US 20010056581A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2001/0056581 A1

(43) Pub. Date: Dec. 27, 2001

Gaitanaris

(54) VECTORS AND METHODS FOR THE MUTAGENESIS OF MAMMALIAN GENES

(76) Inventor: **George A. Gaitanaris**, Frederick, MD (US)

Correspondence Address:
CLARK & ELBING LLP
176 FEDERAL STREET
BOSTON, MA 02110-2214 (US)

(21) Appl. No.: **09/931,957**

(22) Filed: **Aug. 16, 2001**

Related U.S. Application Data

(60) Continuation of application No. 09/847,090, filed on May 1, 2001, which is a division of application No. 09/002,046, filed on Dec. 31, 1997, now Pat. No. 6,228,639, which is a non-provisional of provisional application No. 60/034,094, filed on Dec. 31, 1996.

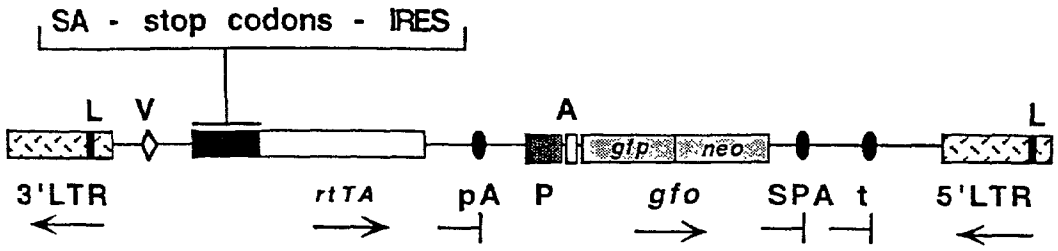
Publication Classification

(51) **Int. Cl.**[7] **A01K 67/027**; C12N 15/867

(52) **U.S. Cl.** **800/14**; 435/456

(57) **ABSTRACT**

Disclosed herein are methods for mutagenizing a mammalian gene, the methods involving introducing into a mammalian cell a retroviral vector which includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences, the introducing step being carried out under conditions which allow the vector to integrate into the genome of the cell. Also disclosed are retroviral vectors for use in these methods as well as methods for the use of mutagenized cells.



LTR, long terminal repeats

L, loxP

V, VDE recognition sequence

SA, splice acceptor

IRES, internal ribosome entry site

rtTA, reverse tetracycline dependent transactivator

pA, polyA addition signal

P, promoter

A, adenovirus tripartite leader sequences

gfo, *gfp-neo* hybrid gene

gfp, green fluorescent protein

neo, neomycin phosphotransferase gene

SPA, synthetic polyA site

t, transcription terminator

Figure 1

Insertional Mutagenesis

A. target gene

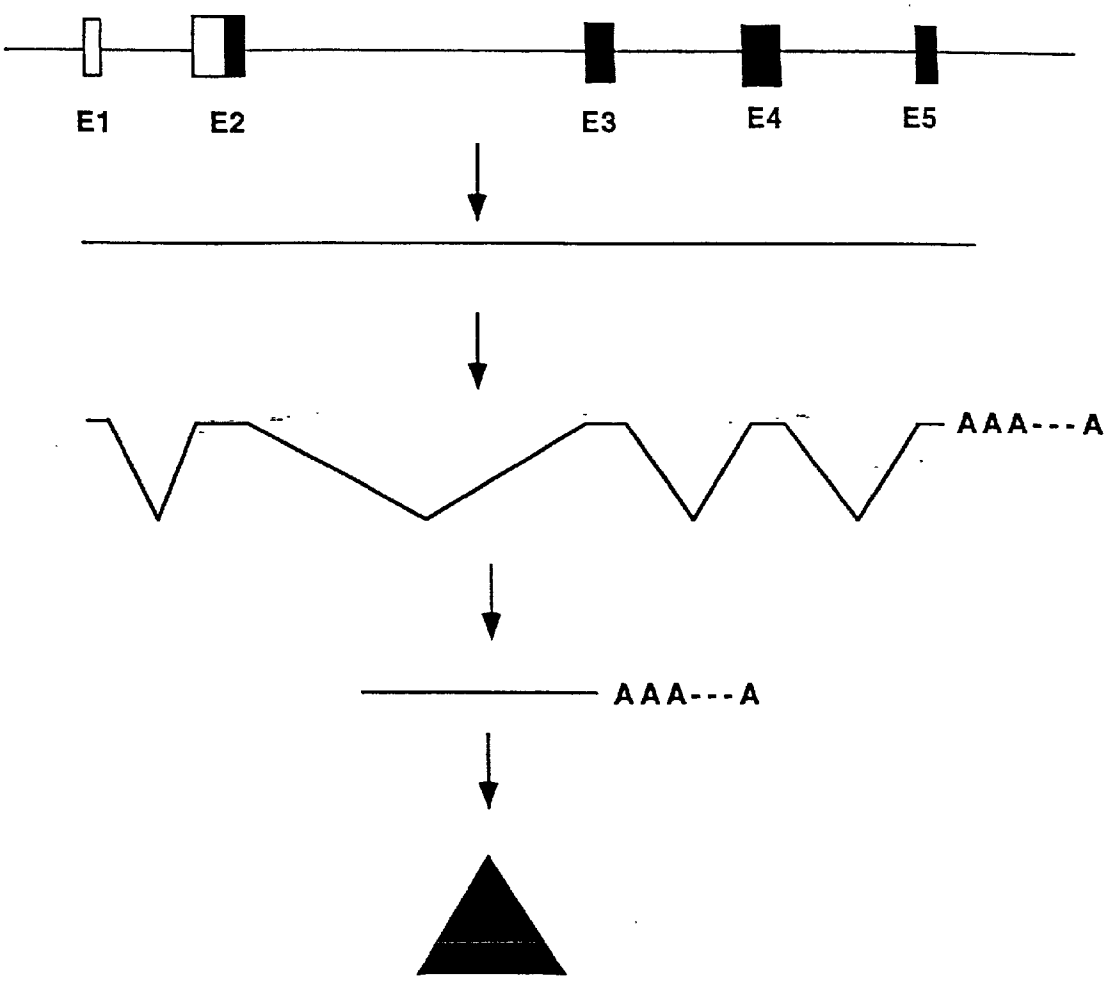
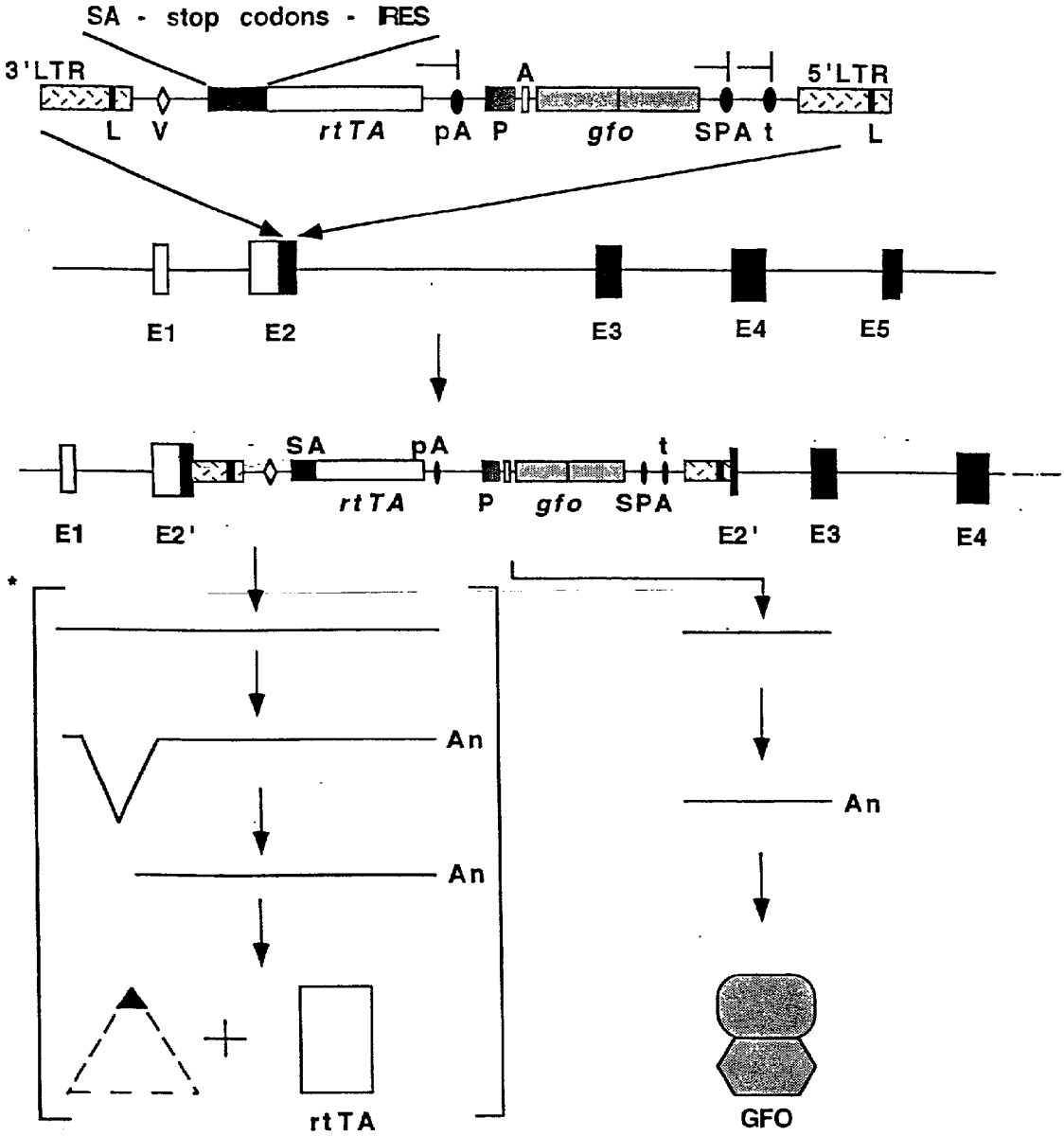


