RECESSIVE GENETIC SCREEN FOR MISMATCH REPAIR COMPONENTS IN *BLM*-DEFICIENT ES CELLS

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

by *Ge Guo*

The Wellcome Trust Sanger Institute
University of Cambridge

August, 2004

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specially indicated in the text. None of the material presented herein has been submitted previously for the purpose of obtaining another degree.

Ge Guo

ACKNOWLEDGEMENT

I would like to give my deepest thanks to my mentor Dr. Allan Bradley for letting me join his lab, his guidance and his continued support during past five years. I would acknowledge the members of my thesis committee at Baylor College of Medicine, Dr. John Belmont, Dr. Craig Chinault, Dr. Hagop Youssoufian and Dr. lawrance Donehower for their guidance during my first two years Ph.D. study in Baylor. I want to express my deepest gratitude to Lillie Tanagho (Baylor) for the support and care she has freely given to me. She will stay in my memory forever as one of the kindest and warmest person I have ever met. Also my gratitude goes to both Baylor and The Wellcome Trust Sanger Institute for the educational opportunity they gave me.

I would like to thank previous Bradley lab members in Houston and members now in England, especially Guangbin Luo and Xiaozhong (Alec) Wang for the instructional discussion at the early stage of my study. I would like to thank Jos for his help in SpPCR experiment. His enthusiasms in science are very inspiring. I would like to thank Wei Wang for his collaborating in Dnmt1 project and his helps in taking care of my cells when I am away. I would express my gratitude to Pentao for reading my paper and the discussion about the future job hunting. I would also like to express my gratitude to David, Madhuri, Deb, Patrick and Antony for reading my thesis. I owe thanks to so many people--too many to mention in a short acknowledge page. Thanks for those sweet memories you have gave me.

Last but not least, I would like to thank my family, my parents, my bother and sister and my husband for their love and support.

ABSTRACT

RECESSIVE GENETIC SCREEN FOR MISMATCH REPAIR COMPONENTS IN *BLM*-DEFICIENT ES CELLS

Phenotype-driven recessive genetic screens in diploid organisms require a strategy to render the mutation homozygous. Although homozygous mutant mice can be generated by breeding, a reliable method to make homozygous mutations in cultured cells has not been available, limiting recessive screens in culture. Cultured embryonic stem (ES) cells provide access to all of the genes required to elaborate the fundamental components and physiological systems of a mammalian cell, as well as genes involved in differentiation. It has been established that in *Blm*-deficient cells, homozygous daughter cells can be readily segregated from cells carrying heterozygous mutations, presumably through mitotic recombination between non-sister chromatids. In this study, I have exploited the high rate of mitotic recombination in *Blm*-deficient ES cells to generate a genome wide library of homozygous mutant cells from heterozygous mutations induced with a revertible gene trap retrovirus. This library is composed of nearly 10,000 individual gene trap clones. To further investigate the use of this library, a recessive genetic screen has been carried out to identify cells with defects in DNA mismatch repair (MMR) that exhibit resistance to 6-thioguanine. Multiple homozygous mutants in mismatch repair homologue 6 (*Msh6*) were recovered, providing confirmation of the effectiveness of this recessive genetic screen. *Dnmt1* was recovered as a novel MMR gene from this screen. It was verified that Dnmt1-deficient ES cells exhibit micro-satellite instability. *Dnmt1* mutant mice are predisposed to certain type of cancers. The finding that *Dnmt1* is a novel MMR gene provides new insights into the mechanistic role of *Dnmt1* in cancer. Importantly, the combination of insertional mutagenesis in *Blm*-deficient ES cells opens a new approach for phenotype based recessive genetic screens in ES cells.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENT

1.1 The mouse as a genetic model

Mice are similar to humans in both anatomy and physiology. As a mammalian model system, the mouse has advantages of a small body size and short generation time. In addition to studies of basic biological processes such as DNA metabolism, mice have served a model for studying mammalians aspects of development, immunology and behavior.

Mice and humans diverged from a common ancestor about 65 millions years ago. Mice have a genome of $2.5x10^9$ bases, which is 14% smaller than the genome of humans, $2.9x\ 10^9$ bases. 99% of human genes are represented by an identifiable mouse homologue, and 80% of mouse genes have a single human orthologue. More than 90% of the mouse and human genomes can be clustered into chromosomal segments of conserved synteny, reflecting the conservation of gene organization (Waterston et al., 2002). Based on the analysis of 67,000 mouse cDNA sequences and the comparative study of the human and mouse genomes, it is predicted that the mouse and human genomes contain about 30,000 protein coding cDNA and 15,000 non-coding cDNA including alternative spliced products (Okazaki et al., 2002).

The wild spread use of the mouse for biomedical research is due to the development of many genetic and genomic tools. One of the landmarks in mouse genetics was the isolation of pluripotent mouse embryonic stem cells (ES) (Evans and Kaufman, 1981) and the demonstration that cultured ES cells can be reintroduced into host blastocysts and repopulate somatic tissues as well as germ line during embryogenesis (Bradley et al., 1984). Importantly, cultured ES cells maintain their pluripotency after modification of their genome enabling these modifications to be established in mice (Robertson et al., 1986). Further more, targeted mutations could be introduced into ES cells through homologous recombination (Capecchi, 1989). These findings initiated a new era in mouse genetics where precise loss or gain of function mutations can be established in

the mouse through *in vitro* manipulation of ES cells. These approaches, together with the traditional transgenic technique via zygote injection, are classified as reverse genetics (Landel et al., 1990). Many new genomic tools have been subsequently developed to help to decipher the functions coded in the mouse genome, for example balancer chromosomes, chromosome deletions and duplications (Yu and Bradley, 2001).

1.2 Mouse ES cells as a genetic tool

Whereas the mouse has unique advantages as a model for humans, it has limitations both technically and economically in studies that require large numbers of animals, for example, genetic screens. A saturating genome screen covering all 30,000 genes is extremely difficult to conduct *in vivo*. An alternative approach is to conduct assays and screens on cultured cells. Mouse ES cells offer unique advantages in cell-based screens. ES cells exhibit unlimited growth in culture, which allows genetic and molecular manipulation of the cells and the manifestation of the phenotypic consequences. In contrast to other cells lines that are capable of long periods of growth such as somatic cell lines established from tumors and transformed cell lines, ES cells more precisely reflect a normal biological and physiological status. It is notable that homologous recombination in ES cells is much more efficient than that in the somatic cell lines, which allows the use of gene-targeting approaches. More than 10,000 genes are expressed in ES cells (Sharov et al., 2003). These genes are required to elaborate the fundamental components required for a mammalian cell such as structural components and physiological systems for essential functions like metabolism, cell division and DNA repair (Sharov et al., 2003). Another important aspect is that ES cells can differentiate in a defined cell culture system. *In vitro* ES cell differentiation recapitulates many *in vivo* developmental processes such as adipogenesis, cardiogenesis, haematopoiesis, myogenesis, neurogenesis and chondrogenesis (Wobus, 2001). Thus, many developmentally regulated genes and processes can be studied in ES cell. In brief, cultured ES cells can provide

access to a significant fraction of the genes in mouse genome and can serve as a genetic tool for a large variety of *in vitro* studies.

1.3 Approaches to generate homozygous mutations in ES cells

1.3.1 Sequential gene-targeting

Loss of function mutations are generated most frequently through gene-targeting in ES cells. Gene-targeting is achieved by homologous recombination, a process in which a DNA sequence recombines with its endogenous homologous genomic locus. For gene-targeting, the vectors are built to contain a drug resistance marker along with the homologous arms so that ES cells with an integrated gene-targeting vector can be selected by drug resistance. To generate ES cells with homozygous autosomal mutations, two alleles can be disrupted sequentially by gene-targeting (te Riele et al., 1990). Abuin et al (1996) reported the construction of an ES cell line carrying homozygous mutations at two different genes by sequential gene-targeting. He used a marker-recycling method, in which the neomycin drug selection marker (*Neo*) is flanked by two *loxP* sites, so that it can be removed by Cre-mediated recombination. Then, the same gene-targeting vector could be used to target the second allele (Fig. 1-1 a). Although gene-targeting allows the generation of defined mutations precisely at any gene locus, the sequential gene-targeting method is relatively time-consuming and can only be applied on a gene-by-gene basis.

1.3.2 High concentration G418 selection

It has been known for several decades that homozygous mutated cells and/or wild type cells can be generated spontaneously from cultured mammalian somatic cells containing a heterozygous mutation, a phenomena known as loss of heterozygosity (LOH). LOH can occur by many mechanisms including regional or whole chromosome loss, mitotic recombination and gene inactivation.

Evidence suggested that the mechanism causing LOH varies in different cell lines. For example, in CHO (Chinese hamster ovary) cells, LOH occurs most commonly through loss of a whole chromosome followed by chromosomal duplication (Campbell et al., 1981, Wasmuth and Vock Hall, 1984). In murine lymphoid cell lines, mitotic recombination between homologous non-sister chromatids is believed to be the major cause of LOH (Nelson et al., 1989, Potter et al., 1987, Rajan et al., 1983). Mortensen et al (1992) explored the potential of producing homozygous gene-targeted mouse ES cells from heterozygous gene-targeted ES cells via LOH. He created four different heterozygous gene-targeted ES cell lines carrying the *Neo* cassette. By culturing these heterozygous gene-targeted ES cells in high concentrations of G418, many homozygous gene-targeted cells were recovered containing two copies of the *neo* cassette. The existence of two copies of *Neo* cassette suggests that LOH has occurred either via mitotic recombination between homologous non-sister chromatids or by chromosomal loss followed by chromosomal duplication. The LOH rate in these studies was estimated to be about $1x10^{-5}$ per locus/ cell/ generation. Compared to the sequential targeting method, this high drug concentration selection approach doesn't require two or more cycles of gene-targeting, therefore, providing an easy method for obtaining homozygous mutations in ES cells (Dufort et al., 1998, Carmeliet et al., 1996, Reaume et al., 1995). Lefebvre et al (2001) investigated the mechanism of LOH in ES cells by gene-targeting the *Neo* cassette into a hybrid ES cell line (R1 cell line) that was established from F1 hybrid embryos obtained from a cross between mice of two different inbred 129 substrains. Use of a hybrid cell line allows tracking the origin of two homologous chromosomes by analyzing polymorphic DNA markers. In this study, they targeted the *Neo* cassette into six different genomic loci on four different chromosomes, Chr 2, Chr 5, Chr10 and Chr17, and showed that all of the homozygous gene-targeted clones recovered by high concentration G418 selection had lost the heterozygosity of distant linked DNA markers, which is consistent with a mechanism of chromosomal loss and duplication. Recovery of homozygous mutants from various chromosomes implied that homozygous

mutations could be generated on a genome-wide basis through high concentrations of G418 selection.

1.3.3 Induced mitotic recombination

Mitotic recombination has been used extensively in *Drosophila* to generate "genetic mosaics", a term for an individual with cells with more than one genotype (Perrimon, 1998). Mitotic recombination is also known as somatic recombination, during which chromosomal crossover occurs between two homologous non-sister chromatids during mitosis. A crossover in the G2 phase of the cell cycle between two homologous non-sister chromatids can be segregated either in a way that the recombinant chromatids segregate to opposite poles, so they separate to two daughter cells (X-segregation) or in a way that the recombinant chromatids segregate to the same pole in the same daughter cell (Z-segregation). Genetic mosaic occurs as a result of G2 crossover followed by X-segregation, in which single allelic genetic variation is localized to two sister-chromatids in the G2 phase and segregated into the same daughter cell (Stern, 1936).

