Title: COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

Abstract: Disclosed are novel compositions and methods useful for modulating expression of a target gene in a cell by insertion of exogenous DNA sequence into the target gene. The compositions and methods of the invention are useful for generation of knockout animals including mammals.
COMPOSITIONS AND METHODS FOR ALTERING GENE EXPRESSION

TECHNICAL FIELD

The present invention is directed generally to the biological sciences, including recombinant genetics and immunology. More particularly, there are described herein non-human knockout animals, preferably mammals, in which the expression of one or more genes has been altered. Also provided herein are xenograft transplants in which the expression of one or more genes has been modulated to prevent or reduce the likelihood of rejection by the transplant recipient.

BACKGROUND

The immune response of mammals, including humans, against invading pathogens, toxins, and other foreign substances involves many specialized cells that act together. Lymphocytes are a class of white blood cells responsible for the specificity of the immune system. Two important classes of lymphocytes are T cells and B cells. T cells develop in the thymus, and are responsible for cell mediated immunity. There are many types of specialized T cells, such as for example, helper T cells (which enhance the activity of other types of white blood cells), suppressor T cells (which suppress the activity of other white blood cells), and cytotoxic T cells (which kill cells). B cells develop in the bone marrow and exert their effect by producing and secreting antibodies.

A key to the coordinated immune response is complement, which, as described in U.S. Patent No. 5,679,345, is involved in the pathogenesis of tissue injury observed in many immunologically mediated diseases, such as systemic lupus, erythematosis, rheumatoid arthritis,
and immune-hemolytic anemia. Complement is also involved in rejection of transplanted organ grafts. Complement is responsible for much of the tissue injury in transplantation due to inflammatory conditions resulting from rejection or superimposed by infection, ischemia, and thrombosis of vessels in the graft, as well as tissue injury due to inflammation from similar causes in patients who have not received an organ transplant. In particular, complement attack on cells is central to the rapid onset phase of immune mediated graft rejection (hyperacute rejection), where complement activation and subsequent tissue damage occur within hours.

Graft rejection may occur through a number of different mechanisms, with the time course of rejection being characteristic of the particular mechanism. Early rejection (hyperacute rejection), occurring within minutes or hours of transplantation, involves complement activation by components that are present at the time of the transplant operation. Activation may occur via the classical pathway by preformed antibodies that are reactive with the "foreign" or non-self markers of the graft or via the alternative pathway in response to tissue damage in the graft as a result of, for example, ischemic damage to the organ during storage before transplantation.

Acute rejection occurs days to weeks after transplantation, and is caused by sensitization of the host to the foreign tissue that makes up the graft. Once the host's immune system has identified the transplanted tissue as foreign, all the resources of the immune system are marshaled against the graft, including both specific (antibody and T cell-dependent) responses and non-specific (phagocytic and complement-dependent) responses. Chronic rejection will usually only occur when the graft recipient is immune-suppressed. Then the graft may survive long enough for tissue to undergo changes which ultimately affect survival of the graft. Such changes include hyperplasia and tissue hypertrophy, and endothelial cell damage leading to narrowing of the vascular lumen and potentially impairing the oxygen supply of the graft tissue.
Xenograft rejection of pig tissue is triggered by natural human antibodies that recognize carbohydrate xeno-antigens, such as Gal α(1,3) galactose, which is expressed on pig endothelial cells that line blood vessels. Weiss, *Science*, 285(20):1221-1222 (August 20, 1999). U.S. Patent No. 5,821,117 describes inhibiting xenotransplant rejection by disrupting the wild type porcine Gal α(1,3) galactosyl transferase gene with a cloned mutant porcine Gal α(1,3) galactosyl transferase sequence specifically within an exon of the wild type gene. The resultant mutant gene does not encode a functional galactosyl transferase, with the expected result that rejection of the transplanted xenograft by the patient’s immune system is avoided.

In such so called "knockout" mammals, expression of an endogenous gene has been altered (typically, suppressed) through genetic manipulation. Preparation of knockout mammals typically has required introducing into an undifferentiated cell type (termed an embryonic stem cell) a nucleic acid construct to suppress expression of a target gene. This cell is introduced and integrated into a mammalian embryo. The embryo is implanted into a foster mother for the duration of gestation. For example, Pfeffer et al. (Cell, 73:457-467 [1993]) describe mice in which the gene encoding the tumor necrosis factor receptor p55 has been disrupted by mutation utilizing homologous recombination. The mice showed a decreased response to tumor necrosis factor signaling. Fung-Leung et al. (Cell, 65:443-449 [1991]; J. Exp. Med., 174:1425-1429 [1991]) describe knockout mice lacking expression of the gene encoding CD8. These mice were found to have a decreased level of cytotoxic T cell response to various antigens and to certain viral pathogens such as lymphocytic choriomeningitis virus.

Typical prior methods, however, describe manipulation of an exon region of the target gene. There is thus a need in the art for new and improved methods for modulating gene
expression in animals including mammals, particularly for overcoming xenograft transplant rejection. It is to these, as well as other, important ends that the following is addressed.

**SUMMARY**

It has been discovered that the expression of a particular gene in an animal may be modulated by introducing into the genomic DNA of the animal a new DNA sequence that results in the disruption of at least some portion of the DNA sequence of the gene to be modulated. The methods described herein are of general utility for altering gene expression in animals including mammals. In contrast to prior methods, it has surprisingly been found that gene expression may be suppressed in part or in total by inserting new DNA sequence into the intron of the target genomic DNA.

The versatility of the methods described herein for generating "knockout" animals is illustrated by the following general description of a preferred embodiments, including the examples. It is to be understood that while the remaining discussion is directed largely to the utility of Gal α(1,3)galactosyl transferase knockout pigs, the utility of the methods described herein is not limited to solely this protein. Rather, the following discussion is provided merely for exemplification of their versatility and preferred use.

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIG. 1 shows the nucleotide sequence of introns of the porcine Gal α(1,3) galactosyl transferase gene from within intron 3 to the end of intron 8. Dashes indicate nucleotides within an exon region. Thus, nucleotide sequence numbering represents the number of bases in the entire porcine Gal α(1,3) galactosyl transferase gene relative to nucleotide position 1 of the insert isolated from the lambda-2 phage clone.
FIG. 2 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal α(1,3) galactosyl transferase gene (see Example 1). This vector is designed to contain a sequence with homology to the 5′ region of intron 3 of the Gal α(1,3) galactosyl transferase gene, a promoterless neomycin phosphotransferase gene engineered to contain multiple stop codons (engineered exon), an engineered splice acceptor site, the 5′ region of intron 4 sequence for splicing the engineered exon to the downstream exon 4, and a sequence with homology to the 3′ region of intron 3 to aid with annealing to the porcine Gal α(1,3) galactosyl transferase gene. Arrows indicate location of primers used for PCR.

FIG. 3 shows the nucleotide sequence of the gene targeting vector used for inactivation of the porcine Gal α(1,3) galactosyl transferase gene (see Example 1). This vector is designed to contain (A.) a sequence with homology to the 5′ region of intron 3 of the Gal α(1,3) galactosyl transferase gene, (B.) an intron 4 splice acceptor sequence, (C.) a promoterless neomycin phosphotransferase gene engineered to contain multiple stop codons (engineered exon), (D.) an intron 4 splice donor signal sequence, and (E.) a 3′ intron 3 sequence to aid with annealing to the porcine Gal α(1,3) galactosyl transferase gene. All underlined sequences correspond to restriction sites in the primer sequences. Bold type indicates primer regions used for PCR. Normal type indicates PCR fragment sequences.

FIG. 4 shows the nucleotide sequence for the neomycin phosphotransferase gene (the neomycin resistance gene). Bold type indicates the location of gene start and stop codons. The underlined sequence corresponds to primer sequences. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 5 shows the nucleotide sequence of the puromycin/bovine growth hormone poly A. The underlined sequences correspond to the puromycin gene start codon.
FIG. 6 shows a schematic representation of the gene targeting vector used for inactivation of the porcine Gal α(1,3) galactosyl transferase gene (see Example 2). This vector is designed to contain a sequence with homology to the Gal α(1,3) galactosyl transferase gene 3’ intron 3 sequence including the 3’ intron splice acceptor sequence, a Kozak consensus sequence, a promoterless puromycin gene engineered to contain a bovine growth hormone poly A sequence (engineered exon), and a sequence with 5’ intron 4 sequence homology including the 5’ intron splice donor sequence. Arrows indicate location of primers used for PCR.

FIG. 7 shows the nucleotide sequence of the gene targeting vector shown schematically in FIG. 6 (see Example 2). The underlined sequences correspond to the primer sequences used. Bold type indicates the intron regions used for homology. The AG and GT splice consensus sequences at the 3’ end of intron 3 and the 5’ end of intron 4 are in upper case.

FIG. 8 shows the nucleotide sequence of the ricin A toxin gene. Nucleotides which are capitalized are within the coding region of this gene.

FIG. 9 shows a schematic representation of the collision construct used for inactivation of the porcine Gal α(1,3) galactosyl transferase gene (see Example 3). This vector is designed to contain a sequence with homology to the Gal α(1,3) galactosyl transferase gene 3’ intron 3 sequence including the 3’ intron splice acceptor sequence, a reverse orientation puromycin gene engineered to contain a bovine growth hormone poly A sequence under the control of a phosphoglycerate kinase (PGK) promoter, and a sequence with 5’ intron 4 sequence homology including the 5’ intron splice donor sequence, and a ricin A toxin gene under the control of a cytomegalovirus (CMV) promoter and containing a SV40 poly A sequence located outside the regions of homology. Arrows indicate location of primers used for PCR.
DESCRIPTION OF THE PREFERRED EMBODIMENTS

Technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art, unless otherwise defined herein. Although any methods and materials similar or equivalent to those described herein may be used in the practice or testing of the described methods, preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection herewith.

As used herein and in the appended claims, the singular forms "a," "an," and "the" are intended to include the plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" is intended to include a plurality of such host cells, reference to "an antibody" is intended as a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth. It is to be understood that the appended claims are not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, which those of skill will appreciate may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which is to be limited only by the appended claims.

The term "knockout" refers to the modulation of the expression of at least a portion of a protein encoded by the target gene. The term "knockout construct" refers to a nucleic acid sequence that is designed to modulate a protein encoded by endogenous DNA sequences in a cell. The nucleic acid sequence used as the knockout construct is typically comprised of DNA from some portion of the gene or genes (including, but not limited to, the exon sequence, intron
sequence, and/or promoter sequence) to be modulated and a sequence marker used to disrupt and select for the presence of the knockout construct in the cell. The nucleic acid sequence of the knockout construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to prevent or interrupt protein expression from the native gene. Such insertion usually occurs by homologous recombination (i.e., regions of the knockout construct that are homologous to endogenous DNA sequences hybridize to each other when the knockout construct is introduced into the cell and recombines so that the knockout construct is incorporated into the corresponding position of the endogenous DNA).

The knockout construct nucleic acid sequence may comprise a full or partial sequence of one or more exons and/or introns of the gene to be modulated, a full or partial promoter sequence of the gene to be modulated, or combinations thereof. In one embodiment of the invention, the nucleic acid sequence of the knockout construct comprises a first nucleic acid sequence region homologous to a first nucleic acid sequence region of the gene to be modulated, and a second nucleic acid sequence region homologous to a second nucleic acid sequence region of the gene to be modulated. The orientation of the knockout construct should be such that the first nucleic acid sequence is upstream of the second nucleic acid sequence and the sequence marker should be therebetween.

A suitable nucleic acid sequence region(s) should be selected so that there is homology between knockout construct sequence(s) and the gene of interest. Preferably, the knockout construct sequences are isogenic sequences with respect to the target sequences. The nucleic acid sequence region of the knockout construct may correlate to any region of the gene provided that it is homologous to the gene. A nucleic acid sequence is considered to be “homologous” if it is at least about 90% identical, preferably at least about 95% identical, or most preferably, about
100% identical to the nucleic acid sequence. Furthermore, the 5' and 3' nucleic acid sequences flanking the selectable marker should be sufficiently large to provide complementary sequence for hybridization when the knockout construct is introduced into the genomic DNA of the target cell. For example, homologous nucleic acid sequences flanking the selectable marker gene should be at least about 500 bp, preferably, at least about 1 kilobase (kb), more preferably about 2-4 kb, and most preferably about 3-4 kb in length. In a preferred embodiment, both of the homologous nucleic acid sequences flanking the selectable marker gene of the construct should be should be at least about 500 bp, preferably, at least about 1 kb, more preferably about 2-4 kb, and most preferably about 3-4 kb in length.

Another suitable DNA sequence includes cDNA sequence provided the cDNA is sufficiently large. Each of the flanking nucleic acid sequences used to make the construct is preferably homologous to one or more exon and/or intron regions, and/or a promoter region. Each of these sequences is different from the other, but may be homologous to regions within the same exon and/or intron. Alternatively, these sequences may be homologous to regions within different exons and/or introns of the gene. Preferably, the two flanking nucleic acid sequences of the knockout construct are homologous to two sequence regions of the same or different introns of the gene of interest. In addition, it is preferred that isogenic DNA is used to make the knockout construct of the present invention. Thus, the nucleic acid sequences obtained to make the knockout construct are preferably obtained from the same cell line as that being used as the target cell.

