5.1 Introduction

5.1.1 Screen strategy

Blm-deficient ES cells exhibit a high LOH rate that allows segregation of homozygous mutants from single allelic autosomal mutations. Potential mismatch repair mutants have been recovered by 6TG selection from a pool of *Blm*-deficient ES cells mutated with EMS (Chapter 3). However, the difficulty in identifying the single nucleotide mutations induced by chemical mutagenesis limits the application of chemical mutagenesis in genetic screens *in vitro*. Revertible retroviral gene trap vector (RGTV-1) has been developed (Chapter4). In this chapter, the results of screens for 6TG-resistant mutants induced by RGTV-1 are described. The overall screen strategy is illustrated (Fig. 5-1). *Blm*-deficient ES cells were infected with RGTV-1 to generate single allele gene trap mutants, which were selected with G418 (180 μg/ml). Gene trap mutants were cultured over 14 population doublings to allow the generation of homozygous mutants via LOH events. These cells were then selected at high cell density ($0.5x10^7$ cells per 10 cm tissue culture plate) in 6TG (2 μm) to select out potential MMR mutants. 6TG resistant ES cell colonies were expanded for further molecular analysis.

5.1.2 Approaches to identify gene trap mutations

The insertion of gene trap vector provides a sequence tag, which allows rapid identification of the molecular basis of the mutation. In a gene trap, a fused mRNA composed of part of the normal endogenous transcript and the gene trap reporter is expressed. The splice junction of the fused gene trap transcript can be identified by a reverse transcription based PCR method, 5'RACE (rapid amplification of cDNA end) (Materials and Methods, 2.4.4) (Fig. 5-2 b).

The length of proviral/host junction is very useful for identifying the clonal relationships between gene trap mutations. Because LOH events occur randomly, a parental gene trap mutant in a pool could have produced many homozygous mutated daughter cells. If an LOH event occurs early, one mutant will dominate the screen. By inspecting the proviral/host junctions, gene trap mutants originating from one clone can be grouped. The Southern-blot analysis scheme used a unique proviral restriction enzyme site (*Eco*RI for RGTV-1). This allows the proviral/host junctions at both the 5'LTR and 3' LTR sides of the provirus to be identified using two viral probes (*Neo* and *LacZ* probes) (Fig. 5-2 a).

The retroviral integration site can be identified by PCR-based methods. Splinkerette PCR (SpPCR) was used in this study to identify the 5'LTR proviral/host junctions (Fig. 5-2 c). To do SpPCR, genomic DNA was digested by a restriction enzyme. An annealed oligo adaptor (Splinkerette) was then ligated to the digested genomic DNA. The ligated Splinkerette oligo provides an anchor sequence so that the flanking genomic fragment can be amplified using a pair of primers for the viral LTR and the Splinkerette oligo. The Splinkerette oligo is specially designed to contain a single strand hairpin structure at the 3' end of the annealed Splinkerette, which can reduce the non-specific amplification by Spinkerette PCR primers (Fig. 5-2c) (Devon et al., 1995, Mikkers et al., 2002). In this study, the SpPCR method has been used as the primary method to amplify the proviral/host junction from RGTV-1 infected ES cells because this method could be easily adapted to handle large numbers of samples (Mikkers et al., 2002). Taken together, the gene trap mutations can be identified using both Southern-blot analysis and PCR-based methods. The proviral/host junctions can be inspected quickly with Southern-blot analysis. SpPCR allows precise mapping of the retroviral insertion site in the host genome. Finally, the 5'RACE method provides an opportunity to access the expression of the fused gene trap transcript.

In this chapter, I describe the generation and use of approximately 10,000 gene trap mutations in *Blm*-deficient cells (NGG5-3) using the RGTV-1 virus, which will be referred to as GT library (gene trap mutation library). Screens for 6TG resistant clones have been performed three times on the GT library (STA, STB and STC screens). These screens use different conditions in either cell population doubling or 6TG dosages used for selection.

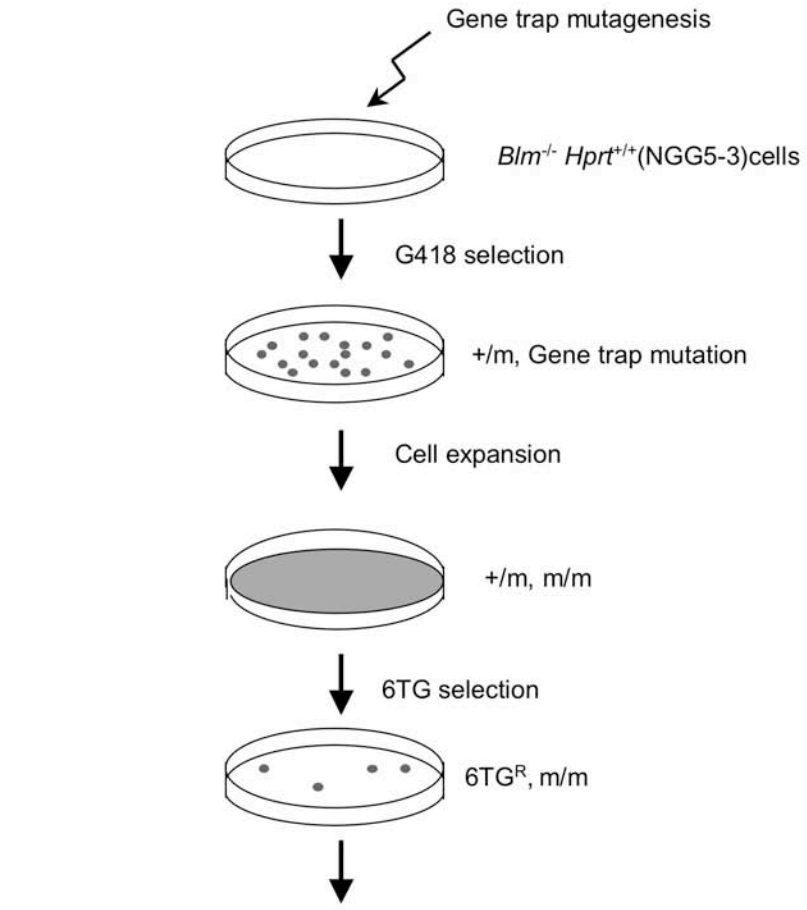


Figure 5-1. Screening strategy for 6TG resistant ($6TG^R$) gene trap mutants

Step 1: Gene trap mutagenesis to generate gene trap mutations

Step 2: The gene-trapped clones are pooled and expanded to allow the segregation of homozygous mutants through LOH.

Step 3: 6TG selection for resistant clones

Step 4: $6TG^R$ mutants are picked and the mutations are identified.

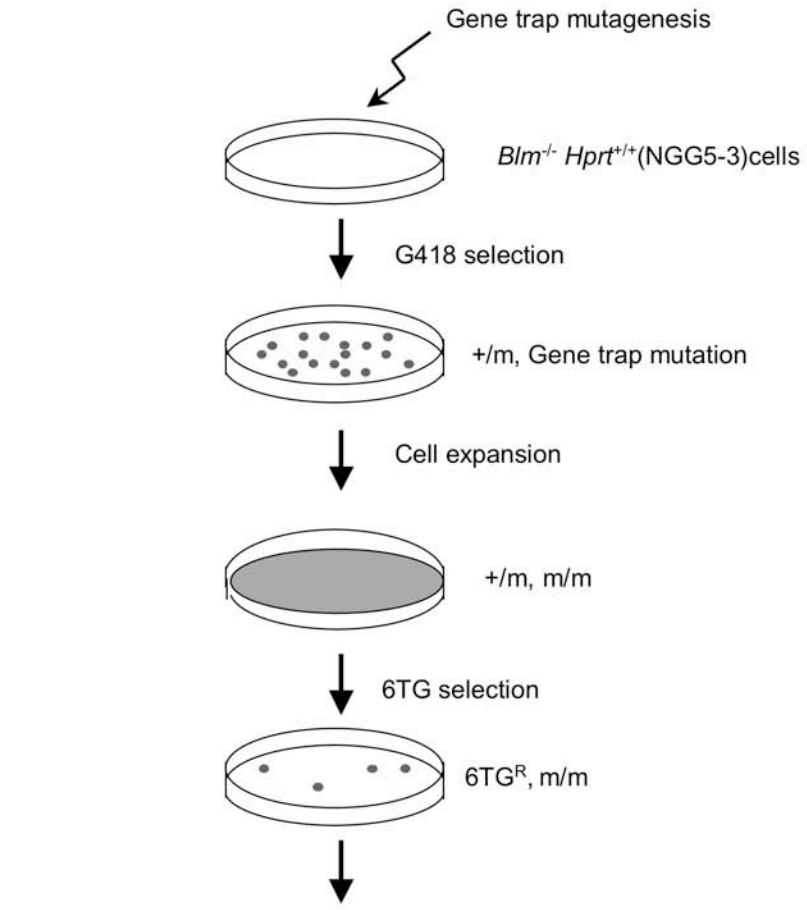


Figure 5-1. Screening strategy for 6TG resistant ($6TG^R$) gene trap mutants

Step 1: Gene trap mutagenesis to generate gene trap mutations

Step 2: The gene-trapped clones are pooled and expanded to allow the segregation of homozygous mutants through LOH.

Step 3: 6TG selection for resistant clones

Step 4: $6TG^R$ mutants are picked and the mutations are identified.

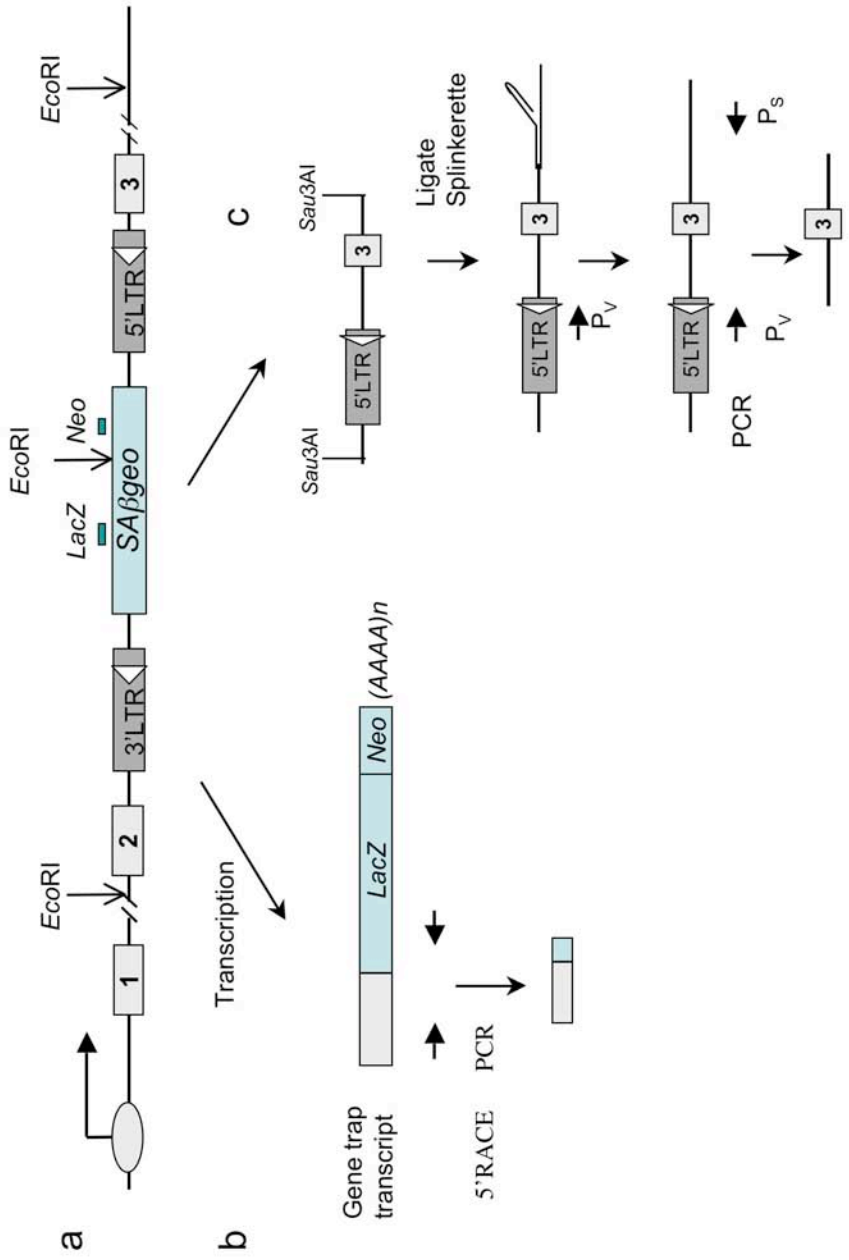

a
EcoRI
1
2
3'LTR
LacZ
Neo
SAβgeo
5'LTR
3
EcoRI
EcoRI
b
Transcription
Gene trap transcript
LacZ
Neo
(AAAA)n
5'RACE
PCR
c
Sau3AI
5'LTR
3
Sau3AI
Ligate Splinkerette
5'LTR
3
PV
5'LTR
3
PS
PV
PCR
3

Figure 5-2. Approaches to identify gene trap mutations

a. The length of proviral/host junction fragments can be identified by Southern-blot analysis on *Eco*RI digested genome and probed with either *lacZ* or *Neo* probes. RGTV-1 contains unique *Eco*RI restriction site. **b.** 5'RACE identifies the fused gene trap transcript. **c.** Splinkerette PCR (SpPCR) amplifies the proviral/host junction fragment at 5'LTR. Genomic DNA is digested by *Sau*3AI and ligated with annealed Splinkerette oligos. The annealed Splinkerette oligo contains a single strand loop at the 3' end and lacks the annealing site for the Splinkertee PCR primer(P_S). SpPCR requires that the first PCR cycle initiates from the virus using viral primer (P_V), which generates the annealing sites for P_S for the following PCR reaction. This design prevents non-specific PCR amplification from Splinkerette Oligos. Open triangle represents the *loxP* site.

5.2 Results

5.2.1 Gene trap mutant library (GT library) on *Blm*-deficient ES cells

RGTV-1 virus was produced by transient transfection of Phoenix viral packaging cells and used to infect the NGG5-3 cells that were cultured on seventeen 90 mm tissue culture plates. The gene trap mutants were selected with G418 (180 μg/ml) for 8 days. ES cells clones growing on one plate were stained and the number was determined. Gene trap clones from pairs of plates were combined to create eight pools. Each pool contains a mixture of 1,200 primary gene trap mutants. In total, about 10,000 primary gene trap mutants are represented in eight pools of this GT library.