In *Drosophila* the mitotic recombination system has been combined with site-specific recombination systems, such as the FLP/FRT system. The mitotic recombination in *Drosophila* can be induced by FLP-mediated recombination between two FRT sites that have been inserted into the same genomic locus on homologous chromosomes. By controlling the expression of FLP enzyme, mitotic recombination can be induced with spatial and temporal specificity. Notably, it has been shown that after FLP-mediated homologous recombination in the G2 phase of the cell cycle, recombinant chromatids are directed to consistently segregate opposite to each other (X-segregation) (Beumer et al., 1998). Recently, Liu et al (2002) adopted the concept of mitotic recombination and demonstrated that mitotic recombination could be induced in mouse ES cells via Cre-mediated recombination between targeted *loxP* sites. The mitotic

recombination frequency varies between different genomic loci and chromosomes with a range from 10^{-5} to 10^{-2} after transient Cre expression. In Liu's study, X-segregation was also observed as the major event followed by G2 crossover compared to Z-segregation, which is consistent with the results shown by FLP/FRT induced mitotic recombination in *Drosophila*. This pioneering work signals that induced mitotic recombination will be a powerful tool in mouse genetics for generating homozygous mutations in ES cells and for mosaic analysis in mice.

1.3.4 Elevated mitotic recombination in *BLM*-deficient cells

Mitotic recombination can occur spontaneously, leading to the LOH in some cell lines, for example, in cultured murine lymphoid cell lines (Nelson et al., 1989). However, the spontaneous LOH rate is very low in normal cells and cannot be used as an efficient tool for generating homozygous mutations. Recently, it has been shown that mitotic recombination rate is increased in human and mouse cells that lack the function of a DNA helicase, BLM. This opens up the possibility of generating homozygous mutations based on the enhanced LOH rate in *BLM*-deficient cells (Fig. 1-1 d). *BLM*-deficient cells constitute the foundation of my study. I used *BLM*-deficient cells as a tool to generate homozygous recessive mutations in order to conduct a genetic screen. The current knowledge about *BLM* gene and its' functions are presented in the following sections.

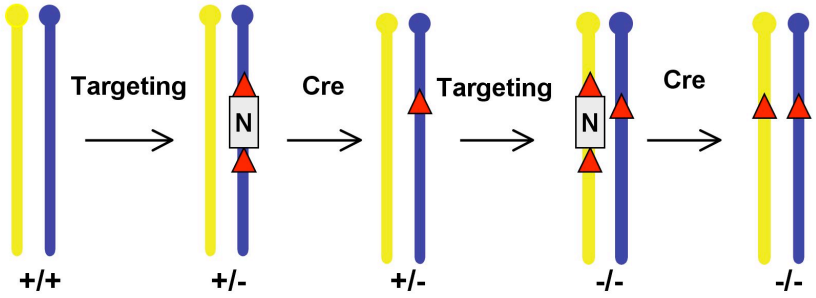


Figure 1-1 a. Sequential gene-targeting, showing the targeting of a genomic locus with *loxP* flanked *Neo* selection cassette. Cre-mediated recombination enables the construction of a marker free homozygous mutated cell. The red triangle represents the *loxP* site; "N" represents the *Neo* cassette.

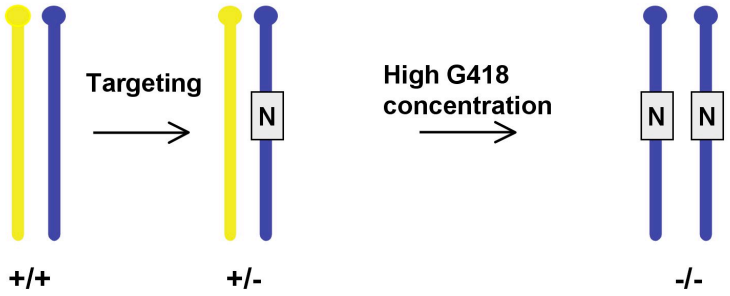


Figure 1-1 b. High concentration G418 selection, showing that homozgous mutations can be obtained by LOH involving loss and gain of whole chromosomes. "N" represents the *Neo* cassette.

Figure 1-1. Approaches to generate homozygous mutations in ES cells

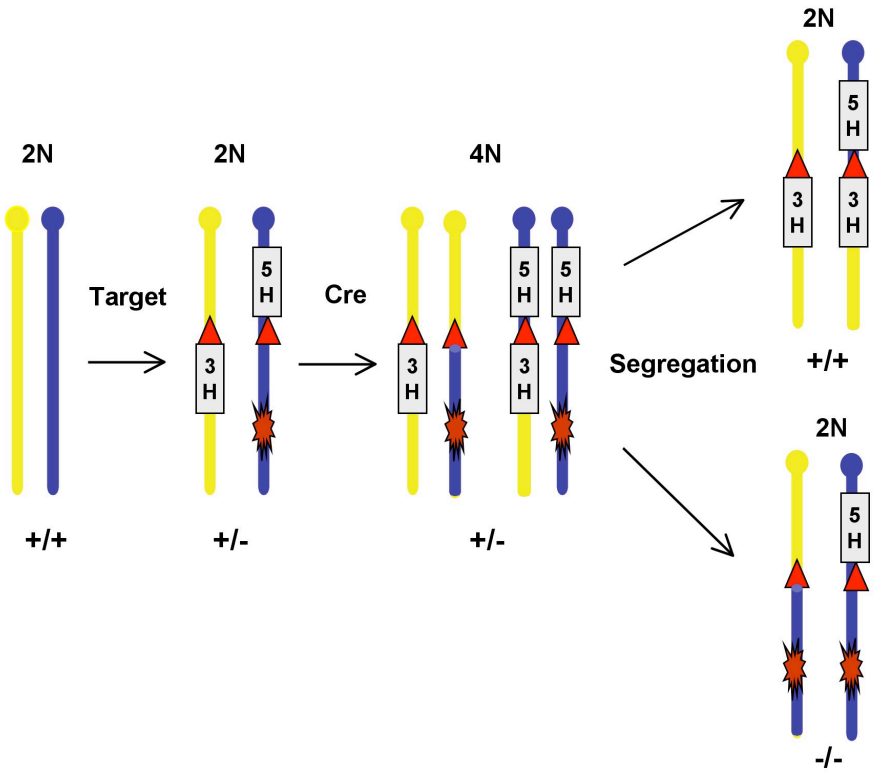


Figure 1-1 c. Induced mitotic recombination between two *loxP* sites that are targeted into two homologous chromosomes. The Cre-*loxP* mediated recombination occurs predominantly at the G2 phase of a dividing cell. "5H" represents the 5' *Hprt* cassette; 3'H represents the 3' *Hprt* cassette. The Red star represents a mutation.

Figure 1-1. Approaches to generate homozygous mutations in ES cells

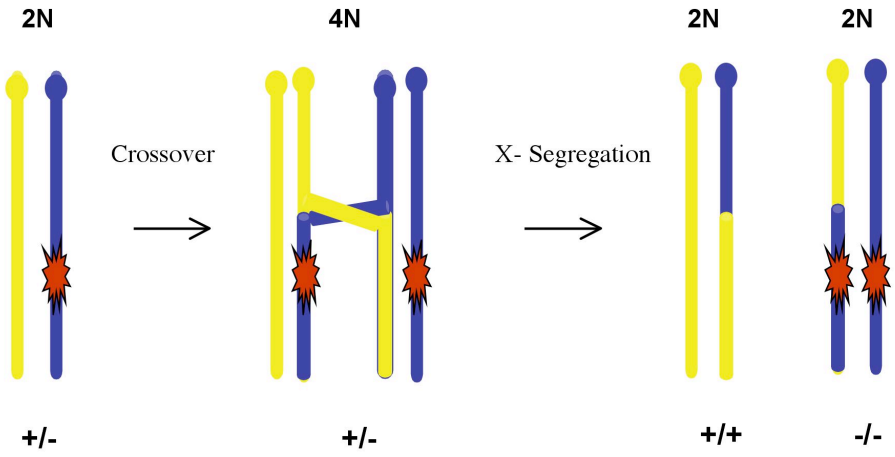


Fig. 1-1 d. *Blm*-deficient cells.

The deficiency in the *Blm* gene leads to increased LOH, which occurs predominantly via mitotic recombination. G2 cross-over between non-sister homologous chromatids followed by X-segregation results in homozygous mutant or wild-type cells.

Figure1-1. Approaches to generate homozygous mutations in ES cells

1.3.4.1 Bloom's syndrome

BLM is the gene responsible for the human disease Bloom's syndrome (BS). BS is a rare autosomal recessive transmitted disorder. German (1993) reviewed the major clinical features of Bloom's syndrome, which includes proportional dwarfism, sun-sensitive erythematous lesions of facial skin, immunodeficiency, a broad spectrum of neoplasm of multiple tissue types with early incidence and reduced fertility. Cells from BS patients exhibited excessive genomic instability including an increased spontaneous mutation rate at specific genomic loci and large microscopically visible genomic rearrangements such as chromosomal gaps, breaks and translocations. The hallmark feature of BS cells is the enormously elevated rate of sister-chromatid exchange (SCE), which is illustrated in bromodeoxyuridine (Brdu)-labeled BS cells. The increase in SCE has been shown to represent the loss of suppression of homologous recombination events in BS cells (Groden et al., 1990, Sonoda et al., 1999).

1.3.4.2 *BLM*, Bloom's syndrome gene

BLM was mapped by genetic linkage analysis to a position about 1 cM away from the gene, *FES,* on human chromsome 15 (German et al., 1994). Ellis et al (1995 a) localized and cloned the *BLM* gene using an unusual method, somatic crossover point (SCP) mapping method. The SCP mapping method was based on the observation that in some BS patients a small population of low SCE lymphocytes exists in the blood in spite of the fact that somatic cells from BS patients exhibited the characteristic high SCE rate. By examination of polymorphic DNA markers around the *BLM* locus, it was revealed that low SCE lymphocytes arose from somatic recombination within the *BLM* locus in individuals who had inherited *BLM* alleles mutated at different sites. Somatic recombination in such compound heterozygotes may reconstitute a functional *BLM* gene. Lymphocytes derived from stem cells which have undergone this

recombination event will then show a low SCE rate (Ellis et al., 1995). Therefore, the position of *BLM* gene can be precisely mapped by identification of the somatic crossover site in low SCE lymphocyte cell lines through DNA polymorphic marker analysis.

BLM encodes a member of the RecQ family of DExH box DNA helicases, comprising 1417 amino acids. The central region of BLM, the helicase domain, contains seven conserved motifs (Ellis et al., 1995). RecQ helicases have DNA-dependent ATPase activity and ATP-dependent DNA helicase activity with a 3' to 5' polarity (Wu et al., 2001). In general, DNA helicases are required to alter DNA topology in processes, in which a single strand DNA needs to be generated such as for DNA replication, repair and transcription. Besides mammalian BLM helicase, other members of RecQ famlily have been identified including the RecQ helicase in *E.coli*, the *BLM* homologue in yeast, fruit fly, worms and mammalian *RECQL1*, *WRN*, *RECQL4* and *RECQL5*. It is notable that *E.coli* and the unicellular eukaryotes, *Saccharomyces cerevisiae*, contain a single member of RecQ helicase, but in humans and mice multiple RecQ homologues exist. The mutations of *WRN* and *RECQL4* in humans lead to Werner's and Rothmund-Thomson syndromes respectively. In addition to the conserved helicase domain, BLM and other mammalian RecQ helicases contain extended N-terminal and C-terminal regions that vary greatly in length with a low degree of sequence conservation (Nakayama, 2002). This sequence variation in mammalian RecQ helicase implies the functional specificity of each protein.