In accordance with the present invention, the integration of the knockout construct nucleic acid sequence into at least one gene of interest results in the modulation of the expression of the gene product. "Modulating" the expression of a gene includes suppressing the expression
of the gene, disrupting the expression of the gene, eliminating the expression of the gene, altering
the expression of the gene, or decreasing the expression of the gene relative to expression of the
wild-type gene. Preferably, the integrated knockout construct results in reduced protein function
relative to native protein function. Most preferably, the integrated knockout construct results in
the production of a non-functional protein. Complete or absolute non-functionality of the protein
is not required.

The phrases "disruption of the gene" and "gene disruption" refer to insertion of a nucleic
acid sequence into at least one region of the native DNA sequence (usually one or more exons or
one or more introns) and/or the promoter region of a gene so as to modulate expression of that
gene in the cell as compared to the wild-type or naturally occurring sequence of the gene. By
way of example, a nucleic acid construct may be prepared containing a DNA sequence encoding
an antibiotic resistance gene which is inserted between the DNA sequence complementary to the
target gene DNA sequence (promoter and/or coding region) to be disrupted. When this nucleic
acid construct is then transfected into a cell, the construct will integrate into the genomic DNA
either randomly or into the target gene by homologous recombination. It has been found that
selection for drug resistant cells in the population of transfectants enhance the probability of
obtaining a homologous gene knockout. Thus, many progeny of the cell will no longer express
the gene, or will express it at a decreased level, as the DNA is now disrupted by the antibiotic
resistance gene.

In some instances, such as, for example, where the methods described herein are used to
produce cells, tissues or organs suitable for xenotransplant into humans, it may not be necessary
to completely eliminate the production of functional protein. Rather, it will be satisfactory to
reduce the production of functional protein only to a level that will, in conjunction with other
therapeutic regimens, prevent or reduce the patient’s immune response and the likelihood of rejection. Thus, for example, a knockout achieved according to the methods described herein may preferably reduce the biological activity of the polypeptide normally encoded therefrom by at least about 70%, preferably at least about 80%, relative to the unmutated gene.

The knockout construct may be inserted into any suitable target cell for integration into its genomic DNA that may be maintained in culture. Suitable cells include, but are not limited to, fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell. In one embodiment, the knockout construct is inserted into an embryonic stem cell (ES cell) and is integrated into the ES cell genomic DNA. ES cells comprising the integrated knockout construct are then injected into, and integrate with, a developing mouse embryo. In another embodiment, the knockout construct is inserted into a nuclear transfer donor cell. Suitable nuclear transfer donor cells include fibroblasts, epithelial cells, and cumulus cells. In this embodiment, the knockout construct is inserted into the nuclear transfer donor cell, and the donor cells comprising the knockout construct are fused with an enucleated oocyte. The resultant fused oocyte is then transferred to a surrogate female.

Furthermore, where the target cell is intended to be used to produce a knockout mammal, it is preferred that the target cell be derived from the same species as the knockout mammal to be generated. Thus, for example, pig embryonic stem cells or pig fibroblasts will usually be used for generation of knockout pigs.

The nucleic acid sequence of the knockout construct may be integrated into the genomic DNA of the host cell using any suitable method. In one preferred embodiment, integration is achieved by the process of homologous recombination. Homologous recombination has been described previously, for example, in Kucherlapati et al. (1984) Proc. Natl. Acad. Sci. USA

The methods described herein may be used to produce a mammal in which one, two, or more genes have been knocked out. Such mammals may be generated by repeating the procedures set forth herein for generating each knockout construct, or by breeding two mammals, each with a different single gene knocked out, to each other, and screening for those with the double knockout genotype.
The term "marker sequence" or "selectable marker" refers to a nucleic acid sequence that is used as part of the knockout construct to modulate the expression of the gene of interest, and as a means to identify those cells that have incorporated the knockout construct into the genome. The selectable marker may be any sequence that serves these purposes. For example, the selectable marker may encode a protein that confers a detectable trait on the cell, such as an antibiotic resistance gene, or an assayable enzyme not typically found in the target cell. The selectable marker gene may be any nucleic acid sequence that is detectable and/or assayable, which is used to recover transformed cell lines. One having skill in the art will be capable of determining suitable selectable markers for use in the present invention. For example, suitable selectable markers include, but are not limited to, \( \beta \)-lactamase (ampicillin resistance), kanamycin resistance, gentecin resistance, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. For example, the herpes simplex virus thymidine kinase (tk) (Wigler, M. et al. (1977) Cell 11:223-32) or adenine phosphoribosyltransferase (aprt) (Lowy, I. et al. (1980) Cell 22:817-23) genes, which may be employed in tk or aprt cells, respectively, may be used as the selectable marker. Also, antimitabolite, antibiotic or herbicide resistance may be used as the basis for selection; for example, dihydrofolate reductase (dhfr), which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); neomycin phosphotransferase (npt), which confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described and include, for example, tryptophan synthetase (trpB), which allows cells to utilize indole in place of tryptophan, or histidinol dehydrogenase (hisD), which allows cells to utilize histidinol in place of histidine (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047- 51). Recently, the
use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase (GUS), and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C. A. et al. (1995) Methods Mol. Biol.

5 55:121-131). In the present invention, it is preferred that the selectable marker gene is an antibiotic resistance gene, such as the neomycin resistance gene or puromycin resistance gene.

Moreover, when the selectable marker encodes a protein, it may also contain a promoter that regulates its expression, or require expression from an endogenous promoter, preferably the target gene promoter. Thus, the selectable marker gene may be operably linked to its own promoter or be promoterless. The selectable marker gene may be inserted into the knockout construct without its own promoter attached as it may be transcribed using the promoter of the gene to be suppressed. In addition, the marker gene may have a polyA signal sequence attached to the 3’ end of the gene, which serves to terminate transcription of the gene and process the transcript with the addition of adenine residues at the 3’ end to stabilize the mRNA.

15 In one embodiment a target gene (e.g., Gal α(1,3) galactosyl transferase) is modulated by insertion of an engineered exon or active gene within an intron of the target gene. In this embodiment, the target gene is prevented from being translated by insertion of an in-frame, promoterless engineered exon (e.g., an antibiotic resistance gene) that contains multiple stop codons within an intron of the target gene. Using this ‘promoter-trap’ strategy, the engineered exon is spliced in frame upstream of the exon comprising the start codon. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the target gene due to the presence of multiple stop codons downstream of the drug resistance gene. As described
herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon.

Using the 'promoter trap' strategy, a gene targeting construct is designed which contains a sequence with homology to an intron sequence of the target gene (e.g., the intron 3 sequence of the Gal α(1,3) galactosyl transferase gene), a downstream intron splice acceptor signal sequence comprising the AG dinucleotide splice acceptor site (e.g., the intron 4 splice acceptor signal sequence of the Gal α(1,3) galactosyl transferase gene), a promoterless selectable marker engineered exon (e.g., drug resistance gene) engineered to contain multiple stop codons, the intron splice donor signal sequence comprising the GT dinucleotide splice donor site (e.g., the intron 4 splice donor sequence of the Gal α(1,3) galactosyl transferase gene) for splicing the engineered exon to the immediate downstream exon (e.g., exon 4 of the Gal α(1,3) galactosyl transferase gene), and additional sequence with homology to the intron sequence of the target gene (e.g., intron 3 sequence homology of the Gal α(1,3) galactosyl transferase gene) to aid with annealing to the target gene. It will be appreciated that the method may be used to target any intron within target gene of interest.

In another embodiment, the 'promoter trap' strategy is used to modulate target gene expression by replacing an endogenous exon with an in-frame, promoterless engineered exon (e.g., an antibiotic resistance gene). The engineered exon is spliced in frame and results in the expression of the drug resistance gene and concomitant inhibited expression of the full-length target gene.

This 'promoter trap' gene targeting construct may be designed to contain a sequence with homology to the target gene 3' intron sequence upstream of the start codon (e.g., the Gal α(1,3) galactosyl transferase gene 3' intron 3 sequence), the upstream intron splice acceptor sequence comprising the AG dinucleotide splice acceptor site (e.g., the intron 3 splice acceptor sequence), a
Kozak consensus sequence, a promoterless selectable marker gene containing e.g., a poly A termination sequence (the engineered exon), a splice donor sequence comprising the GT dinucleotide splice donor site from an intron region downstream of the start codon (e.g., the 5' intron 4 splice donor sequence), and a sequence with 5' sequence homology to the downstream intron (e.g., 5' intron 4). It will be appreciated that the method may be used to target any exon within the Gal α(1,3) galactosyl transferase gene or any other gene of interest. A representative construct useful for targeting the pig Gal α(1,3) galactosyl transferase gene was deposited with the American Type Culture Collection (ATCC) on 28 September 2000 with accession number _____ and is described herein in Example 2.

In yet another embodiment, the selectable marker may be inserted into the knockout construct in a reverse orientation to the targeted gene. In this embodiment, a strong promoter is used with the selectable marker all in the reverse orientation, which drives transcription in the reverse direction and therefore, modulates the expression of the targeted gene. The target gene is modulated using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene, such as a selectable marker gene, is under the control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario et al., (1996) Nat. Biotech.14:1592-1596). The selectable marker gene is further engineered to contain a transcription termination sequence. Insertion of the engineered gene may be made to replace any exon, within any intron, or portions thereof to result in a truncated transcript which modulates the expression of a functional target gene product. It will be appreciated that this method may be used to target any intron or exon of interest of the target gene. Positive selection for transfected cells in which the construct has been integrated may be accomplished via expression of the selectable marker gene. As described herein, it will be appreciated
that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene (e.g., Ricin A toxin) is preferably engineered into the collision construct inserted to eliminate random integration events. A representative collision construct useful for targeting the pig Gal α(1,3) galactosyl transferase gene has been deposited with the ATCC on 28 September 2000 with accession number ________ and is described herein in Example 3.

The integrated selectable marker nucleic acid in the cell is capable of modulating the expression of the gene of interest. Expression of the selectable marker allows for selection of the cells which comprise the integrated sequence. Modulation of the expression of the gene of interest is accomplished by disruption of the endogenous gene by an engineered exon in forward or reverse orientation with the endogenous gene.

The term "animal," as used herein, is intended to include any multicellular eukaryotic organism, preferred among which are mammals. When used in the context of a xenograft donor, the term "mammal" preferably includes, but is not limited to, pigs, sheep, goats, cows, deer, rabbits, hamsters, rats, mice, horses, cats, dogs, and the like. Preferably, humans are excluded.

The term "progeny" refers to any and all future generations derived or descending from a particular mammal, i.e., a mammal containing a knockout construct inserted into its genomic DNA. Thus, progeny of any successive generation are included herein such that the progeny, the F1, F2, F3 generations, and so on indefinitely, are included in this definition.

The terms "immunomodulate" and "immunomodulation" refer to changes in the level of activity of any components of the immune system as compared to the average activity of that component for a particular species. Thus, as used herein, immunomodulation refers to an increase or a decrease in activity. Preferably, in accordance with the present invention, the
integration of the selectable marker into the gene of interest results in a decreased immune response, when the host cell is introduced to a patient. Immunomodulation may be detected by assaying the level of antibody reactivity, complement activity, B cells, any or all types of T cells, antigen presenting cells, and any other cells believed to be involved in immune function.

Additionally or alternatively, immunomodulation may be detected by evaluating the level of expression of particular genes believed to have a role in the immune system, the level of particular compounds such as cytokines (interleukins and the like) or other molecules that have a role in the immune system, and/or the level of particular enzymes, proteins, and the like that are involved in immune system functioning.

The target gene to be knocked out may be any gene, provided that at least some sequence information on the DNA to be disrupted is available to use in the preparation of both the knockout construct and the screening probes. It is not necessary that the entire genomic sequence of the target gene be known in order to use the methods described herein.

The target gene to be knocked out preferably will be a gene that is expressed in mature and/or immature T cells and/or B cells. It is a further preference that the target gene is expressed in target antigen presenting cell, target endothelial cell, target neuronal cell, or any target cell that may be attacked by the humoral or cellular immune system of the recipient. The target gene is further preferably involved, either directly or indirectly, in the activation pathway during inflammation or immunosuppression responses by the immune system, and does not result in lethality when knocked out. In accordance with the present invention, expression of target genes may advantageously be altered according to the methods described herein to produce xenotransplant cells, tissues and organs for use in humans, in order to reduce or prevent immune response and rejection by the patient.
Thus, in accordance with the present invention, any gene may be used provided that it can undergo homologous recombination and the expression of which may be modulated by insertion of the knockout construct of the present invention. Suitable genes include, but are not limited to, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1, CD28, CD80, CD86, CD154, major histocompatibility complex class I, β-2-microglobulin, invariant chain (Ii), caspase-1, caspase-3, and Gal α(1,3) galactosyl transferase gene. This list is not intended to be exhaustive. One having ordinary skill in the art would be capable of ascertaining suitable genes to be modulated. Preferably, the gene is implicated in the immunoresponse system of a patient. More preferably, the target gene is a porcine target gene selected from the group consisting of CD 80, CD 86, B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2 or VCAM-1. A presently preferred porcine target gene is the Gal α(1,3) galactosyl transferase gene.