5.2.2 STA screen

5.2.2.1 *Msh6*, a most frequently identified STA clones

For the STA screen, 2.5 x10^8 gene trap cells that have been cultured about 14 population doublings were plated in 6TG (2 μM) for 8 days. Twenty five 6TG resistant clones were recovered. Gene trap mutations in these clones were identified using SpPCR and/or 5' RACE methods (Table 5-1). The most frequently identified mutation was *Msh6* (MutS homologue 6) (Palombo et al., 1995). Fused transcripts between *ßgeo* and exon1 of *Msh6* were identified in 10 STA clones by 5'RACE (Fig. 5-3 a). The proviral/host junctions from these *Msh6* gene trap clones were cloned by SpPCR. Sequences of the proviral/host junctions revealed that retrovirus inserted into six different positions in the first intron of *Msh6* (Fig. 5-3 b). Therefore, these are six independent mutated clones. One gene trap clone STA4.1, was originally identified by 5'RACE as a novel transcript located on mouse chromosome 1. SpPCR on *Sau*3AI digested STA4.1 genomic DNA amplified a fragment less than 100bp. However, Blast search against Ensemb and NCBI database didn't yield any significant hits. In order to

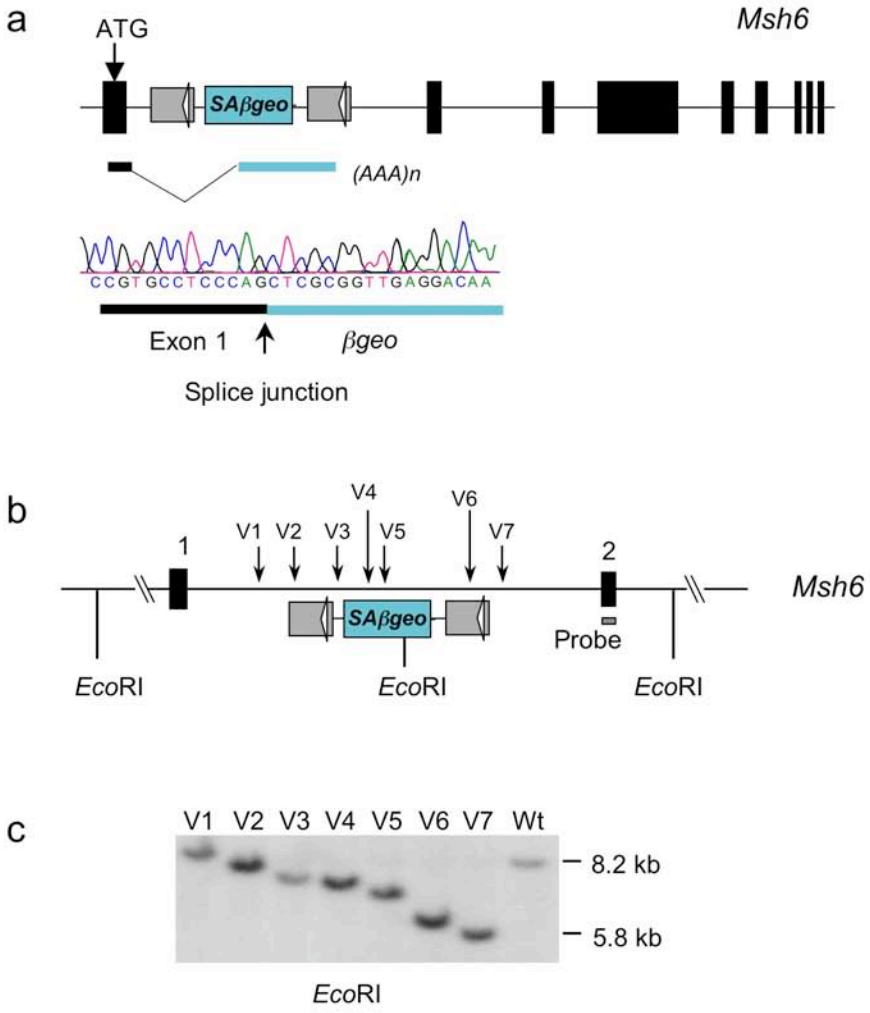


Figure 5-3. Identification of *Msh6* gene trap mutations

a. Shematic representation of insertion of a provirus in the first intron of *Msh6* locus, generating a fusion transcript of *Msh6* exon 1 with *βgeo*. *Msh6* exon1contains the translation initiation codon (ATG). **b.** Schematic demonstration of retroviral insertion sites in independent gene trap *Msh6* mutants. **c.** Various proviral/host junctions were demonstrated by Southern-blot analysis using a *Msh6* exon 2 probe in independent *Msh6* gene trap clones (V1 to V7). Note the *Msh6* gene trap mutants are homozygous, lacking the wild type allele. For Southern-blot analysis, the genomic DNA was digested with *Eco*RI.

obtain a longer flanking genomic sequence, Splinkerette oligos were designed so that SpPCR could be performed on genomic DNA digested with restriction enzymes, *Eco*RI*, Bam*HI*, Hind*III *and Xba*I. Compared to *Sau*3AI(four base pair cutter), these 6 base pair cutters generate longer restricted genomic fragments. SpPCR on *Xba*I and *Hind*III digested STA4.1 genomic DNA amplified a 1.3 kb and 800 bp flanking genomic DNA respectively, both of which were mapped to intron1 of *Msh6*.

Southern-blot analysis were carried out on *Eco* RI restricted genomic DNA from seven independent gene trap *Msh6* mutants, including STA4.1, using a *Msh6* exon2 probe. This probe revealed a 8.2 kb *Eco*RI fragment from the wild type *Msh6* locus, whereas the insertion of the gene trap virus resulted in different sized proviral/*Msh6* junction fragments. This result confirmed the SpPCR analysis. Importantly, all the seven independent *Msh6* gene trap mutants contain only the gene trap alleles. None of the clones retained the wild type *Msh6* allele, suggesting that all of the insertions were homozygous (Fig. 5-3 c).

5.2.2.2 Expression of *Msh6* is reduced in gene trap mutants

A *Msh6* cDNA probe spanning exon 2, exon 3 and exon 4 was PCR amplified from AB2.2 cDNA and Northern-blot analysis was performed to study the expression of *Msh6* in five gene trap *Msh6* mutants (Fig. 5-4 a). Compared to AB2.2 and the parental NGG5-3 cells, the gene trap *Msh6* mutants expressed only a trace level of *Msh6,* which suggests that the *SAßgeo* gene trap cassette in RGTV-1 can efficiently block the expression of host genes. Moreover, AB2.2 and NGG5-3 cells exhibited a similar level of *Msh6* expression, suggesting that *Blm* mutation doesn't affect *Msh6* expression in ES cells.

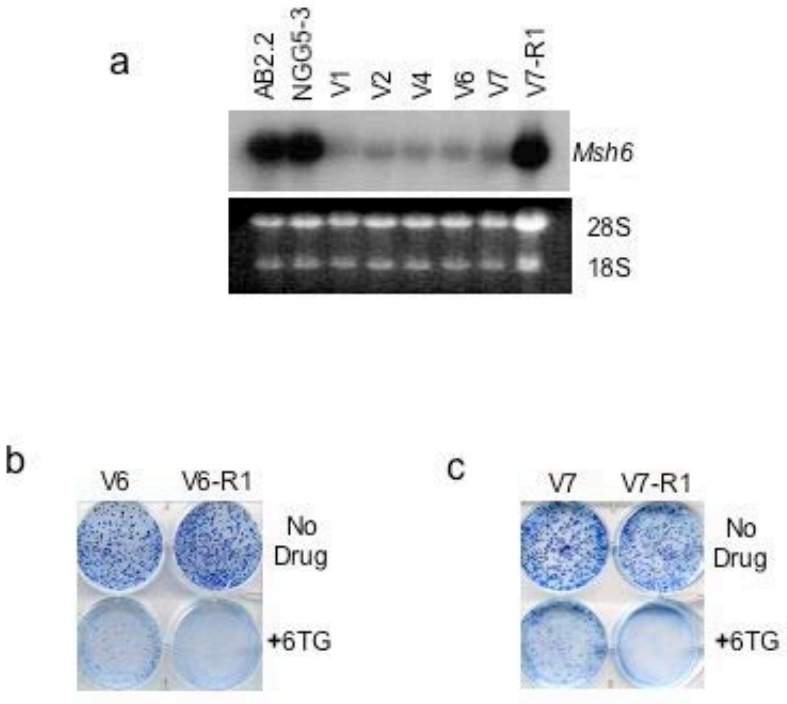


Figure 5-4. Molecular analysis of *Msh6* mutations

a. Northern-blot analysis of *Msh6* expression. Compared to *Msh6* wild type AB2.2 and NGG5-3 cells, *Msh6* gene trap clones (V1, V2, V4, V6, V7) exhibited trace level of *Msh6* expression. The excision of the provirus in Cre-revertant (V7-R1) restored the expression. NGG5-3 is the parental *Blm*-deficient cell line used for gene trap mutagenesis. 28S and 18S RNAs are loading controls. **b & c**. Colony form ability of cells with(+6TG) and without 6TG selection(No drug), demonstrating the 6TG sensitivity in Cre revertants (V6-R1 and V7-R1), in contrast to the 6TG resistance in *Msh6* gene trap mutants (V6 and V7).

5.2.2.3 Cre-mediated reversal of gene trap mutations

The reversibility of gene trap mutations recovered from the STA screen was tested by Cre-mediated recombination. Cre-expressing plasmid was transfected into all 25 STA clones by electroporation. Cells that have lost the inserted provirus, the revertants, were identified by PCR using a pair of *lacZ* primers for loss of both copies of the *SAßgeo* gene trap cassettes. These revertants were sib-selected in G418 and verified to be G418 sensitive. Two to three revertants from each gene trap clones were plated at low density to test the colony forming ability with and without 6TG selection. Two non-revertants from each cell line were plated as controls. 12 out of 25 tested STA clones exhibited recovery of sensitivity to 6TG after Cre-mediated removal of the retrovirus. These clones belong to the seven independent Msh6 gene trap mutants (Table 5-1) (Fig. 5-4 b & c). The *Msh6* expression in one of the revertants was examined by Northern-blot and confirmed that it returned to the normal level despite the presence of a LTR in the intron (Fig. 5-4 a). The recovery of homozygous gene trap mutants of known components of the mismatch repair machinery validates the establishment of the recessive genetic screen. It is notable that the Cre-reversal assay was performed before SpPCR identified the gene trap mutations. The fact that the Cre-reversal assay was able to recover all *Msh6* mutants demonstrated its efficiency. In addition to seven independent *Msh6* gene trap mutants, nine individual STA gene trap mutants were also identified (Table 5-1). The 6TG resistance could not be reverted in these clones after removal of the integrated virus.

Table 5-1. Gene trap mutations in STA clones

Gene trap Clones[a]	Gene[b]	Chromosome	Reversal[b]	Viral insertion[d]
STA1.1	*Ctbp2*	Chr7	N	S
STA1.2 (3)	*Ctbp2*	Chr7	N	S
STA2.1	*Clasp2*	Chr9	N	S (Q)
STA2.2 (4)	*Msh6*	Chr17	Y	D
STA2.4	ESTT00000014070	Chr11	N	S (Q)
STA4.1	*Msh6*	Chr17	Y	D
STA5.1 (3)	*Clasp2*	Chr9	N	S (Q)
STA6.1	*CUGbp1*	Chr2	N	S (Q)
STA6.2	ENSMUSG00000020794	Chr11	N	S (Q)
STA6.3 (2)	*Msh6*	Chr17	Y	D
STA6.4	*Msh6*	Chr17	Y	D
STA7.1	*Eno1*	Chr4	N	S (Q)
STA7.2 (2)	*Msh6*	Chr17	Y	D
STA8.1	*Msh6*	Chr17	Y	D
STA8.2	*Ctbp2*	Chr7	N	S
STA8.3	*Msh6*	Chr17	Y	D

Table 5-1. Gene trap mutations in STA clones.

Gen trap mutations identified by SpPCR and 5'RACE methods in the STA screen.

a: The number in parenthesis represents the number of daughter cells.

b: Gene names were given as either Ensembl gene symbol or Ensemble ID if a gene symbol is not available.

c: Cre revertible clones were designated as "Y" and Non-reversible clones were designated as "N"

d: "S" represents single allelic retroviral insertion. "D" represents bi-allelic retroviral insertion. QTSouthern was used to inspect the copy number of viral insertions, which was designated as "Q" in parenthesis.

5.2.2.3 The copy number of gene trap insertions

In order to examine if the gene trap mutations were homozygous, a quantitative Southern-blot analysis (QTSouthern) was employed to investigate the copy number of the integrated retrovirus. QTSouthern compares the Southern hybridization intensity between the SA*ßgeo* gene trap cassette and a X-linked single copy gene, Adrenoleukodystrophy Protein Homolog (*Aldp*). Because homozygous mutations in this screen are derived predominantly through mitotic recombination, the majority of homozygous gene trap clones are expected to contain two copies of retrovirus (bi-allelic mutants). QTSouthern revealed that 5 out of the 7 *Msh6* gene trap clones had two copies of gene trap insertions. The other two clones appeared to have single gene trap insertions (Fig. 5-5 a). These single allele *Msh6* mutants may have a deletion mutation encompassing *Msh6* exon 2 in the other *Msh6* allele, which cannot be seen in Southern-blot analysis with the *Msh6* exon 2 probe (Fig. 5-3 c). Apart from the *Msh6* gene trap clones, all other STA clones appeared to contain single copy gene trap insertions (Fig. 5-5 b). Genomic flanking probes for two clones, STA1.2 and STA8.2, were generated from the cloned SpPCR product and Southern-blot analysis demonstrated that STA1.2 and STA 8.2 contained both wild type and the gene trap alleles, which confirms the result of QTSouthern analysis (Fig. 5-6). Although it is formally possible that the wild type allele displayed in the Southern-blot may contain mutations that cannot be identified by Southern-blot analysis, removal of the single allele gene trap mutations from these clones will generate heterozygous mutants, which should be 6TG sensitive. The fact that all nine single allelic gene trap STA clones were not revertible argues against this possibility and suggests that they are false positive clones.