1.3.4.3 Enzyme activity, an untypical DNA helicase

The studies of enzymatic activity of BLM helicase and its homologue in bacteria and yeast reveal that BLM can unwind a variety of DNA constructs *in vitro*. In contrast to a typical DNA helicase that binds and unwind a standard B-form DNA duplex, BLM protein does not bind or unwind blunt-ended, fully duplex DNA molecules. It prefers some atypical substrates. One of the preferred substrates of

BLM *in vitro* is a synthetic X-structured DNA molecule, which is used as a model for Holliday junction formed during homologous recombination. This finding suggests a role of BLM in homologous recombination (Karow et al., 2000). Moreover, BLM, WRN and the yeast RecQ helicase homologue, SGS1 can efficiently unwind an unusual G-G paired G-quadruplex structure *in vitro*. A G-quadruplex structure forms *in vivo* within G-rich DNA sequences, for example in G-rich telomeres and rDNA. This activity suggests a potential role of RecQ helicase in DNA replication during which the G-quadruplex formed in some specific G-rich sequence needs to be disrupted to allow the progression of DNA replication forks.

1.3.4.4 BLM in DNA replication

Evidence from biochemical and genetic studies in both prokaryotes and eukaryotes suggest that BLM is a multifunctional protein, which has a major role in DNA replication. For example, BS cells exhibit a protracted S phase and accumulate DNA replication intermediates of abnormal sizes (Lonn et al., 1990). *S.cerevisiae* or *S. pombe* RecQ helicase mutants fail to execute normal cell cycle progression following recovery from a S-phase cell cycle arrest caused by DNA replication inhibitor HU (Hydroxyurea), and these mutant strains are hypersensitive to HU (Frei and Gasser, 2000, Stewart et al., 1997). Consistent with this result, it has been shown that BS cells are hypersensitive to HU. BLM has also been shown to be a substrate of the protein kinase, ATR (ataxia telangiectasia and rad3+ related), which is activated in response to a DNA replication block. Blocking of BLM phosphorylation results in the failure of recovery from HU-induced replication inhibition (Davies et al., 2004). A role of BLM in DNA replication is also supported by the studies of the expression pattern of BLM and the cellular localization of BLM protein. BLM localizes to promyelocytic leukemia protein (PML) nuclear bodies and accumulates during the late S and G2 phase during the cell division cycle (Dutertre et al., 2000, Bischof et al., 2001). Following the inhibition of DNA replication by HU, BLM is

found to be localized to the DNA replication foci in S-phase (Sengupta et al., 2003). These results suggest that BLM is translocated to arrested replication forks to assist the progression of DNA replication.

The exact role of BLM in DNA replication is unclear. BLM may act in two non-exclusive processes. One possible role of BLM is to function as a "roadblock remover", in which BLM removes unusual DNA secondary structures, such as the G-Quadruplex or other obstacles, to prevent potential replication stalling or the collapse of the replication fork. This view is supported by the observation that BLM protein can promote branch migration of a Holiday junction (Karow et al., 2000, Hickson, 2003). Another possible role is that BLM is involved in re-starting DNA replication after the demise of a replication fork, a process involving homologous recombination-mediated double strand break (DSB) repair. It has been shown that BLM interacts with proteins required for DNA replication, for example, RPA (replication protein A), the major single strand DNA binding protein which is required in DNA replication and recombination (Brosh et al., 2000). BLM forms a complex with RAD51. RAD51 catalyzes DNA strand invasion and exchange in homologous recombination (Wu et al., 2001).

1.3.4.5 A model for sister-chromatid exchange caused by *BLM*-deficiency

The characteristic phenotypic consequence of BLM mutation is excessive somatic recombination and SCE. It has been shown that the occurrence of SCE requires the homologous recombination system. Cells deficient in homologous recombination protein RAD51 and RAD54 exhibited a significantly reduced level of SCE (Sonoda et al., 1999). A model involving DNA replication and homologous recombination has been proposed to explain why *BLM*-deficiency during DNA replication will lead to SCE (Nakayama, 2002). In brief, DNA strand breaks or gaps may exist under various physiological conditions. When the replication fork encounters a single-strand nick or a gap on the template strand, it will collapse and a double strand break is created. Then, a repair process is

initiated, leading to the formation of a Holiday junction through homology-directed strand invasion and exchange, which is a process mediated by the homologous recombination machinery. The Holliday junction can be simply unwound by a RecQ helicase (BLM in mammals) to re-establish the replication fork (Karow et al., 2000). In this case, the repair process is error-free and no SCE is generated. In the absence of the RecQ helicase, the Holliday junction may be resolved by recombination pathways that cause chromatid exchanges.

1.3.4.6 Proteins interacting with BLM

The proteins interacting with BLM may provide additional clues to the functions of BLM. BLM has been shown to direct interact or co-localize with many proteins. Topoisomerase IIIα is one of the BLM interacting proteins. This interaction is a direct one and exists in both prokaryotes and eukaryotes. The function of this interaction is not clear yet. It has been suggested that topoisomerase IIIα is required for resolving abnormal recombination intermediates (Wu et al., 1999). BLM also associates with RAD51 and RPA, which is consistent with the role of BLM in DNA replication (discussed above).

Recently, BLM has been found to be a component of a large protein complex including BRCA1, which is referred to as BASC (BRCA1-associated genome surveillance complex) (Wang et al., 2000). This complex includes many proteins involved in DNA repair or DNA damage response such as MSH2, MSH6, MLH1, ATM, BLM, the RAD50-MRE11-NBS complex and DNA replication factor C (RFC), a protein complex that facilitates the loading of PCNA (proliferating cell nuclear antigen) onto DNA. MSH2, MSH6 and MLH1 are major components of the DNA mismatch repair (MMR) system. MMR system removes mismatched nucleotides generated during DNA duplication. MMR also plays a role in the genome surveillance process, in which certain types of DNA lesion are recognized and signal cell death or cell cycle arrest. ATM (Ataxia-telangiectasia mutated) is a serine/threonine protein kinase that plays a central role in sensing

and transducing cellular signals in response to DNA damage. The RAD50-MRE11-NBS (RMN) complex and BRCA1 are critical in repairing DNA double strand breaks (DSB) via homologous recombination. The precise roles of BASC complex haven't been established yet. The existence of multiple proteins involved in DNA damage repair and signaling processes suggest that BASC complex plays a role in the DNA damage response. Current studies suggest that the function of BRCA1 in DSB repair doesn't require BLM. BS cells don't exhibit obvious sensitivity to γ-irradiation that induces DSB, while the deficiency of BRCA1 leads to a hypersensitive γ-irradiation response (Franchitto and Pichierri, 2002).

p53 has also been reported to be a binding partner of BLM and another mammalian RecQ helicase WRN (Blander et al., 1999, Spillare et al., 1999, Wang et al., 2001). p53 is a transcription factor that plays a central role in cell cycle arrest and apoptosis (Oren and Rotter, 1999). However, the interaction between BLM and p53 is not required in either cell cycle arrest or apoptosis, the majority of BS cells appear to have normal p53 accumulation and undergo cell cycle arrest and apoptosis in response to certain type of DNA damage (Lu and Lane, 1993, Ababou et al., 2002). In contrast, p53 may have a role in repairing stalled replication forks, a process involving BLM. This view is based on a recent finding that p53 modulates the frequency of homologous recombination and SCE in *BLM*-deficient cells (Sengupta et al., 2003). In this study, it was reported that p53, BLM and RAD51 co-localized to sites of stalled DNA replication forks in response to DNA replication inhibition induced by HU treatment. Loss of p53 function enhanced synergistically the homologous recombination and SCE frequency in *BLM*-deficient cells derived from Bloom's syndrome patients. Consistent with this observation, it has been reported that p53 can bind to Holiday junctions and facilitates their resolution (Lee et al., 1997). p53 can modulate the procession of Holiday junctions by BLM *in vitro* (Yang et al., 2002). Mutation in p53 also results in elevated homologous recombination (Susse et al., 2000, Slebos and Taylor, 2001, Saintigny and Lopez, 2002), and this activity of

p53 appears to be independent of its transcriptional activation function (Willers et al., 2000).

1.3.4.7 Mouse models of Bloom's syndrome

The BLM homologue in mice is located on chromosome 7. The gene is approximately 88 kb in length and consists of 23 exons. The first exon is non-coding 5'UTR and is represented differently in testis and somatic cells as the result of alternative splicing (McDaniel et al., 2003). Six different *Blm* knockout alleles have been described, including $Blm^{tm1ches}$, $Blm^{tm1/Brd}$, $Blm^{tm2/Brd}$, $Blm^{tm3/Brd}$, $Blm^{tm1/Gos}$ *and* $Blm^{tm3/ches}$ (Chester et al., 1998, Luo et al., 2000, Goss et al., 2002, McDaniel et al., 2003). Four *Blm* alleles, $Blm^{tm1ches}$, $Blm^{tm1/Brd}$, $Blm^{tm1/Gos}$, $Blm^{tm3/ches}$, were generated by gene-targeting with replacement targeting vectors, resulting in deletion of coding exons. These alleles have been shown to be null by Western-blot analysis of Blm protein expression and homozygous knockout mice with these alleles appear to be embryonic lethal. *Blm*-deficiency doesn't have overt effect on the growth and survival of ES cells. Two alleles, $Blm^{tm2/Brd}$ and $Blm^{tm3/Brd}$, were generated by insertional gene-targeting events, which results in the duplication of exon3. This duplication caused a frame-shift mutation. The $Blm^{tm2/Brd}$ allele is homozygous lethal while the derived $Blm^{tm3/Brd}$ is viable. The homozygous mice ($Blm^{tm2/Brd}$, $Blm^{tm3/Brd}$) exhibited genomic instability and tumor susceptibility, a phenotype mimicking the human Bloom's syndrome. Thus $Blm^{tm3/Brd}$ mice serve as a better animal model for human Bloom's syndrome. Moreover, a significant increase in the SCE was observed in $Blm^{tm1/Brd}$ / $Blm^{tm3/Brd}$ ES cells (Luo et al., 2000). Recently, it was shown that $Blm^{tm3/Brd}$ could rescue the embryonic lethality of $Blm^{tm3/ches}$ alleles. *Blm*-deficient cells carrying Blm^{tm3Brd} and $Blm^{tm3/ches}$ alleles have been reported to have a SCE rate about two fold lower than the cells with two $Blm^{tm3/ches}$ alleles, which the authors suggest reflecting the hypomorphic activity of the Blm^{tm3Brd} allele (McDaniel et al., 2003).

1.3.4.8 Elevated LOH rate in *Blm*-deficient mouse ES cells

The direct phenotypic consequence of increased somatic recombination is loss of heterozygosity (LOH) of single allelic mutations. Luo et al (2000) determined the LOH rate in $Blm^{tm1/Brd}$ / $Blm^{tm3/Brd}$ ES cells by measuring the loss of a single copy *Hprt* minigene that was gene-targeted into an autosomal genomic locus. Cells that have lost the *Hprt* minigene become resistant to the drug, 6-thioguanine. By Luria-Delbruck fluctuation analysis the rate of LOH was determined to be 2.3X 10^{-5} (locus/cell/generation) in wild type ES cells and 4.2X10^{-4} (locus/cell/generation) in *Blm*-deficient ES cells, respectively (Luo et al., 2000). Although the *Hprt* gene can be lost by several mechanisms, for example, loss of whole chromosome, spontaneous mutation and deletion, mitotic recombination between homologous chromosomes is believed to be the major cause of LOH in *Blm*-deficient cells (Sonoda et al., 1999).