The DNA sequence to be used to knock out a selected gene may be obtained using methods well known in the art such as those described by Sambrook et al. (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Such methods include, for example, screening a genomic library with a cDNA probe encoding at least a portion of the same gene in order to obtain at least a portion of the genomic sequence. Alternatively, if a cDNA sequence is to be used in a knockout construct, the cDNA may be obtained by screening a cDNA library with oligonucleotide probes or antibodies (where the library is cloned into an expression vector). If a promoter sequence is to be used in the knockout construct, synthetic DNA probes may be designed for screening a genomic library containing the promoter sequence. Another method for obtaining the DNA to be used in the knockout construct is to manufacture the DNA sequence synthetically, using a DNA synthesizer.
In another embodiment, porcine genomic DNA encoding the Gal α(1,3) galactosyl transferase gene is isolated from a lambda phage clone library. A pig genomic library is screened using a cDNA corresponding to exon 4 of the Gal α(1,3) galactosyl transferase gene. Phage are screened and unique clones, which contain exon 4 sequences are isolated using standard library screening methods. (Sambrook et al.) Clones obtained by this procedure contain inserts 15-40 kb in length. These clones, were designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal α(1,3) galactosyl transferase gene from within intron 3 through intron 8 have been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers ________________, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1. These sequences may be used to determine regions of sequence homology in design of targeting constructs for modulation of the pig Gal α(1,3) galactosyl transferase gene.

The DNA sequence encoding the knockout construct is preferably generated in sufficient quantity for genetic manipulation and insertion the target cell. Amplification may be accomplished by known methods, such as by placing the sequence into a suitable vector and transforming bacterial or other cells that may rapidly amplify the vector, by PCR amplification, or by synthesis with a DNA synthesizer.
The DNA sequence to be used in producing the knockout construct is digested with a restriction enzyme selected to cut at a location(s) such that a new DNA sequence encoding a selectable marker gene may be inserted in the proper position within this DNA sequence. The proper position for a selectable marker gene insertion is that which will serve to modulate expression of the native gene. This position will depend on various factors such as the restriction sites in the sequence to be cut, and whether, for example, an intron sequence, an exon sequence or a promoter sequence is (are) to be modulated. In other words, the precise location of insertion of the selectable marker into the DNA sequence is that which will result in the modulation of promoter function or of synthesis of the native exon. For example, the knockout construct may be engineered to insert the selectable marker entirely within a single intron of the target gene. In this manner, the first nucleic acid sequence would comprise a region of the selected intron upstream from the second nucleic acid sequence and the second nucleic acid sequence would be selected comprising a region of the selected intron located downstream of the first nucleic acid sequence. The selectable marker would be introduced between the first and second nucleic acid sequences. When the construct is then introduced to the cell, the construct nucleic acid sequence is integrated into the target gene and the selectable marker is inserted entirely within the targeted intron.

Similarly, the construct may be engineered to insert the selectable marker within any desired and suitable region of the gene provided that expression is modulated. For example, the construct may be engineered to insert the selectable marker between two adjacent introns and thereby completely remove an endogenous exon of the target gene, to span over a region comprising at least a portion of an intron and at least a portion of an adjacent intron of the targeted gene, to span over a region comprising at least a portion the promoter for the targeted
gene to an adjacent intron, to span over a region encompassing more than one targeted gene, and combinations thereof.

After the genomic DNA sequence has been digested with the appropriate restriction enzymes, the selectable marker gene is ligated into the genomic DNA sequence using methods well known to the skilled artisan and described in Sambrook et al., supra. The ends of the DNA fragments to be ligated must be compatible; this is achieved by either cutting all fragments with enzymes that generate compatible ends, or by blunting the ends prior to ligation. Blunting is done using methods well known in the art, such as for example by the use of Klenow fragment (DNA polymerase I) to fill in sticky ends.

The ligated knockout construct may be introduced directly into the target cell, or it may first be placed into a suitable vector for amplification prior to insertion. Preferred vectors are those that are rapidly amplified in bacterial cells such as the pBluescript II SK vector (Stratagene, San Diego, Calif.) or pGEM7 (Promega Corp., Madison, Wis.).

In another embodiment of the invention, embryonic stem (ES) cells are used as the target cell for their ability to integrate into and become part of the germ line of a developing embryo so as to create germ line transmission of the knockout construct. Thus, any ES cell line that is believed to have this capability is suitable for use herein. For example, one mouse strain that is has been used for production of ES cells is the 129J strain. The cells are cultured and prepared for DNA insertion using methods well known to the skilled artisan such as those set forth by Robertson (Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, E. J. Robertson, ed. IRL Press, Washington, DC (1987)), Bradley et al. (Current Topics in Devel. Biol., 20:357-371 (1986)), and Hogan et al. (Manipulating the Mouse Embryo: A Laboratory Manual, Cold

Insertion of the knockout construct into the target cells may be accomplished using a variety of transfection methods well-known in the art. For example, suitable transfection methods include electroporation, microinjection, and calcium phosphate treatment (see Lovell-Badge, in Robertson, ed., supra). A preferred method of transfection is electroporation. If the cells are to be electroporated, the targeted cells and knockout construct DNA are exposed to an electric pulse using an electroporation machine and following the manufacturer’s guidelines for use. After electroporation, the cells are allowed to recover under suitable incubation conditions. The cells are then screened for the presence of the knockout construct.

Each knockout construct DNA to be introduced into the cell must first be linearized if the knockout construct has been inserted into a vector. Linearization is accomplished by digesting the DNA with a suitable restriction endonuclease selected to cut only within the vector sequence and not within the knockout construct sequence.

For introduction of the DNA sequence, the knockout construct DNA is added to the target cells under appropriate conditions for the insertion method chosen. Where more than one construct is to be introduced into the target cell, DNA encoding each construct may be introduced simultaneously or one at a time.

Screening may be done using methods known in the art or combinations thereof. Where the selectable marker gene is an antibiotic resistance gene, the cells are cultured in the presence of an otherwise lethal concentration of the antibiotic. Those cells that survive have presumably integrated the knockout construct. If the selectable marker gene is other than an antibiotic resistance gene, the genomic DNA of the target cell may be extracted from the cells using
standard methods such as those described by Sambrook et al., supra. The DNA may then be
probed on a Southern blot with a probe or probes designed to hybridize only to the selectable
marker sequence. If the selectable marker gene is a gene that encodes an enzyme whose activity
may be detected (e.g., beta-galactosidase), the enzyme substrate may be added to the cells under
suitable conditions, and an appropriate assay for enzymatic activity may be conducted. In
addition, the genomic DNA may be amplified by polymerase chain reaction (PCR) with probes
specifically designed to amplify DNA fragments of a particular size and sequence (i.e., only
those cells containing the knockout construct in the proper position will generate DNA fragments
of the proper size). PCR may be used in detecting the presence of homologous recombination
338:153-156). Primers may be used which are complementary to a sequence within the
construct and complementary to a sequence outside the construct and at the target locus. In this
way, one may only obtain DNA duplexes having both of the primers present in the
complementary chains in which homologous recombination has occurred. By demonstrating the
presence of the primer sequences or the expected size sequence, the occurrence of homologous
recombination is supported.

Upstream and/or downstream from the target gene knockout construct may be inserted a
gene which provides for identification of whether a double crossover has occurred. For this
purpose, any suitable marker may be used for as described herein. Preferably, the selectable
marker used to identify double crossovers is different than the selectable marker used to identify
the integration of the target gene knockout construct. In one preferred embodiment, the herpes
simplex virus thymidine kinase gene is employed, since the presence of the thymidine kinase
gene may be detected by the use of nucleoside analogs, such as Acyclovir or Gancyclovir, for
their cytotoxic effects on cells that contain a functional HSV-tk gene. The absence of sensitivity to these nucleoside analogs indicates the absence of the thymidine kinase gene and, therefore, where homologous recombination has occurred, a double crossover event has also occurred.

The knockout construct may be integrated into several locations in the target cell genome, and may integrate into a different location in each cell’s genome, due to the occurrence of random insertion events. Notwithstanding random multiple integration sites, the desired location of the insertion is in a complementary position to the DNA sequence to be knocked out. It has been found that less than about 1-5% of the targeted cells that take up the knockout construct will actually integrate the knockout construct in the desired location. Identification of those cells with proper integration of the knockout construct is described herein.

In one embodiment of the present invention, suitably transfected target cells containing the knockout construct in its proper location are inserted into an embryo. Insertion may be accomplished in any suitable method known in the art. Preferably, the cells are introduced into the embryo by microinjection. Most preferably, the cells are ES cells for injection into mouse embryos. For microinjection, about 10-30 cells are collected into a micropipet and injected into embryos that are at the proper stage of development to integrate the transfected cell into the developing embryo. The suitable stage of development for injecting into the embryo is prior to the formation of the germinal layer of the developing embryo as one having ordinary skill in the art may readily determine. Preferably, the embryo is in the early blastocyst stage. By way of example, mice embryos may be introduced to the transfected cells in about 3.5 days. The embryos are obtained by perfusing the uterus of pregnant females by methods known to the skilled artisan (e.g., Bradley (in Robertson, ed., supra)). Preferably, the embryos are male.
After the transfected target cell having proper integration of the target gene has been introduced into the embryo, the embryo is implanted into the uterus of a pseudopregnant foster mother. While any foster mother may be used, selection of the foster mother is based upon its ability to breed and reproduce well, and to care for its young. Such foster mothers are typically prepared by mating with vasectomized males of the same species. The stage of the pseudopregnant foster mother is important for successful implantation, and is species dependent. For mice, this stage is about 2-3 days pseudopregnant.

In another embodiment, the suitable transfected target cells are nuclear transfer donor cells. Nuclear transfer donor cells may be virtually any somatic cell type and include fibroblasts, epithelial cells, cumulus cells, etc. Nuclear transfer donor cells are cultured in vitro and targeted using the constructs and techniques described herein via homologous recombination. Cells are grown in the appropriate medium to allow for selection of cells comprising the having properly integrated the knockout construct. PCR may also be done for confirmation of correctly targeted integration. Thereafter, an unfertilized oocyte of an animal is enucleated using known methods. The enucleated unfertilized oocyte is then fused to the selected knockout nuclear transfer donor cell. Fusion may be conducted by electrical stimulation, chemical stimulation, insertion by injection, or other known methods. The fused product is then cultured, assessed for viability and transferred to a surrogate recipient female. For reference and methods, see e.g., Campbell et al. (1996) Nature 380:64; Wilmut et al. (1997) Nature 385:810; WO00/25578; WO97/07669; WO99/36510; WO00/42174; WO99/53751; WO99/45100, which are incorporated herein by reference.

Offspring or progeny that are born to the foster mother or surrogate recipient female are screened (e.g., by PCR) for genomic DNA comprising the knockout construct. This step is
particularly important for selecting for progeny of foster mothers that carried embryos in which the transfected target cell was injected. On the other hand, the progeny of surrogate recipient females that carried the transfected target cell fused with the enucleated unfertilized oocyte will typically have the knockout construct inserted into its genome.

Any suitable selection method may be used. For example, if a coat color selection strategy has been used, the offspring may be screened for a coat color indicative of proper integration of the targeted gene into the offspring. Other methods include obtaining DNA from the offspring and screening for the presence of the knockout construct using Southern blots and/or PCR as described herein. Other means of identifying and characterizing the knockout offspring include the use of Northern blots and Western blots. For example, Northern blots may be used to probe the mRNA for the presence or absence of transcripts encoding either the gene knocked out, the marker gene, or both. In addition, Western blots may be used to assess the level of expression of the gene knocked out in various tissues of these offspring by probing the Western blot with an antibody against the protein encoded by the gene knocked out, or an antibody against the marker gene product, where this gene is expressed. *In situ* analysis (such as fixing the cells and labeling with antibody) and/or fluorescence activated cell sorting (FACS) analysis of various cells from the offspring may be conducted using suitable antibodies to look for the presence or absence of the knockout construct gene product.

Offspring that appear to contain the integrated knockout construct in its genome may then be out-crossed to generate multiple offspring if they are believed to carry the knockout construct in their germ line to generate F1 offspring heterozygous for the knockout construct. F1’s will then be crossed to generate homozygous knockout animals.
The heterozygotes may then be crossed with each other to generate homozygous knockout offspring. Homozygotes may be identified by any screening method as described herein. For example, the homozygotes may be identified by Southern blotting of equivalent amounts of genomic DNA from the host animal(s) that is (are) the product of this cross, as well as host animals that are known heterozygotes and wild-type host animals. Probes to screen the Southern blots may be designed as set forth herein.

The knockout mammals described herein will have a variety of uses depending on the gene or genes that have been modulated. Where the targeted gene or genes modulated encode proteins believed to be involved in immunosuppression or inflammation, the knockout mammal may be used to screen for drugs useful for immunomodulation, i.e., drugs that either enhance or inhibit these activities. Screening for useful drugs may involve administering the candidate drug over a range of dosages to the knockout mammal, and assaying at various time points for immunomodulatory effects of the drug on the immune disorder being evaluated. Such assays may include, for example, looking for increased or decreased T and B cell levels, increased or decreased immunoglobulin production, increased or decreased levels of chemical messengers such as cytokines (e.g., interleukins and the like), and/or increased or decreased levels of expression of particular genes involved in the immune response.