Figure 2

Insertional Mutagenesis

B. insertion into an exon

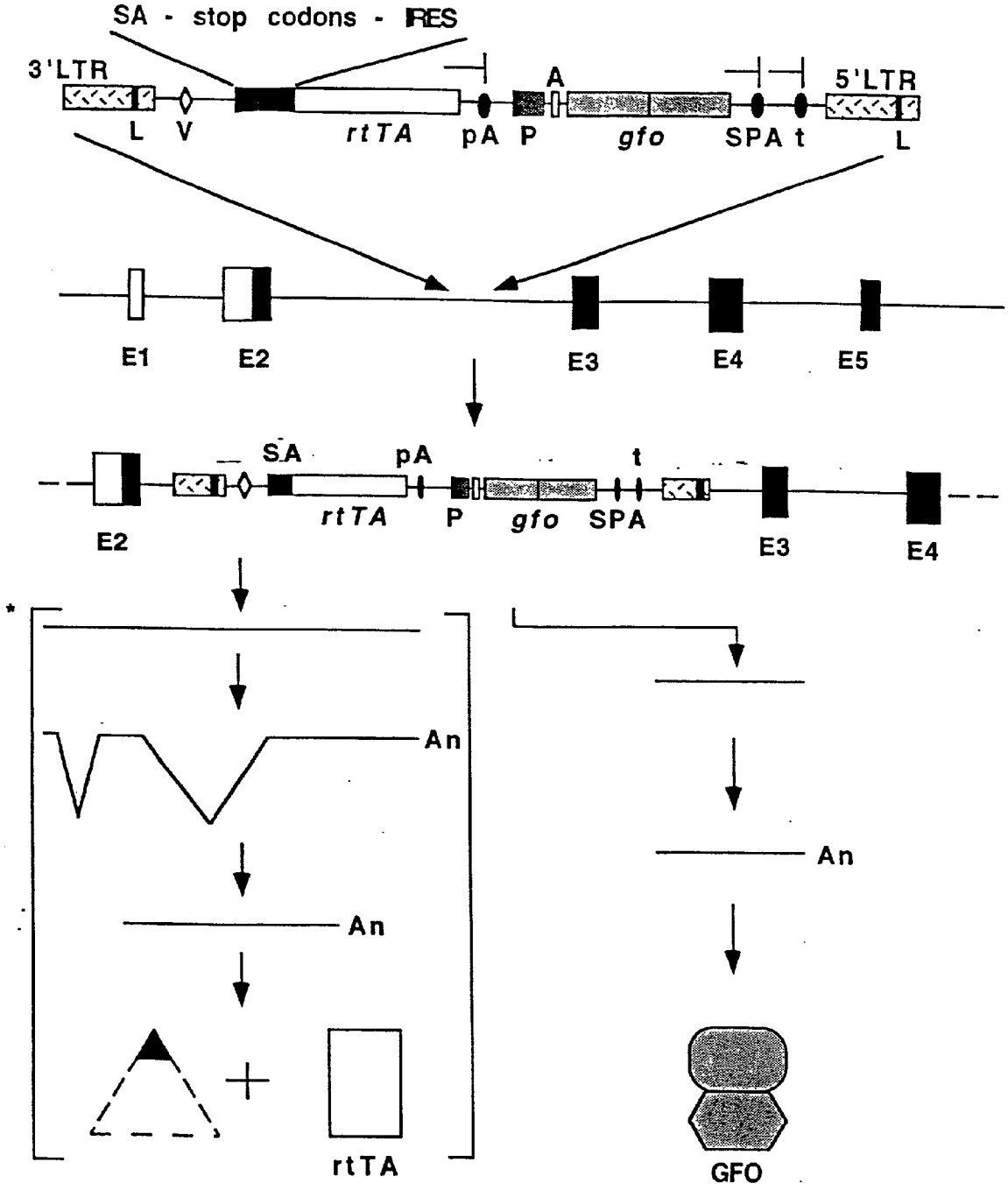


*pathway active only when gene X is expressed

Figure 3

Insertional Mutagenesis

C. insertion into an intron



*pathway active only when gene X is expressed

Figure 4

VECTORS AND METHODS FOR THE MUTAGENESIS OF MAMMALIAN GENES

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application is a continuation of and claims priority from U.S. Pat. application 09/847,090, filed May 1, 2001 which is a divisional of U.S. utility application U.S.S.N. 09/002,046, filed Dec. 31, 1997 (now U.S. Pat. No. 6,228,639), which claims benefit from U.S. provisional application U.S.S.N. 60/034,094, filed Dec. 31, 1996 (now abandoned).

BACKGROUND OF THE INVENTION

[0002] This invention relates to retroviral vectors and their use in methods of mammalian gene mutagenesis.

[0003] Eukaryotic genomes are estimated to contain 6,000-80,000 genes (Collins, Proc. Natl. Acad. Sci. USA 92:10821-10823 (1995)). Even in the best characterized organisms, the function of the majority of these genes is unknown. In addition, relatively little information is available concerning the fraction of the genome that is expressed in particular cell types or the cellular processes in which specific gene products participate.

[0004] In an attempt to decipher genes' functions, large scale mutagenesis screens have been developed and have proven instrumental in unraveling the roles of certain genes in organisms such as *Drosophila melanogaster* (Nusslein-Volhard and Wieschaus, Nature 287:795-801 (1980); Ballinger and Benzer, Proc. Natl. Acad. Sci. USA 86:9402-9406 (1989); Kaiser and Goodwin, Proc. Natl. Acad. Sci. USA 87:1686-1690 (1990); and Spradling et al., Proc. Natl. Acad. Sci. USA 92:10824-10830 (1995)), *Caenorhabditis elegans* (Hirsh and Vanderslice, Dev Biol. 49:220-235 (1976); and Zwaal et al., Proc.

[0005] Natl. Acad. Sci. USA 90:7431-7435 (1993)), *Zebrafish* (Solnica-Krezel et al., Genetics 136:1401-1420 (1994); and Riley and Grunwald, Proc. Natl. Acad. Sci. USA 92:5997-6001 (1995)), *Arabidopsis* (Jurgens et al., Development Suppl. 1:27-38 (1991); Mayer et al., Nature 353:402-407 (1991); and Sundaresan et al., Genes Dev. 9:1797-1810 (1995)), *Maize* (Scanlon et al., Genetics 136:281-294 (1994); and Osborne and Baker, Curr. Opin. Cell Biol. 7:406-413 (1995)), and *Saccharomyces cerevisiae* (Burns et al., Genes Dev. 8: 1087-1105 (1994); and Chun and Goebl, Genetics 142:30-50 (1996)). In mammals, however, these approaches have generally been limited by the large genome size and the development of the embryo inside a mother's uterus.