In summary, the biochemical and genetic studies point out that BLM plays a critical role in repairing DNA replication fork abnormalities. BLM facilitates the smooth progressing of DNA replication by preventing stall of the replication forks or facilitating stalled DNA replication forks to restart in an error-free way. The increased rate of SCE and mitotic recombination exhibited in *BLM*-deficient cells is the result of the switching from BLM-dependent error-free repair to BLM-independent error-prone repair of the stalled replication forks. BLM is not required for cell growth or survival in culture. *Blm*-deficiency in mouse ES cells caused a 20-fold increase in the rate of LOH, which provides the basis for deriving homozygous autosomal mutations from a single allele mutation.

1.4 Recessive genetic screens in mammalian cells

Recessive genetic screens in a diploid genome require a strategy to generate homozygous mutations. In the early 70's, it has been shown that recessive mutations could be recovered from cultured mammalian cells that are partially

hemizygous (Siminovitch, 1976). The most frequently used cell line for deriving recessive mutations were CHO (Chinese hamster ovary) cells. Functional hemizygosity could be caused by several possible mechanisms, such as gene inactivation and genomic rearrangements, which results in loss of function of one copy of certain autosomal genes, therefore rendering phenotypic hemizygosity to these cells. Since genomic rearrangement and gene inactivation occur randomly, each CHO cell line may have accumulated mutations in different sets of genes (Deaven and Petersen, 1973, Worton, 1978, Gupta et al., 1978). It has been shown that multiple recessive mutations from CHO-CHO hybrids segregated randomly, suggesting that the functional hemizygosity in CHO cells is not only restricted to one or a few chromosomal regions but appears to be wildly distributed (Gupta, 1980). In the early 80's, CHO cells were used to isolate recessive mutants that are sensitive to killing by ultraviolet radiation (UV) (Thompson et al., 1980, Busch et al., 1980). 44 UV sensitive mutant clones were classified into 4 different complementation groups (Thompson et al., 1981). The genes mutated in the second complementation group were determined by a genetic rescue method using cloned human genomic DNA, which led to the identification of the important nucleotide excision repair gene, *ERCC1* (excision repair complementing defective repair in Chinese hamster cells) (Westerveld et al., 1984). These data demonstrated the application of genetic aneuploidy in a recessive genetic screen. However, this strategy is greatly restricted by the fact that CHO cell lines contain partial functional hemizygous genomes. Therefore, recessive mutations located in the functional diploid regions can't be recovered. In this regard, *Blm*-deficiency has an overt advantage. Since homozygous mutations in *Blm*-deficient cells are generated preferentially by mitotic recombination, in principle all genes on an autosomal chromosome can be accessed. Therefore, *Blm*-deficient cells will allow broader genome coverage in a genetic screen than CHO cells.

1.5 DNA mismatch repair

The proper functions of DNA metabolism processes including DNA replication and DNA repair are crucial for the integrity of the genetic materials. The integrity of DNA is constantly challenged by many environmental or physiological factors. Accordingly, many proteins and DNA repair systems have been identified acting coordinately to prevent and eliminate the errors in DNA. The DNA mismatch repair (MMR) system plays a critical role in guarding genome integrity in virtually all organisms from bacteria to human. The primary function of MMR system is to recognize and correct base-base mismatches and small insertion and deletion (I/D) loops that arise during DNA replication (replicative mismatch repair). Defects in the MMR system will result in an elevated spontaneous mutation rate, a mutator phenotype and expansion or deletion of simple repeat sequences, known as microsatellite instability (MSI). The importance of MMR in guarding genome stability has been highlighted by the association of defects in MMR with cancer. Besides its function in repairing DNA replication errors, the MMR system has been linked to general DNA recombination processes including meiosis and homologous recombination. Moreover, a role of MMR in processing chemically damaged DNA, also known as DNA damage surveillance has been documented. The MMR system has been extensively reviewed by others (Modrich, 1991, Modrich and Lahue, 1996, Buermeyer et al., 1999, Hsieh, 2001). In this section I would like to provide an overview of the basics of the MMR system and emphasize the functions of eukaryotic MMR.

1.5.1 DNA mismatch repair in bacteria

The first studies of the MMR system started more than three decades ago in bacteria when genetic screens were conducted to isolate mutants with elevated spontaneous mutation rates. This research resulted in the identification of four central components of MMR system, MutS, MutL, MutH, and MutU. MMR in bacteria has been most thoroughly investigated and serves as the model for other organisms (Modrich, 1991). For the simplicity, it is separated into three major steps (Fig. 1-2): Step 1: Mismatch recognition, in which MutS proteins form

a homodimer complex which binds to the mismatched nucleotides. Step 2: Strand discrimination and excision. In an ATP dependent manner, the MutS homodimer complexes with dimerized MutL protein and stimulates the endonuclease activity of MutH. Consequently, a nick is generated in the newly synthesized DNA strand by the activated MutH using the semi-methylated GATC as the strand discrimination signal. Then, MutU, a DNA helicase, is loaded to the MutH induced nick to unwind the duplex DNA molecule. Then, with the help of exonucleases, the newly synthesized DNA strand containing the mismatched nucleotide is removed to leave a single strand DNA gap. Step 3: Resynthesis and ligation. Single-strand DNA binding protein (SSB), DNA polymerase III and DNA ligase are required for the resynthesis and ligation, which fills in the gap created by strand excision.

1.5.2 DNA mismatch repair in eukaryotes

The MMR system has been highly conserved throughout evolution. Compared to the MMR system in bacteria, MMR systems in higher eukaryotes have evolved more specificity and functions, reflected by the existence of multiple MutS and MuL homologues in yeast and mammals (Table 1-1). In yeast and mammals, mismatch recognition is conducted by three MutS homologues, MSH2, MSH3 and MSH6. MSH2 can complex with either MSH6 or MSH3, forming two protein heterodimers, MutSα (MSH2/MSH6) or MutSβ (MSH2/MSH3). MutSα and MutSβ exhibit different binding preferences for DNA substrates. MutSα predominantly binds single base mismatches and single insertion/deletion loops while MutSβ binds single and larger insertion and deletion loops (Fig. 1-3 a) (Acharya et al., 1996, Drummond et al., 1995). The function of MutSα and MutSβ overlaps with respect to the recognition of small insertion/deletion mismatches. Consistent with the role of MSH2 in both complexes, MSH2 mutations cause the highest level of mutator and MSI phenotypes, while MSH6 and MSH3 mutants exhibit mild or modest ones (Fig. 1-3 b). The functional homologue of bacterial MutL in yeast and humans is MLH1 (MutL homologue). In mammals, MLH1 forms a heterodimer protein complex with

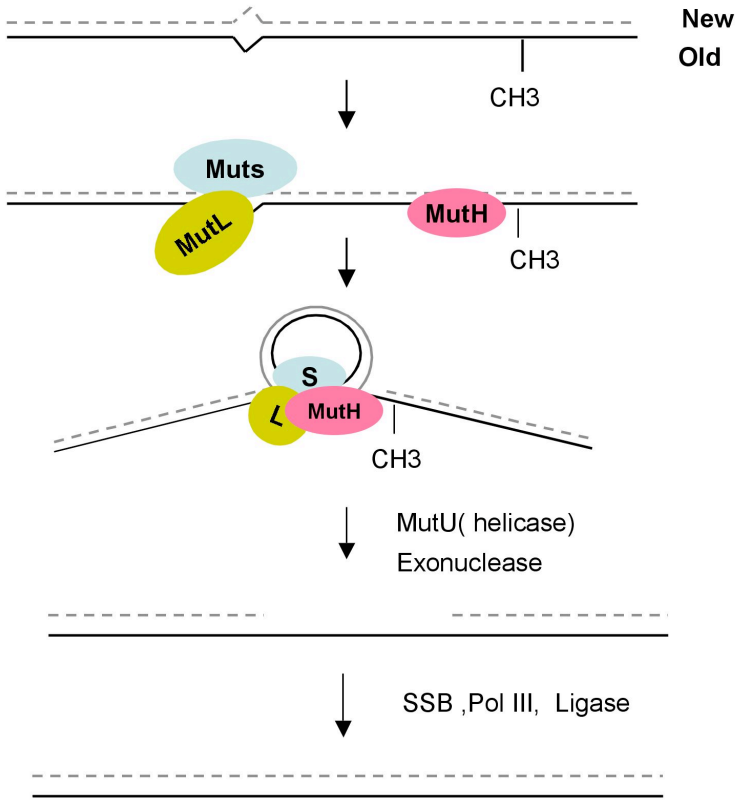


Figure 1-2. DNA mismatch repair in *E.coli*

DNA Mismatch repair (MMR) initiates when a homodimer of MutS protein recognizes mismatched nucleotides. Then, the MutS homodimer complexes with a homodimer of MutL and stimulates the endonuclease activity of MutH. A nick is made in the newly synthesized DNA strand by activated MutH using the semi-methylated GATC as the strand discrimination signal. Then, the MutU DNA helicase is loaded to the MutH induced nick to unwind the duplex DNA molecule. With the help of exonucleases, the newly synthesized DNA strand with the mismatched nucleotides is removed to leave a single strand DNA gap. Finally, the single strand gap is filled in with the help of single-strand DNA binding protein (SSB), DNA polymerase III and DNA ligase.

E.coli	*S.cerevisiae*	Human/mouse
MutS	*MSH1*	-
	MSH2	*MSH2*
	MSH3	*MSH3*
	MSH6	*MSH6*
	MSH4	*MSH4*
	MSH5	*MSH5*
MutL	*MLH1*	*MLH1*
	PMS1	*PMS2*
	MLH2	*PMS1*
	MLH3	*MLH3*
MutH	-	-
MutU(UvrD)	-	-

Table 1-1. Homologues of bacterial MMR genes in *S.cerevisiae* and mammals.

Multiple *MutS and MutL* homologs exist in eukaryotes.

Mammalian *MSH2, MSH3, MSH6, MLH1, PMS2* (yeast *PMS1*), *PMS1*(yeast *MLH2*) are involved in replication repair.

MSH4, MSH5, MLH1 and *MLH3* function in meiotic processes.

MSH2 has a role in homologous recombination.

MSH2, MSH3, MSH6, MLH1,PMS2 are invovled in DNA damage surveillance. *MSH4* and *MSH5* also exhibit a minor role.

MSH1 in *S. cerevisiae* is required for normal mitochondria function in yeast. A mammalian homologue of *MSH1* has not been identified.

Mut H and *MutU* homologues have not been identified in eukaryotes.

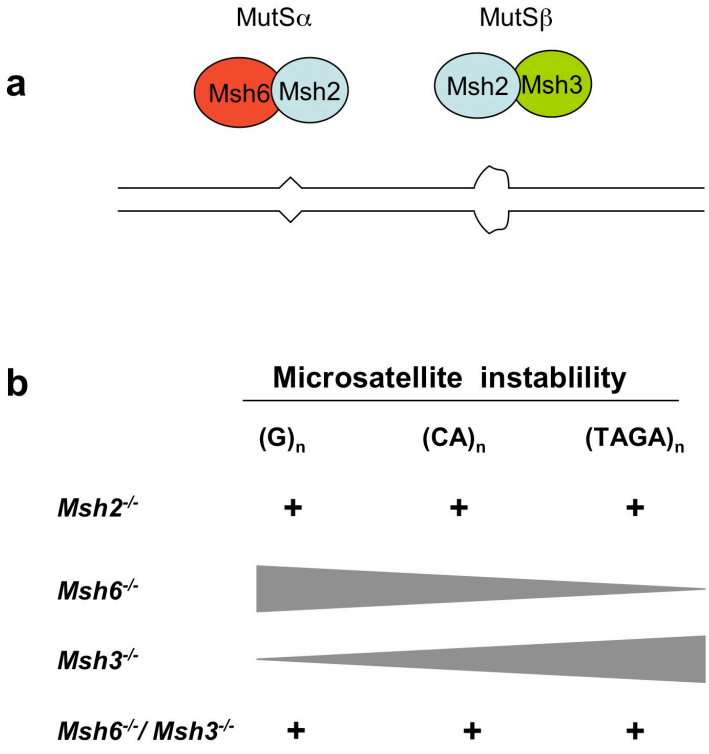


Figure 1-3. DNA mismatch recognition in Eukaryotes

a. Three MutS homologues, Msh2, Msh3 and Msh6 are involved in mismatch recognition in eukaryotes. They form two protein complexes, MutSα (Msh2/Msh6) and MutSβ (Msh2/Msh3). MutSα and MutSβ exhibit different mismatch binding abilities. MutSα predominantly binds single base mismatches and single insertion/deletion loops while MutSβ binds single and larger insertion and deletion loops. **b**. Defects in different MMR recognition proteins results in a different MSI phenotype. *Msh2* deficiency and *Msh3/Msh6* double mutants cause MSI in single nucleotide repeats ($(G)_n$), dinucleotide repeats ($(CA)_n$) and tetranucleotide repeats ($(TAGA)_n$), whereas *Msh6* or *Msh3* single mutants exhibit mild or modest MSI that varies depending on the repeat unit length. The gradient triangles represent the variation of the MSI activity.