For example, patients undergoing chemotherapy often experience immunosuppression. It would be desirable to activate the immune system of such individuals by administering to the patient a therapeutic agent capable of producing such an effect. A knockout mammal as described herein could be used to screen a variety of compounds, either alone or in combination, to determine whether partial or total restoration or activation of the immune response results.
Similarly, the same strategy could be applied to find compounds that would be useful in suppressing the inflammatory response observed in many patients with arthritis, or useful in suppressing the autoimmune phenomenon observed in patients with rheumatoid arthritis and lupus. In addition, mammals may be useful for evaluating the development of the immune system, and for studying the effects of particular gene mutations.

In a preferred embodiment, the knockout mammals described herein are used for xenograft transplantation into human patients. The xenograft tissue may be from any mammal, preferably a pig. The xenotransplanted tissue may be in the form of an organ including, for example, a kidney, a heart, a lung, or a liver. Xenotransplant tissue may also be in the form of parts of organs, cell clusters, and glands including, for example, lenses, pancreatic islet cells, skin, corneal tissue, and the like.

In yet another aspect of the present invention, the target gene is the Gal α(1,3) galactosyl transferase gene in pigs. The Gal α(1,3) galactosyl transferase is an attractive target for knockout in the pig. This enzyme is responsible for the addition of a carbohydrate residue, Gal α(1,3) Gal, that is recognized by human IgM and IgG antibodies in pig-to-human xenotransplanted tissue and leads to subsequent hyperacute rejection. Knockout pigs, which lack the Gal α(1,3) galactosyl transferase gene, may thus potentially serve as a rich source for xenotransplanted organs. Nucleic acid sequences encoding Gal α(1,3) galactosyl transferase and mutants thereof are disclosed. Preferably, the nucleotide sequence encodes pig Gal α(1,3) galactosyl transferase. Nucleotide sequences may be in the form of DNA, RNA or mixtures thereof. Nucleotide sequences or isolated nucleic acids may be inserted into replicating DNA, RNA or DNA/RNA vectors (as are well known in the art), such as plasmids, viral vectors, and

Nucleotide sequences encoding Gal $\alpha(1,3)$ galactosyl transferase may include promoters, enhancers and other regulatory sequences for expression, transcription and translation. Vectors encoding such sequences may include restriction enzyme sites for the insertion of additional genes and/or selection markers, as well as elements necessary for propagation and maintenance of vectors within cells.

Targeting constructs comprising nucleotide sequences, and mutants thereof, of the Gal $\alpha(1,3)$galactosyl transferase are particularly preferred as they may be used to inactivate wild type Gal $\alpha(1,3)$ galactosyl transferase genes according to the methods of the present invention.

Mutant Gal $\alpha(1,3)$ galactosyl transferase nucleotide sequences include, but are not limited to, nucleotide deletions, insertions, substitutions and additions to wild type Gal $\alpha(1,3)$ galactosyl transferase, such that the resultant mutant does not encode a functional galactosyl transferase. These nucleotide sequences may be utilized in the methods of modulating expression of galactosyl transferase of the present invention. In this manner, mutant sequences are recombined with wild type genomic sequences in the target cells.

In a most preferred embodiment, knockout pigs are produced in which the Gal $\alpha(1,3)$galactosyl transferase gene produces a non-functional protein. By producing a non-functional protein, the human antibody that would otherwise bind to the Gal $\alpha(1,3)$Gal epitope expressed on the xenotransplanted tissue does not bind, so that immune responses which give rise to tissue rejection are prevented. In this embodiment, any knockout construct capable of
modulating the interaction between antibodies directed to the Galα(1,3) galactosyl transferase linkage may be used.

**EXAMPLES**

The following examples are for purposes of illustration only, and are not intended to limit the scope of the disclosure or claims.

**EXAMPLE 1**

Inactivation of the Gal α(1,3) galactosyl transferase gene by insertion of an engineered active gene in the form of an engineered exon within intron 3.

In this example, the Gal α(1,3) galactosyl transferase protein is prevented from being translated by insertion of an in-frame, promoterless engineered exon (e.g., an antibiotic resistance gene) that contains multiple stop codons within an intron of the Gal α(1,3) galactosyl transferase gene. Using this ‘promoter-trap’ strategy, the engineered exon is spliced in frame upstream of exon 4 of the Gal α(1,3) galactosyl transferase gene. This results in the expression of the drug resistance gene prior to the gene of interest and concomitantly inhibits expression of the transferase gene due to the presence of multiple stop codons downstream of the drug resistance gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, genticin, neomycin phosphotransferase, puromycin-N-acetyltransferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. The present example employs neomycin.

A gene targeting construct is designed which contains a sequence with homology to the Gal α(1,3) galactosyl transferase gene 5′ intron 3 sequence, an intron 4 splice acceptor signal sequence, a promoterless neomycin phosphotransferase gene engineered to contain multiple stop
codons (engineered exon), the intron 4 splice donor sequence for splicing the engineered exon to the downstream exon 4, and additional intron 3 sequence homology to aid with annealing to the porcine Gal α(1,3) galactosyl transferase gene. Although this example describes targeting intron 3, it will be appreciated that the method may be used to target any intron within the Gal α(1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α(1,3) galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 2 and 3.

The gene targeting construct is generated by ligating 5 distinct DNA fragments (1-5 below) together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The PCR reactions use the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to the manufacturer’s instructions. In the present example, a 50 ul final reaction volume is used, with 2 ul of DNA template, 1 ul of ELOGASE Enzyme Mix, 60 mM Tris-So4 (pH9.1) 18mM (NH4)2SO4, 1.2 mM MgSO4, 200mM dNTP mix, 10% DMSO and 200nM of each primer. The reaction is hot started at 95°C for 1 minute and followed by 30-40 cycles in a standard PCR thermocycler (GeneAmp PCR System 2400; PE Applied Biosystems, Foster City, CA).

1. A polymerase chain reaction (PCR) product consisting of intron 3 sequences as listed in Figures 1 and 3, nucleotide numbers 10-4020, is generated using standard PCR conditions for long range PCR of genomic fragments. Primers used include a 5' primer containing a NotI restriction site and intron 3 sequences 10-23 (GGCGGCGCGAGCCTCACCTGGC), and a 3′ primer containing a SalI restriction site and sequences homologous to intron 3 sequences 3999-4020 (GGTCGACGGATGCTGGTGGTGGAATAACAGG), where underlined sequence
indicates restriction sites and bold type indicates homology to endogenous sequences. An additional guanine nucleotide is added to the 5' end of all probes in this example to balance out 1 bp deletions that sometimes occur during cloning.

2. A PCR product is generated consisting of intron 4, the 3' splice sequence (the pyrimidine rich lariat and Gal α(1,3) galactosyl transferase intron 3 dinucleotide acceptor sequences) and 196 bases 5' flanking the ag dinucleotide acceptor site (nucleotides 11521-11716 in Figures 1 and 3). Primers used include a 5' primer containing a Sall restriction site (GGTGACCCACCGTTTGATCTGAG); and a 3' primer containing a EcoRI restriction site and the complementary strand homologous to the pyrimidine rich lariat and Gal α(1,3) galactosyl transferase dinucleotide acceptor sequences (GGAATTCCTAAAAGCAAATGGAAAATAAAAAACATATC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

3. A PCR product consisting of a neomycin resistance gene (Genbank Accession #AF081957; Figure 4) is generated using a 5' primer containing an EcoRI restriction site, and homology to the neomycin resistance gene, including the ATG start codon (GGAATTCAATGGATCCCCACCATGG); and a 3' primer containing a HindIII restriction site and complementary strand sequences to the 3' coding region of the neomycin gene, including the natural stop codon followed by two additional engineered stop codons (GAAGCTTCGCTATTAAGTATGGATATCC), where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence (see Figures 3 and 4).
4. A PCR product is generated containing the 5' splice donor sequences for intron 4 of the Gal α(1,3) galactosyl transferase gene, corresponding to sequences 4938-5173 in the claimed sequence comprising intron 4 (Figures 1 and 3). Primers used include a 5' primer containing a HindIII site and sequence identity to intron 4, sequences 4938-4962, including the Gal α(1,3) galactosyl transferase dinucleotide splice site (GAAGCTTGTAATTATGAAAACATGATG); and a 3' primer containing a PstI site and complementary strand sequence from intron 4 corresponding to nucleotide numbers 5152-5173 and includes multiple stop codons (GCTGCAGGCCACAGGTCACGGCAATGCGG); where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

5. A PCR product containing 1150 nucleotides of intron 3, corresponding to nucleotides 4024-4826 of the claimed sequence (Figures 1 and 3). Primers used include a 5' primer containing a PstI site and sequences 4024-4050 of the claimed sequence (GCTGCAGGCCCTCTCTCAACTACAATTTCATGCAGC); and a 3' primer containing a XhoI restriction site and complementary strand sequences to 4801-4826 of the claimed sequence (GCTCGAGAGAAAATTAGATTAATACACCCAGAG);

where underlined sequences indicate restriction sites and bold type indicates sequences with homology to the endogenous sequence.

Each PCR fragment (steps 1-5) is separately amplified. A single PCR fragment is cloned into the pCR2.1 vector (Invitrogen, San Diego, CA) according to the manufacturer's ligation instructions. The recombinant plasmid DNA is transformed into a suitable bacterial host (Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid DNA with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.
Following PCR fragment amplification, a series of ligations is performed to clone the final construct in the bacterial plasmid pBS SK+ (Stratagene, La Jolla, CA).

a. The HindIII-PstI fragment from the intron 4 PCR product (step 4) and the PstI-XhoI 3' homology fragment from intron 3 (step 5 above) are ligated to a pBS SK+ vector DNA following digestion with HindIII and XhoI. The 3 DNA fragments are mixed in equal molar ratios and incubated in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and plasmid DNA is isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.

b. The resulting plasmid (step 5a) is digested with HindIII and EcoRI and ligated with the HindIII-EcoRI Neomycin resistance gene fragment (step 3), that has been previously digested with HindIII and EcoRI. The resulting recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life Technologies, Gaithersburg, MD). The bacteria are cultured, and plasmid DNA is isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, is used to construct the final product.

c. The resulting plasmid (step 5b) is digested with EcoRI and NotI and ligated to the SalI-EcoRI intron 4-3' splice fragment (step 2) previously digested with SalI and EcoRI and the intron 3 4 kb NotI-SalI fragment (step 1) previously digested with NotI and SalI. The 3 DNA fragments are incubated in equal molar ratios in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid DNA is transformed into a suitable bacterial host (DH10B, Life...
Technologies, Gaithersburg, MD). The bacteria are cultured, and the recombinant plasmid DNA is isolated.

This final construct is used to transfect porcine embryonic fibroblasts, transgenic pig fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that are resistant to neomycin are screened by PCR to determine the site of integration. A primer located in the region of intron 4 not incorporated into the final construct (complementary strand of 5407-5427; GGACAATGGCAACATGGCAGG; see Figures 1 and 3) is used in combination with the 5’ neomycin gene primer (step 3). Only targeted insertions yield the appropriate sized PCR fragment. All other integration events produce a negative result.

Cell clones with a targeted insertion are then used to generate transgenic animals using nuclear transfer techniques, or in the case of the stem cells, used to inject into developing blastocysts and produce chimeric offspring.

**EXAMPLE 2**

Inactivation of the Gal α(1,3) galactosyl transferase gene by replacement of exon 4 with an active gene in the form of an engineered exon.

In this example, the Gal α(1,3) galactosyl transferase protein was prevented from being translated by replacing an endogenous exon (exon 3) with an in-frame, promoterless engineered exon (an antibiotic resistance gene) that contained a bovine growth hormone poly A sequence attached to the 3’ end of the gene, which served to terminate transcription of the engineered exon.

The engineered exon was spliced in frame, so as to take advantage of the endogenous promoter typically used by the Gal α(1,3) galactosyl transferase gene (‘promoter-trap’ strategy). This resulted in the expression of the drug resistance gene and concomitantly inhibited expression of the full-length
Gal α(1,3) galactosyl transferase gene. As described herein, any gene that confers survival of the targeted cells under appropriate selection conditions may be used as the engineered exon, including, but not limited to, ampicillin, kanamycin, genticin, neomycin phosphotransferase, puromycin-N-acetyl-transferase, hygromycin b-phosphotransferase, thymidine kinase, and tryptophan synthetase. The present example utilizes puromycin.

