4.1 Introduction

4.1.1 Gene trap mutagenesis, *SAβgeo* gene trap cassette

The gene trap approach has been described in detail in Chapter 1. Gene trap mutagenesis is unique in that it can efficiently cause loss of function mutations as well as tag the mutated gene. In a high throughput genetic screen, this feature is extremely valuable. For this reason, the gene trap approach was chosen in this study as a mutagen. One of the most frequently used gene trap cassettes is *SAβgeo* (Friedrich and Soriano, 1991). *SAβgeo* is a promoter gene trap cassette, containing the consensus adenovirus major late transcript splice acceptor (SA) from the intron1/exon2 boundary and followed by the *βgeo* reporter. The bacteria initiation codon in *βgeo* was replaced by the protein translation initiation sequence from the Moloney murine leukemia virus (MoMuLV) Env gene. The *βgeo* reporter gene is a fused gene consisting of the *E.coli LacZ* gene at N-terminus and the neomycin phosphotransferase gene (*Neo*) at C-terminus. *βgeo* encodes a fused protein with both bacteria β-galactosidase and neomycin phosphotransferase activities. Cells with gene trap mutations are resistant to G418 because of the expression of neomycin phosphotransferase. *LacZ* expression can be used to display the expression of the trapped endogenous gene. The efficiency of the *SAβgeo* gene trap cassette has been tested using both electroporation and retrovirus based gene transfer methods. These experiments show that 95% of G418-resistant ES cells resulting from the integration of the *SAβgeo* gene trap cassette also express β-galactosidase, which is detectable by X-gal staining. 60 mouse lines were generated from gene-trapped ES cells and half of them exhibited obvious phenotypes, indicating that the insertions of *SAβgeo* cassette are mutagenic at most genomic loci (Friedrich and Soriano, 1991).

4.1.2 A revertible retroviral gene trap vector design

The recombinant retroviral gene transfer system has been developed to transfer exogenous genes into mammalian cells. The vector designs and applications of retroviral based gene trap approaches have been discussed in detail in Chapter 1.

Phenotype driven genetic screens require an approach to identify the isolated mutation. For example, loss of function mutations identified in cultured cells can normally be identified by genetic rescue experiments in which, a cDNA or genomic DNA fragment containing the candidate genes are introduced into the mutated cells. This method requires the construction of an expression vector for each candidate genes. If the gene is unknown, then a library of expression clones can be used. Therefore, this approach cannot be applied on a large for a high throughput genetic screen.

To provide a confirmation for a gene trap mutation, revertible retroviral gene trap vectors have been designed (Fig. 4-1). The basic design of this type of vector consists of the *SAβgeo* gene trap cassette between viral 5' and 3' LTRs in a reversed transcription orientation in relation to viral transcription. To make it revertible, a *loxP* site is inserted into the viral 3'LTR, which replaces part of the viral enhancer region, resulting in a self-inactivating (SIN) retrovirus (Ishida and Leder, 1999). During reverse transcription and integration, the *loxP* site will be duplicated from the 3'LTR to the 5'LTR, resulting in a provirus flanked by two *loxP* sites. By Cre-*loxP* mediated recombination, the provirus with the gene trap cassette can be removed, leaving a single LTR fragment at the viral integration site. Because vectors containing gene trap cassettes with a splice acceptor such as *SAβgeo* often insert into genes' introns (Hansen et al., 2003), the remaining LTR in the intron is less likely to be able to disrupt a gene's expression. With this method, recessive gene trap mutations can be verified by observing the reversal of a phenotype by Cre-*loxP* mediated recombination in gene-trapped ES cells.

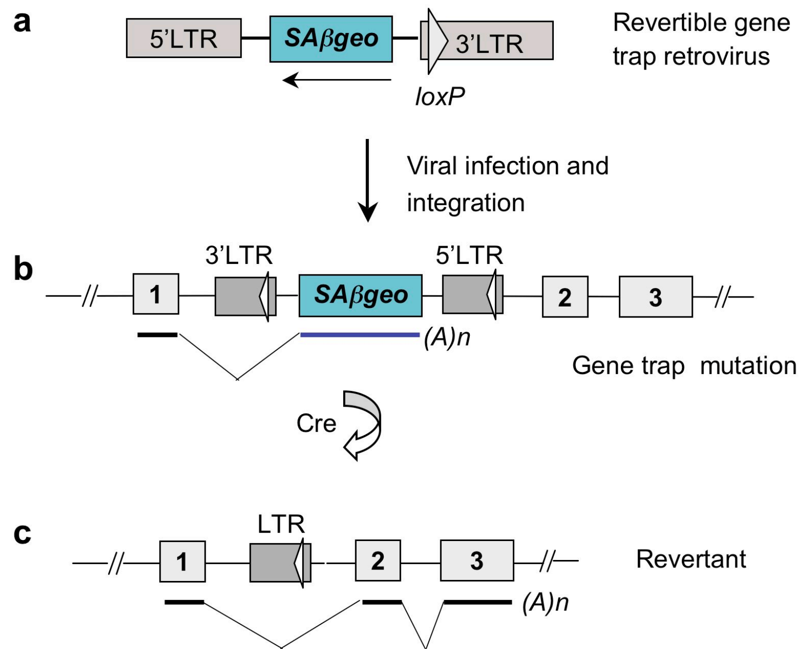


Figure 4-1. Revertible gene trap virus

a. The structure of a revertible gene trap retrovirus, containing a *loxP* site in the viral 3'LTR and the *SAβgeo* gene trap cassette. The arrow indicates the transcription orientation of *βgeo* . **B.** The structure of an integrated provirus in cells, showing the duplicated *loxP* sites in both LTRs and the transcription of a fused gene trap messenger RNA. **c**. By Cre-*loxP* mediated recombination, the gene trap cassette is removed, leaving a single LTR at the viral integration site.