[0006] Some progress has been made in understanding mammalian gene function as a result of the development of mouse embryonic stem (ES) cell technology. This technology has significantly altered the field of mammalian genetics by allowing the bulk of genetic manipulations to be executed in vitro (Evans and Kaufman, Nature 292:154-156 (1981); Bradley et al., Nature 309:255-256 (1984); and Robertson, Trends Genet. 2:9-13 (1986)). This is possible because mouse ES cells are pluripotent, that is, they have the ability to generate entirely ES cell-derived animals. Accordingly, gene inactivation in mouse ES cells and subsequent generation of "knock-out" (KO) mice is a powerful method for gaining information about the function of a gene in a whole animal system. If desired, genetic alterations, such as gene KOs which inactivate genes, may be introduced into these cells, and their consequences may be studied in the whole animal (Jaenisch, Science 240:1468-1474 (1988); and Rossant and Nagy, Nat. Med. 1:592-594 (1995)).

[0007] Currently, the available mouse mutagenesis methodologies are somewhat limited in their general utility as gene function screening systems. Gene targeting, the most widely used approach, is laborious and time consuming (Capecchi, Science 244:1288-1292 (1989)). And gene trap and chemical/radiation induced mutagenesis are generally restricted in their targets (Gossler et al., Science 244:463-465 (1989); Friedrich and Soriano, Genes Dev. 5:1513-1523 (1991); Skarnes et al., Genes Dev. 6:903-918 (1992); von Melchner et al., Genes Dev. 6:919-927 (1992); Reddy et al., Proc. Natl. Acad. Sci. USA 89:6721-6725 (1992); Takeuchi et al., Genes Dev. 9:1211-1222 (1995); and Takahashi et al., Science 264:1724-1733 (1994)). The gene trap approach is limited to genes expressed in ES cells, although variations of the method have been developed for targeting specific subclasses of genes expressed in early embryonic stages (Wurst et al., Genetics 139:889-899 (1995); Skames et al., Proc. Natl. Acad. Sci. USA 92:6592-6596 (1995); and Forrester et al., Proc. Natl. Acad. Sci. USA 93:1677-1682 (1996)). And the chemical/radiation induced mutagenesis technique is generally limited to genes that can result in dominant phenotypes when mutated. None of these approaches, as currently exploited, may be readily streamlined or automated, nor can they be readily adapted to carry out saturated mutagenesis of the mouse genome.

SUMMARY OF THE INVENTION

[0008] In general, the invention features a method for mutagenizing a mammalian gene, the method involving introducing into a mammalian cell (for example, a stem cell, such as an embryonic stem cell) a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences, the introducing step being carried out under conditions which allow the vector to integrate into the genome of the cell.

[0009] In preferred embodiments, the retroviral vector includes packaging and integration sequences derived from a Moloney murine leukemia virus sequence; the retroviral vector further includes a reporter gene whose expression is under the control of a mammalian cell promoter, the promoter being operably linked to the reporter gene upon integration of the vector into the genome of the mammalian cell; the reporter gene encodes a regulatory protein, the regulatory protein being capable of modulating the expression of a detectable gene; the regulatory protein is a tetracycline repressor fused to an activator protein (for example, VP16); the retroviral vector further includes a DNA sequence encoding a constitutively expressed marker gene, the marker gene being detectable in a mammalian cell; the marker gene is a green fluorescent protein (for example, a green fluorescent having increased cellular fluorescence relative to a wild type green fluorescent protein); the green fluorescent protein is fused to a mammalian selectable marker; the mammalian selectable marker encodes neomycin resistance; the retroviral vector further includes a recognition sequence derived from a yeast VDE DNA endo-

nuclease; the retroviral vector further includes a sequence which is recognized by a recombinase enzyme (for example, a loxP sequence); the mammal is a mouse; and the cell is an embryonic stem cell.

[0010] In a related embodiment, the invention features a retroviral vector which includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences. In preferred embodiments, the retroviral vector includes packaging and integration sequences derived from a Moloney murine leukemia virus sequence; the retroviral vector further includes a reporter gene whose expression is under the control of a mammalian cell promoter, the promoter being operably linked to the reporter gene upon integration of the vector into the genome of the mammalian cell; the reporter gene encodes a regulatory protein, the regulatory protein being capable of modulating the expression of a detectable gene; the regulatory protein is a tetracycline repressor fused to an activator protein (for example, VP16); the detectable gene includes an operably linked tetracycline operator; the retroviral vector further includes a DNA sequence encoding a constitutively expressed marker gene, the marker gene being detectable in a mammalian cell; the marker gene is a green fluorescent protein (for example, a green fluorescent protein having increased cellular fluorescence relative to a wild type green fluorescent protein); the green fluorescent protein is fused to a mammalian selectable marker; the mammalian selectable marker encodes neomycin resistance; the retroviral vector further includes a recognition sequence derived from a yeast VDE DNA endonuclease; and the retroviral vector further includes a sequence which is recognized by a recombinase enzyme (for example, a loxP sequence).

[0011] In other related embodiments, the invention includes a cell containing a retroviral vector of the invention; a transgenic non-human mammal (for example, a mouse) which includes a retroviral vector of the invention; a library (that is, having at least 100 members) of mutagenized mammalian genes produced by the methods of the invention;

[0012] and cells (for example, stem cells) which include a library of mutagenized mammalian genes produced by the methods of the invention.

[0013] In a related method, the invention features a method for identifying a cell (for example, a stem cell) which includes a retroviral vector, the method involving: (a) introducing into a mammalian cell population a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, retroviral packaging and integration sequences, and a constitutively expressed detectable marker gene, the introducing step being carried out under conditions which allow the vector to integrate into the genomes of the cells; and (b) identifying the cell which includes the retroviral vector by detecting expression of the marker gene.

[0014] In preferred embodiments, the marker gene is a green fluorescent protein; and the green fluorescent protein has increased cellular fluorescence relative to the wild-type green fluorescent protein.

[0015] In a second related method, the invention features a method for identifying a mutagenized mammalian gene, the method involving: (a) introducing into a mammalian cell (for example, a stem cell) population a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences, the introducing step being carried out under conditions which allow the vector to integrate into the genomes of the cells; (b) isolating the genomic DNA from the population of cells; (c) amplifying the genomic DNA using amplification primers based at least in part on the retroviral sequence; and (d) identifying the mutagenized mammalian gene by sequence homology with a wild-type nucleic acid sequence. In a preferred embodiment, the sequence homology is identified using a hybridization technique.

[0016] In a third related method, the invention features a method of conditionally ablating a cell lineage, the method involving: (a) providing a first transgenic non-human mammal which includes an activator protein expressed only in the cell lineage; (b) providing a second transgenic non-human mammal which includes a nucleic acid sequence encoding a cell ablation factor, the nucleic acid sequence being under the control of the activator protein and the activator protein being capable of binding to and regulating the nucleic acid sequence only upon induction; (c) mating the first and the second transgenic mammals to produce offspring in which the cell ablation factor is expressed under the control of the activator protein, the cell ablation factor being capable of destroying cells in which it is expressed; and (d) inducing binding and regulation by the activator protein.

[0017] In preferred embodiments, the activator protein is introduced into the transgenic non-human mammal on a retroviral vector that includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences;

[0018] the activator protein is a tetracycline repressor fused to VP16 and the nucleic acid sequence encoding a cell ablation factor is operably linked to a tetracycline operator; the cell ablation factor is chosen from the group consisting of a toxin, a thymidine kinase, or an apoptotic protein; the conditional induction occurs by administration of tetracycline or a tetracycline derivative to the transgenic mammal; and the mammal is a mouse.