A gene targeting construct was designed which contained a sequence with homology to the Gal α(1,3) galactosyl transferase gene 3’ intron 3 sequence, an intron 3 splice signal sequence (splice acceptor sequence), a Kozak consensus sequence, a promoterless puromycin N-acetyl transferase gene linked to a bovine growth hormone poly A sequence (poly A) (engineered exon), the 5’ intron 4 splice signal sequence (splice donor sequence), and a sequence with 5’ intron 4 sequence homology. Exon 4 of the Gal α(1,3) galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Although this example describes targeting introns 3 and 4, it will be appreciated that the method may be used to target any exon within the Gal α(1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α(1,3) galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

The gene targeting construct was generated by ligating two distinct DNA fragments together to form the final gene targeting construct using standard molecular biology techniques well known to those skilled in the art. The first DNA fragment was obtained from the 3’ end of intron 3 containing the 3’ splice sequence (the pyrimidine-rich branch site used in forming the lariat during splicing and the AG dinucleotide splice acceptor sequence). The second DNA fragment was obtained from the 5’ end of intron 4 containing the GT dinucleotide splice donor sequence. The fragments were ligated into the pBlueScript vector containing a Kozak consensus sequence in-frame with the coding sequence of a promoterless puromycin gene linked to the
bovine growth hormone poly A sequence (Figure 5) to form the final gene construct. A schematic diagram of the targeting vector and corresponding nucleotide sequence are shown in Figures 6 and 7. Additionally, this construct has been deposited with ATCC on 28 September 2000 with accession number _________.

1. Generation of the first DNA fragment.

The first DNA fragment was a polymerase chain reaction (PCR) product consisting of intron 3 sequence as shown in Figure 7 (nucleotide numbers 235-4851, positions relative to nucleotide position 1 of the insert isolated from the lambda phage clone) and generated using standard PCR conditions as described by Randolf et al., (1996) for long range PCR of genomic fragments. The 5’ primer, consisting of intron 3 sequences 235-260, was

5’-AAGATTATAAATAGCCTCGTGTCAGG-3’. The 3’ reverse primer sequence was complementary to sequence 4827-4851 at the extreme 3’ end of intron 3 and containing the AG splice acceptor site, and was 5’-CTCCTGGGAAAAAAGAAAGAGAAGG-3’.

PCR reaction conditions to generate the 4.616 kb intron 3 sequence were performed using the ELONGASE Enzyme Mix (Life Technologies, Gaithersburg, MD) according to manufacturer’s conditions. In the present example, a 50 µl final reaction volume was used, with 2 µl of DNA template, 1µl of the ELONGASE Enzyme Mix, 60 mM Tris SO4 (pH 9.1), 18 mM (NH4)2SO4, 1.2 mM MgSO4, 200 mM dNTP mix, 10% DMSO and 200 nM of each primer. The reaction was hot started at 95°C for 1 min, followed by 30-40 cycles in a standard PCR machine (e.g., Gene Amp PCR Systems 2400; PE Applied Biosystems, Foster City, CA).

2. Preparation of the PCR2.1 cloning vector.

The Not I site of the PCR2.1 cloning vector (Invitrogen, San Diego, CA) was destroyed to avoid carrying over a second Not I site into the final construct. The Not I site was unique and
used to linearize the final plasmid construct. The PCR2.1 vector was digested with Not I and the overhangs filled-in using the Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) according to the manufacturer’s specifications. The plasmid was re-ligated using T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer’s recommendations.

5 Plasmid DNA was transformed into a suitable bacterial host (Top 10 F’, Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated and incubated with Not I enzyme to confirm loss of this site by restriction analysis.

3. Insertion of the first DNA fragment into the PCR2.1 vector.

Following PCR, the 4.616 kb fragment was ligated into the modified PCR2.1 vector using T4 DNA ligase according to the manufacturer’s specifications. Plasmid DNA was transformed into a suitable bacterial host (e.g., Top 10 F’, Invitrogen, San Diego, CA). The bacteria are cultured and plasmid DNA is isolated. Plasmid with the correct insert in the proper orientation, as determined by restriction analysis and sequence analysis, was used to construct the final product.

4. Preparation of the second DNA fragment.

A 2.084 kbp PCR product consisting of the intron 4 homology sequence containing the GT dinucleotide donor consensus splice sequence was constructed using standard PCR conditions as described above in step 1. The 5’ primer consisting of sequence 4938-4961 at the extreme 5’ end of intron 4 was 5’-GTAATTATGAAACATGATGAAATG-3’. The 3’ primer was homologous to the complementary strand of intron 4 at position 6997-7021 and has the sequence 5’-AGCCAGCGCTTACTAAGTACGTTGC-3’

5. Insertion of the second DNA fragment into the PCR2.1 vector.
Following PCR, the 2.084 kb fragment was ligated into the pCR2.1 vector from Invitrogen (San Diego, CA) using the manufacturer’s ligation conditions. Following ligation, the recombinant plasmid DNA was transformed into a suitable bacterial host (Top 10 F', Invitrogen, San Diego, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, and orientation in the plasmid, as determined by restriction analysis and sequence analysis, was used to construct the final product.

6. Preparation of a synthetic oligonucleotide linker sequence.

A synthetic oligonucleotide linker containing a Kozak consensus sequence and relevant restriction enzyme sites was prepared for in-frame cloning of the promoterless puromycin gene:

Xho I | Kozak seq.--- | HpaI | Hind III | Bgl II | Sal | Eco RV | Eco RI
TCGAGCCACCATGGTTAAACAAGCTTAGATCTGTCGACGATATCG
CGGTGGTACCAATTGTTCGAATCTAGACAGCTGCTATAGCTTAA

7. Assembly of the gene targeting construct.

The following ligations were performed to generate the final construct in the bacterial plasmid pBS KS+ (Strategene, La Jolla, CA). The final construct is illustrated in Figure 6:

a. The oligonucleotide linker containing the Kozak consensus sequence (step 6) was ligated to the pBS KS+ vector DNA following digestion with Xho I and Eco RI. Ligation was carried out using at least a 3:1 molar ratio of linker to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer’s recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Strategene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Restriction enzyme analysis was performed to confirm successful ligation using unique
restriction sites within the linker (Bgl II or Hpa I). This plasmid containing the linker was then used to construct the final product.

b. The resulting “mother” plasmid (step 7a) was then digested with Eco RV and Spe I to clone in the 3’ arm of the targeting construct. The 2.084 kb PCR fragment cloned into the PCR 2.1 vector (step 5) was digested with Eco RV and Spe I, isolated away from vector DNA by agarose gel electrophoresis and purified. The 2.084 kb fragment was ligated between the EcoRV and Spe I sites of the mother plasmid (step 7a). Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer’s recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF’, Stratogene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used to construct the final product.

c. The next fragment cloned into the mother plasmid (step 7b) was the cassette containing the promoterless puromycin gene coding sequence with the bovine growth hormone gene polyA signal sequence attached to its 3’ end following the TGA stop codon (Figure 5). The PGK puromycin bpolyA plasmid (used as a positive control for puromycin resistance of transformed cells) was digested with Hind III and Xho I. The puromycin bpolyA fragment was separated away from the rest of the vector DNA containing the PGK promoter by electrophoresis on a 0.7% agarose gel and purified. The mother plasmid (step 7b) was digested with Hind III and Sal I. The Hind III/Xho I puromycin bpolyA cassette was ligated to the Hind III and Sal I sites of the mother plasmid. Ligation was carried out using a 3:1 molar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the
manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was used to construct the final product.

d. The final cloning step involved ligating the 5' arm of the construct, which was the 4.616 kb intron 3 insert from the PCR2.1 vector (step 3). The PCR2.1 vector (step 3) was digested with Kpn I and Xho I. The 4.616 kb PCR fragment was isolated away from vector DNA by agarose gel electrophoresis and purified. The 4.616 kb Kpn I/Xho I insert was ligated into the mother plasmid (step 7c) that was digested with Kpn I and Xho I. Ligation was carried out using equimolar ratio of insert to vector in the presence of T4 DNA ligase (New England Biolabs, Beverly, MA) according to the manufacturer's recommendations. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated by standard molecular biology techniques. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was used as the final product.

e. The final construct may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones that are resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in a region of intron 4, which is not incorporated into the final construct, may be used in combination with a 5' puromycin gene primer. Only targeted insertions will yield the appropriate size PCR fragment. All other integration events will produce a negative result.
Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

Example 3

Inactivation of the Gal α(1,3) galactosyl transferase gene by replacement of exon 4 with an a reverse orientation active gene.

In this example, the Gal α(1,3) galactosyl transferase gene was functionally inactivated by using a "collision construct" to insert an active gene in place of an exon and at least part of the flanking introns, including the splice donor and splice acceptor sites. The inserted gene is under the control of a highly active promoter such as the phosphoglycerate kinase I (PGK) gene promoter, such that transcription of this gene causes the termination of transcription of the endogenous gene (Rosario et al., (1996) Nat. Biotech.14:1592-1596). Exon 4 of the Gal α(1,3) galactosyl transferase gene codes for ATG start codon and the N-terminal portion of the protein. Thus, the insertion was made to replace exon 4 as well as a portion of the flanking introns 3 and 4, resulting in a truncated transcript that did not code for a functional enzyme. Although this example describes targeting introns 3 and 4, this method could be used to target any introns within the Gal α(1,3) galactosyl transferase gene or any other gene of interest. A sequence listing of the introns in the Gal α(1,3) galactosyl transferase gene (from within intron 3 to the end of intron 8) is provided in Figure 1.

In this example, the PGK promoter was inserted driving the expression of the puromycin resistance gene with the bovine growth hormone poly A (polyA) transcription termination sequence. This gene replaced the Gal α(1,3) galactosyl transferase exon 4 as well as a portion of the flanking intron 3 and 4 sequences by standard homologous recombination techniques utilizing intron 3
and 4 sequences for homology flanking the inserted gene. Intron 3, which separates exons 3 and 4, is greater than 5 kb in length, and a construct was built such that there was at least about 4.6 kb of homologous sequence on one end of the gene. Intron 4, which separates exons 4 and 5 is about 6.8 kb in length, and the construct was built such that there is at least about 2.2 kb of homologous sequence on the other end of the gene. Positive selection for transfected cells in which the construct has been integrated was accomplished via expression of the puromycin resistance gene. As described herein, it will be appreciated that any selection marker gene that confers survival of the targeted cells under appropriate selection conditions may be driven by the strong PGK promoter. Additionally, a toxin gene was inserted to eliminate random integration events.

The collision construct was generated using standard molecular biology techniques well known to those skilled in the art. The 4.616 kb intron 3 homology fragment and the 2.084 kb intron 4 homology fragment were generated using PCR and cloned into the PCR2.1 cloning vector as described in Example 2, steps 1-5 above for the replacement targeting construct. The generation of the collision construct first involved ligating the 2.084 kb intron 4 homology fragment into the pBS KS+ vector as the 3' arm of the collision construct, followed by the PGK-puromycin-bovine polyA cassette in the opposite orientation to the coding sequence of the GT gene. The 4.616 kb intron 3 homology fragment, as the 5' arm, was cloned in next. This generated the targeting construct for homologous recombination. The ricin A toxin gene was also added to the plasmid outside the region of homology, which will effectively kill a percentage of the cells in which random integration has occurred. The ricin A toxin gene was PCR amplified and cloned based upon the published sequence (Figure 8). A schematic diagram of the final construct is shown in Figure 9. Additionally, this collision construct has been deposited with ATCC on 28 September 2000 with accession number ____________.
1. The 2.084 kb 3' arm of the construct (intron 4 homology fragment) was the first fragment to be ligated into the pBluscript cloning vector, which was modified to contain the XhoI – EcoRI linker (see Example 2, step 6) within its multiple cloning site (pBS KS+). The ligation of the linker into the vector is described above in Example 2, step 7a, and ligation of the 2.084 kb 3' arm into the Eco RV and Spe I sites of the vector is described above in Example 2, step 7b.

2. The next step involved ligation of the PGK puromycin bovine polyA cassette into the pBS KS+ vector, which contained the 2.084 kb 3' arm. The PGK-puro-bPA cassette was digested with Eco RI which was immediately 5' of the PGK promoter. The Eco RI overhangs were blunted by filling in with Klenow enzyme (Roche Molecular Biochemicals, Indianapolis, IN) using the manufacturer's specifications. The PGK-puro-bPA cassette was then released from the vector by digestion with Xho I, which was immediately 3' of the bovine polyA sequence. The blunted PGK-puro-bPA-Xho I cassette was separated from vector DNA by agarose gel electrophoresis and purified. The pBS KS+ vector (step 1) was digested with Hpa I and Xho I, and the blunted PGK-puro-bPA-Xho I fragment was ligated between the Hpa I and Xho I sites of the vector. Following ligation, the recombinant plasmid was transformed into a suitable bacterial host (XL1-Blue MRF', Stratagene, La Jolla, CA). The bacteria were cultured, and plasmid DNA was isolated. Plasmid with the correct insert, as determined by restriction analysis and sequence analysis, was then used as the final product.

3. The 4.616 kb intron 3 homology fragment was ligated into the pBS KS+ mother plasmid (step 2) and represented the 5' arm of the collision construct. Isolation of this fragment from the PCR2.1 cloning vector and ligation into the Kpn I and Xho I sites of the mother plasmid is described above in Example 2, step 7d.
4. The ricin A toxin gene (Figure 8) was inserted into a commercially available mammalian expression vector, e.g., pcDNAI/Amp (Invitrogen). The insert was then excised with the CMV promoter and the SV40 poly A site and cloned into the Not I site of the recombinant plasmid by blunt end ligation, following Klenow fill in reactions on both the insert and vector.