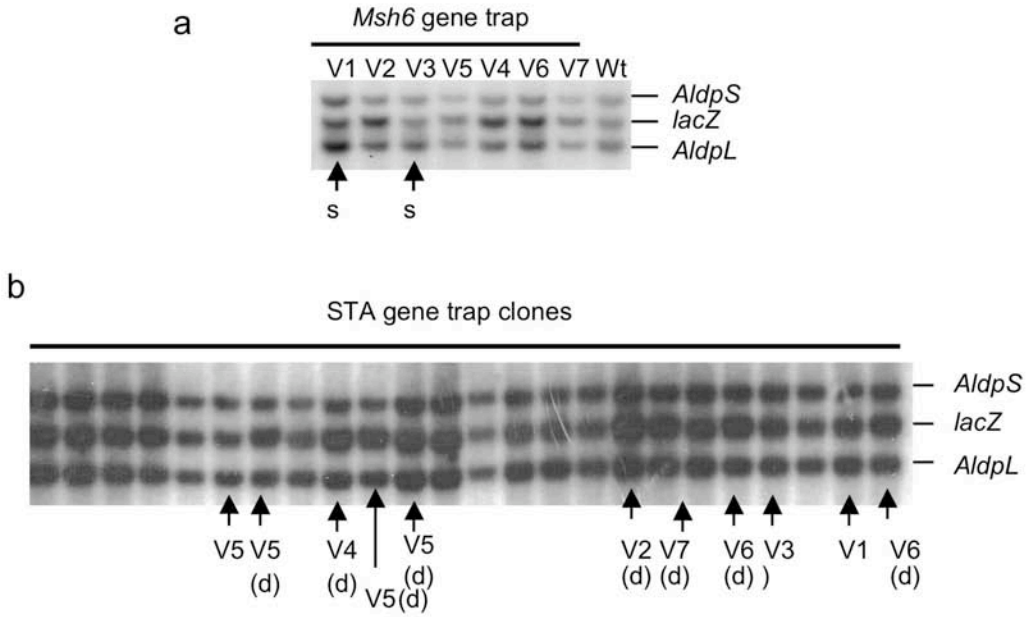


Figure 5-5. Quantitative Southern-blot analysis (QTSouthern) of STA gene trap mutants

a. QTSouthern analysis of seven *Msh6* gene trap clones (V1 to V7) from the STA screen. Clones containing a single copy of the viral insertion are indicated by arrows (s). **b.** QTSouthern analysis of 24 gene trap clones from STA screen, demonstrating single allele gene trap insertions. *Msh6* mutants are indicated by arrows. "d" indicates bi-allelic viral insertion. Genomic DNA was digested with *Eco*RV and *Hind*III and probed with *AldpS, AldpL and LacZ probes*. *Aldp*L and *AldpS* are two reference probes for the X-linked single copy gene, Adrenoleukodystrophy Protein Homolog. *LacZ* probes the inserted virus.

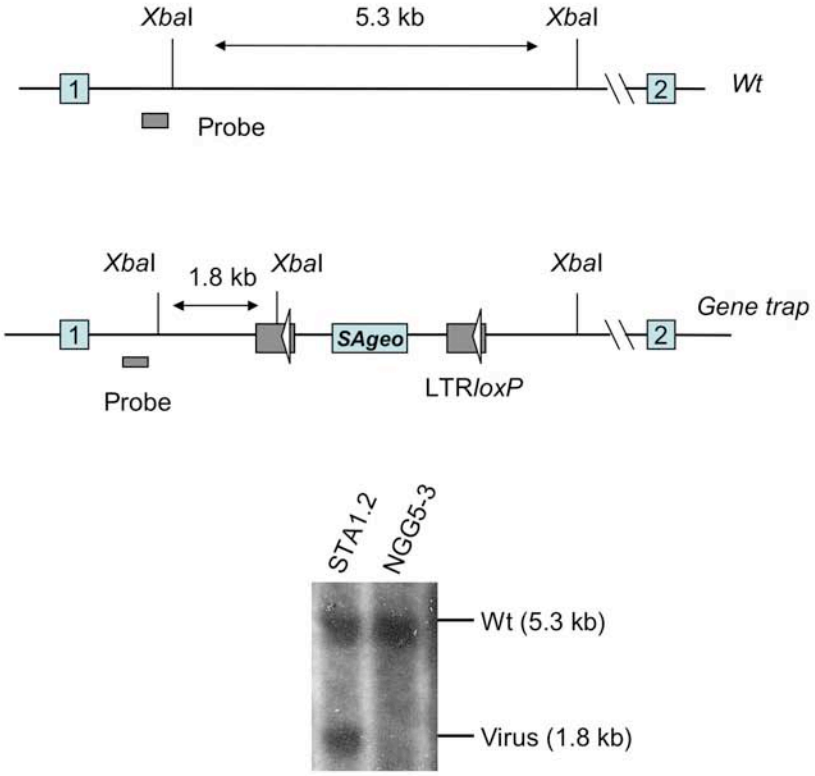


Figure 5-6. Southern-blot analysis of gene trap clone STA1.2,

demosntrating a single allele mutation. Both the wild type (*Wt*) and the gene trap allele (*Gene trap*) are present in this clone. The provirus inserted into the first intron of C-terminal binding protein 2 (*Ctbp2*) in this clone.

5.2.3 STB screen

Derivation of homozygous mutations by LOH is a stochastic process which is related to the number of population doublings. To allow more gene trap mutations to be converted to homozygous mutations, the gene trap mutants were expanded for four more population doublings for the STB screen. The 6TG screen was performed under the same conditions as the STA screen. In total 104, 6TG resistant clones were recovered in the STB screen. Based on the analysis of STA screen, it was expected that a portion of the 6TG resistant clones would be single allelic gene trap mutations. To identify the potential homozygous mutations, QTSouthern analysis was performed. From this analysis 24 clones were identified as potential bi-allelic mutants and 58 clones were identified as single allelic mutants (Fig. 5-7). The other clones cannot be determined either because of bad Southern-blot signals or the cells were lost during expansion or were slow growing. Retroviral integration sites were identified in 24 potential bi-allelic gene trap mutants (Table 5-2 a). 12 of these are *Msh6* mutants, which is consistent with the result of STA screen. Three new genes were also identified, including *Dnmt1* (DNA (cytosine 5) methyltransferase), *Tgif* (5'-TG-3' interacting factor) and a complex locus with a genomic rearrangement involving *Parp-2* (poly (ADP-ribose) polymerase-2) and *Rbpsuh* (Recombining binding protein suppressor of hairless). Gene trap mutations were identified in some of the single allelic gene trap mutants and listed in table 5-2 b.

5.2.3.1 *Dnmt1* gene trap mutant

QTSouthern identified three bi-allelic mutants from GT library pool 8, and Southern-blot analysis of the provirus/host junctions using *lacZ* probe suggested that they were daughter cells. SpPCR analysis of two clones revealed that RGTV-1 inserted in the first intron of the *Dnmt1* locus. This gene trap mutation was named *Dnmt1-V1*. A Cre-revertant clone, *Dnmt1-V1-R1,* was obtained from *Dnmt1-V1*. Southern-blot analysis was performed on *Nde* I digested genomic

DNA using a *Dnmt1* probe, which was PCR amplified from AB2.2 genomic DNA. The Southern-blot analysis revealed the predicted *Dnmt1* wild type (1.7 kb), *Dnmt1-V1* (4.0 kb*)* and *Dnmt1-V1-R1* (2.3 kb*)* allele (Fig. 5-8 a & b). Importantly, This Southern analysis demonstrated that *Dnmt-V1* and another *Dnmt1* gene trap mutant, *Dnmt1-V2* are homozygous mutants, containing only gene trap alleles. *Dnmt1-V2* was recovered from the STC screen (discussed later) (Fig. 5-8 b).

The expression of *Dnmt1* and the fused gene trap transcripts were examined by RT-PCR in the *Dnmt1* gene trap mutant and the revertant. Because the retrovirus inserted into the first intron of *Dnmt1*, a fused transcript composed of exon 1 of *Dnmt1* and the *ßgeo* reporter should be expressed in the *Dnmt1* gene trap mutant. RT-PCR using *Dnmt1* Exon1 and the *LacZ* primers revealed the expression of the fused transcript in *Dnmt1-V1* cells, but not in *Dnmt1* wild type NGG5-3 cells and the Cre-revertant, *Dnmt1-V1-R1* cells. RT-PCR using *Dnmt1* Exon 1 and Exon 6 primers didn't detect *Dnmt1* expression in *Dnmt1-V1* cells and *Dnmt1* expression was reverted to normal level in *Dnmt1-V1-R* cells, compared to that in NGG5-3 cells (Fig. 5-9 a). The expression of *Dnmt1* was further investigated by Northern-blot analysis using a *Dnmt1* cDNA probe spanning *Dnmt1* exon 1 to exon 6. This experiment revealed that the expression of *Dnmt1* was totally blocked in *Dnmt1-V1* cells, suggesting that the gene trap mutation produced a null allele. AB2.2, NGG5-3 and *Dnmt1-V1-R1* cells exhibited similar level of *Dnmt1* expression (Fig. 5-9 b). *Dnmt1-V1-R1* and *Dnmt1-V1* cells were plated at low density in 6-well tissue culture plate to test the colony forming ability with and without 6TG selection, which showed that *Dnmt1-V1-R1* cells recovered 6TG sensitivity (Fig. 5-9 c).

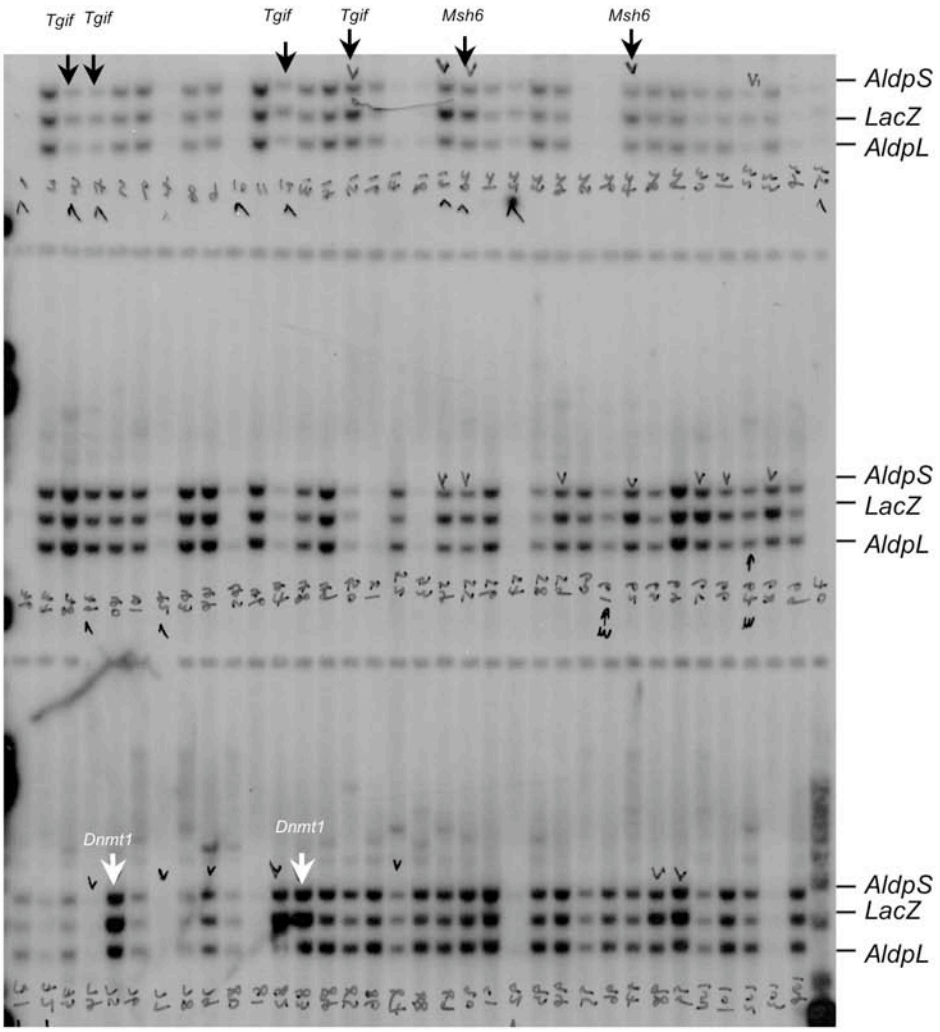


Figure 5-7. QTSouthern analysis of gene trap mutants obtained in the STB screen.

Genomic DNA was extracted from gene trap clones cultured in 96 well tissue culture plates, digested with *Eco*RV and *Hind*III and probed with *AldpS, AldpL* and *LacZ* probes. Arrows indicate some picked bi-allelic gene trap insertions in *Tgif, Msh6*, *Dnmt1* loci, which were identified by SpPCR.

Table 5-2 a. Genes mutated in bi-allelic gene trap STB clones

Clone	Gene trap mutations	Chromosome	Reversal	Hom/Het
STB1 (8)	*Tgif*	Chr17	N	Hom
STB20 (12)	*Msh6*	Chr17	Y	Hom
STB77 (3)	*Dnmt1*	Chr9	Y	Hom
STB60	*Parp-2/Rbpsuh*	Chr14/Chr5	N	Het

Table 5-2 a. Genes mutated in bi-allelic gene trap STB clones

24 potential bi-allelic STB clones were identified by QTSouthern.
The number in parenthesis represents the number of gene trap clones with insertions in the identified gene. STB1 represents one *Tgif* gene trap mutation with 8 daughter cells. STB77 represents one *Dnmt1* mutant with 3 daughter cells. STB20 represents *Msh6* mutants that are recovered with several independent gene trap mutations. The exact number of independent gene trap *Msh6* mutations was not determined in the STB screen.
Cre revertible clones were designated as “Y” and Non-reversible clones were designated as “N”
“Hom” represents homozygous mutant; “Het” represents heterozygous mutant
Note that STB60 was revealed to be a heterozygous mutant at the *Rbpsuh* locus by Southern-blot analysis using a *Rbpsuh* probe.