[0019] In a fourth related method, the invention features a method for conditional ectopic expression of a gene of interest, the method involving: (a) providing a first transgenic non-human mammal which includes an activator protein expressed under the control of the promoter of an endogenous gene of the mammal; (b) providing a second transgenic non-human mammal which includes a nucleic acid sequence encoding the gene of interest, the nucleic acid sequence being under the control of the activator protein and the activator protein being capable of binding to and regulating the nucleic acid sequence only upon induction; (c) mating the first and the second transgenic mammals to produce offspring in which the gene of interest is expressed under the control of the activator protein; and (d) inducing expression of the activator protein.

[0020] In preferred embodiments, the activator protein is introduced into the transgenic non-human mammal on a retroviral vector that includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences;

[0021] the activator protein is a tetracycline repressor fused to VP16 and the nucleic acid sequence encoding a cell ablation factor is operably linked to a tetracycline operator; the induction occurs by administration of tetracycline or a tetracycline derivative to the transgenic mammal; and the mammal is a mouse.

[0022] In a fifth related method, the invention features a method of generating a non-human transgenic mammal having a conditional malignancy, the method involving: (a) providing a first transgenic non-human mammal which includes an activator protein expressed under the control of the promoter of an endogenous gene of the mammal; (b) providing a second transgenic non-human mammal which includes a nucleic acid sequence encoding a neoplastic factor, the nucleic acid sequence being under the control of the activator protein and the activator protein being capable of binding to and regulating the nucleic acid sequence only upon induction; (c) mating the first and the second transgenic mammals to produce offspring in which the neoplastic factor is expressed under the control of the activator protein, the neoplastic factor being capable of promoting the development of the malignancy; and (d) inducing binding and regulation by the activator protein.

[0023] In preferred embodiments, the activator protein is introduced into the transgenic non-human mammal on a retroviral vector that includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences; the activator protein is a tetracycline repressor fused to VP16 and the nucleic acid sequence encoding a cell ablation factor is operably linked to a tetracycline operator; the neoplastic factor is an oncogene; the induction occurs by administration of tetracycline or a tetracycline derivative to the transgenic mammal; and the mammal is a mouse.

[0024] The invention also features a cell line derived from one of these transgenic non-human mammals, as well as transgenic mosaic non-human mammals generated by the methods of the invention and uses therefor.

[0025] In a final related method, the invention features a method for conditional tissue-specific inactivation of a gene of interest, the method involving: (a) providing a first transgenic non-human mammal which includes an activator protein expressed under the control of the promoter of the endogenous gene of interest; (b) providing a second transgenic non-human mammal which includes a ribozyme gene under the control of the activator protein, the ribozyme being capable of specifically interfering with expression of the gene of interest and the ribozyme being produced only upon induction; (c) mating the first and the second transgenic mammals to produce offspring in which the ribozyme is expressed under the control of the activator protein; and (d) inducing expression of the activator protein, whereby the gene of interest is inactivated in cells in which it is endogenously expressed.

[0026] In preferred embodiments, the activator protein is introduced into the transgenic non-human mammal on a retroviral vector that includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences; the activator protein is a tetracycline repressor fused to VP16 and the nucleic acid sequence encoding a ribozyme is operably linked to a tetracycline operator; induction occurs by administration of tetracycline or a tetracycline derivative to the transgenic mammal; and the mammal is a mouse.

[0027] The present invention provides a number of advantages. For example, it combines versatile retroviral vectors for ES cell mutagenesis with powerful detection methods for rapid identification of mutant cells of interest. In addition, the method permits mutagenesis in a large number of mammalian genes, in a short period of time, and at a significant reduction in cost. Moreover, the method may be readily streamlined, and many genes may be processed in parallel. Considering that every gene is a potential mutagenesis target, the proposed approach facilitates the generation of extensive libraries of mutated mammalian genes, as well as libraries of pluripotent stem cells carrying those gene mutations.

[0028] Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

DETAILED DESCRIPTION

[0029] The drawings will first briefly be described.

[0030] **FIG. 1** is a schematic representation of a Moloney murine leukemia virus (MoMLV)-based vector for use in the MAGEKO process.

[0031] **FIG. 2** is a schematic representation of an insertional mutagenesis event.

[0032] **FIG. 3** is a schematic representation of the MAGEKO process of insertional mutagenesis in an exon sequence.

[0033] **FIG. 4** is a schematic representation of the MAGEKO process of insertional mutagenesis in an intron sequence.

[0034] The present invention involves vectors and a process, termed "MAGEKO" (or "massively parallel gene knock out")which permits the mutagenesis of large numbers of mammalian genes, the creation of libraries containing those mutant genes, and the ready selection from that library of stem cells carrying mutant genes of interest. Although this process is applicable to any mammalian system, it is now described for the generation of mutations and libraries in a mouse system. The following examples are presented for the purpose of illustrating the invention, and should not be construed as limiting.

[0035] The MAGEKO Process

[0036] The MAGEKO process involves retroviral insertional mutagenesis, on average every 1 Kb in the mouse genome, to create a comprehensive library of KO ("LOK") embryonic stem (ES) cells, and a gene KO identification system ("KIS"). The LOK generally includes mutations in every mouse gene, and the KIS allows the rapid isolation of desired mutant ES cells. The LOK and KIS facilitate the large scale automated search for KO cells potentially corresponding to any desired gene.

[0037] Once appropriate ES cells are identified, ES cell-derived embryos are generated in vitro, by aggregation with tetraploid or morulae stage embryos (for example, by the method of Wood et al., Nature 365:87-89 (1993)). These embryos are subsequently implanted into foster mothers for the generation of heterozygotic mice with a KO in the gene

of interest. Conventional blastocyst injection methods can also be employed, if appropriate (see, for example, Robertson, Trends Genet. 2:9-13 (1986)). Heterozygotic mice are converted to homozygotes through mating.

[**0038**] In parallel, the heterozygotic mutant ES cells may also be converted to homozygotic cells invitro, according to published protocols (for example, Mortensen et al., Mol. Cell. Biol. 12:2391-2395 (1992)), and used to generate homozygotic mice with the above described techniques. The homozygotic mice obtained by either method may be analyzed to determine the function of the knocked out gene of interest.

[**0039**] The MAGEKO Components

[**0040**] The MAGEKO process broadly encompasses three components: (i) the generation of gene mutations in mammalian genes using retroviral vectors; (ii) the production of libraries of knocked out genes which may be used to generate mutant animals; and (iii) the selection of cells carrying mutations in desired genes. Each of these components is now discussed.

[**0041**] (I) Components of the Retroviral Vectors

[**0042**] Retroviruses are RNA viruses which replicate through a DNA intermediate and which include as an obligatory step of their life cycle integration of the proviral DNA into the host chromosome (Varmus and Brown, Retroviruses. In *Mobile DNA* (ed. Berg, D. E. and M. M. Howe), pp. 53-108. American Society for Microbiology, Washington, D.C. (1989)). Following integration, the provirus is maintained as a stable genetic element in the infected cell and its progeny. Most or possibly all regions of the host genome are accessible to retroviral integration (Withers-Ward et al., Genes Dev. 8:1473-1487 (1994)), and the above properties make retroviruses invaluable as both potent mutagens and chromosomal markers.

[**0043**] The MAGEKO process employs one or more retroviral vectors as mutagens. The principal vector is preferably based on the Moloney murine leukemia virus (MoMLV) (Varmus and Brown, supra). Secondary vectors are of different retroviral origin and include, for example, lentiviral (Vannus and Brown, supra) or avian leukosis-sarcoma virus (ALV) (Varmus and Brown, supra) based vectors. Different retroviral backbones are utilized in the MAGEKO technique to increase the number of genes that are affected by insertional mutagenesis, on the theory that different retroviruses may have different genomic targeting preferences. Furthermore, in the case of the lentiviral vectors, it is known that this retrovirus is capable of transducing nondividing cells (Naldini et al., Science 272:263-267 (1996)), thus allowing for earlier detection of infected cells. Integration of vectors involving MoMLV depends on mitosis (Roe et al., EMBO J. 12:2099-2108 (1993)).