5. Following ligation, the recombinant plasmid DNA was used to transform a suitable bacterial host (XLI-blue, Stratagene). The bacteria were cultured, and plasmid DNA isolated. This final construct DNA was then linearized with Kpn I (a unique enzyme site in the plasmid MCS outside of the construct sequence). Linearized plasmid may be used to transfect porcine embryonic fibroblasts, transgenic porcine fibroblasts, or porcine embryonic stem cells, or porcine primordial germ cells. Cell clones resistant to puromycin may be screened by PCR to determine the site of integration by methods well known to those of skill in the art. A primer located in the region of intron 4 not incorporated into the final construct may be used in combination with a 5’ puromycin gene primer. Only targeted insertions yield the appropriate size PCR fragment. All other integration events produce a negative result.

10 Cell clones with a targeted insertion may then be used to generate transgenic animals using nuclear transfer techniques, or in the case of stem cells, used to inject into developing blastocysts and produce chimeric offspring.

EXAMPLE 4

Isolation of porcine genomic DNA encoding the Gal α(1,3) galactosyl transferase gene from a Lambda phage clone library.

In this example, a pig genomic library was screened using a cDNA corresponding to exon 4 of the Gal α(1,3) galactosyl transferase gene using molecular biology techniques that are well
known to those skilled in the art (e.g., Sambrook et al., supra). A pig genomic library (Clontech, Palo Alto, CA) was obtained and screened with a PCR fragment derived from exon 4 of the porcine Gal α(1,3) galactosyl transferase gene. Exon 4 was labeled with ³²P dCTP using the Random Prime Kit (Stratagene, La Jolla, CA) according to the manufacturer’s instructions. Approximately 4 million phage forming units were screened and unique clones that contain exon 4 sequences as determined by Southern blotting were isolated. Clones obtained by this procedure contained inserts 15-40kb in length. These clones, designated pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2. Five vectors comprising unique, overlapping nucleotide sequences which span the entire the pig Gal α(1,3) galactosyl transferase gene from within intron 3 through intron 8 have been deposited with the ATCC: (1) a 1.6 kb insert within intron 3 of the extreme 5' end of the 18.275 kb lambda-2 phage clone, (2) a 6.7 kb HindIII fragment spanning intron 3 to intron 4 of the 18.275 kb lambda-2 phage clone, (3) a 4 kb HindIII fragment following the 6.7 kb fragment 2 of the 18.275 lambda-2 phage clone, (4) a 6 kb HindIII-SalI fragment at the 3' most portion of the 18.275 lambda-2 phage clone, and (5) a 13 kb fragment of the lambda-2 phage clone spanning exon 7 to exon 9. These five vectors were deposited with ATCC on 29 September 2000 with accession numbers ________________________, respectively. Subclones of the various inserts were used to generate the claimed intron sequences from within intron 3 to intron 8 as provided in Figure 1 using molecular biology techniques well-known to those skilled in the art (see e.g., Sambrook et al., supra). These sequences may be used to determine regions of sequence homology in design of targeting constructs for modulation of the pig Gal α(1,3) galactosyl transferase gene

Although the compositions and methods provided herein have been set forth in detail, one skilled in the art will recognize that numerous changes and modifications may be made, and that
such changes and modifications may be made without departing from the spirit and scope thereof.
We Claim:

1. A method of modulating the expression of a eukaryotic gene in a cell, comprising
   transfecing said cell with a nucleic acid construct, said construct comprising a first
   construct sequence homologous to a first gene sequence, a sequence encoding a selectable
   marker, and a second construct sequence homologous to a second gene sequence, wherein said
   first and second gene sequences independently comprise at least a portion of one or more intron
   regions of said eukaryotic gene, and
   integrating said selectable marker into said eukaryotic gene,
   wherein expression of said selectable marker results in modulation of expression of said
   eukaryotic gene in said cell.

2. The method of claim 1, wherein said first construct sequence and said second construct
   sequence are each homologous to at least a portion of an intron region of the gene.

3. The method of claim 1, wherein said sequence encoding a selectable marker is integrated
   into said eukaryotic gene by homologous recombination, wherein said first construct sequence
   recombines with said first gene sequence and said second construct sequence recombines with
   said second gene sequence to insert the selectable marker into the gene.

4. The method of claim 1, further comprising screening said cell for expression of said
   selectable marker.

5. The method of claim 2, wherein said first construct sequence and said second construct
   sequence are homologous to different regions from within the same intron of the gene.
6. The method of claim 2, wherein said first construct sequence and said second construct sequence are homologous to regions of different introns.

7. The method of claim 1, wherein said selectable marker gene is an antibiotic resistance gene.

8. The method of claim 1, wherein said sequence encoding a selectable marker is a nucleotide sequence which, when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentecin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphoribosyltransferase activity, dihydrofolate reductase activity, and histidinol dehydrogenase, anthocyanin, bet-glucuronidase and luciferase.

9. The method of claim 8, wherein said sequence encoding a selectable marker confers neomycin resistance or puromycin resistance.

10. The method of claim 1, wherein said eukaryotic gene is selected from the group consisting of genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28, CD80, CD86, CD154, major histocompatibility complex class I B-2-microglobulin, invariant chain, caspase-1 caspase-3, and Gal α(1,3) galactosyl transferase.

11. The method of claim 10, wherein said eukaryotic gene encodes Gal α(1,3) galactosyl transferase.

12. The method of claim 11, wherein said Gal α(1,3) galactosyl transferase gene is a porcine gene.
13. The method of claim 12, wherein said first construct sequence and said second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal α(1,3) galactosyl transferase gene.

14. The method of claim 13, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1.

15. The method of claim 13, wherein intron 5 has the nucleotide sequence of nucleotides 11753-13748 in Figure 1.

16. The method of claim 13, wherein intron 6 has the nucleotide sequence of nucleotides 13810-14358 in Figure 1.

17. The method of claim 13, wherein intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1.

18. The method of claim 13, wherein intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.

19. The method of claim 13, wherein said first construct sequence and said second construct sequence are homologous to different regions within the same intron of the eukaryotic gene.

20. The method of claim 19, wherein said intron is intron 3 of the porcine Gal α(1,3) galactosyl transferase gene.

21. The method of claim 13, wherein said first construct sequence and said second construct sequence are homologous to different introns of porcine Gal α(1,3) galactosyl transferase gene.
22. The method of claim 21, wherein said first construct sequence is upstream of said second construct sequence.

23. The method of claim 21, wherein said first construct intron region is homologous to an intron 3 region and said second construct intron region is homologous to an intron 4 region of porcine Gal α(1,3) galactosyl transferase.

24. The method of claim 2, wherein said sequence encoding a selectable marker is a promoterless gene.

25. The method of claim 2, wherein said sequence encoding a selectable marker further comprises a promoter.

26. The method of claim 25, wherein said promoter is a phosphoglycerate kinase (PGK) promoter.

27. The method of claim 2, wherein said sequence encoding a selectable marker is transcribed in the opposite orientation relative to the orientation of said eukaryotic gene.

28. The method of claim 27, wherein said sequence encoding a selectable marker further comprises a promoter sequence.

29. The method of claim 1, wherein said cell is selected from the group consisting of a fibroblast, epithelial cell, endothelial cell, transgenic embryonic fibroblast, embryonic stem cell, and primordial germ cell.
30. The method of claim 2, wherein said cell is a porcine cell.

31. The method of claim 2, wherein said construct further comprises an AG dinucleotide splice acceptor site.

32. The method of claim 2, wherein said construct further comprises a GT dinucleotide splice donor site.

33. A nucleic acid construct comprising a first construct sequence homologous to a first gene sequence, a sequence encoding a selectable marker, and a second construct sequence homologous to a second gene sequence, wherein said first and second gene sequences independently comprise at least a portion of one or more intron regions of a eukaryotic gene.

34. The nucleic acid construct of claim 33, further comprising an AG dinucleotide splice acceptor site.

35. The nucleic acid construct of claim 33, further comprising a GT dinucleotide splice donor site.

36. The nucleic acid construct of claim 33, further comprising a Kozak consensus sequence.

37. The nucleic acid construct of claim 33, wherein said sequence encoding a selectable marker is a nucleotide sequence, which when expressed, confers a phenotype selected from the group consisting of ampicillin resistance, kanamycin resistance, gentamycin resistance, neomycin resistance, puromycin resistance, hygromycin b resistance, thymidine kinase activity, tryptophan synthetase activity, adenine phosphoribosyltransferase activity, dihydrofolate reductase activity, and histidinol dehydrogenase, anthocyanin, beta-glucuronidase and luciferase.
38. The nucleic acid construct of claim 37, wherein said sequence encoding a selectable marker confers puromycin resistance or neomycin resistance.

39. The nucleic acid construct of claim 33, wherein said eukaryotic gene is selected from the genes encoding B7.3, P-selectin, E-selectin, ICAM-1, ICAM-2, VCAM-1, CD28, CD80, CD86, CD154, major histocompatibility complex class I & beta;2-microglobulin, invariant chain, caspase-1, caspase-3, and Gal \( \alpha(1,3) \) galactosyl transferase.

40. The nucleic acid construct of claim 39, wherein said eukaryotic gene is porcine Gal \( \alpha(1,3) \) galactosyl transferase.

41. The nucleic acid construct of claim 40, wherein said first construct sequence and said second construct sequence are independently selected from homologous regions of the intron selected from the group consisting of intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, and intron 9 of the porcine Gal \( \alpha(1,3) \) galactosyl transferase gene.

42. The nucleic acid construct of claim 41, wherein intron 4 has the nucleotide sequence of nucleotides 4938-11716 in Figure 1, intron 5 has the nucleotide sequence of nucleotides 11753-13748 in Figure 1, intron 6 has the nucleotide sequence of nucleotides 13810-14358 in Figure 1, intron 7 has the nucleotide sequence of nucleotides 14463-21627 in Figure 1, and intron 8 has the nucleotide sequence of nucleotides 21766-27048 in Figure 1.

43. The nucleic acid construct of claim 41, wherein said first construct sequence and said second construct sequence are homologous to different regions within the same intron of the eukaryotic gene.
44. The nucleic acid construct of claim 43, wherein said intron is intron 3 of the porcine Gal α(1,3) galactosyl transferase gene.

45. The nucleic acid construct of claim 41, wherein said first construct sequence and said second construct sequence are homologous to different introns of porcine Gal α(1,3) galactosyl transferase gene.

46. The nucleic acid construct of claim 45, wherein said first construct sequence is homologous to an intron 3 region and said second construct sequence is homologous to an intron 4 region of porcine Gal α(1,3) galactosyl transferase.

47. A cell transfected with the nucleic acid construct of claim 33.

48. A cell transfected with the nucleic acid construct of claim 41.

49. A cell transfected with the nucleic acid construct of claim 44.

50. A cell transfected with the nucleic acid construct of claim 46.

51. A bacterial cell transformed with the nucleic acid construct of claim 33.

52. A bacterial cell transformed with the nucleic acid construct of claim 41.

53. A bacterial cell transformed with the nucleic acid construct of claim 44.

55. A bacterial cell transformed with the nucleic acid construct of claim 46.

56. A nucleotide sequence of intron 4 of the Gal α(1,3) galactosyl transferase gene having nucleotides 4938-11716 in Figure 1.
57. A nucleotide sequence of intron 5 of the Gal α(1,3) galactosyl transferase gene having nucleotides 11753-13748 in Figure 1.

58. A nucleotide sequence of intron 6 of the Gal α(1,3) galactosyl transferase gene having nucleotides 13810-14358 in Figure 1.

59. A nucleotide sequence of intron 7 of the Gal α(1,3) galactosyl transferase gene having nucleotides 14463-21627 in Figure 1.

60. A nucleotide sequence of intron 8 of the Gal α(1,3) galactosyl transferase gene having nucleotides 21766-27048 in Figure 1.

61. A lambda phage clone derived from a porcine genomic library comprising at least a portion of the Gal α(1,3) galactosyl transferase gene, wherein the lambda phage clone is selected from the group consisting of pgGT, lambda 1, lambda 2, lambda 4-1 and lambda 8-2.

62. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 33, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

63. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating said embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.
64. A method of making a transgenic mammal comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, and generating offspring from said impregnated host mammal.

65. A transgenic mammal made according to the method of claim 62.

66. A transgenic mammal made according to the method of claim 63.

67. A transgenic mammal made according to the method of claim 64.

68. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 32, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

69. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 44, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

70. A method of reducing transplant rejection comprising transfecting a nuclear donor cell with the nucleic acid construct of claim 46, selecting for transfected cells comprising the nucleic acid of the construct, introducing said selected cells into an embryo, impregnating embryo into an appropriate host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.
host mammal, generating offspring from said impregnated host mammal, harvesting cells, tissue, or organs from said offspring, and transplanting said harvested cells, tissue, or organs into a patient in need thereof.