Table 5-2 b. STB gene trap clones

Clone	Ensembl ID	Gene	Chromosome	Copy number
STB8	ENSMUSG00000002379	2010012C24Riken	Chr17	S
STB9	ENSMUSG00000027572	Death inducer-obliterator-1; Apoptosis	Chr2	S
STB10	ENSMUSG00000024097	Splicing factor, arginine/serine-rich 7	Chr17	–
STB12	ENSMUSG00000049397	Pk3; Pyruvate kinase 3	Chr9	–
STB14	Novel	No description	Chr4	–
STB26	ENSMUSG00000050498	E3 ubiquitin protein ligase	Chr10	S
STB29	ENSMUSG00000034017	Msi2h; RNA-binding protein MUSASHI2-L homologue(Drosophila)	Chr11	S
STB30	ENSMUSG00000002302	Atf1; Cyclic-AMP-dependent transcription factor	Chr15	S
STB36	ENSMUSESTT00000024752		Chr6	S
STB39	not determined		Chr13	–
STB43	ENSMUSESTT00000043659		Chr8	S
STB45	ENSMUSESTG00000027010		chr7	S
STB50	ENSMUSG00000008450	NTF2; neclear transporter factor2)	Chr8	–
STB53	ENSMUSESTT00000006383		unlocalized fragment	
STB55	EMSMUSG00000026439	Retinoblastinoma-binding protein	Chr1	S
STB58	ENSMUSG00000020794	Ube2g2; Ubiquitin-conjugating enzyme E2 G1	Chr11	S
STB60	ENSMUSG00000036023	Parp-t2 (Poly(ADP-ribose) polymerase2) (5'RACE product)	Chr14	–
STB60	ENSMUSG00000039191	Rbpsuh (Splinkerette PCR product)	Chr5	S
STB68	ENSMUSESTT00000046619		chr10	–
STB73	Msh6		Chr17	–
STB78	ENSMUSG00000005732	Ranbp1; RAN binding protein 1	chr16	S
STB83	ENSMUSESTG00000017407		chr13	–
STB86	ENSMUSG00000004563	Heterogenous nuclear ribonuclear proteins C1/C2	Chr14	–
STB87	Msh6		Chr17	S
STB93	not determined		Chr10	–
STB95	ENSMUSG00000011960	cyclin T1	Chr15	S
STB96	ENSMUSESTT00000038626		Chr18	–

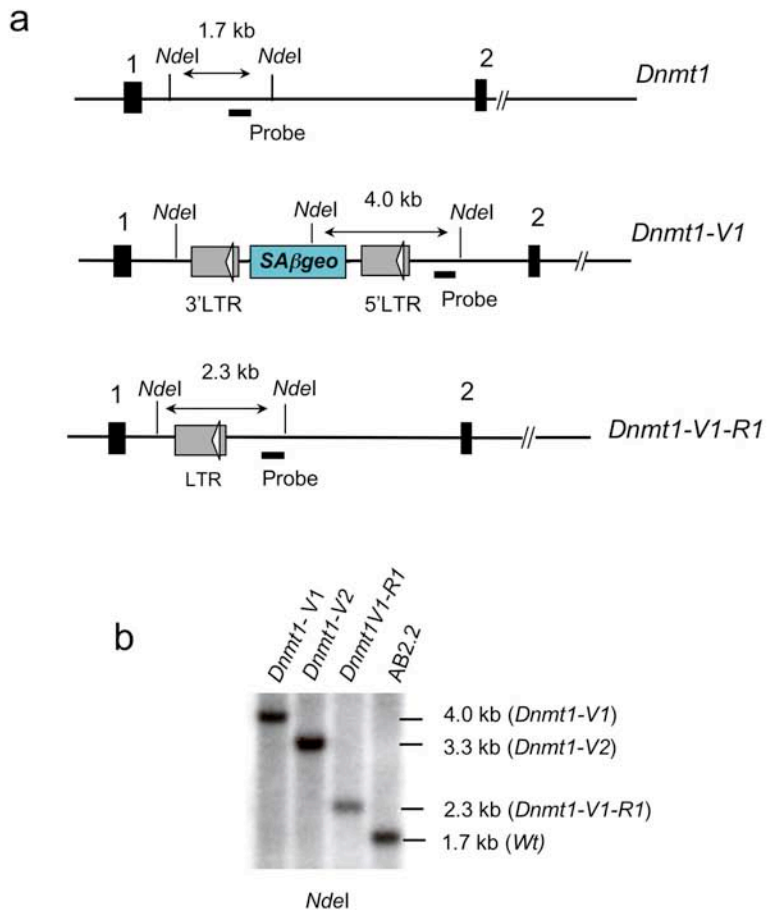


Figure 5-8. Southern-blot analysis of *Dnmt1* gene trap mutations

a. Schematic of the structure of the 5' protion of *the Dnmt1* genomic locus, demonstrating the Southern-blot analysis strategy that distiguishes the wild type, gene trap and the Cre-reverted *Dnmt1* alleles. **b.** Southern- blot showing two gene trap *Dnmt1* mutants (*Dnmt1-V1, Dnmt1-V2*), *Dnmt1-V1* derived Cre-revertant, *Dnmt1-V1-R1* and wild type cells AB2.2, using *Dnmt1* flanking probe (Probe) on *Nde*I digested genomic DNA. Note that *Dnmt1-V2* was recovered from the STC screen.

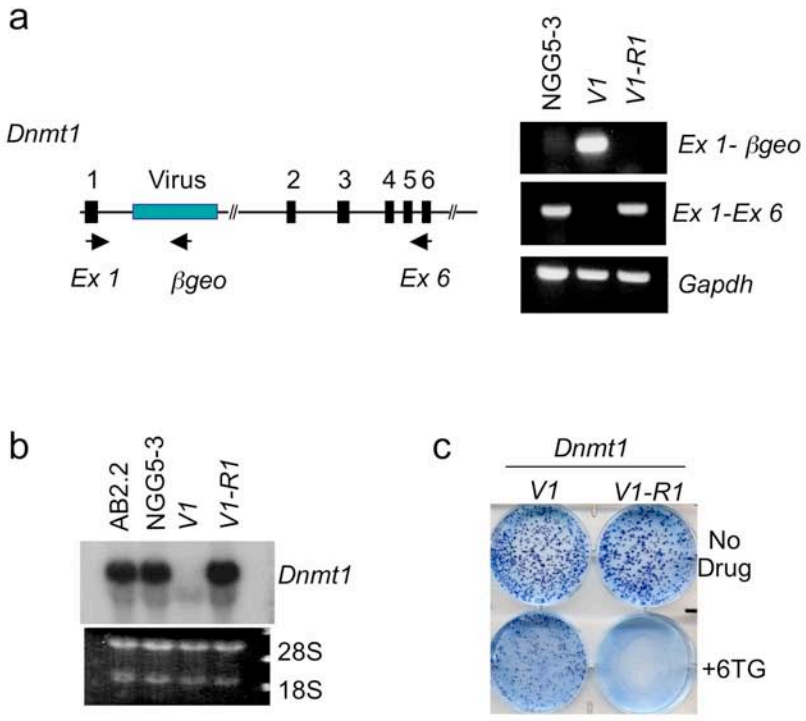


Figure 5-9. Molecular analysis of gene trap *Dnmt1* mutants

a. RT-PCR analysis of *Dnmt1* transcription, showing the expression of the fused *Dnmt1*exon1/ *βgeo* transcript in gene trap *Dnmt1-V1* but not in the Cre-revertant *Dnmt1-V1-R1* cells (PCR amplification by *Ex 1 and βgeo* primers). NGG5-3 cell line was included as a control. The expression of *Dnmt1* was blocked in the gene-trap clone and reverted in the *Dnmt1-V1-R1* (PCR amplification by *Ex 1 and Ex 6 primers).* **b.** Northern-blot analysis showing the absence of *Dnmt1* expression in *Dnmt1-V1* cells and the reversal of *Dnmt1* expression in *Dnmt1-V1-R1* cells . **c .** Cre reversal assay, showing the conversion from 6TG resistance of *Dnmt-V1* to 6TG sensitivity of *Dnmt1-V1-R1*.

5.2.3.2 *Tgif* gene trap mutant

QTSouthern and the proviral/host junction analysis identified a bi-allelic mutant consisting of 8 daughter clones from GT library, pool 1. SpPCR analysis revealed that retrovirus had inserted into 5' UTR region of *Tgif* genomic locus. This gene trap clone was named *Tgif-V1.* Southern-blot analysis on *Xba* I digested genomic DNA using a *Tgif* flanking probe revealed the predicted 5.2 kb wild type allele in NGG5-3 cells and the 3.8 kb gene trap band in three *Tgif-V1* daughter clones, confirming that *Tgif-V1* was a bi-allelic gene trap mutant (Fig. 5-10 a & b). One Cre-revertant clone, *Tgif-V1-R1*, was generated by Cre-mediated recombination. PCR analysis using *LacZ* primers identified the deletion of the inserted provirus (Fig. 5-11 a). However, when *Tgif-V1-R1* cells were plated in 6TG, they exhibited the same level of resistance to 6TG as the parental *Tgif-V1* cells (Fig. 5-11 b). To exclude variation between individual clones, three more Cre-revertants were derived from *Tgif-V1*, and the colony forming ability was tested. Consistent with previous results, they all exhibited resistance to 6TG (data not shown).

By 5'RACE, the splice junction of the fused gene trap transcript was cloned. Sequence analysis of the 5'RACE product revealed that *ßgeo* was spliced with an exon located about 1 kb upstream of the retroviral insertion site. Database searches against mouse Ensembl, NCBI as well as human Ensembl did not identify any known transcripts or ESTs. This novel transcript was named *Tgif-*γ. Based on NCBI and ensemble databases, two other *Tgif* transcripts exist, which share the common exon 2 and exon 3 and with the alternatively spliced first exon. These two transcripts were referred to as *Tgif-*α (ENSMUST00000059775) and *Tgif-*β (ENSMUST00000055383) respectively (Fig. 5-12 a).

RT-PCR was performed to inspect the expression of *Tgif-*α, *Tgif-*β and *Tgif-*γ in *Tgif-V1*, *Tgif-V1-R1* and the parental NGG5-3 cells. RT-PCR using *Tgif* alternative exon1 primers and an exon3 primer detected the expression of *Tgif-*α

and *Tgif-γ* in NGG5-3 ES cells (Fig. 5-12 b & c), but no expression of *Tgif-β* could be detected. The fused gene trap transcript was amplified using *Tgif exon1-γ* and *lacZ* primers in the gene trap *Tgif-V1* cells, but not in the NGG5-3 and *Tgif-V1-R1* cells (Fig. 5-12 b). The expression of *Tgif-γ* could not be detected in *Tgif-V1* cells, but this was reverted to normal in the *Tgif-V1-R1*, showing that the expression of *Tgif-γ* is fully blocked by the gene trap insertion and reverted to normal in the Cre-revertant (Fig. 5-12 b). RT-PCR analysis using *Tgif-α* exon1 and exon3 primers also revealed a reduced expression of *Tgif-α* in *Tgif-V1* cells and the expression returned to normal in*Tgif-V1-R1* cell (Fig. 5-12 c). These results suggested that the 6TG resistance phenotype exhibited *Tgif-V1* cells was not caused by the gene trap *Tgif* mutation since 6TG resistance didn't reverted to 6TG sensitivity in *Tgif-V1-R1* cells. The real mutation that causes the 6TG resistance phenotype in *Tgif-V1* cells is thus unknown. It is possible that the retroviral insertion affects the function of a novel gene, which has not been identified yet. Or mutations have occurred randomly in other mismatch proteins, which cause 6TG resistance. No change in the expression of *Msh6* was detected in *Tgif-V1* cells by RT-PCR analysis using *Msh6* exon1 and exon 3 primers (data not shown).

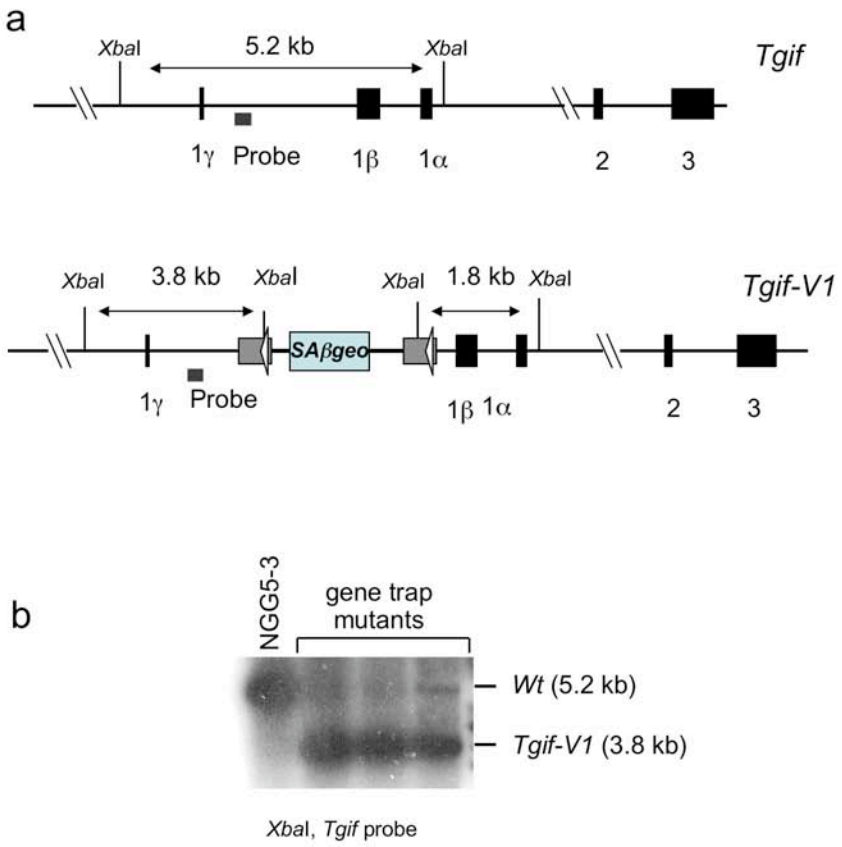


Figure5-10. Southern-blot analysis of *Tgif* gene trap mutants.

a. Schemaic of *Tgif* genomic locus, demonstrating the Southern-blot analysis strategy that displays *Tgif* wild type and gene trap *Tgif-V1* alleles. 1α, 1β and 1γ are three alternative spliced forms of exon 1. **b.** Souhern-blot shows three bi-allelic gene trap *Tgif* mutants.

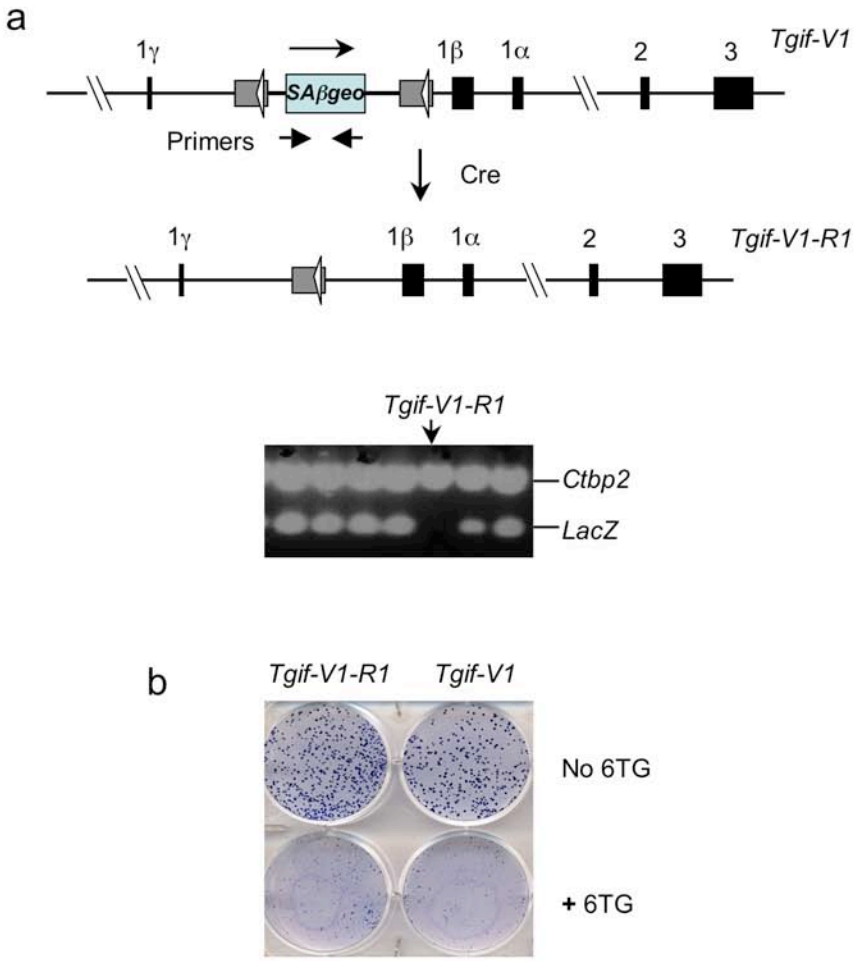


Figure 5-11. Cre reversal assay of *Tgif* gene trap mutation

a. Cre-mediated removal of inserted retrovirus from *Tgif-V1* to generate *Tgif-V1-R1*. *Tgif-V1-R1* was screened by PCR for absence of amplification of gene trap cassette *SAβgeo*. *Ctbp2* (C-terminal binding protein 2) primers were included as a PCR control (Materials and Methods 2.3.4). The arrow above the *SAβgeo* cassette indicates the direction of its transcription. **b.** Colony forming ability assay, showing the 6TG resistance of both *Tgif-V1* and the Cre-revertant *Tgif-V1-R1.*