4.2 Results

4.2.1 Construction of revertible retroviral backbones

The retroviral vectors developed in present study were based on the highly efficient pBabe retroviral vectors. pBabe retroviral vectors were derived from Moloney Murine Leukemia Virus. The high efficiency of pBabe vectors was achieved by including the mutated virus splicing donor and mutated ATG -minus *gag/pol* sequences in addition to the Ψ viral packaging signal. The viral titer of pBabe-based retrovirus can be as high as $5x10^6$ (cfu/ml) on NIH3T3 cells (Morgenstern and Land, 1990). To make a revertible retroviral vector, a *loxP* site was synthesized with flanking *Xba* I restriction sites and cloned into *Nhe*I/*Xba*I restricted U3 region of 3'LTR which was subcloned from pBabe. Consequently, the 267 bp *Nhe*I/*Xba*I fragment of viral enhancer in U3 region was deleted, resulting in a SIN retroviral vector (Fig. 4-2 a). The cloned *loxP* site in the viral LTR was sequenced to confirm the correct *loxP* sequence. A portion of the sequence of the viral 3'LTR is illustrated, showing the deleted viral enhancer (Fig. 4-2 b), and the inserted *loxP* site (Fig. 4-2 c) in the modified 3'LTR (3'LTR*loxP*).

Three revertible retroviral vectors were constructed, containing the modified 3'LTR*loxP*, which replaces the original 3'LTR in pBabe based retroviral vectors. **pBaER** (pBabeEGFPRevertible) was derived from pBabeEGFP by replacing the 3'LTR in pBabeEGFP with the 3'LTR*loxP* (Fig. 4-3 a). pBabeEGFP contains a fluorescence reporter gene, *SV40/EGFP* (enhanced green fluorescent protein) driven by the SV40 early promoter. *SV40/EGFP* can be used to monitor the presence of the retrovirus by examining the expression of EGFP in live cells. This feature enables monitoring the transfection efficiency of the retroviral vectors in viral packaging cell lines. The SV40 early promoter can also function as a DNA replication origin which allows the replication of an episomal DNA molecule in mammalian cells expressing SV40 T antigen. The replication ability of a

transiently transfected retroviral vector in viral packaging cells is expected to increase the viral production. **pBaOR**(pBabeOligoRevertible) was derived from pBabeOligo. To make pBabeOligo, the *SV40EGFP* cassette in pBabeEGFP was deleted and replaced with a multiple cloning sites. Because the maximal viral packaging limit is near 8kb, the smaller pBaOR retroviral backbone will have more room for the cloning of the gene trap cassette. To allow the replication of the retroviral vector cassette in viral packaging cells, an SV40 origin for DNA replication was inserted into the plasmid backbone outside the virus (Fig. 4-3 b). **pCBaOR** (pCMVBabeOligoRevertible) was constructed from pBaOR, by removing the 5'LTR including the viral enhancer and part of the promoter fragment including the CCAAT box which were replaced by Human cytomegalovirus (CMV) immediate early promoter(Fig. 4-4)(Boshart et al., 1985). To do this, the viral promoter region between *Xba*I/*Sac*I restriction sites in 5'LTR of pBaOR was deleted by insertion of an oligonucleotide containing *Hind*III restriction sites. Using this promoter-less pBaOR as template, the whole virus from the R region (before the *Sac*I restriction site until the end of the 3'LTR) was PCR amplified, digested with *Hind*III and *Apa*I and cloned into *Hind*III/*Apa*I sites in the multiple cloning site of a construct containing the CMV promoter, pcDNA3EGFP, resulting in a revertible, SIN retroviral vector transcribed from CMV promoter. pcDNA3EGFP contains the SV40EGFP cassette in the plasmid backbone.