71. The nucleic acid construct of claim 43, further comprising a nucleic acid sequence encoding a gene which is toxic to said eukaryotic cell

72. The nucleic acid construct of claim 71, wherein said gene which is toxic to said eukaryotic cell is the ricin A toxin gene.
Sequence from within Intron 3 to the end of intron 8

<table>
<thead>
<tr>
<th>Strain</th>
<th>First Few Nucleotides</th>
<th>Remaining Nucleotides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source 1</td>
<td>gtcgactctaa ggcctccagt gctcataact</td>
<td>gggagctcga ggtactaata</td>
</tr>
<tr>
<td>Source 2</td>
<td>gaggccaccc atctccttc gcctcacta</td>
<td>tggagtaatt tagaatattt</td>
</tr>
<tr>
<td>Source 3</td>
<td>aatatagtcc tcttgctaat</td>
<td>aactgctcga cggaggtcct</td>
</tr>
<tr>
<td>Source 4</td>
<td>aatgtacact caaatattag</td>
<td>ttaacactagt ctaactaag</td>
</tr>
<tr>
<td>Source 5</td>
<td>atccagattc tctggctagc</td>
<td>gggagctcga ggtactaata</td>
</tr>
<tr>
<td>Source 6</td>
<td>agttgggaatt gcgtgctagt</td>
<td>aactgctcga cggaggtcct</td>
</tr>
<tr>
<td>Source 7</td>
<td>aatgtacact caaatattag</td>
<td>ttaacactagt ctaactaag</td>
</tr>
<tr>
<td>Source 8</td>
<td>atccagattc tctggctagc</td>
<td>gggagctcga ggtactaata</td>
</tr>
<tr>
<td>Source 9</td>
<td>aatgtacact caaatattag</td>
<td>ttaacactagt ctaactaag</td>
</tr>
<tr>
<td>Source 10</td>
<td>atccagattc tctggctagc</td>
<td>gggagctcga ggtactaata</td>
</tr>
</tbody>
</table>

Figure 1
**Figure 2**

Promoter trap Construct: Insertion of Engineered exon into the intron of a gene.

Endogenous alpha-Gal transferase gene diagram

![Gene Diagram]

- **PCR Primers**
- **sa** = splice acceptor
- **sd** = splice donor
- **pBSK** = pBluescript cloning vector
- **Puro-bGHpolyA** = puromycin resistance with the bovine Growth hormone gene ploy A acceptor signal sequence
Figure 3

A.

10  20  30  40  50
aggccccggAGGCCCTACTGGCCTAATACGACTCATAATTAGGAGCTCGA
CAGCTGAGATCCCGAGTTATGCTGATATCCCTGCAGCT

60  70  80  90  100
GGATCATAATTAGGTCACCATCCCTCTTTCCATGCAATGCTTATAGGAGATTAC
CTAGTTAACATTCTCAGGTTAGGAAAGGACTATTAGGATTTCCGC

110  120  130  140  150
TGTGGAATTAGAATTTTTCAATGCAATATTTTTCATATTTGCTTATATTATGC
ACAACCTATAATTTAAAAAGTTTAGCTTTAATATAATAGGATATAAAC

160  170  180  190  200
AAGGCCTTAAGTGCAATTCATATACACAAATGCTTATTTATTTATG
TTGCAGGGGTCAACAGTCTGAGTTGTTCTCAGTGTATTTGTTAACATGTTATATAAC

210  220  230  240  250
CAAGAAAATATTATAAAGCTCTCACTAAGTAGTTAAGATTATAAATAGGCC
GTTCCTTTAAATTTGTCACGAGTGATTCCTATTACATATATTTTGTCG

260  270  280  290  300
TCGTTACGGGCTCTGCAATGCTAAATTAAAACAGGCTTTTGTTTTT
AGACAGTCCCGGACCTACGTTGACTTTATATTTTGTCGCAAGAAAACCAAA

310  320  330  340  350
CAGAGCTTTGGAGTGGATGAGGTGTGGATGAAAATGCTCTGCTTGATCGTGGGATG
GTCTGACACTGCACTTCACTCAAATCTCAGCGAATGGTCAGGAGTTCAAGGTCC

360  370  380  390  400
TACCCAGTCTGTGTTGGTCTCAGAGCCACCCACCTCTCCTCCAAAGCTCCTTC
ATGGGTCAAGACACCACACAGAGGTGTGGAGAGGAAGGGTTCCGAGAGG

410  420  430  440  450
ACCACACACATCTCTCTGAGATAGCAAAAACGAGAGAGGCAATGCGC
TGAGTTGTTGAGAAGACCTACTTTCTTTGTTGGCGGTCTCTCCCTGAC

460  470  480  490  500
CAGTGCCAGCATCTCCACCCGCTCTGCTTGGCTCCAAACCAGCCCTT
GTCACCGGTTGTCAAGAGGTGGGGCTAGACAGCAGGAGGGTTTGTTGCTGGGAA

510  520  530  540  550
TCCTCGCACTCATGCTTACGCTGTTGATGTTACCAGTCTACATCTCCAGTTACCCGAA

560  570  580  590  600
CCAGCATCTTGGCCGAGTGCATGCTGAGCCTCGTGTATTTCTGTAACCTTTGC
<table>
<thead>
<tr>
<th>610</th>
<th>620</th>
<th>630</th>
<th>640</th>
<th>650</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGTCTAGACCCGGCTCGAGCTATCGGCGGCACTGAAAGCTATGAAACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>660</td>
<td>670</td>
<td>680</td>
<td>690</td>
<td>700</td>
</tr>
<tr>
<td>GGGCGGCGCGGGCGTATGATTGCGACAGAGGTGTGCTACAGGAAAGA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>710</td>
<td>720</td>
<td>730</td>
<td>740</td>
<td>750</td>
</tr>
<tr>
<td>AGCATGTTCAATTCCTAACACCTTGCTCTGACTGAGAGCTACTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>760</td>
<td>770</td>
<td>780</td>
<td>790</td>
<td>800</td>
</tr>
<tr>
<td>TTCTTCACACCTAATTTTGACATGGCATGATGTCAGATATTTTGAGAATACCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>810</td>
<td>820</td>
<td>830</td>
<td>840</td>
<td>850</td>
</tr>
<tr>
<td>TCTTCTTACTGCGATTTGCAACAGAGATGGGTAGAGAGGACAGATCTTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>860</td>
<td>870</td>
<td>880</td>
<td>890</td>
<td>900</td>
</tr>
<tr>
<td>AGCCCTGGAAAGTTGCTAACCTAGAAGCTCTTTCTCTGACATGGAGATG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>910</td>
<td>920</td>
<td>930</td>
<td>940</td>
<td>950</td>
</tr>
<tr>
<td>GTCCCCACCAAGGACACATCTCAGTGCCATATTCACCTTTACATCAGACATC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>960</td>
<td>970</td>
<td>980</td>
<td>990</td>
<td>1000</td>
</tr>
<tr>
<td>CAGGTTATGCTGGCTCAGGGCCGTATGAGGCTGAGTGGAGGTCAGACATTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1060</td>
<td>1070</td>
<td>1080</td>
<td>1090</td>
<td>1100</td>
</tr>
<tr>
<td>GGTAGCTGCAAGGCCTACGCAGGCAACAGACACACCCGGATCTGAGCCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1110</td>
<td>1120</td>
<td>1130</td>
<td>1140</td>
<td>1150</td>
</tr>
<tr>
<td>TCTGGAAACACCTACACACGTACGCGCAACCCGGATCTGAGCAGATACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1160</td>
<td>1170</td>
<td>1180</td>
<td>1190</td>
<td>1200</td>
</tr>
<tr>
<td>AGCAAGGCAGGGTGGAAACCACACACTTCTTCTGAGTCTGAGATTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1210</td>
<td>1220</td>
<td>1230</td>
<td>1240</td>
<td>1250</td>
</tr>
<tr>
<td>TTAACCCAGAGGCAACACCTTGGACCTTGGTACAGTCCAGGTGGGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AATTGTTGTCTCGGTGGGTTCGGCTCTTACCTGCTGACATCAGTCAGAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1050</td>
<td>1060</td>
<td>1070</td>
<td>1080</td>
<td>1090</td>
</tr>
<tr>
<td>CAGGTTATGCTGGCTCAGGGCCGTATGAGGCTGAGTGGAGGTCAGACATTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1110</td>
<td>1120</td>
<td>1130</td>
<td>1140</td>
<td>1150</td>
</tr>
<tr>
<td>TCTGGAAACACCTACACACGTACGCGCAACCCGGATCTGAGCAGATACG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1160</td>
<td>1170</td>
<td>1180</td>
<td>1190</td>
<td>1200</td>
</tr>
<tr>
<td>AGCAAGGCAGGGTGGAAACCACACACTTCTTCTGAGTCTGAGATTCG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1210</td>
<td>1220</td>
<td>1230</td>
<td>1240</td>
<td>1250</td>
</tr>
<tr>
<td>TTAACCCAGAGGCAACACCTTGGACCTTGGTACAGTCCAGGTGGGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AATTGTTGTCTCGGTGGGTTCGGCTCTTACCTGCTGACATCAGTCAGAC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TTGTCTAATCTTCTTTATGTCCTTTCTTTATTATTCGTACCTTTTCAGT

3310 3320 3330 3340 3350
AATATATAGTCGACTCTAACAACCTTAAATAAACTAGTACAGAT
TTATATATCCTCCACACAGTGGTTGTGAATTTTATGTCATATGCTA

3360 3370 3380 3390 3400
GTTAAAAAGACTAAATTATATAGTGGAAGGATAGCTGTGAAGTGAATACCT
CAATTTTCTGATTTTAAATATATACACTTCTACCTAGATATTGA

3410 3420 3430 3440 3450
ATGACATCTAAAAACACAAAAATGTGCGGTCCGCCGTCAGAGCTGGAA
TACTGTAGATTTTGTGTTTAAACCCGAAAGGCACTGCGGCCTGCACTTT

3460 3470 3480 3490 3500
ACGAATCCCCGACTAGGAACCATTAGGTTCCGCTCTCTCTCCCTTGC
TGCTTAGCCTGATCCTCTTGTGACTCCCAACGTCAGTTAAGGACGGAAAGC

3510 3520 3530 3540 3550
TCAGTGCTGGTTAAGGATCCCGGTGTTGCCGCTGAGCTGTGTAGTGGTAGCA
AGTCAACCCCATTTCTAGGCCACAAAGGCACTGCCGACTCACCATCATCGGT

3560 3570 3580 3590 3600
ATGAGGGTTGGATCCCCTTCTTGCTCGTCCTGTGGCTGTGGTAGGCGGTGGCTA
TACTCCCGAACTTGGGCGCAACGCAACCGAGACCCGACATCCGGCCACCGAT

3610 3620 3630 3640 3650
CAGCTCCGATTCCCGACCTCGCTGGGAACTCCACTATGCAGGACGGAGCG
GTCGAGGGCTAAGCTGGGGATCCGACCCTTTGGAGTTATACGGCGCCCTCGC

3660 3670 3680 3690 3700
GGCCCTTAAAAAGACGAAAAAGCCTAGCCTAGAAAAACCAACAAAAACCCAC
CCGGGAATTTTTTCTGTTTTTCTGGTTTTTTTTTTTTTTTTTTTTTGGTG

3710 3720 3730 3740 3750
AAAAAGTGGGAAATACGTCCTCCCTCTACTAGTATTATGTATTGTCACAGTTTT
TTTTACAACCCCTTCTGCGGAGATGATCAATATAATACAATACGTTAA

3760 3770 3780 3790 3800
CCTTTTATGTCCTGTTAAATATTGCTTCTAGATATGTAGTGCTCTGATATC
GGAAAAATACGACAGAAATTATATACAAACAGGAAGATATCTACATCCAGGAGACTTAG

3810 3820 3830 3840 3850
GTGTGCCATATATGTTAACCACATTGTATGTAGTGCTCTGATCCCT
CACAGCTATATATACATTATATGTTAACATACAGAAAGGAGACCATACACTAGGGA

3860 3870 3880 3890 3900
TTTTATATGTTAATGCTACTTTTATTTTATGTATTTGTTACATTCTTTTTGGTTAT
AAACAAATAATACATTACGGGATGAAATAGAAAAAATGTAAGAAACAAATA

3910 3920 3930 3940 3950
GAGTTTTCATGAATATGTCGCTACCGTCCACACCTTTTCTTGTCCCTTAT
CTCACAACGACTATAACCCGTGCAGGAGGTTGTGGAAGAAGAACAAGGAAAGGTA
B.
Ligated to the following intron 4 splice acceptor sequence

Intron 4 splice acceptor sequence

\[
\begin{align*}
3960 & \quad 3970 & \quad 3980 & \quad 3990 & \quad 4000 \\
TTACAATAATATTTTCTATCTCCACCCAAATTTAAGTACTCCGCAACC & \\
AATGTTATTTTATAGAAAGATAGAGGTTGGTTTTAATTTCATGAGCGTGG & \\
\end{align*}
\]

\[
\begin{align*}
4010 & \quad 4020 \\
TGTATATCCACCCAGCATCCgtcgcG & \\
ACAATAAGGTGGTTGACTAGGcagcctg & \\
\end{align*}
\]

\[
\begin{align*}
11530 & \quad 11540 & \quad 11550 \\
\text{gtcgcG} & \text{CCACCGTTTGAGTCTGAATTTCTGAAATG} & \\
\text{cagcctg} & \text{GCTGGCAACTAGACTCATTAAGACTTTAC} & \\
\end{align*}
\]

\[
\begin{align*}
11560 & \quad 11570 & \quad 11580 & \quad 11590 & \quad 11600 \\
ACGAGAGTCCCCGATATCATTTTTTTCTCTGATCTCGAAGGTIGAGTTCAACCTGG & \\
TGCTCTCACGGGCACTATAGTTAAAAAAGCTAGAGTCCCTACCCTTTGAGG& \\
\end{align*}
\]

\[
\begin{align*}
11610 & \quad 11620 & \quad 11630 & \quad 11640 & \quad 11650 \\
AGTAGCCCAAACCAGGCTCTCAGTCGAGCCTAGGGTTTTCAATGATAATG & \\
TCATCGGTGTGGGTCCGAGAGTCGACTCGGATCCCCAAAGTTACTTTAC & \\
\end{align*}
\]

\[
\begin{align*}
11660 & \quad 11670 & \quad 11680 & \quad 11690 & \quad 11700 \\
ATTGCAAATAGCTTTTCCTCAGGATCCAACTGATAGATGTTTTTTAT & \\
TAACGTCTTTTATCGAAAGAGACGTGCCTTGATCTATACAAAAATA & \\
\end{align*}
\]

\[
\begin{align*}
11710 \\
TTCCATTTGCTTTTAGgaatc & \\
AAGGTAACGAAAATCcttaag & \\
\end{align*}
\]

C.
Ligated to Neomycin Resistance gene

\[
\begin{align*}
\text{gaatcAATGGATCCCACCCATGG-NEO} & \quad \text{-GGATATCCACTACTTAGAATAGGCGaagctt} & \\
\text{CCTATAGGTGATGAATCATTATCGGtctgaa} & \\
\end{align*}
\]
D.