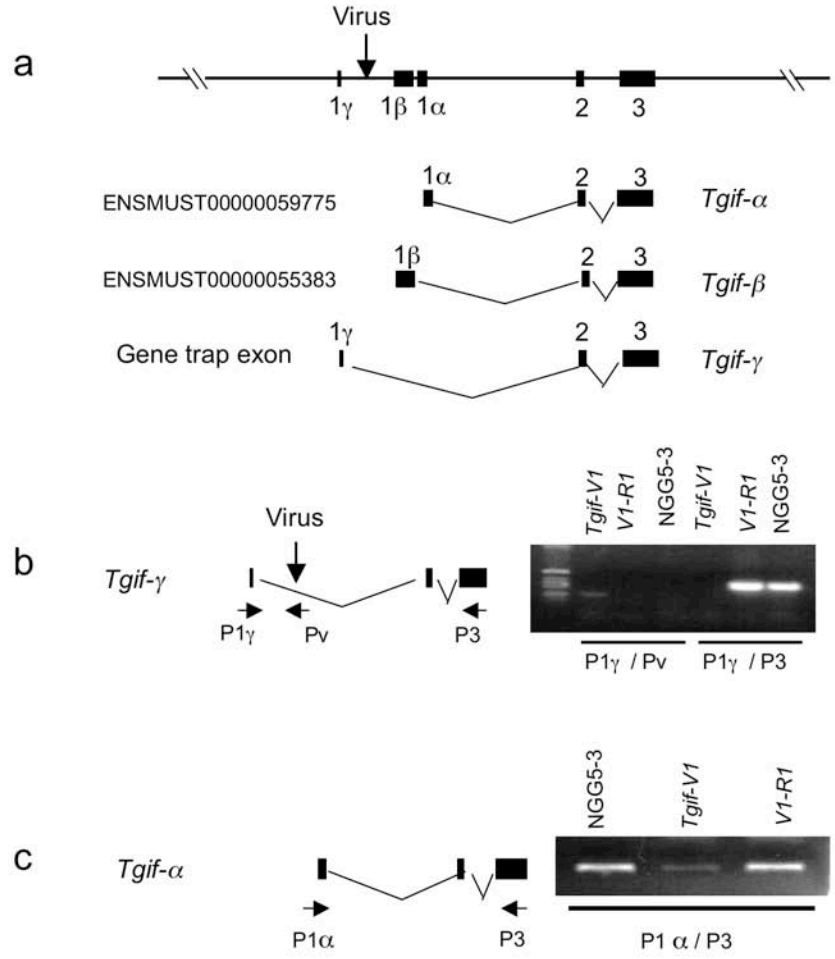


Figure 5-12. RT-PCR analysis of *Tgif* expression

a. Schematic representation of three alternative *Tgif* transcripts, *Tgif-α*, *Tgif-β* and *Tgif-γ*. *Tgif-γ* was identified by 5'RACE in the gene trap *Tgif-V1* mutant. **b**. RT-PCR analysis showing the expression of *Tgif-γ / βgeo* fused transcript in *Tgif-V1* mutant, but not in the Cre-revertant, *Tgif-V1-R1* cells. The expression of *Tgif-γ* was interrupted in *Tgif-V1* cells, indicated by the lack of PCR amplification of the *Tgif-γ* transcript in the *Tgif-V1* cells, but the transcript was restored in the *Tgif-V1-R1* cells. NGG5-3 cells are used as control. **c.** Expression analysis of *Tgif- α* by RT-PCR, which is reduced in the *Tgif-V1* mutant and reverted in Tgif-V1-R1. Note that the expression of *Tgif-β* can not be detected by RT-PCR in ES cells (data not shown).

5.2.3.3 Identification of a complex locus, Parp-2/Rbpsuh

SpPCR identified the viral insertion site in one of the potential bi-allelic mutants (STB60), which revealed that the retroviral inserted into the first intron of *Rbpsuh*, a gene located on mouse chromosome 5. 5'RACE identified the fusion transcript, which revealed that the *ßgeo* reporter was spliced to *Parp-2* exon1, which is on mouse chromosome 14 according to Ensembl (Fig. 5-13 a). The discrepancy between the 5'RACE result and the SpPCR result may come from cross contamination between two gene trap cell lines that carry mutations in *Parp-2* and *Rbpsuh.* However, such cell-to-cell contamination was excluded because this clone had been single cell cloned by seeding cells at low density before 5'RACE and SpPCR were performed. Based on this evidence, a reciprocal chromosomal translocation may have occurred that places the retrovirus that inserted in the *Rbpsuh* locus under the transcription control of *Parp-2* (Fig. 5-13 b). This translocation event will place both *Parp-2* and *Rbpsuh* out of frames. Southern-blot analysis using a *Rbpsuh* probe revealed that STB60 contained the predicted gene trap allele as well as the wild type *Rbpsuh* allele. Therefore, it is a heterozygous gene trap mutant (Fig. 5-13 c).

Two Cre-reverted clones were obtained from STB60 and both exhibited the 6TG resistant phenotype as the parental STB60 cells (data not shown). Thus the mutations in STB60 cannot be reverted by Cre-mediated removal of the inserted retrovirus. Because of the complexity of this locus, the real molecular lesion that causes 6TG resistance is not clear. *Parp-2* may be a better candidate. *Parp-2* encodes ADP ribose polymerase 2, one member of the poly (ADP ribose) polymerase family, which includes three genes, *Parp-1, Parp-2, and Parp-3* (Johansson, 1999, Ame et al., 1999). Parp-1 and Parp-2 proteins are activated by DNA strand breaks and catalyze the post-translation modification of some nuclear proteins by adding a ADP-ribose moiety, which has functional implications in DNA repair, cell cycle regulation and cell death. MEFs from a *Parp-2* knockout mouse exhibited increased post-replicative genomic instability, G2/M cell cycle arrest following exposure to alkalizing agents (Menissier de

Murcia et al., 2003). Moreover, *Adprtl1 (Parp-1)* and its homolog were identified in the screen in *C.elegans* for genes that protect *C.elegans* genome against mutations. This screen also identified other mismatch repair genes (Pothof et al., 2003). *Rbpsuh* is also referred to as recombination signal sequence-binding protein *J-kappa* (*Rbp-J*). *Rbpsuh* (*Rbp-J*) encodes a transcription factor that is involved in embryonic and adult development (Schroeder et al., 2003).

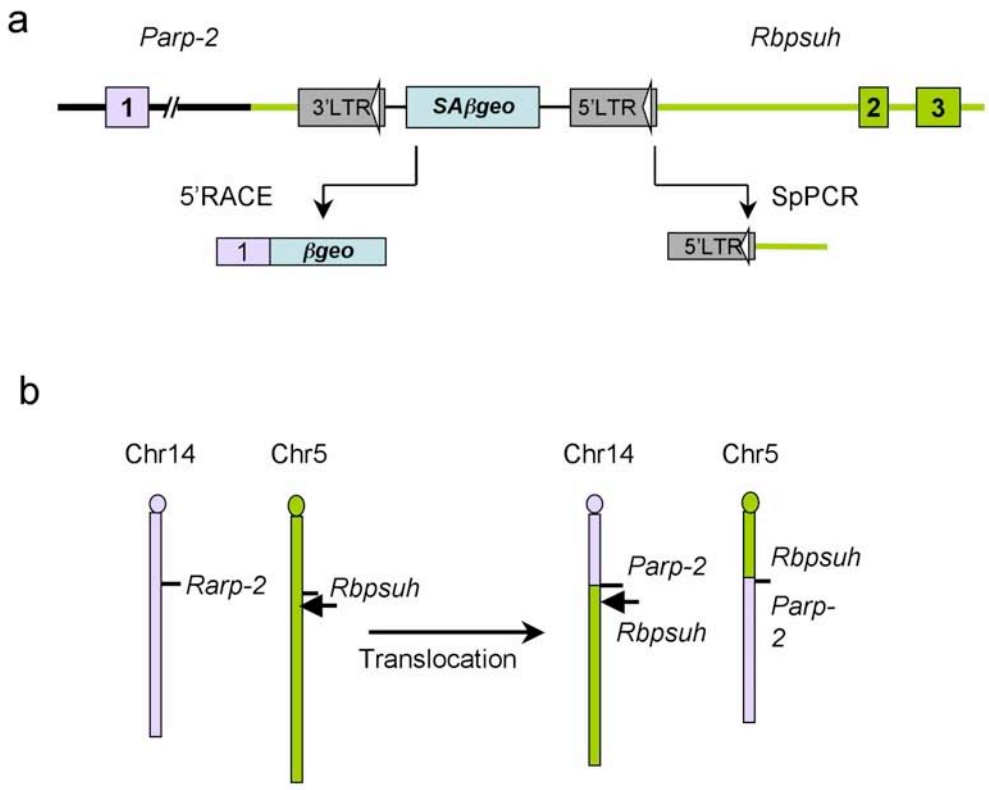
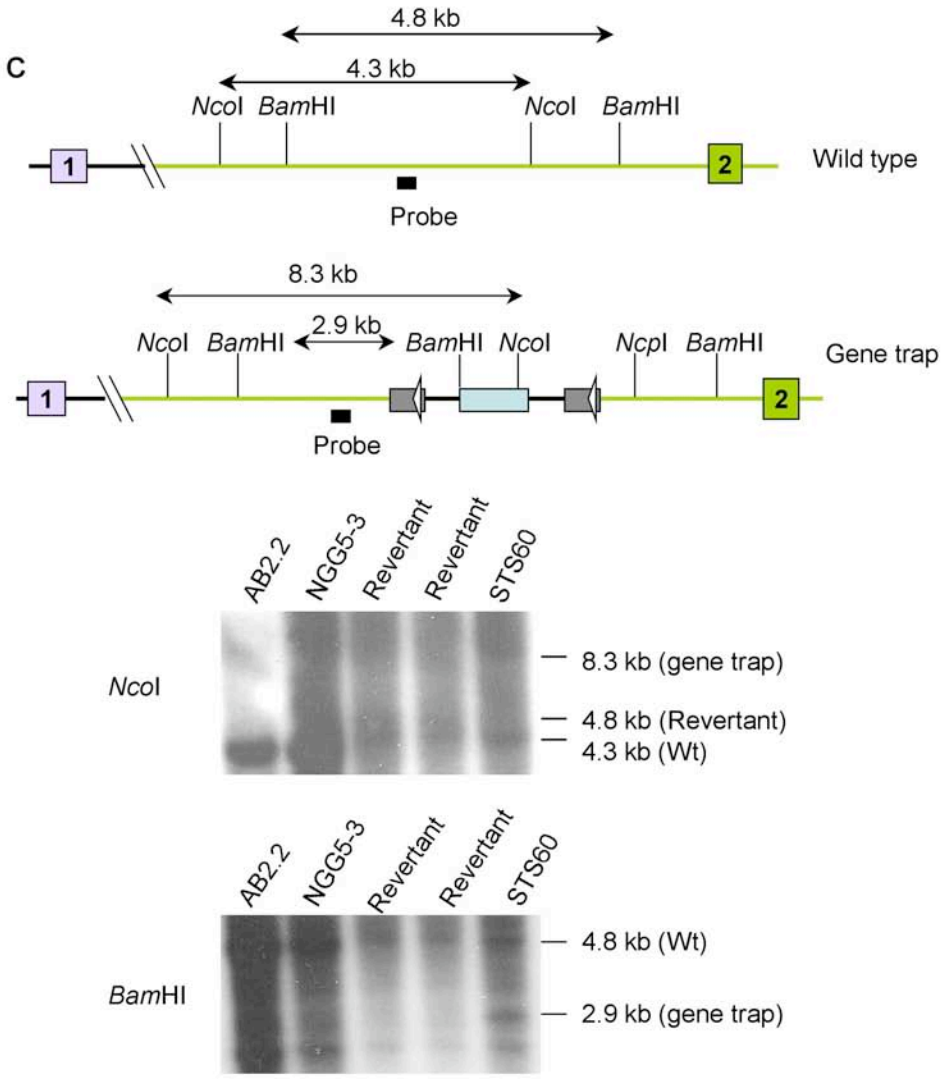


Figure 5-13. Characterization of a translocated *Parp-2/Rbpsuh* locus

a. The schematic representation of a complex genomic locus involving both *Parp-2* and *Rbpsuh* genomic loci. SpPCR identified the viral insertion site in the first intron of *Rbpsuh*. 5'RACE identified the splice junction of the gene trap transcript including *Parp-2* exon 1. **b.** A model showing the formation of *Parp-2/Rbpsuh* locus by reciprocal translocation. Arrow indicates the viral insertion site.



c. Southern analysis of gene trap clone STB60 (*Parp-2/Rbpsuh*) using a *Rbpsuh* probe. The gene trap mutation was identifed as the predicted 8.3 kb *Nco*I or 2.9 kb *Bam*HI fragments. Two Cre revertants showed the 4.8 kb *Nco*I fragment. STB60 is a heterozygous gene trap mutant containing the wild type *Rbpsuh* allele.