4.2.2 Efficiency of retroviral vectors for transferring autonomous genes

To assess the efficiency of these revertible retroviral vectors, a *neo* reporter driven by a constitutive PGK promoter was cloned between viral LTRs. Cells with integrated proviruses can be selected by G418 resistance. The retroviral vectors containing the *PGK-neo* cassette were named as pBaERneo, pBaORneo and pCBaORneo after the different retroviral backbones, pBaER, pBaOR and pCBaOR. Because the construction of pCBaOR involved PCR amplification,

a

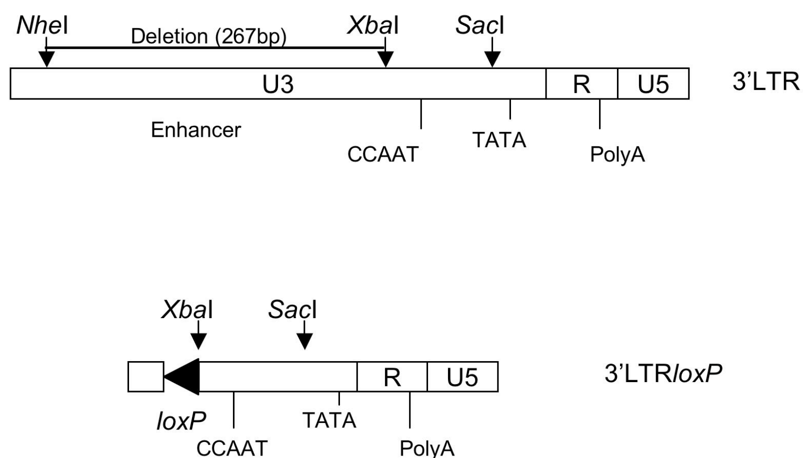


Figure 4-2. Schematic the construction of 3'LTR for revertible retroviral vectors

a.The structure of the 3'LTR of MoMuLV and modifed 3'LTR*loxP*, illustrating the viral enhancer, promoter(CCAAT and TATA boxes) and Poly adeylation(PolyA) site. The Viral enhancer (267 bp) is deleted and replaced with a *loxP* site in the 3'LTR*loxP*. **b.** Sequence of a portion of the MoMulV 3'LTR, showing the sequence of deleted 267 bp *Nhe*I/*Xba*I fragment. **c.** Sequence of a portion of the modified 3'LTR*loxP*, showing the insertion of the *loxP* site.

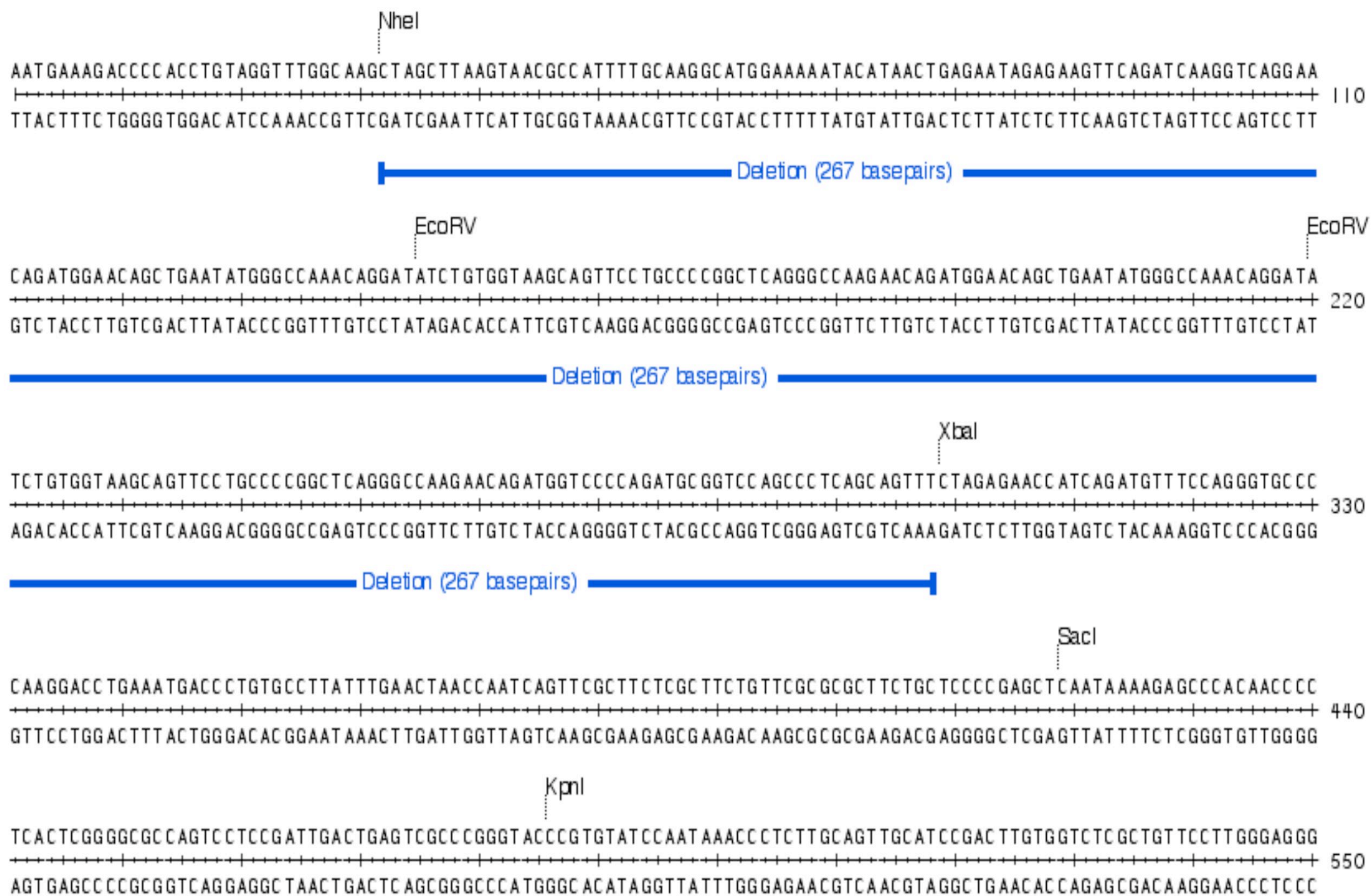


Figure 4-2 b. Sequence of the part of the MoMulV 3'LTR showing the sequence of the deleted viral enhancer