[**0044**] Each vector used in the mutagenesis procedure is quite similar, differing significantly only in the retroviral backbone sequence. Otherwise, the vectors carry in common several unique features essential to the subsequent functional characterization of the iractivated genes of interest. In particular, as discussed in more detail below, each of the vectors is highly mutagenic, each allows rapid identification of infected cells, and each allows specific detection of cells expressing the gene with the retroviral insertion. Moreover, the vectors specifically mark cells expressing the mutant gene, allow temporal and spatial analysis of the phenotype of the disrupted gene, provide for conditional tissue-specific gene inactivation, facilitate the conditional ablation of cell lineages expressing the mutant gene, and facilitate conditional ectopic expression of any gene in any desired tissue. Other important attributes of these vectors include the ability to generate animals with conditional tumors of any cell origin as well as the ability to establish conditional immortal cell lines of any cell type.

[**0045**] **FIG. 1** depicts the MoMLV-based retroviral vector and its various features. The orientation of the transcriptional units is indicated by arrows. Vectors of other retroviral origin are quite similar to the MoMLV-based vector, differing only in the sequences of the retroviral backbone. The origin and importance of the elements are as follows.

[**0046**] (a) Retroviral Sequences. The retroviral sequences are necessary for packaging and random integration of the incoming DNA into the host genome. The MoMLV sequences are substantially similar to the sequences found in the vector pGen (Soriano et al., J. Virol. 65:2314-2319 (1991)), a vector which lacks the viral enhancer sequences and which contains the bacterial supF gene positioned in the **3'** long terminal repeats (LTR). Upon integration into the genome, the **5'**LTR enhancer sequences are also deleted, and the supF sequences are copied to the **5'**LTR. As described below, the viral LTRs of the parental vector are modified to contain loxP sequences. In addition, the trancriptional orientation of all non-retroviral vector sequences are inverted relative to the transcriptional orientation of the **5'**LTR promoter (**FIG. 1**). Production of high titer stocks from this vector are accomplished following published procedures (for example, Soneoka et al., Nucleic Acids Res. 23:628-633 (**1995**); Yee et al., Proc. Natl. Acad. Sci. USA 91:9564-9568 (1994); and Mann et al., Cell 33:153-159 (1983)). Alternative retroviral sequences may, for example, be derived from or based upon any lentiviral or ALV vector, and appropriate standard techniques may be used for viral propagation.

[**0047**] (b) LoxP. The loxP sequence is the recognition sequence of the bacteriophage PI CRE recombinase, and its use is described in Sauer, Meth. Enzymol. 225:890 (1993). This sequence mediates recombinational excision of the retroviral insertion in the presence of CRE. It also facilitates targeted chromosomal rearrangements, such as translocations and deletions (Ramirez-Solis et al., Nature 378:720-724 (1995)) in cells containing more than one provirus. Such cells may be obtained through mating of mice, each carrying a different loxP-tagged retroviral insertion. Alternatively, FRT, the recognition sequence of the *Saccharomyces cerevisiae* FLP recombinase (Dymecki, Proc. Natl. Acad. Sci. 93:6191-6196 (1996)) may be used for this purpose, and recombinational excision may be mediated by the FLP protein.

[**0048**] (c) V. V, or VDErs, is the recognition sequence of the VDE DNA endonuclease from Saccharomyces cerevisiae (Bremer et al., Nucleic Acids Res. 20:5484 (1992)). This sequence provides a unique chromosomal marker. Other chromosomal markers may also be utilized for this purpose.

[**0049**] (d) Splice Acceptor. As shown in **FIG. 1, a** consensus splice acceptor sequence is also included in the retroviral vectors. This sequence is required for fusion of the retroviral transcripts to the endogenous gene transcript in

situations where the retroviral integration occurs in an intron. The splice acceptor site prevents the retroviral transcript from being inadvertently spliced out of the genome, thereby maximizing the likelihood that an insertion is mutagenic for the endogenous gene (Gossler et al., Science 244:463-465 (1989); Friedrich and Soriano, Genes Dev. 5:1513-1523 (1991); Skames et al., Genes Dev. 6:903-918 (1992); Takeuchi et al., Genes Dev. 9:1211-1222 (1995); Wurst et al., Genetics 139:889-899 (1995); Forrester et al., Proc. Natl. Acad. Sci. USA 93:1677-1682 (1996); and Brenner et al., Proc. Natl. Acad. Sci. USA 86:5517-5521 (1989)). A preferable consensus splice acceptor is derived from the Adenovirus major late transcript (Robberson et al., Mol. Cell. Biol. 10:84-94 (1990)), but any other splice acceptor sequence may be utilized in the vectors of the invention.

[0050] (e) Stop Codons. Nonsense codons in all three reading frames ensure translational termination in the gene with the retroviral insertion. Any nonsense codon or set thereof may be used for this purpose.

[0051] (f) IRES. The internal ribosome entry site provides for translation initiation of the tag gene (described below). As shown, a preferred IRES is derived from the Encephalomyocarditis virus (Morgan et al., Nucleic Acids Res. 20:1293-1299 (1992)). Other appropriate ribosome entry sites may also be used in the present vectors.

[0052] (g) rtTA. The sequence indicated as "rtTA" in **FIG. 1** is preferably a hybrid protein composed of a mutant tetracycline repressor and the VP16 transcription activation domain (Gossen et al., Science 268:1766-1769 (1995)). rtTA possesses the ability to stimulate expression of genes placed under the control of the tetracycline operator in the presence of tetracycline derivatives (Gossen et al., Science 268:1766-1769 (1995)). In the present invention, rtTA is expressed under the control of the promoter of the endogenous cellular gene which has been mutated by the retroviral insertion (**FIGS. 3 and 4**). Conditionally expressed rtTA is a key component to functional characterization of genes facilitated by the MAGEKO approach.

[0053] (h) pA. As shown in **FIG. 1**, the vectors of the invention also include a polyA addition signal. This signal is required for the processing and expression of the rtTA mRNA. One preferred pA sequence is derived from the bovine growth hormone gene (Goodwin and Rottman, J. Biol. Chem. 267:16330-16334 (1992)), although any other polyadenylation signal may be used. Examples of other useful pA sequences include, without limitation, the insulin and SV40 pA sequences.

[0054] (i) P. P is the constitutively expressed mouse phosphoglycerate kinase-1 (PGK) promoter (Adra et al., Gene 60:65-74 (1987)). This promoter is required for the expression of GFO (described below). Other constitutive mammalian promoters may be used in place of the PGK sequence.

[0055] (j) ATL. ATL, or the adenovirus tripartite leader sequence (Sheay et al., BioTechniques 15:856-862 (1993)), is included in the vector as a cis-acting inducer of gene expression. This sequence enhances production of GFO. Other leader sequences may be substituted for ATL.

[0056] (k) gfo. A hybrid gfo gene is included in the vectors. This gene is composed of a mutant GFP at the **5'** end and neomycin (NEO) coding sequences at the **3'** end. GFP mutants are derivatives of the *Aequorea Victoria* GFP, an autofluorescent protein widely used as a reporter of gene expression (Chalfie et al., Science 263:802-805 (1994); and Palm et al., Nature Structural Biology 4:361-365 (1997)). Preferred mutants encode a green fluorescent protein with increased cellular fluorescence and include, without limitation, a GFP sequence which is based on the sequence of Heim et al. (Current Biology 6:178-182 (1996)) but which includes at least one of the following mutations: P4-3 (Y66H, Y145F), W7 (Y66W, N146I, M153T, V163A, N212K), SG11 (F64L, I167T, K238N), SG25 (F64L, S65C, 1167T, K238N), or SG50 (F64L, Y66H, V163A). The gfo sequence is used for fluorescence activated cell sorting (FACS) of infected ES cells, an important step in the generation of LOKs. The neo gene codes for bacterial neomycin phosphotransferase (Southern and Berg, J. Mol. Appl. Gen. 1:327-341 (1982)). Expression of this sequence renders ES cells which contain the provirus resistant to G418. Neomycin resistance is used in the methods of the invention to select ES cells which are homozygotic for the proviral insertion; this is accomplished by increasing the concentration of G418 in the cell culture medium, as previously described (Mortensen et al., Mol. Cell. Biol. 12:2391-2395 (1992)). Other detectable and selectable markers may also be utilized in the invention.

[0057] **(1)** SPA. A synthetic polyA addition signal is also included in the vector to facilitate processing and expression of the gfo MRNA (Levitt et al., Genes Dev. 3:1019-1025 (1989)). Other synthetic or natural poly A sequences may be utilized.