5.2.4 STC screen

Although several independent *Msh6* mutants have been identified in the STA and STB screens, other known mismatch repair genes, *Msh2*, *Mlh1*, *Pms2* were not identified. It was also observed that the *Msh6* gene trap mutants were more resistant to 6TG treatment than many other gene trap mutants. This raises the concern that the 6TG concentration used in the STA and STB screens might be too high for mutants that only have weak 6TG tolerance. Compared to genes involved in 6TG metabolism, it is likely that most mutants that modify the mismatch repair process or genome surveillance have a modest tolerance to 6TG. In an effort to recover these genes, the 6TG concentration was titrated using the gene trap *Parp-2/Rbpsuh* clone as a control for 6TG resistance because this clone exhibited a weak 6TG resistance phenotype in a colony forming ability assay. Based on this pilot experiment, a new 6TG screen (STC screen) was performed with 6TG selection at 0.5 μM for10 days. A total of $5x10^8$ gene-trapped cells that have been passaged about 18 population times were plated for this screen. Roughly, 800 6TG tolerant clones were picked into 96 well tissue culture plate. These clones (STC clones) were composed of a variable number of daughter cells from independent mutations represented in the primary pools. To establish relationships between clones, Southern-blot analysis was performed to inspect the proviral/host junction fragments at both 5'LTR and 3'LTR sides using *LacZ* and *Neo* probes on *Eco* RI digested genomic DNA (Fig. 5-3 a). With this method, daughter cells, exhibiting the same hybridization pattern could be grouped (data not shown). Many gene trap *Msh6* mutants were identified by Southern-blot analysis using *Msh6* probe and were excluded from further analysis (data not shown). QTSouthern identified 119 potential bi-allelic mutants. Sequence information was obtained from 82 clones by SpPCR and/or 5'RACE (Table 5-3). Genes that have been recovered as homozygous mutants in the STB screen were also identified in this screen. Sequence analysis revealed that *Msh6, Tgif* and *Dnmt1*, account for 18 out of 82 identified gene trap mutations in STC clones. Two *Msh6* clone (STC3-D4 and STC3-G9) were

identified from GT library pool 3, in which *Msh6* mutants were not recovered in the STA and STB screens. This mutation was therefore counted as a new *Msh6* gene trap allele. A new *Dnmt1* allele was also recovered (which is presented as *Dnmt1* allele B in table 5-3. This new allele was named *Dnmt1-V2.* Southern-blot analysis using a *Dnmt1* probe revealed that the *Dnmt1-V2* was a homozygous mutant (Fig. 5-8 a).

The Cre reversal assay were performed on 44 STC clones. Three revertants of each clone were plated at low density in 24 well tissue culture plates to test the colony forming ability in 6TG at various 6TG concentrations, 0.15 μM, 0.3 μM or 0.5 μM. This assay demonstrated that 6TG tolerance could be reverted in two clones, *Dnmt1-V2* mutant and a clone STC4-F11 (Fig. 5-14 a). In STC4-F11, the retrovirus inserted in a novel gene (ENSMUSG00000032361, Ensembl) on mouse chromosome 9, which is a member of a family of genes related to MORF4 (mortality factor on chromosome 4) (Bertram et al., 1999). The human homolog (MRG15) is functionally implicated in cell cycle progression (Pardo et al., 2002). This gene is named as *mMRG9* for mouse MORF related gene on chromosome 9. Southern-blot analysis using a *mMRG9* probe revealed that STC4-F11 is a heterozygous mutant, containing both the gene trap and the wild type alleles (Fig. 5-14 b). The 6TG tolerance phenotype may be a result of haploinsufficiency. However, it cannot be excluded that the “wild type” allele detected by Southern-blot analysis may carry point mutations or small deletion /insertion mutation.

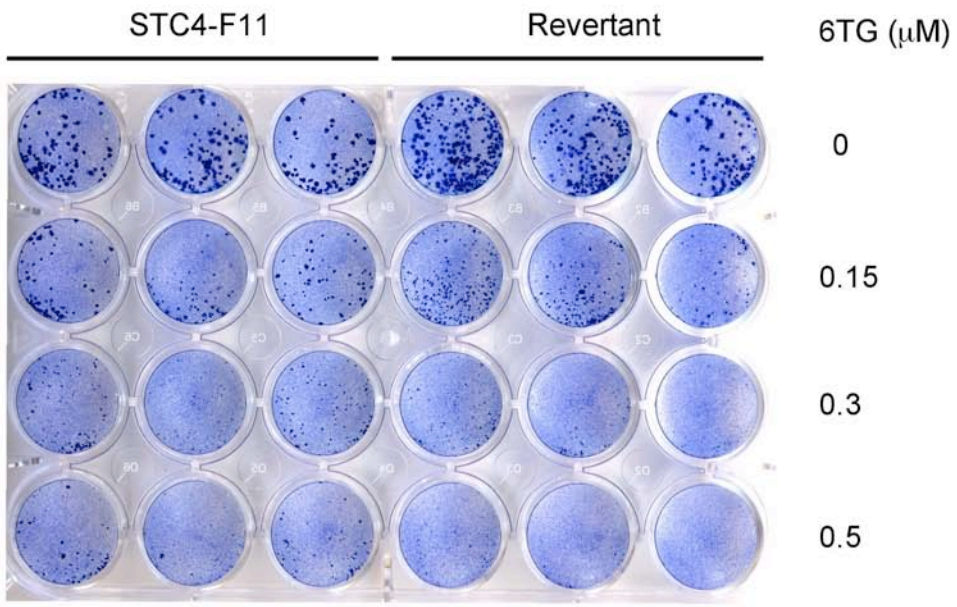


Figure 5-14. Characterization of the gene trap mutation in STC4-F11

a. Cre-reversal assay. Three STC4-F11single cell clones (left three columns) and three revertants (right three columns) were plated at low density in 24 well tissue culture plates and cultured at various 6TG dosages.

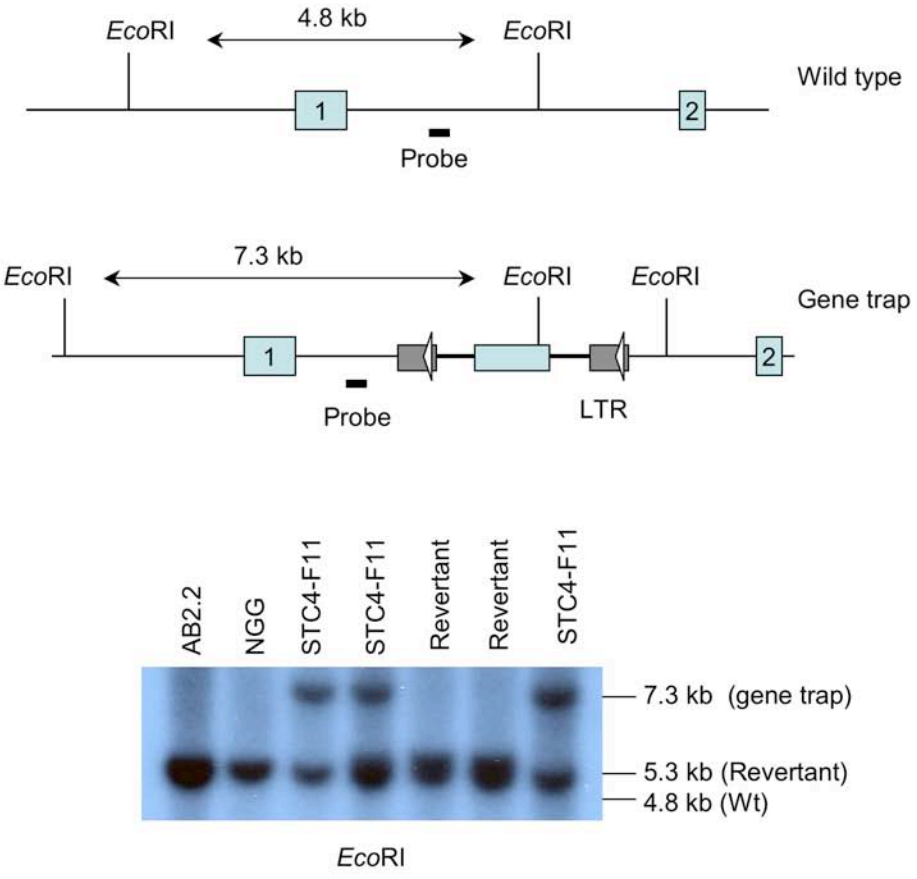


Figure 5-14. Characterization of the gene trap mutation in STC4-F11

b. Southern analysis of STC4-F11 showing the insertion of the provirus in the first intron of *mMRG9,* a member of a family of genes related to MORF4. A *mMRG9* probe revealed the 4.8 kb *Eco*RI fragment from the wild type locus and the 7.3 kb fragment from the gene trap insertion. Two Cre-revertants showed a 5.3 kb *Eco*RI fragment.

5.2.5 Single allelic or non-revertible gene trap mutants

Many gene trap mutations were identified, which contain only a single gene trap allele and/or the phenotype was not revertible. Loss of the wild type allele can occur through many mechanisms. Apart from mitotic recombination, loss of the wild type allele can occur by single nucleotide changes and insertion/deletion mutations. At least two *Msh6* clones were identified as “homozygous” mutations, but contained single gene trap allele, implying that loss of the wild type allele may have occurred by chromosomal deletion. Single nucleotide changes or small insertion and deletion mutations would not be identified by the Southern analysis strategies used in this study, therefore these mutants will appear as a single allele gene trap mutations.

The reversibility of the gene trap virus was enabled by the Cre-*loxP* mediated removal of inserted retrovirus. After Cre-*loxP* mediated recombination, the *βgeo* gene trap cassette is deleted, but a single LTR remains in the genome. Although, it is common that retrovirus gene trap vectors insert into introns, in a few cases the virus inserts into an exon or an UTR region. In such cases, the remaining LTR may disrupt the gene’s expression; for example, clone STC2-E3 (Table 5-3), in which the provirus has inserted into the 5’UTR region of the transacting transcription factor 1 (*Sp1*). It has been reported that *Msh6* contains seven functional Sp1 binding sites and binding of Sp1 and the transacting transcription factor 3 (Sp3) to these sites contribute to *Msh6* promoter activity (Gazzoli and Kolodner, 2003).

Other than the complexity caused by the various mechanisms of loss of the other allele, the virus insertion sites or some times the complicated (rearranged) host gene structure, it is expected that most of gene trap clones should be homozygously mutated and be revertible. The abundance of non-revertible and heterozygous mutants in the 6TG screens implies that many of these clones are false positive clones, in which the 6TG resistant phenotype is not caused by the

gene trap mutation. Compared to the STA and STB screens, the portion of false positive clones is extremely high in the STC screen. The low efficiency of the STC screen is caused predominantly by the low stringency 6TG selection used in the screen, which causes high selection background. Many STC clones recovered were not “real” 6TG resistant. 50% of the STA clones and nearly 70% of the STB clones are either non-revertible and/or heterozygously mutated, but are resistant to high 6TG concentrations. A possible explanation for these is that mutations may have accumulated in mismatch repair genes. To investigate this further, Southern-blot strategies were designed to detect genomic rearrangements in *Msh2, Msh6, Mlh1* and *Dnmt1* using cDNA probes on 28 non-revertible gene trap clones. A *Msh6* cDNA probe spanning exon 1 to exon 4 revealed a homozygous change in exon 3 in two clones, STA 5.1 and STA7.1 (Fig. 5-15). Southern-blot analysis with *Msh6* exon2 probe revealed that STA1.2 might contain a deletion in *Msh6* exon2 (data not shown). No obvious genomic rearrangements were observed in non-revertible gene trap clones in *Msh2*, *Mlh1*, and *Dnmt1* loci (Fig. 5-16, Fig. 5-17, Fig. 5-18). It must be pointed out that single nucleotides changes, small insertions and deletions are unlikely to be detected by this method.

Recovery of gene trap clones with homozygous genomic rearrangement in mismatch repair genes reflect the instability of the *Blm*-deficient genetic background, which allows random mutations occurring at a low frequency to segregate homozygous mutation. The ratio of the positive clones (homozygous/revertible clones versus total clones) decreases from 50% in the STA screen to 30% in STB screen with an extended cell doubling time, implying that more homozygous random mutations were generated during the prolonged cell culture in the STB screen. This process might have been exaggerated by the 6TG selection. 6TG forms mismatched 6-mG/T nucleotides that will affect the coding information if occurring in a gene. Also, the 6-mG/T mismatch can be processed into DNA strand breaks by MMR machinery, which leads to deletion/insertion and chromosome translocations.

Table 5-3. STC gene trap clones. Page 1

Clone	Chr	5'RACE	SpPCR	Gene/Function domain
1-A5	Chr17		ENSMUSG00000047407	Tgif; TALE family homeobox; transcription factor
1-A11	Chr11		ENSMUSG00000005209	Tnfsf12; TNF-ralated weak inducer of apoptosis
1-B1	Chr14	ENSMUSG00000004563		Heterogenous nuclear ribonuclear proteins C1/C2
1-B3	Chr17		ENSMUSG00000047407	Tgif; TALE family homeobox; Transcription factor
1-B10	Chr17	ENSMUSG00000047407		Tgif; TALE family homeobox; Transcription factor
1-B11	Chr17		ENSMUSG00000047407	Tgif; TALE family homeobox; Transcription factor
1-C2	Chr8	ENSMUSG00000036180		p66 ahpha homologue
1-C5	Chr8	ENSMUSG00000036180		p66 ahpha homologue
1-E5	Chr10	ENSMUSESTT00000005442		No description
1-E6	Chr8	ENSMUSESTG00000017323		No description
1-F5	Chr9	ENSMUSG00000032449		Mitochondrial carrier protein - Rim2p/Mrs12p
1-F8	Chr17	ENSMUSG00000047407	ENSMUSG00000047407	Tgif; TALE family homeobox; transcription factor
1-G3	Chr10	ENSMUSG00000020235		FZR1 protein (Drosophila); cell division related protein
1-G7	Chr8	ENSMUSG00000036180		p66 ahpha homologue
1-G9	Chr5	Novel transcript	Novel transcript	Predicted by genescan; alleleA; no description for function
1-H6	Chr18	ENSMUSG00000034391	ENSMUSG00000034391	Cyclin-like F-box
2-A11	Chr11	ENSMUSG00000013415	ENSMUSG00000013415	Igf2bp1; insulin-like growth factor 2, binding protein 1
2-E3	Chr15	Chr15/Novel	ENSMUSG00000001280	UTR region of Sp1; Trans-acting transcription factor 1
2-F1	Chr4	ENSMUSG00000028580	ENSMUSG00000028580	PUMILIO 1(DROSOPHILA); RNA binding; translation regulation
2-F3	Chr6	ENSMUSG00000042079		RNA binding motif
2-F6	Chr9		ENSMUSG00000032397	Timeless-interacting protein; Cell cycle control
3-D3	Chr5	ENSMUSG00000029267		Metal response element binding transcription factor 2; Transcription factor
3-D4	Chr17	Chr17/close to Msh6	ENSMUSG00000005370	New allele of Msh6 (alternative splicing?)
3-G1	Chr2		ENSMUSG00000026794	2900073H19Rik; Ubiquitin-like protein
3-G9	Chr17		ENSMUSG00000005370	Msh6; DNA mismatch repair protein
4-A10	Chr6	ENSMUSG00000042079	ENSMUSG00000042079	RNA binding motif
4-B4	Chr15		ENSMUSG00000002302	Atf1; Cyclic-AMP-dependent transcription factor
4-9B	Chr11	ENSMUSG00000018362		Kpna2; Intracellular trafficking, secretion, and vesicular transport
4-C3	Chr13	ENSMUST00000042517	ENSMUST00000042517	Novel transcript