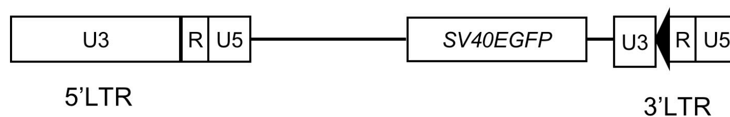
```
                                                                        XbaI
AATGAAAGACCCCACCTGTAGGTTTGGCAAGCTAGATAACTTCGTATAGCATACATTATACGAAGTTATCTAGAGAACCATCAGATGTTTCCAGGGTGCC  100
TTACTTTCTGGGGTGGACATCCAAACCGTTCGATCTATTGAAGCATATCGTATGTAATATGCTTCAATAGATCTCTTGGTAGTCTACAAAGGTCCCACGG
                            |---------- LoxP ----------|

                                                                                         SacI
CCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTCCCCGAGCTCAATAAAAGAG  200
GGTTCCTGGACTTTACTGGGACACGGAATAAACTTGATTGGTTAGTCAAGCGAAGAGCGAAGACAAGCGCGCGAAGACGAGGGGCTCGAGTTATTTTCTC

                                                  KpnI
CCCACAACCCCTCACTCGGGGCGCCAGTCCTCCGATTGACTGAGTCGCCCGGGTACCCGTGT  262
GGGTGTTGGGGAGTGAGCCCCGCGGTCAGGAGGCTAACTGACTCAGCGGGCCCATGGGCACA
```

Figure 4-2 c. Sequence of the modified 3'LTR*loxP*, showing the inserted *loxP* site.

a



b

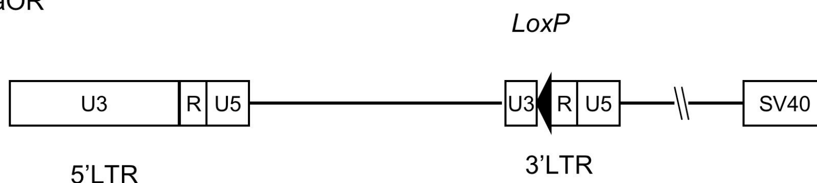


Figure 4-3. Retroviral gene trap vectors. pBaER and pBaOR are MoMulV based retroviral vectors, containing the MoMulV 5'LTR and the modified 3'LTR*loxP*. **a.** pBaER contains a fluorescent marker *EGFP* (enhanced green fluorescent protein)driven by the SV40 early promoter, which can be used to monitor the presence of the retrovirus construct by examining the expression of EGFP. The SV40 early promoter allows the replication of the retroviral vector cassette in mammalian cells expressing SV40 T antigen. **b.** pBaOR contains the minimal cis-elements for viral packaging between the LTRs. A SV40 origin fragment is inserted into the plasmid backbone to allow DNA replication.