[0058] (m) t. Transcriptional termination sequences are an important feature of the retroviral vectors. These sequences terminate transcription from both the PGK and the cellular promoters. Appropriate transcription termination results in a considerably increased mutagenic potential of the retroviral insertion and a decrease in the abnormal expression of genes adjacent to the provirus; this eliminates potential complications in the phenotypic characterizations of KO mice, as has been observed in some instances (Olson et al., Cell 85:1-4 (1996)). As shown in **FIG. 1, a** preferred termination sequence is derived from the human complement gene (Ashfield et al., EMBO J. 10:4197-4207 (1991)), but any other appropriate transcription termination sequence may be utilized.

[0059] (II) Unique Properties of and Uses for the Retroviral Vectors of the Invention

[0060] The vectors of the invention possess a number of unique properties, making them useful for various types of gene disruption methods and types of analyses. Examples of these unique properties and uses now follow.

[0061] The retroviral vectors are highly mutagenic. One significant advantage provided by the present retroviral vectors is the fact that these vectors are highly mutagenic. This property arises, at least in part, because the vectors contain a combination of a consensus splice acceptor and transcriptional termination sequences. The splice acceptor has been previously described (Gossler et al., Science 244:463-465 (1989); Friedrich and Soriano, Genes Dev. 5:1513-1523 (1991); Skarnes et al., Genes Dev. 6:903-918 (1992); Takeuchi et al., Genes Dev. 9:1211-1222 (1995); Wurst et al., Genetics 139:889-899 (1995); Forrester et al., Proc. Natl. Acad. Sci. USA 93:1677-1682 (1996); and

Brenner et al., Proc. Natl. Acad. Sci. USA 86:5517-5521 (1989)), but the combination with termination sequences is novel, and this combination is important for the elimination of read-through transcription which is frequently observed in cellular sequences flanking proviruses (Swain and Coffin, Science 255:841-845 (1992)). The termination sequence also enhances mutagenicity by blocking potential bypassing of the insertion by alternative splicing mechanisms which make use of fortuitous chromosomal splice sites; these sites are inaccessible due to transcription termination at t.

[**0062**] Insertion of the retroviruses into a gene of interest, for example, gene X in FIGS. **2-4**, leads to gene inactivation which is independent of the site of integration. Normal transcription and subsequent translation of gene X **(FIG. 2)** are disrupted, whether or not the retroviral insertion has occurred in an exon sequence **(FIG. 3)** or an intron sequence **(FIG. 4)**. This advantage is quite important. Although gene disruption is generally expected following integration of standard retroviruses into exons, the outcome of retroviral integration into introns is less predictable, and only a small fraction of retroviral insertions have been found to be associated with recessive phenotypes in the mouse (Jaenisch, Science 240:1468-1474 (1988)). Accordingly, the combination of a splice acceptor sequence and a transcriptional terminator is an important feature of the present invention, rendering the presently described vectors highly mutagenic even when integrated at intron locations.

[**0063**] The MAGEKO method allows rapid identification of infected cells. In a second advantage, the invention allows rapid identification of infected cells. As described above, the vectors of the invention include a marker which facilitates the identification of vector-containing cells. In one embodiment, the vectors carry a GFP mutant with increased cellular fluorescence linked to the PGK promoter. This marker allows for the identification of infected cells hours after infection, thus enabling the rapid sorting of transduced cells, for example, by FACS analysis. This is an important element for the generation of LOKs.

[**0064**] The MAGEKO approach provides for specific detection of cells expressing the mutant gene. As described above, the fusion gene, rtTA, is produced only in cells expressing the gene mutated by a retroviral insertion. The conditional nature of rtTA synthesis allows the specific tagging of insertion-containing cells through a binary mammalian system, such as a binary mouse system. According to this technique, mice carrying the retroviral vector of the present invention may be mated to mice containing a marker gene under the control of the rtTA-dependent promoter. In offspring containing both transgenes, that marker will only be produced in cells expressing rtTA, and only in the presence of tetracycline derivatives. As a result, the only cells in the offspring which synthesize the marker are those cells in which the gene mutated by the provirus is expressed. These cells, depending on the nature of the marker, may then be detected and, if desired, separated from the remaining cells using standard techniques. The marker may be any reporter of gene expression. Such reporters include, without limitation, the bacterial lacZ gene (An et al., Mol. Cell. Biol. 2:1628-1632 (1982)), green fluorescent protein, wavelength variations of green fluorescent protein (Heim et al., Proc. Natl. Acad. Sci. USA 91:12501-12504)), luciferase (de Wet et al., Mol. Cell. Biol. 7:725-737 (1987)), and chloramphenicol acetyltransferase (CAT) (Gorman et al., Mol. Cell. Biol. 2:1044-1051 (1982)).

[**0065**] The MAGEKO process facilitates conditional ablation of cell lineages expressing mutant genes. The use of the rtTA construct facilitates the ability to conditionally ablate cell lineages expressing mutant genes. Cell ablation studies are instrumental in assigning function to entire cell lineages, as has been demonstrated in several instances (Breitman et al., Science 238:1563-1565 (1987); Behringer et al., Genes Dev. 2:453-461 (1988); Landel et al., Genes Dev. 2:1168-1178 (1988); Breitman et al., Development 106:457-463 (1989); Heyman et al., Proc. Natl. Acad. Sci. USA 86:2698-2702 (1989); Borrelli et al., Nature 339:538-540 (1989); Breitman et al., Mol. Cell. Biol. 10:474-479 (1990); Kunes and Steller, Genes Dev. 5:970-983 (1991); Moffat et al., Development 114:681-687 (1992); Nirenberg and Cepko, J. Neurosci. 13:3238-3251 (1993); and Dzierzak et al., Intern. Immunol. 5:975-984 (1993)). The retroviral vectors of the present invention are designed to utilize this powerful approach.

[**0066**] According to this aspect of the invention, conditional cell ablation is accomplished through a binary transgenic mouse system. In this system, a mouse that contains the "weapon" transgene in a silent form is mated to a mouse that expresses the activator. In the offspring that inherit both transgenes, the "weapon" is activated, and it exerts its killing effects only in cells expressing the activator. In the context of the rtTA system, mice expressing rtTA under the control of the endogenous mouse gene promoter synthesize rtTA only in cells expressing the mutant gene **(FIGS. 3 and 4)**. These mice are mated with mice carrying conditionally produced "cell ablation factors" which are themselves synthesized only in the presence of both rtTA and tetracycline derivatives. Offspring containing both transgenes are subjected to cell ablation studies following administration of tetracycline derivatives and resultant destruction of cells expressing the gene with the retroviral insertion. Examination of these offspring provides a functional characterization of the ablated cell lineage.

[**0067**] Conditionally produced "cell ablation factors" useful in the invention include, but are not limited to, wild-type and mutant toxins (Borrelli et al., Nature 339:538-540 (1989); Frankel et al., Mol. Cell. Biol. 9:415-420 (1989); and Frankel et al., Mol. Cell. Biol. 10:6257-6263 (1990)), wild-type and mutant herpes simplex virus thymidine kinases (HSV-tk) (Salomon et al., Mol. Cell. Biol. 15:5322-5328 (1995); and Black et al., Proc. Natl. Acad. Sci. USA 93:3525-3529 (1996)), and apoptotic proteins such as the *Drosophila reaper* gene product (White et al., Science 271:805-807 (1996)). If an HSV-tk gene is utilized, gancyclovir, in addition to tetracycline derivatives, is administered to trigger cell killing. In another example, conditionally produced β-galactosidase may also be used to facilitate cell ablation, as shown for various cell types in the nervous system (Nirenberg and Cepko, J. Neurosci. 13:3238-3251 (1993)).