Table 5-3. STC gene trap clones. Page 2

Clone	Chr	5'RACE	SpPCR	Gene/Function domain
4-E3	Chr2	ENSMUSG00000026917		Wdr5; WD40 repeat-containing protein
4-E6	Chr2	ENSMUSG00000027572		Death inducer-obliterator-1; Apoptosis
4-F3	Chr10		ENSMUSESTG00000004368	No description
4-F4	Chr2	ENSMUSG00000026917		Wdr5; WD40 repeat-containing protein
4-F11	Chr9	ENSMUSG00000032361	ENSMUSG00000032361	MORF-related gene 15 protein; Ageing
4-G6	Chr9		ENSMUSG00000004099	Dnmt1; C-5 cytosine-specific DNA methylase
6-5A	Chr17	ENSMUSG00000034868		Mylc2b; Myosin regulatory light chain-like
6-A7	Chr9	ENSMUSG00000004099	ENSMUSG00000004099	Dnmt1(alleleB); C-5 cytosine-specific DNA methylase
6-B2	Chr5		ENSMUSG00000056421	Gtf2ird1; Transcription factor GTF3 gamma 2
6-B4	Chr17	ENSMUSG00000005370	ENSMUSG00000005370	Msh6; DNA mismatch repair protein
6-B5	Chr11		ENSMUSG00000020794	Ube2g2; Ubiquitin-conjugating enzyme E2 G1
6-B11	Chr1	ENSMUSG00000054221		Retinoblastoma-binding protein 5
6-B12	Chr1		ENSMUSG00000026174	RCD1 homologue; required for cell differentiation
6-C2	Chr9		ENSMUSG00000004099	Dnmt1(alleleB); C-5 cytosine-specific DNA methylase
6-C10	Chr11	ENSMUSG00000020149		Rab1; Ras related protein
6-C11	Chr18	ENSMUSG00000024231		Cullin2; E3 ubiquitin ligase; protein turn over
6-D4	Chr5		ENSMUSG00000056421	Gtf2ird1; Transcription factor GTF3 gamma 2
6-D11	Chr15	BY731911		Novel EST
6-E6	Chr9	ENSMUSG00000004099	ENSMUSG00000004099	Dnmt1(alleleB); C-5 cytosine-specific DNA methylase
6-E12	Chr1	ENSMUSESTG00000003813		
6-G2	Chr18	ENSMUSG00000025420		ATP-binding; protein turnover
6-G7	Chr7	ENSMUSG00000030970	ENSMUSG00000030970	C-terminal binding protein 2;Transcription corepressor
6-G9		Chr11/Rab1	Chr9/DNMT1(alleleB)	Dnmt1(alleleB); C-5 cytosine-specific DNA methylase
6-H1	Chr9	ENSMUSG00000004099	ENSMUSG00000004099	Dnmt1(alleleB); C-5 cytosine-specific DNA methylase
6-H3	Chr5	ENSMUSG00000056421		Gtf2ird1; Transcription factor GTF3 gamma 2
6-H8	Chr5	ENSMUSG00000029267		metal response element binding transcription factor 2
6-H10	Chr5	ENSMUSG00000040731		Wbscr1; Williams-Beuren syndrome chromosome region 1 homolog
6-H11	Chr15	novel EST		No description
7-B10	Chr11	ENSMUSG00000009079	ENSMUSG00000009079	Ewsh; RNA-binding protein EWS

Table 5-3. STC gene trap clones. Page 3

Clone	Chr	5'RACE	SpPCR	Gene/function domain
7-C4	Chr17		AY036118	Positive cell proliferation regulator
7-C6	Chr13		ENSMUSG00000041817	Same gene clone 4-C3, second gene trap allele
7-C9	Chr11	ENSMUSESTG00000010948	ENSMUSESTG00000010948	Novel EST
7-D3		Chr5/ close to Rbpsuh	Chr19/ENSMUSG00000024949	ENSMUSG00000024949/Zinc finger protein 162
7-D6	Chr15	Chr15/novel	ENSMUSESTT00000028621	Chromobox containing protein
7-E2	Chr15	ENSMUSESTG00000022513		No description
7-E9	Chr7		ENSMUSESTG00000013850	No description
7-E10	Chr5	Rbpsuh	Novel transcript	Same gene as clone 1-G9 (allele B); Recombining binding protein suppressor of hairless
7-E11	Chr1	ENSMUSG00000025982		Sf3b1; splicing factor 3b, subunit 1
7-F7	Chr6		ENSMUSG00000029767	Calumenin
7-F9	Chr5	ENSMUSG00000029144	ENSMUSG00000029144	ATP-binding domain; G-protein beta WD-40 repeat
7-F10	Chr5	Novel transcript	23.72-23.73Mb	Genomic insertion is about 6 Mb away from the gene 5'RACE located
8-A1	Chr9		ENSMUSG00000004099	Dnmt1(allele A)
8-A3	Chr9	ENSMUSG00000004099		Dnmt1(allele A)
8-E7	Chr7	ENSMUSG00000041769		Protein phosphatase 2A, regulatory subunit B, delta isoform
8-E2	Chr9		ENSMUSG00000004099	
8-E4	Chr2		ENSMUSG00000027010	Solute carrier family 2; Mitochondrial substrate carrier
8-C6	Chr1	ENSMUSG00000042772		TPR repeat
8-F9	Chr9		ENSMUSG00000004099	Dnmt1(allele A)
8-H2	Chr9	ENSMUSG00000004099	ENSMUSG00000004099	Dnmt1(allele A)
8-H6	Chr17		ENSMUSG00000024002	Brd4; interact with replication factor C, inhibition cell cycle progress to S phase
8-H5	Chr11	Homologue to human unigene cluster Hs.295734		
8-H8	Chr1	ENSMUSESTT00000002455		No description

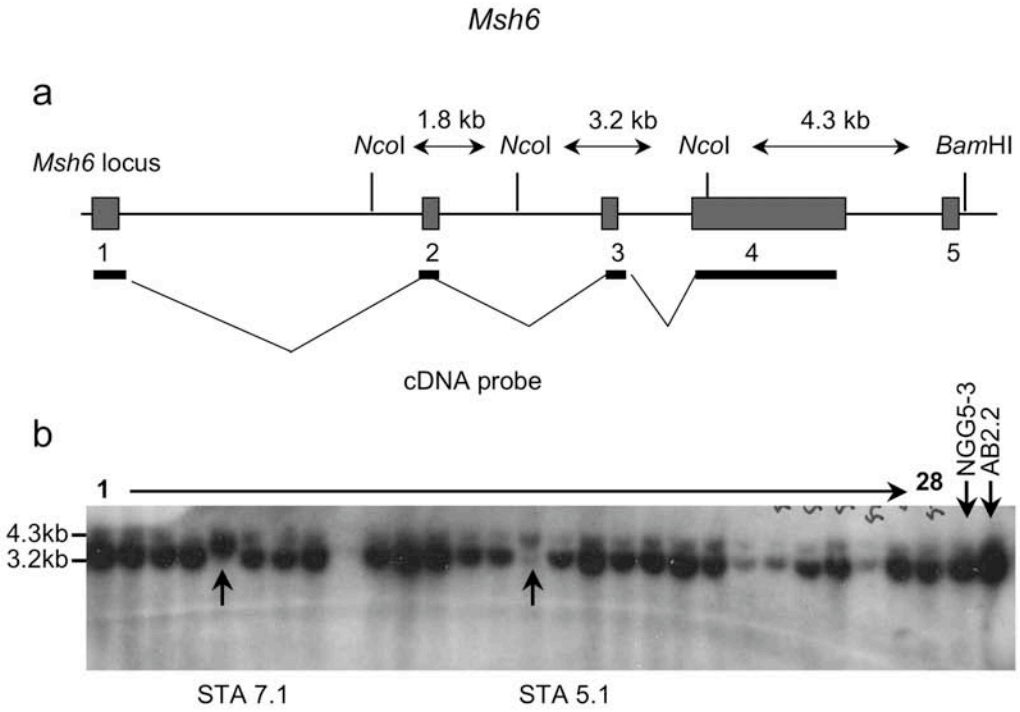


Figure 5-15. Southern analysis of *Msh6* locus in gene trap clones.

a. Schematic representation of 5' portion of *Msh6* locus, showing the Southern analysis scheme. cDNA probe spanning *Msh6* exon1 to exon4 recognizes 1.8 kb, 3.2 kb *Nco*I fragments and a 4.3 kb *Nco*I/*Bam*HI fragments. **b**. Southern blot showing the *Msh6* locus in 28 gene trap clones. NGG5-3 and AB2.2 were included as controls. Note that the 3.2 kb fragment including *Msh6* exon 3 was not detected in two gene trap clones from the STA screen, STA7.1 and STA 5.1.

The order of clones (from 1 to 28): STA1.1, STA1.4, STA8.2, STA6.1, STA7.1 , STC2-F2, STC4-B4, STC4-F11, STC7-A11, STB78, STC7-B10, STB60,Tgif, STB9, STA5.1, STB12, STC6-C10, STB50, STB58, STC1-C5, STC2-F1, STC3-D3, STC6-D4, STC2-E3, STC8-H6, STC2-A11, STC6-C11, STC4-A1

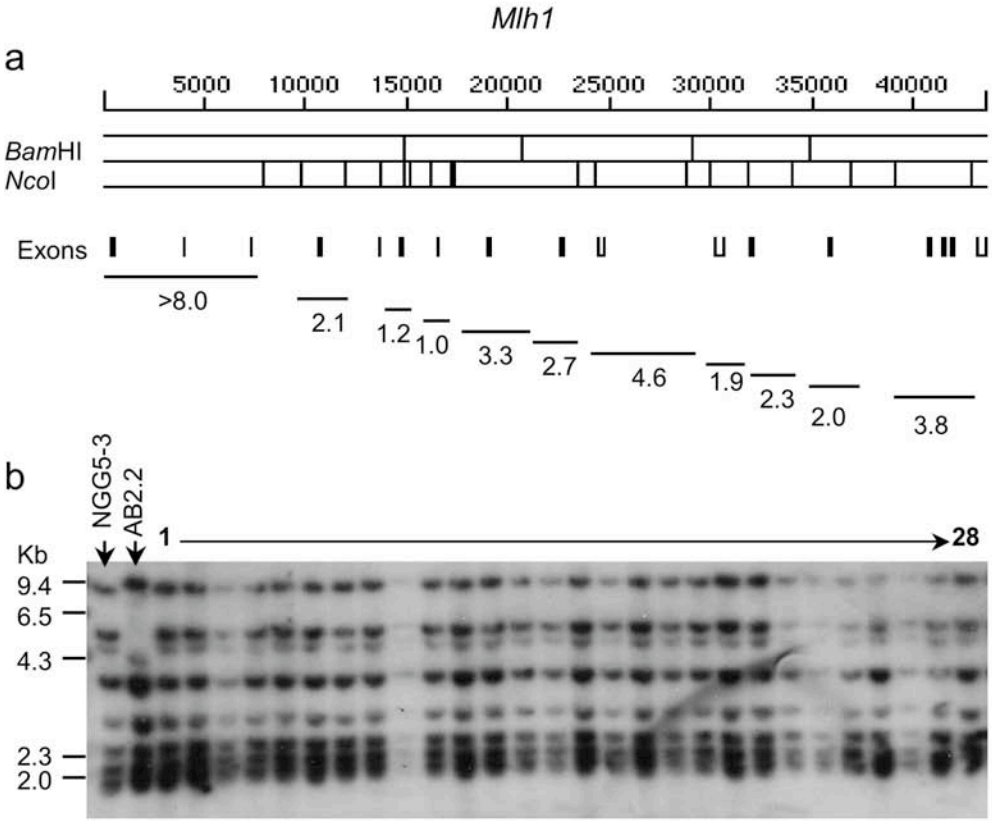


Figure 5-16 Southern analysis of *Mlh1* locus in gene trap clones.

a. Restriction map of *Mlh1*, showing *Bam*HI and *Nco*I restriction digestion sites and the positions of *Mlh1* exons. **b**. Southern-blot showing the *Mlh1* locus in 28 gene-trap clones. Genomic DNA was digested with *Bam*HI/*Nco*I and hybridized with a full length *Mlh1* cDNA probe. NGG5-3 and AB2.2 were included as controls. Note the variation between AB2.2 and NGG5-3 derived cells lines.

The order of clones(from 1 to 28): STA1.1, STA1.4, STA8.2, STA6.1, STA7.1 , STC2-F2, STC4-B4, STC4-F11, STC7-A11, STB78, STC7-B10, STB60,Tgif, STB9, STA5.1, STB12, STC6-C10, STB50, STB58, STC1-C5, STC2-F1, STC3-D3, STC6-D4, STC2-E3, STC8-H6, STC2-A11, STC6-C11, STC4-A1

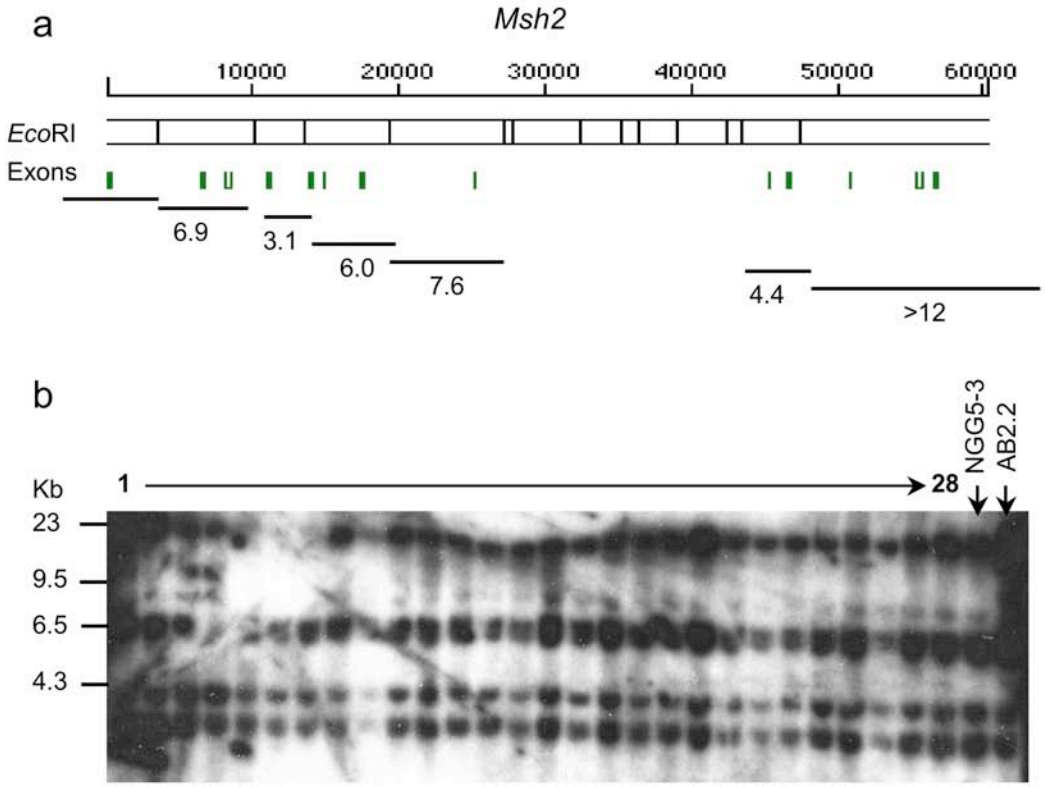


Figure 5-17 Southern analysis of *MSh2* locus in genetrap clones.

a. Restriction map of *Msh2*, showing *Eco*RI restriction digestion sites and the postions of *Msh2* exons. **b**. Southern-blot showing *Msh2* locus in 28 gene trap clones. Genomic DNA was digested with *Eco*RI and hybridized with a full length *Msh2* cDNA probe. NGG5-3 and AB2.2 were included as controls.