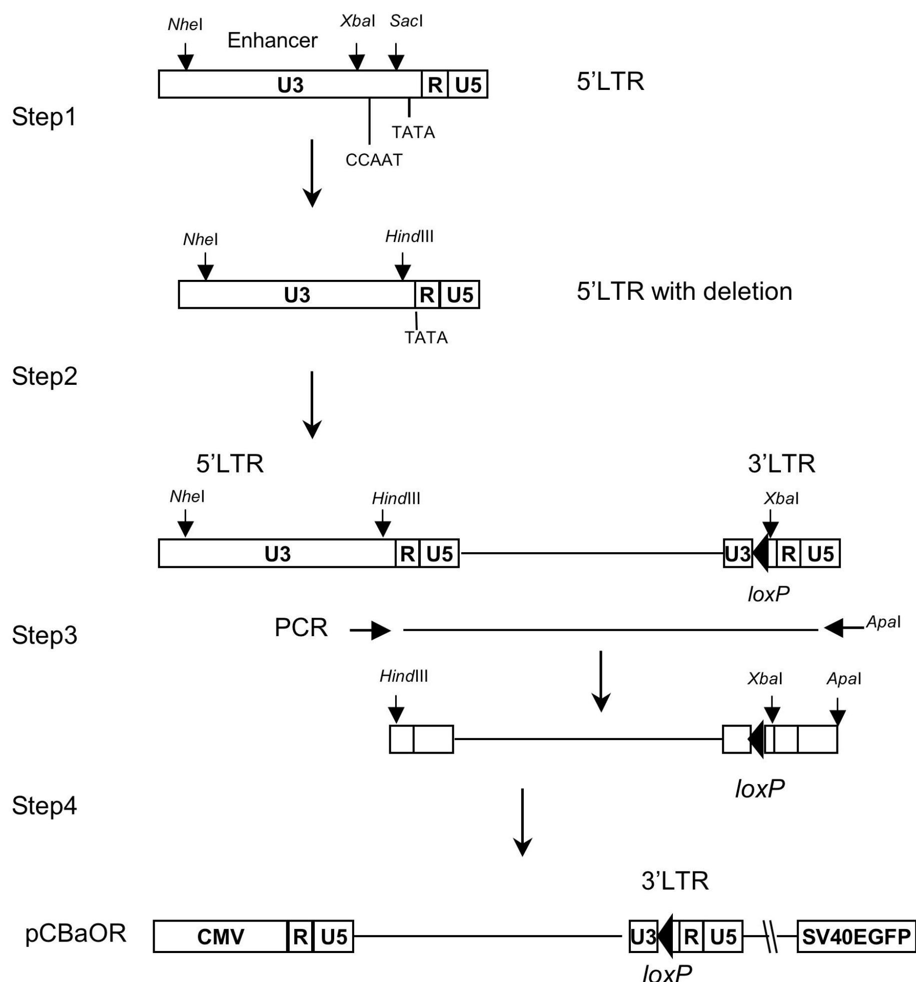


Figure 4-4. Construction of a revertible retroviral vector driven by CMV promoter/enhancer.

Step1: The viral 5'LTR was subcloned and the promoter region between *XbaI/Sac*I restriction sites was deleted by insertion of an oligonucleotide containing a *Hind*III restriction site. **Step2**: The modified 5'LTR was cloned into pBaOR to replace it's original 5'LTR. Then, using it as template, a promoter-less virus was PCR amplified initiating from the R region at 5'LTR until the end of 3'LTR. **Step3**: The PCP product was restricted by *Hind*III *and Apa*I enzymes and cloned into the plasmid backbone, pcDNA3EGFP, to create pCBaOR. The pcDNA3EGFP backbone contains a *SV40EGFP* cassette.

clonal variation may occur because of PCR errors. To test that the backbone still functioned efficiently, the *PGK-neo* cassette was cloned into 4 individual clones of pCBaOR, giving rise to of pCBaOR1 to pCBaOR 4.

To produce virus, a transient transfection strategy was utilized using phoenix viral packaging cells (Hitoshi et al., 1998). Phoenix viral packaging cells were constructed to express Gag/Pol and Env in a high transfectable subclone of 293T (human embryonic kidney derived) cell line expressing SV40 T antigen. The high transfection efficiency of phoenix cells allows viral particles to be produced at high efficiency within a few days following transient transfection (Nolan and Shatzman, 1998). For transient transfection, DNA was prepared by the standard alkaline lysis method and purified by CsCl density gradient ultracentrifugation. Transfection was carried out using calcium phosphate co-precipitation method. Two days after transfection, the viral supernatant was collected and used to infect AB2.2 ES cells. Cells with integrated proviruses were selected in G418. The number of G418 resistant ($G418^R$) cells recovered from one milliliter (ml) virus was determined as viral titer. The viral titer for each retroviral vector is listed in figure 4-5. pBaERneo had the highest titer (1,300 cfu/ml), which is nearly three times the titer produced by pBaORneo (390 cfu/ml). Four individual pCBaORneo clones (pCBaORneo 1 to 4) exhibited titers varying significantly from 740 cfu/ml to zero. This result suggests that PCR error may affect the retroviral construct efficiency.

4.2.3 Revertible retroviral gene trap vectors

To construct gene trap vectors, the *SAβgeo* gene trap cassette (Friedrich and Soriano, 1991) was cloned into pBaER and pCBaOR retroviral backbones, creating pBeGTV (from pBaER) and pCbGTV (from pCBaOR) (Fig. 4-6 a). pBeGTV and pCbGTV were transiently transfected into phoenix packaging cells using calcium phosphate co-precipitation and the viral supernatant was collected and used to infect AB2.2 ES cells. Cells with inserted provirus were selected with

G418 (180 μg/ml). Gene trap titer was determined using the same method as described above. The pBeGTV exhibited a viral titer of 50 cfu/ml, which is slightly higher than the viral titer of pCbGTV (30 cfu/ml) (Fig. 4-6 b). The difference between viral titers from pCbGTV and pBeGTV does not appear to arise from the variation in the DNA transfection efficiency, but is inherited in the structure of the vectors. The transfection efficiency was determined by quantifying the portion of cells expressing EGFP after transfection by flow cytometry. This experiment revealed that both the pCbGTV and pBeGTV vectors had a 60% transfection efficiency, suggesting that the transfection was highly efficient for both vectors (data not shown). This result is consistent with the almost 2 fold difference in viral titer between pBaERneo and pCBaORneo (Fig. 4-5). Note that gene trap titer was about 200 times lower than the titer from the constitutively expressed *PGK-neo* cassette, which suggests that only a small fraction of viral insertions can activate the *βgeo* cassette by a gene trap event.