[**0068**] Use of MAGEKO for temporal and spatial phenotypic analysis of disrupted genes. Use of the methods of the invention and, for example, the rtTA construct, also facilitates the temporal and spatial characterization of the phenotypes of disrupted genes. In many instances, especially if

the insertional mutation in the homozygotic state is lethal or results in a phenotype interfering with further analysis (Copp, Trends Genet. 11:87-93 (1995)), it is preferable to inactivate a gene of interest in an animal in a temporal and spatial manner. In the present invention, this is accomplished through the use of mosaic animals derived from a mixture of ES cells, some of which are heterozygotic and some of which are homozygotic for mutations in the gene of interest. In these mosaic animals, the heterozygotic cells rescue those cells which are homozygotic, as has been generally demonstrated previously (Nagy and Rossant, J. Clin. Invest. 97:1360-1365 (1996); and Robb et al., EMBO J. 15:4123-4129 (1996)), and this leads to the generation of mosaics.

[**0069**] According to this aspect of the invention, mosaic mice are generated from homozygotic mutant ES cells in the gene of interest with mutant ES cells containing the identical proviral insertion in only one of the two alleles of the same gene. The heterozygotic cells (derived from animals generated as described above for the conditional ablation technique) also contain conditionally produced "cell ablation factors." These factors are synthesized only in the presence of both rtTA and tetracycline derivatives, and rtTA, in turn, is produced only in cells expressing the gene with the retroviral insertion (**FIGS. 3 and 4**).

[**0070**] Administration of tetracycline derivatives to mosaic animals leads to the specific obliteration of heterozygotic cells in which the mutant gene is expressed, due to the presence of the "ablation factors" in those cells only. As a result, the cell population of an animal expressing the mutant gene will be exclusively composed of homozygotic mutant cells. Under these conditions, the phenotype associated with the gene of interest may be assessed. This approach is useful for the phenotypic analysis of mutants, particularly when generation of adult mice is compromised in the homozygotic state.

[**0071**] Use of the MAGEKO process for conditional tissue-specific gene inactivation. In some instances, temporal and spatial phenotypic analysis of a disrupted gene may not be adequate to assign gene function. To address this problem, a different but complementary approach, termed conditional tissue-specific gene inactivation, may be employed. According to this approach, a gene of interest is inactivated, when desired, in the cells in which it is expressed. This general technique has been previously used to assign gene functions through the use of tissue-specific gene targeting (Gu et al., Science 265:103-106 (1994); Kuhn et al., Science 269:1427-1429 (1995); and Rajewsky et al., J. Clin. Invest. 96:600-603 (1996)).

[**0072**] Conditional tissue-specific gene inactivation is accomplished through a binary transgenic mouse system, similar in principle to the one described above for conditional ablation of cell lineages. Here, the mating partner carrying the "activator" is derived from heterozygotic mutant ES cells containing a retroviral insertion in one of the two alleles of the gene to be subjected to the conditional tissue-specific inactivation. This mouse produces rtTA only in cells synthesizing the target gene (**FIGS. 3 and 4**). the other mating partner, i.e., the one with the silent "weapon," carries a conditionally expressed ribozyme and a conditionally expressed recombinase.

[**0073**] Ribozymes are molecules capable of catalyzing sequence specific cleavage of targeted RNAs (Altman, Proc. Natl. Acad. Sci. USA 90:10898-10900 (1993)). In this system, the ribozyme is preferably expressed using an RNA polymerase III (Pol III) dependent promoter, such as the U6 small nuclear RNA promoter (Das et al., EMBO J. 7:503-512 (1988)). The Pol III promoter synthesizes the appropriate ribozyme only in the presence of rtTA and tetracycline derivatives. In addition, the constitutive Pol III promoter is preferably separated by transcription terminators from the ribozyme sequences. Each ribozyme is specifically designed to target and inactivate the gene of interest (according to published protocols, for example, by Altman, Proc. Natl. Acad. Sci. USA 90:10898-10900 (1993); and Liu and Altman, Genes Dev. 9:471-480 (1995)). The presence of the terminators blocks downstream transcription (Das et al., EMBO J. 7:503512 (1988)) and thus interferes with the synthesis of the ribozyme. The terminator sequences are flanked by FRT or loxP (i.e., the recognition sequence of either the *Saccharomyces cerevisiae* Flp recombinase (Dymecki, Proc. Natl. Acad. Sci. USA 93:6191-6196 (1996)) or the bacteriophage P1 CRE recombinase (Sauer, Methods Enzymol. 225:890-900 (1993))). Flp or CRE is expressed only in the presence of rtTA and tetracycline derivatives.

[**0074**] In offspring containing both transgenes, Flp or CRE is produced in cells expressing the target gene when tetracycline derivatives are administered to the animal. Production of Flp or CRE leads to recombinational excision of the termination sequences and synthesis of the ribozyme in those cells. As a result, the target gene is subjected to ribozyme action, and the phenotype of this conditional tissue-specific gene inactivation event is amenable to analysis.

[**0075**] Another approach for conditional tissue-specific gene inactivation is based on conditional functional complementation between the disrupted and wild type alleles of the mouse gene or between the disrupted mouse gene and its wild type human homolog. This is a two step procedure that first involves mating of heterozygotic mice carrying the retroviral sequences of the present invention integrated in a particular gene to heterozygotic mice containing an extra copy of the wild type version of this gene under the rtTA-dependent promoter. Crossing F1 offspring containing both transgenes generate mice that are homozygotic in the disrupted gene but that also carry the wild type allele under the rtTA-dependent promoter. As a result, in the F2 mice, the wild type allele is expressed in the presence of tetracycline derivatives in the same cells that express the mutant gene. The presence of the wild type gene rescues the mutant phenotype which, in turn, may be assessed, when desired, upon withdrawal of the tetracycline derivatives. The very same approach may be used to complement the disrupted mouse gene with its human homolog, which is then expressed in the same cells that express the mouse mutant gene. If a human disease state gene is utilized in this technique, the F2 mice obtained may be used as animal models of the human disease, for example, to study the disease or isolate or identify therapeutic compounds.

[**0076**] Use of the MAGEKO process for conditional ectopic expression of the gene of interest in any desired tissue. Targeted gene expression is a powerful method for assigning function to genes, as has been demonstrated in several instances (Balling et al., Cell 58:337-347 (1989); Kessel et al., Cell 61:301-308 (1990); Brand and Perrimon,

Development 118:401-415 (1993); and Halder et al., Science 267:1788-1792 (1995)). The retroviral vectors of the present invention are designed to utilize this powerful approach. According to this aspect of the invention, conditional targeted expression of a gene of interest is accomplished through a binary transgenic mouse system, similar to those described above. Again, in this system, one mating partner expresses rtTA under the control of the promoter associated with the gene having the retroviral insertion; as such, rtTA is synthesized only in cells expressing the mutant gene (**FIGS. 3 and 4**). The other mating partner contains the gene of interest and synthesizes this gene product conditionally, i.e., only in the presence of both rtTA and tetracycline derivatives. In offspring having inherited both transgenes, the gene of interest is specifically expressed only in cells where the gene having the retroviral insertion is expressed, and only in the presence of tetracycline derivatives. The physiological consequences of this conditional targeted gene expression is thereby amenable to analysis in the offspring.

[**0077**] Importantly, this approach provides an unlimited number of different target tissues for analysis; in theory, every tissue in the animal can be selected, if desired, to study the consequences of the conditional ectopic expression of a gene of interest.

[**0078**] MAGEKO allows establishment of animals with conditional tumors in any desired cell type. The binary transgenic mouse system is also useful for the generation of animals with conditionally induced tumors. Here, one mating partner expresses rtTA under the control of the promoter of the gene with the retroviral insertion, and thus synthesizes rtTA only in cells expressing the mutant gene (**FIGS. 3 and 4**). The other mating partner carries conditionally produced "neoplastic factors," such as combinations of oncogenes (Bishop, Cell 64:235-248 (1991); and Hunter, Cell 64:249-270 (1991)) and (if necessary) other facilitating genes, such as telomerase (deLange, Proc. Natl. Acad. Sci. USA 91:2882-2885 (1994); Counter et al., Proc. Natl. Acad. Sci. USA 91:2900-2904; and Sharma et al., Proc. Natl. Acad. Sci. USA 92:12343-12346 (1995)). These factors are synthesized only in the presence of both rtTA and tetracycline derivatives. Accordingly, offspring containing both transgenes develop tumors in the cells expressing the gene with the retroviral insertion upon administration of tetracycline derivatives.