The order of clones(from 1 to 28): STA1.1, STA1.4, STA8.2, STA6.1, STA7.1 , STC2-F2, STC4-B4, STC4-F11, STC7-A11, STB78, STC7-B10, STB60,Tgif, STB9, STA5.1, STB12, STC6-C10, STB50, STB58, STC1-C5, STC2-F1, STC3-D3, STC6-D4, STC2-E3, STC8-H6, STC2-A11, STC6-C11, STC4-A1

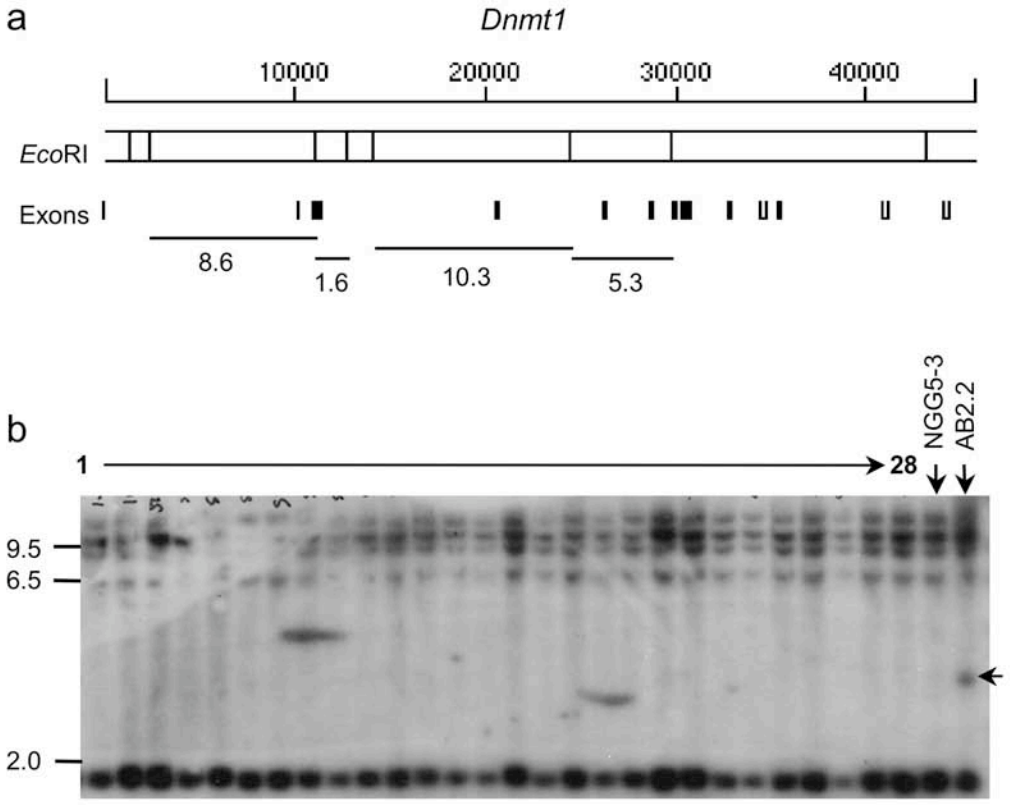


Figure 5-18 Southern analysis of *Dnmt1* locus in gene-trap clones

a. Restriction map of *Dnmt1*, showing *Eco*RI restriction digestion sites and the positions of *Dnmt1* exons. **b**. Southern-blot showing *Dnmt1* locus in 28 gene trap clones. Genomic DNA was digested with *Eco*RI and hybridized with a *Dnmt1* cDNA probe spanning from exon1 to exon6. NGG5-3 and AB2.2 were included as controls. Note the variation between NGG5-3 based cells and the AB2.2 cells (indicated by arrow).

The order of clones(from 1 to 28): STA1.1, STA1.4, STA8.2, STA6.1, STA7.1 , STC2-F2, STC4-B4, STC4-F11, STC7-A11, STB78, STC7-B10, STB60,Tgif, STB9, STA5.1, STB12, STC6-C10, STB50, STB58, STC1-C5, STC2-F1, STC3-D3, STC6-D4, STC2-E3, STC8-H6, STC2-A11, STC6-C11, STC4-A1

5.3 Discussion

5.3.1 Summary

In this chapter, genetic screens were performed to identify gene trap mutants that are resistant to 6TG. These mutants contain potential mutations in MMR mediated DNA damage surveillance. A gene trap mutation library was constructed using the RGTV-1 gene trap retrovirus on NGG5-3 cells, containing10, 000 individual gene trap clones. Three screens have been carried out with various 6TG concentrations and cell doubling times. In total, about a billion cells have been screened and about 900 ES cells exhibiting 6TG tolerance phenotype were picked into a 96-well tissue culture plate and analysed. Southern-blot strategies were designed to inspect the proviral/host junction fragments and the copy number of the inserted virus, so that daughter cells with the same gene trap insertions could be grouped and the potential bi-allelic mutants identified. Gene trap mutations were identified in 121 clones (representing STA, STB and STC screens) by 5'RACE or SpPCR methods. Bi-allelic mutations were identified in three genes, *Msh6, Dnmt1* and *Tgif*, which represent 11 independent gene trap mutations including 8 different *Msh6* insertions and 2 different *Dnmt1* insertions. The 6TG resistant phenotype is revertible in *Msh6* and *Dnmt1* mutants, but not in *Tgif* mutants. A revertible gene trap mutation (*mMRG9*) was identified in a novel gene encoding the mouse homologue of human *MRG15* gene. A complex gene trap mutation (*Parp-2/Rbpsuh*) involved a chromosome translocation, causing mutations in two genes *Parp-2* and *Rbpsuh* was also identified. *Parp-2* is the homolog of the *Parp* genes that were identified in a genetic screen in *C.elegans* for MMR genes (Pothof et al., 2003).

5.3.2 High throughput analysis of gene trap mutations

The molecular tag provided by the inserted retrovirus in the gene trap mutations allows high throughput molecular analysis of the mutations. Southern-blot analysis using viral specific probes can establish the unique proviral/host junction

fragments for each clone, so that related clones can be identified. This analysis can be applied to cells cultured on 96 well tissue culture plates and hundreds of gene trap clones can be studied at one time. This analysis is important in a screen based on *Blm*-deficient cells. Because homozygous mutants cells are segregated at random during cell expansion, early segregation will lead to some mutants (for example *Msh6*) dominating the pool. The unique proviral/host junction will identify these clones. Sub-dividing the screen into several pools also reduces the impact of early segregation from a single clone and provides additional evidence of independent mutations.

Gene trap mutations can be identified by PCR based methods. 5'RACE (Fig. 5-19) and SpPCR (Fig. 5-20) methods were modified in this study to suit the analysis of ES cells cultured on 96 well tissue culture plates (Materials and Methods 2.3.3 and 2.4.4). Although gene trap mutations can be identified by either 5'RACE or SpPCR method, they complement each other, providing information about gene trap expression and the viral integration site. The gene trap expression information can be used to identify transcripts, for example, *Tgif-*γ. Moreover, a complex locus with genomic rearrangement could also be identified, for example, the *Parp-2/Rbpsuh* locus was identified because SpPCR revealed that virus inserted into a genomic locus that was different from the locus predicted by 5'RACE.

The establishment of a revertible retroviral gene trap vector offers a high throughput means to validate the mutations. Cre-mediated reversal assay doesn't require prior knowledge of the mutated gene. Unlike the traditional cDNA rescue or more recently developed BAC rescue method, it doesn't require the construction of individual expression vectors. Cre-mediated recombination can be applied by electroporation of Cre-expression plasmid into ES cells cultured on 6-well plates. The revertants can be identified by PCR-based methods. All these aspects allow many gene trap mutants to be analyzed simultaneously at once.

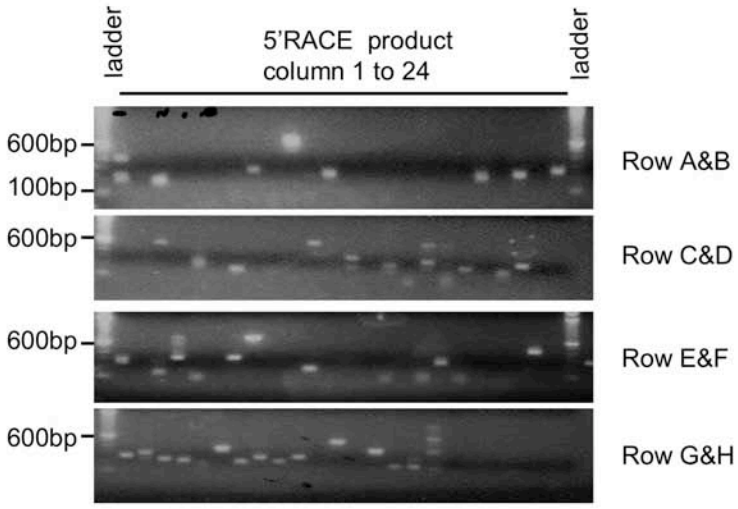


Figure 5-19. 5'RACE amplification of the splicing junction

High throughput 5'RACE reaction performed in 96 well format. Each panel represents 24 samples from rows A,B,C,D,E,F according to the layout of a 96 well plate. The size of PCR product ranges from 700 base pairs to 100 base pairs.

10 μl 5'RACE-PCR product was seperated on 1% agarose gel and stained with ethidium bromide.

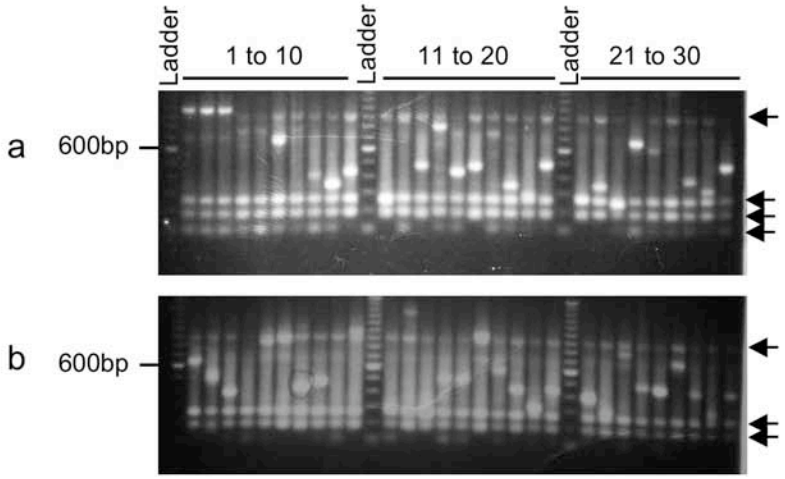


Figure 5-20. Proviral/host junctions identified by SpPCR

SpPCR products amplified from *Sau*3AI digested genomic DNA. Black arrows indicate the non-specific amplification from endogenous retrovirus. The amplified DNA fragments range from 1.6kb to less than 100 bp. Panel **a** and Panel **b** represents 60 samples from a 96 well plate.

30 μl nested SpPCR product was separated on 1% agarose gel and stained with ethidium bromide.

The fact that eight *Msh6* mutants and two *Dnmt1* mutants could be reverted from 6TG resistance to the 6TG sensitivity suggests the reversal is efficient.

Because homozygous mutants are derived preferentially via mitotic recombination in *Blm*-deficiency cells, they are expected to contain two gene trap alleles. This aspect allows the potential bi-allelic mutants be identified from a pool of gene trap mutants with a quick and high throughput Southern-blot based analysis (QTSouthern). The STB screen demonstrates its usage, in which the QTSouthern identified 24 bi-allelic mutants from a total of 104 gene trap clones. 23 clones were later confirmed to be homozygous mutants. 12 single allelic STB mutants identified using QTSouthern method were confirmed to be heterozygous mutants by Southern-blot analysis using flanking genomic probes. These results suggest that the QTSouthern provides a reliable and fast pre-screening method for bi-allelic mutants (data not shown).

5.3.3 LOH efficiency on different genomic locus

Recovery of recessive mutations in *Blm*-deficient cells depends on LOH events, which occurs randomly. In general, the longer the cells were cultured, the more LOH events will occur. Therefore, the STB screens were able to identify more homozygous gene trap mutations compared with the STA screen. In addition to the stochastic nature with which homozygous mutant are generated, there is also likely to be a gradient of mitotic recombination from the centromere to the telomere. A gene located close to the telomere will have a higher rate of LOH than a gene located near the centromere, and is therefore more likely to be over represented in the screen. In fact, *Msh6* gene is located at the tip of chromosome 17, 88 Mb distant from the centromere. Consistently, multiple homozygous gene trap *Msh6* mutations were identified in the STA screen. *Dnmt1* is located about 21 Mb from the centromere on chromosome 9. The *Dnmt1* mutations were only identified in the STB and STC screens with four to five more population doublings.

5.3.4 Genomic coverage of gene trap mutagenesis

Three mouse mismatch repair genes *Msh2*, *Mlh1* and *Pms2* were not recovered in these screens despite the fact that deficiency in these genes causes 6TG-resistance. Thus the gene trap library (GT library) with 10,000 gene trap clones is incomplete in its genome coverage. Gene trap approaches have limited genome coverage, and some genomic loci appear to be preferred by a gene trap mutagen (so called gene trap “hot spot”). The recovery of gene trap mutations relies heavily on the stable expression of the gene trap reporter, which is affected by the host gene structure and gene trap vector design. To achieve better genome coverage, use of various gene trap vector is important (reviewed in Skarnes 2000; Hansen *et al*., 2003).

However, the frequent recovery of *Msh6* in this study cannot be explained as a simple preferred gene trap insertion site. Although *Msh6* is recovered at a frequency of 1 in 1400 gene-trap clone in this study, searches of available gene trap data from Lexicon and German Gene Trap Consortium (GGTC) reveals that the frequency of insertions in *Msh6* locus is less than the average gene trap hit in general, suggesting that *Msh6* does not appear to be a general insertion “hot” spot. It is not clear that whether the abundance of *Msh6* insertion derives from the use of the specific RGTV-1 retroviral vector. However, the RGTV-1 retroviral vector was derived from the commonly used MoMulV based retroviral backbone, which has also been used by both Lexicon and GGTC. Therefore, prominent insertion “hot spots” should be common in all data sets. It cannot be excluded that *Msh6* has a dominant role in MMR-mediated 6TG resistance in ES cells, which hasn’t be identified yet.