4.2.4 The function of *loxP* sites in integrated provirus

To test the function of the *loxP* sites in the integrated provirus, AB2.2 ES cells containing the inserted proviruses were obtained by infection with pCbGTV or pBeGTV viruses. Cre-*loxP* mediated recombination was performed on one gene -trapped clone from the pCbGTV and pBeGTV infections. The Cre-expression vector pOG231 was transiently transfected into each clone and 96 ES cell clones were picked for each clone and sib-selected by G418. Cells, in which the retroviral has been deleted, will be sensitive to G418. For the pCbGTV gene-trapped cell line, 7 out of 96 ES cells were G418 sensitive, suggesting that the *loxP* site in pCbGTV gene trap vector was functional. However, no G418 sensitive clones could be recovered from the pBeGTV gene-trapped clone. This experiment was repeated once for pBeGTV vector in a different gene-trapped cell line, still no G418 sensitive clones could be recovered, suggesting that the *loxP* site is not functional in pBeGT (Fig. 4-6 b). pBeGTV may have acquired mutations in the *loxP* site during cloning. Since pCbGTV contains functional *loxP*

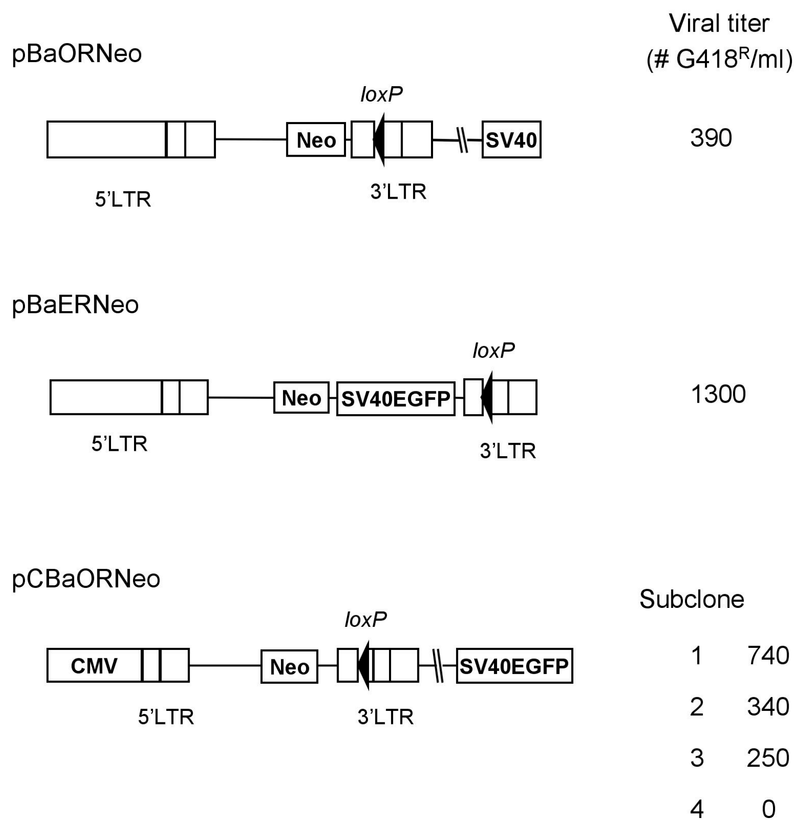
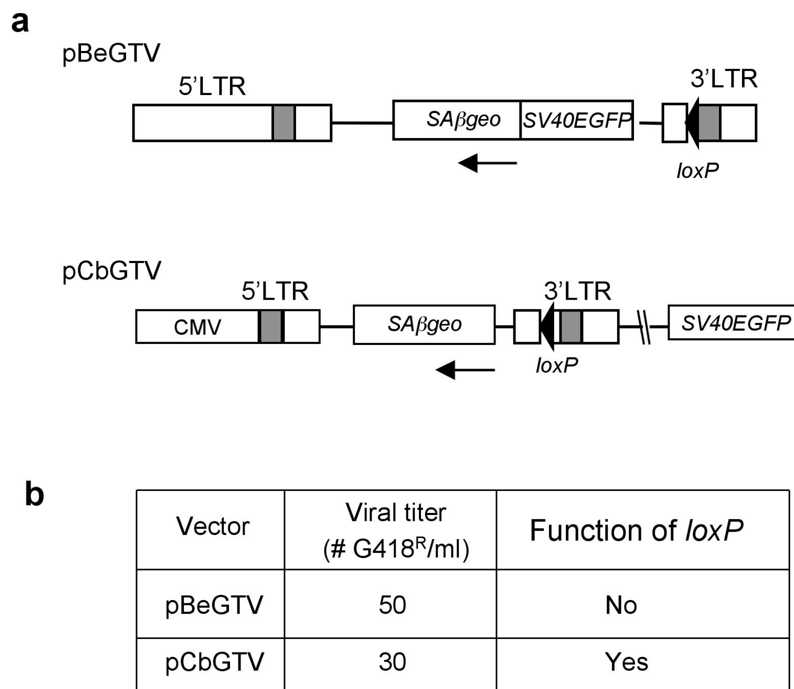


Figure4-5. Viral titer by pBaORNeo, pBaERNeo and four individual pCBaORNeo vectors

Each retroviral construct was transfected into Phoenix viral packaging cells. The transiently produced virus supernatant was used to infect AB2.2 ES cells. G418 resistant ($G418^R$) ES cell colonies were counted and the viral titer was determined as the number of $G418^R$ ES cell colonies per ml of viral supernatant used for infection.

