearly, central institutions that collect and organize the data, as well as maintain and distribute mouse strains, need to be rapidly established. With the availability of all these tools and those currently under development, our understanding of mammalian gene function will dramatically increase in the next years. Hopefully, this knowledge will be increasingly exploited to understand the pathology and progression of human diseases and to discover new drugs, which will act at a defined molecular level. Animal models of human disease can be also useful for testing of new drugs. The development of the gene targeting technology in somatic human cells would be a huge step forward in the treatment of genetic human diseases (33).

Acknowledgments
Nicholas Pullen is a recipient of HFSPO Fellowship. Siniša Volareviæ and Mario Pende are supported by a grant from EEC.

References
9 Withers DJ, Gutierrez JS, Towery H, Burks DJ, Ren JM, Previs S et al. Disruption of IRS-2 causes type 2 diabetes in mice. Nature 1998;391:900-3.

Received: May 24, 1999
Accepted: June 6, 1999

Correspondence to:
Siniša Volarević
Department of Growth Control
Friedrich Miescher Institute
Maulbeerstrasse 66
Basel, Switzerland
sinisa.volarevic@fmi.ch