PMS2 (post-meiotic segregation 2), which is designated as MutLα. MutLα complex binds to either MutSα or MutSβ protein complex (Li and Modrich, 1995). It has been shown that MLH1 plays a central role in MMR similar to MSH2. Mutations of either MSH2 or MLH1 fully abolish mismatch repair (Prolla et al., 1998). Besides mammalian PMS2, MLH1 can also complex with PMS1 and MLH3. These protein complexes appear to have a minor role in DNA mismatch repair. Mutations of PMS1or MLH3 result in less severe mutator or MSI phenotypes compared with MLH1 (Papadopoulos et al., 1994, Lipkin et al., 2000). In yeast, the functional homologue of MutLα is composed of Mlh1 and Pms1 (Prolla et al., 1994). Yeast Mlh1 can also complex with Mlh2, Mlh3. Both complexes have been shown to have a role in inhibition of mutation of simple sequence repeats (Harfe and Jinks-Robertson, 2000, Flores-Rozas and Kolodner, 1998).

Similar to the bacteria MMR system, mismatch repair in eukaryotes is also directed to the newly synthesized DNA strand. However, a functional MutH homologue has not been identified in either yeast or mammals. Instead, strand discontinuity has been hypothesized to serve as the strand discrimination signal in eukaryotes. Nicks or gaps that exist between neighboring Okazaki fragments in the nascent DNA strand during DNA replication could direct MMR to the newly synthesized strand. This hypothesis is supported by an *in vitro* mismatch repair experiment in which MMR process was directed by nicks situated in the proximity of mismatched nucleotides and this substrate can be efficiently corrected in a directional manner in extracts of *E.coli* mutH mutant (Modrich, 1997). Another hypothesis is that DNA mismatch repair is directly coupled to the DNA replication fork by DNA polymerase associated factor, PCNA (proliferating cell nuclear antigen), therefore alleviating the necessity of a MutH homologue in eukaryotes. It has been showed that PCNA associates with eukaryotic MutS and MutL protein complexes and mutation of PCNA causes a mutator and MSI phenotype in yeast (Gu et al., 1998, Johnson et al., 1996, Umar et al., 1996, Kokoska et al., 1999).

Several DNA nucleases are implicated in eukaryotic MMR, which may act in excision of the mismatched nucleotides. The major player is exonuclease 1 (EXO1). EXO1 protein has 5' to 3' exonuclease activity and it interacts with MSH2 in yeast and mammals (Tishkoff et al., 1997a). Mutation of *Exo*1 causes an increased spontaneous mutation rate in yeast (Amin et al., 2001). *Exo1* knockout mice were created recently, and *Exo1* null cells have a MMR deficiency with an increased spontaneous rate and microsatellite instability (Wei et al., 2003). Recently it was shown in an *in vitro* MMR reconstitution experiment that hRPA (human replication protein A), a single strand DNA binding protein, plays multiple roles in MMR, protecting the template DNA strand from degradation *in vitro,* enhancing DNA excision by Exo1 and facilitating the repair synthesis (Ramilo et al., 2002, Genschel and Modrich, 2003).

In yeast, Rad27 (the yeast flap exonuclease homologue (FEN1)), DNA polymerase delta and DNA plolymerase zeta were thought to play a role in DNA mismatch repair based on the mutator phenotypes of mutations in these components (Kolodner and Marsischky, 1999). However, it was shown recently that FEN1 plays a critical role in processing Okazaki fragments and homologous recombination mediated DNA repair processes. Thus, the mutator phenotype exhibited in yeast rad27 mutants and mammalian FEN1 mutants may be an indirect result of abnormalities in DNA replication, which argues against a direct role of Rad27/FEN1 in MMR (Tishkoff et al., 1997b, Ruggiero and Topal, 2004, Liu et al., 2004). Until now the function of DNA polymerase delta and DNA polymerase zeta in mammalian DNA mismatch repair has not been reported.

1.5.3 MMR in homologous recombination

The MMR system also plays a role in DNA recombination. DNA recombination involves annealing of complementary DNA strands, which often will contain imperfectly matched sequences. These strands form heteroduplex DNA intermediates, which are the substrates for MMR. In yeast, it has been shown that

the MMR system can repair the mismatched nucleotides in heteroduplex recombination intermediates. Mutations in yeast *MSH2, MLH1* and *PMS1* (post meiotic segregation 1) genes caused an increase in post meiotic segregation, which is the result of lack of repair of the heteroduplex intermediates generated during the first mitotic division following meiosis (Alani et al., 1994, Prolla et al., 1994).

Studies in bacteria, yeast and mammals have all revealed that the MMR system acts as a barrier to homologous recombination, in which the binding of MMR to the heteroduplex intermediates elicits a yet unclear downstream process that prevents the occurrence or the progression of homologous recombination between diverged DNA sequences. It has long been known that homologous genes in two closely related bacteria, *Escherichia coli* and *Salmonella typhimurium*, generally will not recombine, although their nucleotide sequences are 80% identical. Mutations in mutH, mutL, mutS and mutU result in a 50 to 3000-fold increase in such interspecies recombination (Rayssiguier et al., 1989). Recombination between two 405 bp substrates in *E.coli* is reduced 240 fold when the sequence homology was decreased from 100% to 89%. While in a MutS deficient strain, the decrease was only about 9 fold (Shen and Huang, 1989). The role of yeast *MSH2, MSH3, MSH6, MLH1* and *PMS1* on homologous recombination have been tested in mitotic recombination assays, in which a homologous recombination event was required to reconstitute a functional selection marker gene on a yeast chromosome. These experiments revealed that mutations in *MSH2* significantly increased homologous recombination between diverged DNA sequences. However, mutation of *MSH3, MSH6, MLH1* or *PMS2* exhibited a modest or minor effect in this assay (Selva et al., 1995, Selva et al., 1997, Nicholson et al., 2000). This result suggests that mismatch recognition protein complexes involving *MSH2* play an important role in recombination between diverged sequences. Consistent with this result, *MSH2*-deficient mouse cells exhibited a significant increase in homologous recombination between diverged DNA sequences, while *MSH3*-deficient cells exhibited a minor effect (de Wind et al., 1995, Abuin et al., 2000). The effects of *Mlh1, Msh6* and

other mammalian *MutS* and *MutL* homologs on homologous recombination between diverged sequences have not been directly examined.

1.5.4 DNA mismatch repair in meiosis

In eukaryotic meiosis, each pair of homologous chromosomes physically interacts and forms chromosomal crossovers as a result of homologous recombination. The connection of the aligned homologous chromosomes can be visualized with an eletron microscope as discernable structure called synaptonemal complex. Two MutS homologues, MSH4 and MSH5 play a role in meiosis. In yeast, Msh4 and Msh5 form a heterodimer protein complex. Mutation of either *MSH4* or *MSH5* gene causes a reduction in meiotic crossover and increased levels of meiosis I chromosome nondisjunciton. *msh4* and *msh5* mutant strains display normal DNA mismatch repair function, suggesting they are not involved in replicative DNA repair (Ross-Macdonald and Roeder, 1994, Hollingsworth et al., 1995). Consistent with this observation, Mammalian homologues of MSH4 and MSH5 exhibit the same effect on meiosis. Human MSH4 and MSH5 form a heterodimer (Bocker et al., 1999). Mice lacking Msh4 or Msh5 are sterile in both males and females, and show abnormalities in chromosome pairing and synapsis at the meiosis prophase 1 (Edelmann et al., 1999, Kneitz et al., 2000). The major mismatch recognition protein in replication repair, Msh2, Msh3, and Msh6, are not involved in meiosis (de Wind et al., 1995, Edelmann et al., 1997). In contrast, Mlh1 acts in both replication repair and meiosis. Mutation in *MLH1* gene caused reduced meiotic crossovers in yeast. In mice, *Mlh1*-deficiency leads to sterility in both male and females (Hunter and Borts, 1997). In addition, *Pms2*-deficiency causes infertility in male mice with abnormal chromosome synapsis, suggesting a role of mammalian Pms2 in meiosis (Baker et al., 1995). Recently, it was shown that the eukaryotic MutL homologue, Mlh3, possesses a distinct function in meiosis. *Mlh3* mutant mice are viable but sterile with reduced meiosis crossovers and a meiotic block. Mlh3 protein is required for Mlh1 binding to meiotic chromosomes and is found to localize to meiotic chromosomes. *Mlh3* mutation in mice doesn't cause discernable

microsatellite instability (Lipkin et al., 2002). The exact role of the MMR proteins in meiosis is still unclear.

1.5.5 MMR in DNA damage surveillance

1.5.5.1 MMR deficiency causes DNA methylation damage tolerance

Studies of MMR deficient cell lines have identified altered responses to DNA methylation damages. The MMR system appears to recognize DNA damage and trigger downstream cell cycle arrest (G2/M) and apoptotic cell death. This function prevents the accumulation of mutagenic DNA lesions and is therefore called MMR-mediated DNA damage surveillance. The function of MMR in DNA damage surveillance was first reported in bacteria. The hypersensitivity of *dam*$^-$ bacteria to simple methylating agents, such as methyl-nitrosourea (MNU) and N-methy-N'-nitro-N-nitrosoguanidine (MNNG) could be rescued by additional mutation in mutS or mutL (Karran and Marinus, 1982). Later, it was demonstrated that cell lines which were tolerant to DNA alkylating agents, such as MNNG, were deficient in mismatch recognition *in vitro*, which implies a link between eukaryote MMR system and DNA methylation damage (Kat et al., 1993, Branch et al., 1993). Clear evidence of a link between MMR and methylation damage came from two human cell lines with mutations in MMR genes, *MLH1* or *MSH2*. The human colorectal adenocarcinoma cell line, HCT116, has a MLH1 mutation and displays microstatellite instability and tolerance to MNNG. Transfer of chromosome 3 that contains the *MLH1* gene to this cell line restored the mismatch repair activity and made the cells sensitive to MNNG (Boyer et al., 1995, Koi et al., 1994). Similar chromosome transfer experiments confirmed that mutations in hMSH2 and hMSH6 caused a MNNG tolerance phenotype in two human endometrial adenocarcinomal cell lines (Umar et al., 1997). These observations suggest that the MMR system is required to trigger cell death in response to DNA methylation damage.