Note that the viral titer by four individual pCBaORNeo clones varies.



Vector	Viral titer (# $G418^R$/ml)	Function of *loxP*
pBeGTV	50	No
pCbGTV	30	Yes

Figure 4-6. Schematic of retroviral gene trap vectors and viral titres.

a. pBeGTV and pCbGTV were derived from pBaER and pCBaOR by insertion of the *SAβgeo* gene trap cassette between viral LTRs in a reversed transcriptional orientation to relative to viral transcription. Arrows represent the transcription orientation of *SAβgeo*. Gray boxes represent the R regions in the viral LTR.

b. Virus produced by pBeGTV and pCbGTV was titred on AB2.2 ES cells for G418resistance ($G418^R$) from the expression of *βgeo*. Note that the gene trap titer is about 200 fold lower than the titer from the constitutively expressed *PGK-neo* cassette. The function of the *loxP* site was tested by Cre-*loxP* mediated removal of the integrated provirus from infected ES cells. pCbGTV had functional *loxP* sites, whereas the *loxP* site in pBeGTV is not functional, possibly caused by mutation of the *loxP* site during molecular cloning.

sites, it was used for subsequent studies and renamed as RGTV-1 (revertible gene trap virus 1).

4.2.5 The structure of integrated proviruses

The life cycle of a virus from a retroviral vector to the integrated provirus involves a series of procedures including viral replication, packaging, reverse transcription and integration. Abnormalities in any of these processes may result in an aberrant proviral structure. To ensure that the structure of the integrated provirus is intact and that the viral was able to integrate randomly in the genome, Southern-blot analysis was carried out to reveal the structure of the integrated provirus on DNA from RGTV-1 infected ES cells. Cells with an intact provirus show the predicted a 5.5 kb *Kpn*I fragment derived from provirus, when probed by a retroviral vector specific *LacZ* probe (Fig. 4-7 a & b). The proviral/host junction can be revealed by Southern-blot analysis on *Eco*RI restricted DNA as *Eco*RI has an unique restriction site in the *SAβgeo* gene trap cassette (Fig. 4-7 a). Southern-blot analysis using the same *LacZ* probe revealed that the size of the proviral/host junction from each RGTV-1 infected ES cell varies (Fig. 4-7 c). These results suggest that RGTV-1 is functional and is able to insert into host genome at random.

4.3 Discussion

4.3.1 Construction of recombinant retroviral vectors

In an effort to generate an efficient recombinant retroviral vector, three recombinant retroviral backbones have been constructed, pBaER, pBaOR and pCBaOR. The viral titer of each vector was examined.
These vectors share some common features: 1). They are based on pBabe, a derivative of Moloney murine leukemia virus. 2). They are revertible SIN vectors. 3). All three vectors were engineered to possess the SV40 origin or SV40 early