Ligated to the intron 4 splice donor sequence

\[
\begin{align*}
4940 & \quad 4950 \\
\text{aagctt} & \quad \text{GTAATTATGAAAC} \\
\text{ttcga} & \quad \text{CATTAATACCTTTG} \\
4960 & \quad 4970 \\
\text{ATGATGAAATGATGTTGATGAAAGTCTCTCTCTCTAAATCTCCCTACGGTTATCAGC} \\
\text{TACTACTTTTACTACACATTTTCAGAGGAGATTAATCGAGATCAATAGTGC} \\
5010 & \quad 5020 \\
\text{CAAGTCAAGCTTGCAATTAAAAGTGATTCTACGTACCGGTAAGAAAGA} \\
\text{GTTCAAGGTGTGAGCAGAATTATTTTCATCTAATGACTGTTGGCATTTCTTT} \\
5060 & \quad 5070 \\
\text{GCATTCCAGAGAGTGGTGCTGCTAGGACGGAGCACAACCCCAATTAGG} \\
\text{CGTAAGGTCTCTCAACGGCAACACCGCGTGTTTTGTTTAAATCC} \\
5110 & \quad 5120 \\
\text{ATTTCAAGAGGAAGGTGGTGGTTTGATTCTGGCCTGCTTCTTTGGCTTTAAGGA} \\
\text{TAGTGTCTCTCCACCCCAAACATCGGACGGCAACCAATTTCCT} \\
5160 & \quad 5170 \\
\text{TCCGGCATTGGCGAGCTGTGGGctgcag} \\
\text{AGGCCGTAACGGCACTGGGACACCagcgtc}
\end{align*}
\]

E.

Ligated to intron 3 3' sequence

\[
\begin{align*}
\text{ctgcag} & \quad \text{CCCTCAACTACAATTTTCATGCAAGC} \\
\text{GGGAGAAGTTGATGTTAAGTACTCGG} \\
4060 & \quad 4070 \\
\text{GATCAAGAAATAGAATGATCCGACGTTGCTCCATTGGGGCGATGGAA} \\
\text{CTAGTTCTTTATCTTTATCGCTGACAACCGGTAAACCAGCCCCTACCGTT} \\
4110 & \quad 4120 \\
\text{AAGTGTTGCGAGAGACGTTAGATTATAAGGCACGGGTGAAAGGT} \\
\text{TTCACCCACCTCCCAGCCTGCAATCTAATATTATCCGATCCCACCTCTCAA} \\
4160 & \quad 4170 \\
\text{CCCATTTGTGTCAGCTGAACTGAATCTGACTACCATCGATGACAGGAA} \\
\text{GGTAACACCCAGTGACTTTACATAGACTGAGTTCAGTTACTGCGTCTTC} \\
4210 & \quad 4220 \\
\end{align*}
\]
Figure 4

taataatgga tcccccaccAT GGGCATTGAG CAGGACGGCG TGGCACGCGGG CAGCCCGCGCC
GCGGCTGGTT AGAGACTGTT CGGTAACGAC TGGCCCCAGC AGACCAGCAG CTGCAGCGAC
GCCGCGGTGG TCAGACTGAG GCCCCAGGCG AGACCCGTGC TGGTCTGAA GACCGACCTG
AGGGCCGCCC TGAAGAGGCT GCAGAGGAGG GGGCCGAGC TGGCTGGCT GGGCCACCACC
GGGGTCCCT TCCCCCGGGT CTGGGACGTG GTGACCAGGG GGGCGAGGAG CGGCTGGCTG
CTGGGCGGAG TGCCCCGCGGA CGACCTGCTG AGCAGCCACC TGGCCCCCGC CGAGAAGGTG
AGCATCATGG CCGACGCCCC GAGAAAGACTG CACACCTGCG ACCCGCCAC CTGCCCCCTTC
GACCACCGG CCAAGAACAG ATCGAGAGA GGGAGAACCA GAATGGAGGC CGGCTGGCTG
GACCAGGACG ACCTGAGGGA GGAGGAGGAC GGGCGGCGT GGGCGAGCT GTTGGCCAGA
CTGAAAGCGA GAATGCCCGA CGCAGAGGAT CTTGAGGCTG CCCACGGCGA CGCCTGGCTG
CCCAACGCTG TGGCTGAGAA CGCGAGGATC AGCGGCGTCA TCGACTGGGG CAGACTGGGC
GTCGCGAGA CATCACAGGA CAGGGCTGCT GCGACGAGA AGACTGGCG AGGACGTGGGC
GGGAGGATGG CGCGAGGAGG CCTGAGCTGAT AGGCGAGGCT CGCCCGCGC CGGCGAAGAG
ATGCGCTTTCT ACAGACTGTG GAGAGAGTTC TTCTGAATCT TGCGAGCTGG GGGCTTGTGG
* ATATCCACTA GTTACGccga ttctatat
Figure 5

1 ctgaagcttt ccatgaccga gtacagccccc acgggtcgcct tgcacaccccc gcagcgacgc
61 ccgccggccgc taagcactct ccgcccgcggc ttcgccagact accggccgcccc gcggggcacc
121 gtcagccgct accggccgct cagacggctgc acgggctgcc aagaacctgct cctcagcgcc
181 gtcgctcctc acacggccac aatcggcgtc ccggagccgcc gcggagcgcc gcggctctgg
241 accacgcgct agacgcgcgc aatcggcgtc ccggagccgcc gcggagcgcc gcggctctgg
301 gatgtttagcgc gttcggcgctc ggccggcggc cagagctgctg aagggccccct gcgggcgcac
361 ccgggccccag gcggccgctgc ccggtccggtg ccggtccggtg cctcgcctg ccaccaggcg
421 aagggctctgc gcacggccgct cgtgtctcccc gggttaaggg ccgggccccg ccgggccccg
481 ccagccccct tggagacccc cggccgcggcc accctctcct tctacggcgc gtcgggccctc
541 accgtcactgc cgagctctgc cgtgccgcag ccggagcccga cctgtttcagc gacccgcaag
601 ccgcgtgctc cagccggccgc ccaagcggccg cagcgcggcgc cccaaaggag agcagccgacg
661 cactcatcgc tgcctctagc ccgggtgtacg acgcctgccg ctggcctcctcg ttggccagcc
721 atcttggtgtctgc gcgcctctct ccgggtggtg ctcgggctg gagtctgctgc cctccactctt
781 ccctctctaa taaatgagac aatagggctgc ccagtgtgtgc acttagggact atttcatcat
841 ggggggttggt gcgggtctgc gcagccccgg cagagctggtg cagagcaata gcagagctgc
901 tcgggatgcgc gcggcctctc cggctgtgctgc gcgggaaaaga accagcctcgc a
**Figure 6**

- PCR primer positions
- SA: splice acceptor, AG dinucleotide
- SD: splice donor, GT dinucleotide
- KCS: Kozak consensus sequence
- Puro bPA: puromycin gene coding sequence and bovine growth hormone poly A signal sequence
- pBS: Bluescript cloning vector
<table>
<thead>
<tr>
<th>Figure 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  gtcgactctta gagctctacg gctattata gggagtctga ggtacgaata</td>
</tr>
<tr>
<td>61  gagctccac gtcctcttttc tcctgtctca gaggccgata tgtggaat ttagatataa</td>
</tr>
<tr>
<td>121  caagtcgaca atatttatttc attatttagtt actcactata agaaccccca</td>
</tr>
<tr>
<td>181  aatgtgatc ctataatattg cagacagact caggtatgtt gtttgttgtt</td>
</tr>
<tr>
<td>241  atataatatgc tctgtgaagg gctcgtgatc cagctagttt tataacaggg</td>
</tr>
<tr>
<td>301  cagaggtcgg agttgtagta gggctcgagc aactctccca tgtgtctttc tacccagcctt</td>
</tr>
<tr>
<td>361  tgtgtggttct ctagccccca ctccttccct aagcccttca aaccacccca cctcctttgag</td>
</tr>
<tr>
<td>421  agtggagcag aagccagaga gctcctggca cagttgctgct tgggtctgtc cagctcgctg</td>
</tr>
<tr>
<td>481  tgtgcccctaa ccagccctcc ggagcccttc cagcgtctac gctgtgcttt tcctgtaggt</td>
</tr>
<tr>
<td>541  gttgggtctct gatgcagctcg cagctgctgt tcgcggcttc ggctgatgtc gtaatacagt</td>
</tr>
<tr>
<td>601  cctgggccccc ggctcggttt ttgctcgcag ttggtgcgct acaggtcatc acagctcact</td>
</tr>
<tr>
<td>661  cccacatccttat cccaattttag cccaattttc aagcccttca gctcctggca cagcgtctac</td>
</tr>
<tr>
<td>721  tgggtgatc tgggtgtctt ccctcctttg aagcccttca gctcctggca cagcgtctac</td>
</tr>
<tr>
<td>781  gttgataact ccaacttttt ttcttctttt ggttctttgag gcaacatag cagcgtctac</td>
</tr>
<tr>
<td>841  ctggaactct tggagacgat cagccctcag ccctcctttg aagcccttca gctcctggca</td>
</tr>
<tr>
<td>901  tggcccttcag aagacacata ctagctgcct ttcatttttt ttcctctttt cagcgtctac</td>
</tr>
<tr>
<td>961  ttttttttttt tggcctttt cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>1021  ggtgcctcag ctgacggtct taaaagttgc ccctcttcatt tggggtgtctt cagcgtctac</td>
</tr>
<tr>
<td>1081  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>1141  tggggttgggg ggggtgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>1201  ttaacggtttc ccacaaatct cccctttctt ctagcttttt tggggtgtttc cagcgtctac</td>
</tr>
<tr>
<td>1261  ttggagcttc gacttttttt cccctttttt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>1321  tggagagttc cactccccca cgggggtttt cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>1381  agggtctagc aagacacata ctagctgcct ttcatttttt ttcctctttt cagcgtctac</td>
</tr>
<tr>
<td>1441  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>1501  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>1561  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>1621  ttaacggtttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>1681  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>1741  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>1801  ttaacggtttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>1861  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>1921  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>1981  tggagagttc cactccccca cgggggtttt cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>2041  tggagagttc cactccccca cgggggtttt cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>2101  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2161  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>2221  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>2281  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>2341  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2401  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2461  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2521  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2581  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2641  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>2701  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2761  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2821  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>2881  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>2941  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>3001  aagccaggtg tggcccttcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
<tr>
<td>3061  tggggttgggg ggggctttt cggggtgggg cggggtgggg cggggtgggg cggggtgggg</td>
</tr>
<tr>
<td>3121  cagcgtctttc ccacaaatct cccctttctt ctagcttttt tggggtgtttt cagcgtctac</td>
</tr>
<tr>
<td>3181  ggcagctcag cagcgtctac cagcgtctac cagcgtctac cagcgtctac cagcgtctac</td>
</tr>
</tbody>
</table>
INTRON 3 ------- puromycin bpolyA ------- beginning of intron 4
**Figure 9**

Gal\(\alpha\)1-3Galactosyl Transferase Collision Intron Targeting Construct

Diagram showing the targeting construct with specific markers and distances labeled. The diagram includes symbols for PCR primer positions, splice acceptor (SA), splice donor (SD), PGK-Puro-bpA, pBS KS+, CMV-Ricin-SV40pA.

- PCR primer positions
- SA: splice acceptor, AG dinucleotide
- SD: splice donor, GT dinucleotide
- PGK-Puro-bpA: PGK promoter, puromycin gene coding sequence and bovine growth hormone poly A signal sequence
- pBS KS+: Bluescript cloning vector
- CMV-Ricin-SV40pA: CMV promoter, Ricin gene and SV40 poly A