1.5.5.2 Two models for the function of MMR in DNA damage surveillance

The major cytotoxic activity of MNNG is to methylate guanine (G) at the O^6 position, generating a modified nucleotide, O^6-methyguanine (O^6-meG). O^6-meG is repaired by methyguanine methyltransferase (MGMT), which removes the methy-group from O^6-meG. The MNNG tolerance exhibited in MMR deficient cells is not a result of increased MGMT activity because in MMR deficient cells, O^6-meG persists in cells instead of being cleared by MGMT, and the cells are overloaded with G-A transitional mutations (Karran and Bignami, 1992). It was later demonstrated that O^6-meG can pair with either thymidine (T) or cytosine (C) during DNA replication and form imperfect O^6-meG/T or O^6-meG/C basepairs. Both O^6-meG/T and O^6-meG/C basepairs can be bound by the mismatch repair recognition protein complex, MutSα, in MMR proficient cell extracts but not in MNNG tolerant cell extracts (Griffin et al., 1994, Duckett et al., 1996). This data suggests that the mismatch binding ability of MMR proteins is involved in the cytotoxic pathway of MNNG. The exact mechanistic link between MMR deficiency and DNA methylation damage tolerance has not been fully established. Two models have been proposed. In one model, the binding of the mismatched nucleotides and the subsequent repair are thought to be essential. During the MMR process, the newly synthesized DNA strand containing the mismatched T of the O^6-meG/T base pair is removed by DNA exonuclease. However, a thymidine will again be incorporated and pairs with O^6-meG, which will initiate another round of mismatch repair. This “futile” repair process could stall DNA replication and create double strand breaks, both of which may serve as a signal for cell cycle arrest and cell death. It has been shown that MNNG could only trigger apoptosis in dividing cells and in these cells the apoptosis was preceded by a wave of DNA double strand breaks (Roos et al., 2004). Giving the established function of MMR in repairing mismatched nucleotides, this model provides a simple explanation for the DNA methylation tolerance.

In the other model, it was proposed that MMR components serve as a general DNA damage sensor. The binding of MMR proteins to damaged DNA could trigger a downstream signaling cascade that signals cell death and cell cycle arrest. In this model, DNA mismatch repair is not required and thus the MMR system is expected to be able to sense a broad spectrum of DNA damage besides DNA methylation (Fink et al., 1998, Karran, 2001). Indeed, it has been shown that tumor cells lines with defects in MSH2, or MLH1 exhibited modest but significant tolerance to many chemotherapeutic drugs, which induce various types of DNA damage, for example, Cisplatin and Doxorubincin. Cisplatin forms bulky intra or inter DNA strand crosslinks and Doxorubincin is a DNA topoisomerase inhibitor. MutSα protein complex is able to bind to the DNA lesion caused by cisplatin (Aebi et al., 1996). The depletion of DNA topoisomerase activity by topoisomerase inhibitors will stall DNA replication, and the arrested DNA replication may trigger the MMR system-mediated cell cycle arrest and cell death pathways (Fedier et al., 2001). Consistent with the general DNA damage sensor model, it was shown that overexpression of human *MSH2* or *MLH* genes can trigger apoptosis in either mismatch repair-proficient or -deficient cells without DNA damage (Zhang et al., 1999).

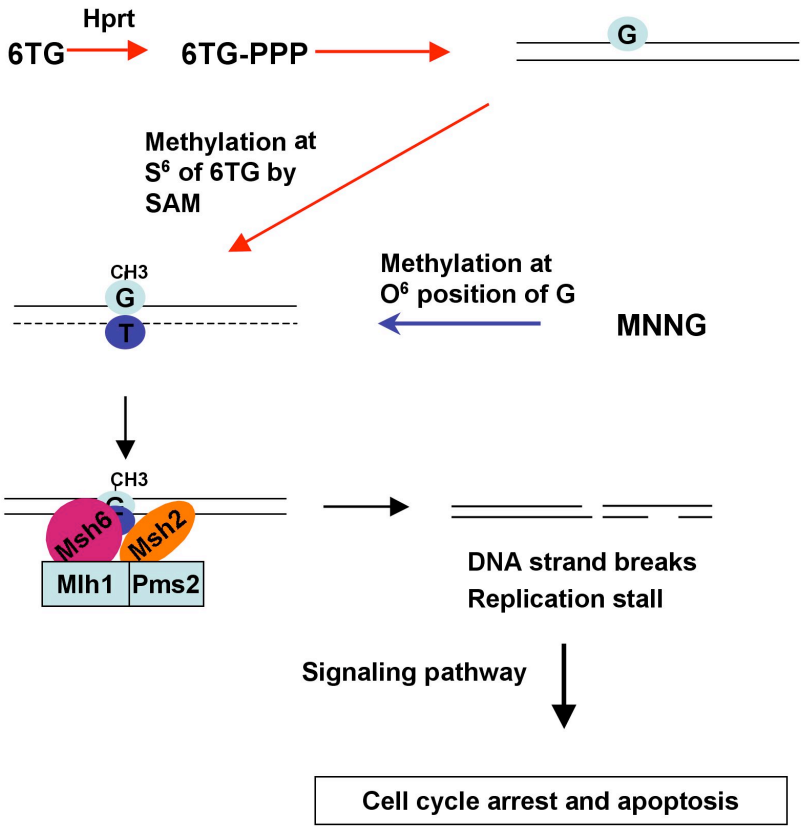


Figure 1-4. DNA mismatch repair in MNNG and 6TG cytotoxity

The cytotoxic mechanism of 6TG is initiated in cells by Hprt (hypoxanthine-guanine phosphoribosyltransferase), which converts 6TG to 2-deoxy-6-thioguanosine triphosphate (6TG-PPP). 6TG-PPP is incorporated into DNA as a guanine analogue during DNA synthesis. S-adenosylmethionine (SAM) methylates 6TG at the S^6 position to form S^6-methylthioguanine (S^6-mG). S^6-mG pairs with T during DNA replication, forming S^6-mG/T mismatched basepairs . MNNG methylates guanine (G) at the O^6 position in DNA to form O^6-mG , which also forms an imperfect basepair with T. The binding of the MutSα to S^6-mG/T or O^6-mG initiates multiple cycles of mismatch repair that causes single and double strand DNA breaks. The DNA strand breaks and the subsequent DNA replication stall signals cell cycle arrest and cell death.

1.5.5.3 MMR deficiency causes tolerance to 6-thioguanine (6TG)

6TG has long been used as a purine anti-metabolite drug in the treatment of acute leukemia (Elion, 1989). The cytotoxity of 6TG requires the activity of hypoxanthine-guanine phosphoribosyl transferase (HPRT), which transfers the sugar phosphate group to 6TG to form 2'-deoxy-6-thioguanosine-triphosphate, the active guanine nucleotide analogue in DNA synthesis. Cells that lack the *HPRT* gene are fully resistant to 6-TG killing. In the mid 90's, it was found that 6TG cytotoxity requires MMR activity. Cells with defects in *MSH2, MSH6, or MLH1* genes are tolerant to 6TG (Aebi et al., 1997). It is believed that the cytotoxic mechanism of 6TG is similar to MNNG. Both drugs show delayed cytotoxity and elevated SCE. Notably, MMR deficient cells that are tolerant to MNNG are also tolerant to 6-TG (Tidd and Paterson, 1974). The direct link betwwn MMR and 6TG cytotoxity was established by two studies. One shows that 2'-deoxy-6-thioguanosine-triphosphate in DNA can be methylated by S-adenosylmethionine (SAM) to form S^6-methylthioguanine (S^6-mG). S^6-mG can pair with either thymidine (T) or cytosine(C) in the growing DNA strand (Swann et al., 1996) and the S^6-mG/T basepair is the binding substrate of mismatch repair complex MutSα (Waters and Swann, 1997). Based on the structural similarity of O^6-methyguanine (O^6-meG, generated by MNNG) and S^6-thioguanine, it is conceivable that MNNG and 6-TG share similar cytotoxic processes involving MMR damage surveillance. The MMR-mediated 6TG cytotoxic mechanism is illustrated (Fig. 1-4).

1.5.5.4 Molecular basis of MMR in DNA damage surveillance

The molecular basis of MMR mediated DNA damage surveillance is poorly defined. The p53 pathway may be involved, which is suggested from the comparison of p53 activity between MMR proficient and deficient cells following DNA methylation damage (O^6-meG) induced by Temozolomide (D'Atri et al.,

1998). The expression of p53 and p21/waf-1 (p21/waf-1 is induced by p53) was up regulated following Temozolomide treatment in MMR proficient lymphoblast cells, which is coincident with a G2/M cell cycle arrest and apoptosis. However, in the MMR deficient cells, the cell cycle arrest and apoptosis phenotype was attenuated and no P53 induction was detected (D'Atri et al., 1998). This relationship was supported by a recent study, which showed that MNNG triggered apoptosis was accompanied by p53 and Fas receptor up regulation. Inhibition of Fas receptor activity attenuated MNNG-induced cell death in a lymphoblast cell line (Roos et al., 2004). Although the p53 pathway appears to be involved in the cell cycle arrest and apoptosis in lymphoblast, the involvement of p53 in other cell lines is less certain. In the human kidney derived fibroblast cell line, 293T, p53 is not essential. Although 293T cells lack p53 activity, they can undergo G2/M cell cycle arrest and apoptosis following MNNG treatment (Cejka et al., 2003, di Pietro et al., 2003). A link between MMR surveillance and the ATR signaling pathway was established recently. ATR (ataxia telangiectasia and rad3+ related) is an important cell cycle checkpoint protein kinase, which is activated in response to a block in DNA replication. MSH2 protein interacts with ATR, which regulates the phosphorylation of downstream effectors including CHK1 and SMC1 (structure maintenance of chromosome 1) (Wang and Qin, 2003). The ablation of ATR or the inhibition of CHK1 attenuates the MNNG and 6TG induced G2/M cell cycle arrest (Stojic et al., 2004, Yamane et al., 2004).

1.5.5.5 MED1/MBD4, a methyl-CpG binding protein involved in DNA damage surveillance

MED1 was identified as a protein interacting with MLH1 in human cells (Bellacosa et al., 1999). MED1 is a member of a group of methyl-CpG binding proteins, which is also referred to as MBD4 (methyl-binding domain 4) in some publications. MBD4 (MED1) binds to fully and hemimethylated DNA but not to unmethylated DNA *in vitro* (Bellacosa et al., 1999). Deamination of 5-methylcytosine (m^5C) to T occurs frequently at CpG sites, which causes T:G

mismatch. Mammalian MBD4 has glycosylase activity that enzymatically removes thymine (T) from a mismatched T:G basepair at CpG sites (Hendrich et al., 1999). MBD4 has been shown to be important in suppressing the mutational load caused by deamination of the m^5C. *Mbd4*-deficient mice have an increased rate of CpG mutability and tumorigenesis (Millar et al., 2002, Wong et al., 2002). However, the function of MBD4 is not limited to repairing T:G mismatches at CpG sites. Transfection of a dominant negative form of MBD4 into cultured cells leads to MSI in an episomal slippage construct that contains tandem CA repeats (Bellacosa et al., 1999). This observation and the fact that MBD4 interacts with MLH1 suggest that MBD4 plays a role in MMR. Further studies have revealed that frameshift mutations in *MBD4* coding sequence occur frequently in colon, endometrial, pancreatic and gastric tumors with high rates of MSI (Riccio et al., 1999, Bader et al., 1999, Yamada et al., 2002). However, *Mbd4*-deficient mice generated by gene-targeting do not exhibit MSI (Millar et al., 2002, Wong et al., 2002). A link between MBD4 and MMR surveillance was demonstrated recently by studies on cultured mouse embryonic fibroblasts (MEF) derived from *Mbd4*-deficient mice. In this study the response to DNA damaging drugs were examined and revealed that *Mbd4*-deficient MEFs failed to undergo G2/M cell cycle arrest and apoptosis in response to the treatment of simple methylating agents like MNNG. Moreover, the cytotoxic response to other DNA damaging drugs such as the DNA crosslinking platinum drugs was also attenuated in *Mbd4*-deficient MEF cells (Cortellino et al., 2003). The drug tolerance exhibited by *Mbd4*-deficient MEFs is similar to the DNA damage tolerance exhibited by cells with deficient MMR, which is characterized by the accumulation of DNA lesions in cells (Cortellino et al., 2003). The function of MBD4 in DNA damage surveillance was also observed in the small intestine in *Mbd4*-deficient mice. Mice deficient for *Mbd4* showed significantly reduced apoptotic responses following treatment with a range of cytotoxic agents including cisplatin and 5-fluorouracil (5-FU), a DNA replication inhibitor. Mice lacking both Mlh1 and Mbd4 functions didn't show synergistic effect on DNA damage induced apoptosis, suggesting that MBD4 and MLH1 act in the same pathway (Sansom et al., 2003).