[**0079**] This approach affords a number of advantages over previous methodologies for the generation of transgenic mouse models for neoplasia (Quaife et al., Cell 48:1023-1034 (1987); Sinn et al., Cell 49:465-475 (1987); Jat et al., Proc. Natl. Acad. Sci. USA 88:5096-5100 (1991); and Sandmoller et al., Cell Growth and Diff. 6:97-103 (1995)). First, the method is not limited by the restricted spectrum of available tissue-specific promoters. And, second, the oncogenic state is not constitutive, but is conditional; the neoplastic transformation of a normal mouse tissue is initiated only in the presence of tetracycline derivatives, making the system more amenable to analysis. Animals generated by this method provide information about the types of oncogenes which play roles in particular cell types, and may also be used as animal models to screen anti-cancer therapies.

[**0080**] MAGEKO allows establishment of conditional immortal cell lines of any desired type. Once available, animals with developed tumors of desired cellular origin (produced as described above) are an immediate source of tumor cell lines. In the alternative, immortal cell lines can be established from these animals prior to tumor development, simply by isolating the desired cells from the animals and culturing in vitro in the presence of tetracycline derivatives.

[**0081**] Such cell lines provide a valuable reagent for high throughput drug screening procedures to identify compounds which affect the gene with the retroviral insertion. In particular, since these cell lines express the target gene and also constitutively synthesize GFP, the cell lines are, for example, GFP^+, β-GAL^+(if β-GAL is the reporter gene). Any drug that specifically affects the gene in question produces GFP^+, β-GAL^+cells.

[**0082**] (III) LOK

[**0083**] The retroviral vectors described above are used to construct libraries of ES cells containing knock outs in endogenous genes or "LOKs". To produce these libraries, the vectors are introduced by infection into ES cells to obtain insertions, on the average, every 1 Kb in the genome. The LOK preferably consists of thirty million such insertions, each carrying an independent provirus. The complexity of an LOK is high enough that most mouse genes should statistically be hit at least once by an independent retroviral integration event.

[**0084**] Following infection of ES cells with the retroviral vectors, transduced cells expressing the visual marker (for example, GFP) are selected by FACS analysis, and the cells are distributed in multi-well plates. The contents of combinations of wells are then pooled, subsequent to duplicate formation and storage of the replica, and appropriate matrices are generated to facilitate assignment of a specific cell to a particular well. Several proposed and established pooling strategies are available for the generation of the desired matrix to screen the LOK (see, for example, Zwaal et al., Proc. Natl. Acad. Sci. USA 90:7431-7435 (1993); Evans and Lewis, Proc. Natl. Acad. Sci. USA 86:5030-5034 (1989); Green and Olson, Proc. Natl. Acad. Sci. USA 87:1213-1217 (1990); Kwiatkowski et al., Nucleic Acids Res. 18:7191-7192 (1990); and Barillot et al., Nucleic Acids Res. 19:6241-6247 (1991)).

[**0085**] (IV) KIS

[**0086**] The present invention also includes a gene Knock out Identification System (or "KIS"). According to this aspect of the invention, genomic DNA, including the integrated nucleic acids of the retroviral vectors, are isolated from the pooled ES cells of the LOK and are fragmented. These fragments are then circularized and amplified by inverted PCR (see, for example, Ochman et al., Genetics 120:621-625 (1988); and Triglia et al., Nucleic Acids Res. 16:8186 (1988)), using primers which hybridize to the retroviral vector sequences but which are not present in the mouse genome; this method has been successfully applied to the detection of retroviral insertions in the Zebrafish genome (Allende, Genes Dev. 10:3141-3155 (1996)), P element insertions in the Drosophila genome (Dalby et al., Genetics 139:757-766 (1995)) and transposon insertions in the Arabidopsis genome (Sundaresan et al., Genes Dev. 9:1797-1810 (1995)). Alternatively, modifications of inverse PCR, such as oligo-cassette mediated PCR (see, for example, Rosenthal and Jones, Nucleic Acids Res. 18:3095-3096

(1990)) or ligation mediated PCR (see, for example, Mueller and Wold, Science 246:780-786 (1989)), may be used.

[0087] Genomic DNA fragments, once amplified, are transferred to hybridization supports, generating an ordered array of genomic DNA flanking the provirus. Labelled DNA from a gene of interest is then hybridized to the pooled genomic DNA, and a positive signal leads to the rapid identification of the desired ES cell clone.

[0088] Alternatively, detection of a retroviral integration site may be accomplished by direct sequencing of the amplified DNA of an ES clone; this approach, however, requires the isolation of single clones of ES cells and is preferably used only for a subset of the generated clones. In another alternative approach, an integration site may be determined by sequence detection using a positional oligonucleotide probing technique (POP), a method which is ideal for the processing of limited sequence information in parallel. According to this technique, all possible oligonucleotides of a specific length are synthesized in a high density array (such as an Affymetrix chip (see, for example, Lipshutz et al., BioTechniques 19:442-447 (1995)) and hybridized to the amplified DNA from ES cells. The POP technique is based on generating sequence information for an unknown region of nucleic acid (i.e., the genomic DNA), which is linked to a known sequence (i.e., a portion of the retroviral vector). Because retroviral integration is precise and results in the integration of a viral LTR within the genomic DNA, the LTR sequence is a preferred sequence for designing oligonucleotide probes. For example, oligonucleotides that contain **8** bases corresponding to the tip of the LTR and nine random bases can probe 4e9=262,144 combinations. This strategy of junction sequencing by oligonucleotide arrays can be used in place of, or in parallel with, the hybridization technique described above. As information about the mouse genome sequence increases, this sequence tag approach will become increasingly useful in identifying insertions in known genes.

[0089] Following identification of ES cell clones with desired mutations, heterozygous and homozygous mutant mice are generated by the procedures described above.

Other Embodiments

[0090] The techniques described herein are applicable to the generation of mutations in any appropriate non-human mammal. In particular examples, the techniques are useful for generating libraries of gene mutations, ES cells, and transgenic animals in any mammal which may be used as a disease model or any domesticated animal including, but not limited to, rodents (for example, mice, rats, and guinea pigs), cows, sheep, goats, rabbits, and horses.

[0091] Other embodiments are within the following claims.

What is claimed is:

1. A transgenic non-human mammal, said mammal comprising a retroviral vector that comprises (i) a splice acceptor sequence, (ii) a transcription terminator, at which site transcription terminates, and (iii) retroviral packaging and integration sequences.

2. The transgenic non-human mammal of claim 1, wherein said mammal is a mouse.

3. The transgenic non-human mammal of claim 1, wherein said retroviral packaging and integration sequences are from a Moloney murine leukemia virus sequence.

4. The transgenic non-human mammal of claim 1, wherein said retroviral vector further comprises a regulatory gene positioned for expression under the control of an endogenous promoter in the genome of said non-human mammal, whereby said promoter is operably linked to said regulatory gene upon integration of said vector into the genome of said mammal.

5. The transgenic non-human mammal of claim 4, wherein said regulatory gene encodes a regulatory protein that modulates the expression of another gene.

6. The transgenic non-human mammal of claim 5, wherein said regulatory protein is a tetracycline repressor fused to an activator protein.

7. The transgenic non-human mammal of claim 6, wherein said activator protein is a VP16 transcription activation domain.

8. The transgenic non-human mammal of claim 1, wherein said retroviral vector further comprises a DNA sequence encoding a constitutively expressed marker gene, wherein said marker gene encodes a marker protein detectable in a mammalian cell.

9. The transgenic non-human mammal of claim 8, wherein said marker protein is a green fluorescent protein.

10. The transgenic non-human mammal of claim 9, wherein said green fluorescent protein has increased cellular fluorescence relative to the wild-type green fluorescent protein.

11. The trans genic non-human mammal of claim 9, wherein said green fluorescent protein is fused to a mammalian selectable marker protein.

12. The transgenic non-human mammal of claim 11, wherein said mammalian selectable marker protein is neomycin phosphotransferase.

13. The trans genic non-human mammal of claim 1, wherein said retroviral vector further comprises a recognition sequence recognized by a yeast VDE DNA endonuclease.

14. The transgenic non-human mammal of claim 1, wherein said retroviral vector further comprises a sequence which is recognized by a recombinase enzyme.

15. The transgenic non-human mammal of claim 14, wherein said sequence is a loxP sequence.

* * * * *