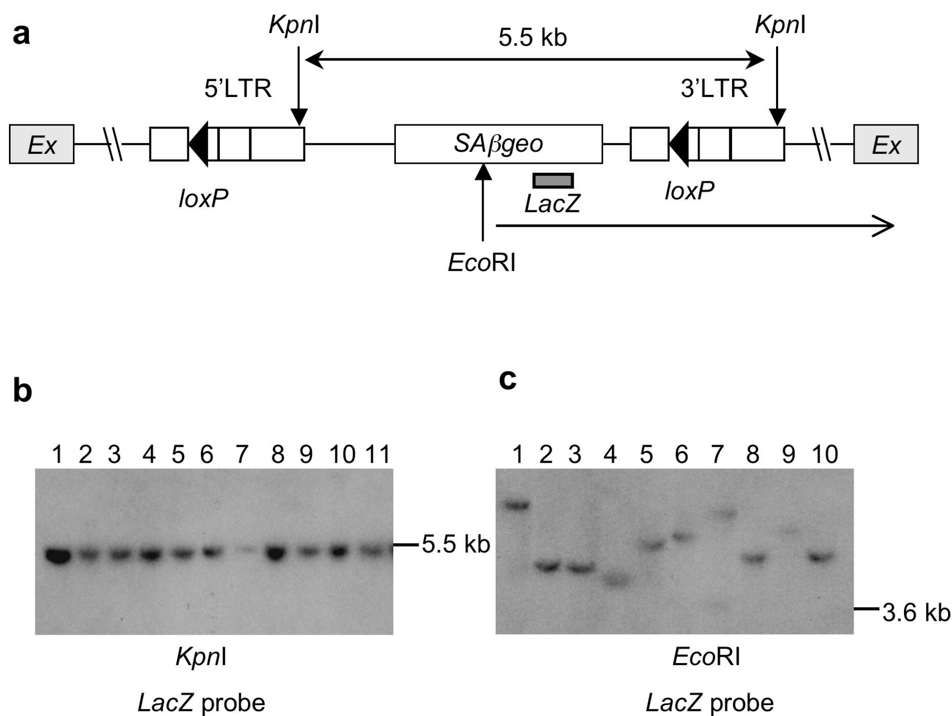


Figure 4-7. Southern-blot analysis of the integrated provirus from RGTV-1

a. Schematic of the RGTV-1 provirus in the intron of a gene. The *LacZ* probe can recognize a 5.5 kb *Kpn*I proviral fragment and the provirus/host junction. RGTV-1 contains an unique *Eco*RI site. **b**. Southern-blot analysis showing the expected 5.5 kb *Kpn*I fragment of integrated proviruses from 11 gene trap clones. Gray boxes represent the exons of an endogenous gene. **c.** The provirus/host junctions of 10 gene trap clones revealed by Southern-blot analysis on *Eco*RI restricted genomic DNA using a *LacZ* probe.

gene promoter (the SV40 promoter overlaps the SV40 origin) to allow the replication of the retroviral vector expression construct in the viral packaging cells. 4). Two vectors pBaER and pCBaOR contain the *SV40EGFP* cassette either between viral LTRs or in the plasmid backbone, which provides a fluorescent reporter for measuring the transfection efficiency. What is unique about each vector is: 1). pBaER contains the *SV40EGFP* cassette between viral LTR, therefore, the infected ES cells can be identified by the expression of EGFP. 2). pBaOR is the smallest construct among those three. It is composed of the original pBabe LTRs and the essential cis-elements for viral function. 3). pCBaOR contains the same viral backbone as pBaOR except that the viral enhancer and promoter in 5'LTR is replaced by a CMV promoter/enhancer.

4.3.2 Factors that affect retroviral vector efficiency

The efficiency of these retroviral backbones has been examined using the expression of the *Neo* reporter gene, which is transcribed either from an autonomous PGK promoter or by a "trapped" endogenous promoter. The CMV is believed to be a very strong promoter, which is able to facilitate viral expression. The viral titer of the pCBaOR backbone (740 cfu/ml) is about two fold higher than the titer by pBaOR (380 cfu/ml), suggesting that CMV promoter/enhancer in pCBaOR is able to increase the viral production, probably by elevating viral expression. Recently, Hlavaty et al. (2004) studied the effect of CMV enhancer on viral titer of a MoMLV based retroviral vector by placing a CMV enhancer in either the 5'or 3'LTR side. They found that a MoMLV-based retroviral vector with a CMV enhancer in the 5'LTR also produced a two fold increase in viral titer.

By comparing the viral titer obtained of pBaER and pBaOR derived retroviral vectors, it was revealed that pBaER-derived retroviral vectors are more efficient than pBaOR-derived vectors. pBaER and pBaOR have a similar structure except that pBaER contains the *SV40/EGFP* cassette between viral LTR, whereas in pBaOR, the SV40 origin is in the plasmid backbone. This may suggest that the

SV40 origin has a positional effect or the PCR amplification of this origin has introduced mutations that affect its function. A position effect of the SV40 origin has not been reported by others. It is worth mentioning that although the CMV promoter/enhancer in pCBaOR vector increased viral titer by two fold, the viral titer of pCBaOR is still nearly two fold lower than the titer of pBaER. The SV40 origin is also located in the plasmid backbone in pCBaOR.

4.3.3 Revertible gene trap vector, a useful tool for genetic screen in *Blm* deficient ES cells

The revertible gene trap vector, RGTV-1, is a useful genetic tools for a genetic screen. This allows quickly a causal link between a recovered phenotype and a mutated gene to be quickly established. RGTV-1 combines the advantages of the retroviral-mediated gene trap method and the Cre-*loxP* technology. The endogenous genes can be mutated by the strong *SAβgeo* gene trap cassette and at the same time be tagged by the insertion of the retrovirus. The consequence of the mutation can be verified by Cre-*loxP* mediated recombination, which removes the *loxP*-flanked provirus, resulting in the deletion of the gene-trap cassette. This method has many advantages. First, the phenotypic reversal provides straightforward genetic verification of a mutation. Compared to the traditional method, such as cDNA rescue or the recently developed BAC rescue methods, the Cre-*loxP* mediated reversal experiment doesn't require the generation of individual expression constructs for each mutations. Second, Cre-*loxP* mediated reversal can be applied to multiple samples simultaneously; therefore, it is suitable for a large scale genetic screen. The limitation of this method is that one LTR fragment with a *loxP* site remains in the host gene after Cre-*loxP* mediated recombination. The remaining LTR is not expected to be mutagenic in most gene-trapped cells because most G418 resistant clones recovered after insertion of this type of gene trap cassette will have the gene trap cassette inserted in an intron. However, if the retrovirus inserts into a 5'UTR region or a promoter region, the remaining LTR may

interfere the expression of the host gene, leading to a non-revertible phenotype. In the latter case, a cDNA rescue or an RNAi mediated gene expression knock down experiment could help to verify the mutation (Brummelkamp et al., 2002).