1.5.6 A genetic screen for genes that protect the *C. elegans* genome against mutations

Research on the MMR system was greatly stimulated when the major mismatch repair components were isolated in genetic screens in bacteria. However, genetic screens in dipoid organism like mice is extremely restricted by the difficulty of obtaining homozygous mutations. Taking advantage of the recently developed RNAi (RNA interference) technology in *C.elegans*, a genetic screen has been conducted for genes that protect the *C. elegans* genome against mutations, which includes MMR genes. For simplicity this screen will be referred to as the *C.elegans* MMR screen (Pothof et al., 2003). RNAi technology in *C. elegans* was developed based on the phenomena that the double strand RNA is able to knockdown the expression of the endogenous genes that are homolgous to it (Hannon, 2002). The *C.elegans* MMR screen was based on the *C.elegans* RNAi library, which contains bacterial strains that each produce double-stranded RNAs (dsRNA) for an individual nematode gene. This library is able to target ~86% of predicted *C. elegans* genes. Loss-of-function phenotypes when performing systemic RNAi on a genome-wide scale is estimated to be ~65% (Fraser et al., 2000). To provide a readout for potential MMR mutations (leading to increased DNA genomic instability), a *gfp-LacZ* reporter construct was put out of frame by an A_{17} mononucleotide DNA repeat cloned directly between the initiation ATG and the *gfp-LacZ* open reading frame. Genomic instability mutations that restore the *gfp-LacZ* reading frame can be identified by inspecting the expression of GFP and/or LacZ. The presence of mono- nucleotide repeat sensitizes the reporter system for frameshift mutations. In this screen, several well-known MMR genes were identified including *C. elegans* homologues of human *MLH1, PMS2, MSH2* and *MSH6*. In addition, many genes were recovered with functions in DNA replication, repair, chromatin organization and cell cycle control (Pothof et al., 2003).

1.5.7 MMR deficiency in Cancer

Carcinogenesis is a multi-step genetic process, during which several mutations must be acquired before a normal cell develops into a tumor. The putative tumor cell has to override normal cell cycle control, genetic programs of differentiation, senescence and apoptosis. Each of these steps requires alteration of one or several genes. It has been hypothesized that genomic instability is fundamental during tumorigenesis because elevated mutation rates facilitate the accumulation of multiple mutations (Schmutte and Fishel, 1999). The finding that MMR deficiency is associated with hereditary non-polyposis colorectal cancer (HNPCC) and several sporadic tumors illustrate the importance of the DNA mismatch repair system in maintaining genomic stability (Peltomaki, 2001). HNPCC accounts for nearly 8% of all colon cancers. HNPCC shows an autosomal dominant mode of inheritance, high penetrance and an early onset of tumorigenesis. The molecular hallmark of HNPCC is a high or low level of microsatellite instability (MSI), a feature that is characteristic of MMR deficiency. Indeed, the first human MutS homologue, *MSH2*, was cloned because of its linkage with HNPCC (Fishel et al., 1993). Besides *hMSH2*, germ line mutations in *hMLH1, hPMS2, hMSH6* and *hPMS1* have also been found in HNPCC patients (Wei et al., 2002). HNPCC patients usually inherit one mutated allele from one parent. The other normal allele is mutated in somatic tissues either by loss of heterozygosity, point mutation or hypermethylation. Cells with homozygous mutated MMR genes will then be predisposed to tumorigenesis. MSH2 and Mlh1 are the core components of mismatch protein complex, MutSα and MutLα. Consistent with this role, mutations in *MSH2* and *MLH1* are responsible for 50% and 40% of HNPCC respectively, while mutations in *MSH6*, *PMS2* and *PMS1* are less frequently found in HNPCC (Peltomaki, 2001).

As a consequence of MSI, those genes having simple repeat sequences in their coding region have a greater chance of acquiring a mutation and could be important targets of the MMR deficiency phenotype. Several genes that regulate

cellular growth, cell cycle control, DNA repair and apoptosis lie in this category, including *TGFβ RII, IGF2R, MSH3, MSH6, p53* and *BAX* (Peltomaki, 2001).

1.6 Mutagenesis in mice and embryonic stem cells

1.6.1 Forward genetics, phenotype-based screens

Mutagenesis followed by phenotypic screening is one of the most powerful genetic approaches to elucidate the molecular basis of complex biological phenomena. Such strategies are referred to as “forward genetics”. In the last several decades, forward genetic screens have been conducted in several “model” organisms, including mice. In mice, mutations occur spontaneously at a low efficiency (about $5x10^{-6}$ per locus per gamete). However, the spontaneous mutation rate is far too low for a genetic screen. Highly efficient mutagenesis can be achieved using DNA damaging agents such as irradiation (X-rays, γ-irradiation) or chemicals such as *N*-ethyl-*N*-nitrosourea (ENU). X-ray and gamma irradiation can induce mutations with a rate 20-100 times higher than the spontaneous mutation rate. However, irradiation can break DNA strands, resulting in chromosome rearrangements such as deletions and translocations. The complexity and size of DNA lesions induced by irradiation limits its use as a mutagenesis method. However, γ-irradiation has been used to generate mice containing regional deletions. These deletion mice have been shown to be useful in combination with single gene mutations (You et al., 1997, Goodwin et al., 2001, Bergstrom et al., 1998, Chao et al., 2003).

Chemical mutagens, such as ENU, are one of the most potent mutagens in mice. ENU mutates mouse spermatogonial stem cells with a frequency of $1.5\text{-}6x10^{-3}$ per locus per gamete (Bode, 1984, Hitotsumachi et al., 1985). ENU alkalizes oxygen of DNA nucleotides, which, if not repaired, causes predominately single nucleotide mutations including A/T to G/C, A/T to C/G, A/T to T/A, G/C to C/G and G/C to T/A transitions and transversions in mice. In a typical ENU

mutagenesis screen, male mice (G_0) are injected with ENU to generate mutated gametes. Mating the ENU-treated founder males to unaffected wild type females will then produce G1 offspring. Each of the G_1 animals will carry a unique set of mutated alleles. These G_1 animals can be used directly to screen for dominant phenotypes, or they can be backcrossed to wild type animals to establish lines of mice that carry the same set of mutations. By inter-crossing mice from the same line, some of the descendants will carry homozygous mutations that can be used to screen for recessive phenotypes. Because of the comparative simplicity of a dominant screen, most of the genetic screens that have already been performed were set up to identify dominant phenotypes. Approximately, 2% of G_1 animals exhibit a heritable dominant phenotype (Justice et al., 1999, Brown and Balling, 2001, Balling, 2001).

However, the majority of mutations induced by ENU are recessive. The requirement for a complicated and expensive breeding program, and the difficulty to genotype mice with point mutations limits the application of genome-wide recessive ENU mutagenesis screens. To circumvent this, mice with defined regional chromosome deletions and inversions have been generated via Cre-*loxP* mediated chromosome engineering techniques. These deletion and inversion mice provide essential genetic tools to maximize the efficiency of ENU mutagenesis because the homozygous mutant mice can be identified and the mutation mapped (Yu and Bradley, 2001). By crossing the ENU mutated G_1 mice with mice carrying chromosomal deletions, recessive phenotypes can be manifested in G_2 animals if the mutation lies in the deletion intervals. The use of chromosome deletion mice has been demonstrated by a few pioneering experiments, which located and identified molecular lesions using mice containing a set of overlapped small deletions generated either by gamma-irradiation or by chromosome engineering techniques (Bergstrom et al., 1998, Su et al., 2000, Lindsay et al., 2001, Chao et al., 2003).

Although mice that carry small deletions are very useful, larger deletions may sometimes lead to reduced fitness, infertility or even lethality because of haploid insufficiency in some genes. Balancer chromosomes were developed originally in *Drosophila* to maintain recessive lethal mutations and have been proved to be an important tool for stock maintenance. A balancer chromosome carries one large inversion or a set of inversions along a chromosome. Productive meiotic crossovers between an inversion chromosome and a normal chromosome are efficiently suppressed because this type of crossover leads to inviable germ cells, harboring dicentric or acentric chromosome. Recently, balancer chromosomes in mice carrying large chromosomal inversions were created with Cre-*loxP* mediated chromosome engineering (Zheng et al., 1999). These balancer chromosomes were engineered to carry a visible dominant marker and a recessive lethal mutation on the chromosomal inversion. Mice with homozygous balancer chromosomes are automatically eliminated from crosses between heterozygous mice with one balancer chromosome because of the recessive lethal mutation. Mutations can be maintained without recombinational loss in the balanced heterozygotes and tracked by the visible dominant marker. The utility of balancer chromosomes has been demonstrated recently in an ENU mutagenesis screen for recessive mutations along a 24 cM balanced chromosomal region on mouse chromosome 11 (Kile et al., 2003).

Because of the lack of an overt molecular tag, the identification of an ENU induced mutation normally starts with linkage analysis in order to locate the mutation of interest within a small chromosomal region of several centimorgens (cM), which often requires analyzing hundreds of meiotic events. Candidate genes within that region can be assessed based on their expression pattern, structure, and functional domains. Sequencing of candidate genes will determine the molecular change. Confirmation of the mutation can be acquired by phenotype-complementation, for example using a cDNA construct or a large genomic DNA fragment such as bacterial artificial chromosomes (BACs) that harbor one or multiple candidate genes (King et al., 1997, Allen et al., 2003,

Floyd et al., 2003, Swing and Sharan, 2004, Zhang et al., 1994). As discussed above, mice with chromosome deletions and inversions provide essential tools for ENU mutagenesis in term of mutation identification and maintenance.

ENU mutagenesis is a useful technique which can quickly create a large number of mutations at random to allow screens for any phenotypic abnormality provided that the phenotype of interest is visible or detectable. Moreover, because ENU induces point mutations in a random fashion, independent mutations in the same gene can be generated which may act as hypermorph and neomorph alleles in addition to the common loss of function alleles. ENU mutagenesis screens have been conducted on many thousands of mutant mice and a number of mutations have been characterized which mimic human diseases.

1.6.2 Reverse genetics, transgenic animals and gene-targeting

In contrast to applying random mutagenesis and conducting forward genetic screen, reverse genetics can be used to directly obtain functional information of a gene by mutating or over expressing the gene and examining the consequence in the resultant transgenic or knockout mice. Transgenic mice were originally generated by directly injecting exogenous DNA into fertilized zygotes or by retroviral infection of early development stage embryos (Jaenisch, 1988). The injected DNA or retroviral vector could stably integrate into the host genome and was transferred into the mouse germ line. These methods usually produce gain of function alleles that are useful in studying the biological effect of over-expressed genes *in vivo* (Berns, 1991). However, these methods have shown some major limitations. First, the integration of injected DNA is random and uncontrollable. The injected DNA often forms head to tail concatemers before integration and integration is often accompanied by chromosomal rearrangements in the flanking DNA around the integration site. Furthermore, the expression of genes varies between different integration sites, cell and tissue types. Expression from a retroviral vector can even be totally abolished by DNA

methylation (Jaenisch, 1988). Recently, gene-targeting technology has been developed as a new version of transgenic technology, which is capable of introducing a single copy DNA fragment into a specific genomic locus in a predicable manner.

Taking advantage of mouse embryonic stem cell (ES) technology and homologous recombination, gene-targeting provides a more powerful means to generate transgenic mice harboring precise mutations in the gene of choice. ES cells are pluripotent cells, established from inner cell mass (ICM) of pre-implantation blastocysts. ES cells in culture maintain unlimited self-renewal ability. Most importantly, ES cells, even after modification in culture, can contribute to all somatic tissues as well as the germ line of chimaeras when they are injected into host blastocysts (Evans and Kaufman, 1981, Bradley et al., 1984, Robertson et al., 1986, Kuehn et al., 1987). Targeted mutations can be achieved by homologous recombination between endogenous genes and a targeting vector (Doetschman et al., 1987, Thomas and Capecchi, 1987). In its simplest form a gene-targeting vector is constructed to carry a DNA fragment homologous to the targeted gene and a positive selection marker. ES cells can be directly selected with the integrated targeting vector and the subset with the engineered mutation at the targeted site can be identified by Southern-blotting or PCR. Since the advent of gene-targeting technology in the late 80's, it has quickly evolved to be one of the most frequently practiced approaches in mouse genetics. With improved molecular cloning technologies like *E.coli* recombineering (Copeland et al., 2001), gene-targeting vectors can be quickly constructed with long homologous arms to obtain better gene-targeting efficiencies. Nowadays, gene-targeting vectors can be engineered at will to target any genes, generating all possible classes of mutations like loss of function, gain of function, point mutations and knockin alleles. Combined with the Cre-*loxP* technology, the "expression" of a mutation can also be made controllable or inducible in a temporal or spatial manner (Ramirez-Solis et al., 1993). Despite the power of creating mutations in targeted genes, the gene-targeting approach

requires the prior knowledge of genes to design a gene-targeting vector. Thus, novel phenotypic information about a gene is often missed. Gene-targeting can only be applied on a gene-by-gene bases and it has not yet been employed for genetic screens.

1.6.3 Insertional mutagenesis, the gene trap approach

Random insertional mutagenesis can also be used to mutate genes. The integrated DNA molecule provides a sequence tag for identifying the mutated gene using a PCR-based method. Retroviruses have been used as insertional mutagens since the late 70's. The integration of a retrovirus may produce a loss of function mutation when it inserts into the coding region of a gene. Retroviruses can also generate gain of function mutations, in which expression of a gene is increased by the viral enhancer element (Jaenisch et al., 1981, Lund et al., 2002, Mikkers et al., 2002). The mutational efficiency of a randomly integrated retroviral vector is very low. Over 95% of mouse genome is non-coding sequences. Retroviral integrations in these regions are often phenotypically "neutral" to cells. The availability of ES cell technology in the mid 80's expedited the design of better insertional mutagens, the gene trap vectors, in the following years. Gene-trap mutagenesis predominantly produces loss of function mutations in a random fashion. The gene trap vector serves as a molecular tag for cloning of the mutation. Combined with the ES cell technology, gene trap offer a valuable tool for rapidly creating large numbers of loss of function mutations for functional genomic studies in mice (Stanford et al., 2001, Evans et al., 1997)

1.6.3.1 Gene trap methods

During past 10 years, various gene trap vectors have been designed for individual experiments. These vectors contain a non-functional reporter gene cassette and the expression of the reporter gene requires the cis-elements of an endogenous gene. The basic gene trap designs include enhancer traps,

promoter traps and polyadenylation signal (PolyA) traps (Fig. 1-5) (Zambrowicz and Friedrich, 1998).

The enhancer trap vectors were originally used to study the effect of host genes on the expression of transgenic reporters in mice. For this purpose, the enhancer trap vectors were built to contain *E.coli lacZ* gene with a minimal promoter sequence. The expression of the reporter requires the vector to insert near a cis-acting enhancer element (Allen et al., 1988, Kothary et al., 1988, Gossler et al., 1989). Similar designs were adopted in *Drosophila* in genetic screens for cis-elements that were able to drive the expression of a minimal promoter fused to a *lacZ* reporter gene. These early experiments established that the expression of the reporter gene is regulated by the flanking cis-elements, and the reporter expression often displays a spatially or temporally restricted pattern (Bellen et al., 1989, Bier et al., 1989). Enhancer trap vectors haven't been extensively used because they are not efficient mutagens. The enhancer elements of a gene are often a large distance away from the coding elements so that the insertion of the enhancer trap vector does not normally disrupt the expression of the gene. Promoter traps and PolyA traps are much better mutagens.

The essential component of a promoter trap is a reporter gene that possesses a strong splicing acceptor (SA) but lacks a promoter. Therefore, the reporter can only be transcribed from the endogenous gene into which the promoter trap reporter integrates, generating a fused transcript containing a 5' portion of the endogenous gene and the coding sequence of the reporter. Consequently, the transcription of the endogenous genes is disrupted, creating a loss of function mutation. The expression of the mutated gene can be assessed by inspecting the expression of the reporter gene. Because of the nature of the promoter gene trap design, these vectors can only mutate genes expressed in the experimental cell line.

A PolyA trap vector utilizes a reporter gene lacking a polyadenylation signal, but possessing a “strong” splice donor (SD). The reporter gene has its own promoter but can only generate a stable transcript if the PolyA trap vector inserts into an endogenous gene and downstream PolyA signal is provided. In contrast to promoter traps, the PolyA trap vector can be used to mutate any gene regardless of its expression status in the experimental cell line since the reporter is expressed from an exogenous active promoter.

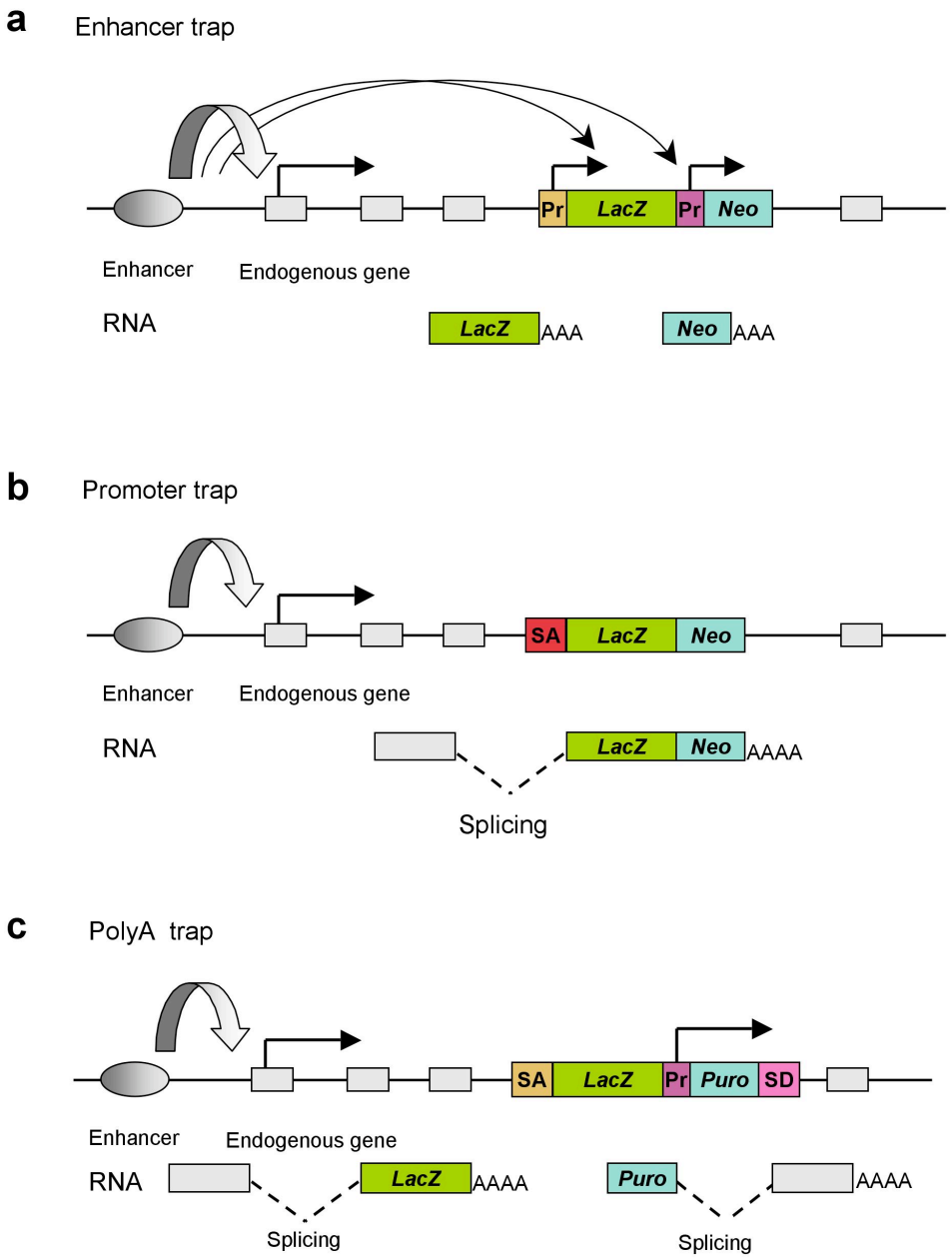

a
Enhancer trap
Pr
LacZ
Pr
Neo
Enhancer
Endogenous gene
RNA
LacZ
AAA
Neo
AAA
b
Promoter trap
SA
LacZ
Neo
Enhancer
Endogenous gene
RNA
LacZ
Neo
AAAA
Splicing
c
PolyA trap
SA
LacZ
Pr
Puro
SD
Enhancer
Endogenous gene
RNA
LacZ
AAAA
Puro
AAAA
Splicing
Splicing

Figure 1-5. Schematic representation of basic gene trap strategies, showing the structure and the expression of the gene trap cassettes integrated in an endogenous gene**. a.** Enhancer trap. *LacZ* and *Neo* reporter genes are driven by minimal promoters (Pr). The exression of reporter genes are enhanced by the endogenous enhancer. **b.** Promoter trap, showing the *SAβgeo* gene trap cassette, a fused *lacZ/Neo* gene with a consensus splicing acceptor (SA). **c**. PolyA trap. *Puro* is transcribed from an autonomous promoter (Pr) and spliced from the splice donor (SD) in the gene trap cassette into endogenous gene. Note that in some cases, a fused *SAβgeo* promoter gene trap cassette is combined with the PolyA trap vector, which can provide a color reporter for monitoring the expression of endogenous genes (Zambrowicz et al*.,* 1998). Gray rectangles represent exons of an endogenous gene.

1.6.3.2 Gene trap mutagenesis in genetic screens

1.6.3.2.1 Expression screens

A genetic screen is an essential approach to establish relationships between genes and functions. Phenotype-driven screens in mice have been extremely restricted because of the difficulty in obtaining homozygous mutations. One fact of a gene's function can often be obtained by assessing its expression. For example, the expression of developmentally important genes often exhibited highly restricted patterns during development. The expression of genes involved in cell signaling pathways can be induced or repressed by physiological molecular signals. Because the promoter trap approach allows a quick examination of the expression of an endogenous gene, this vector type has been used for expression screens. Wurst et al. (1995) performed an expression screen in mouse embryos for genes involved in embryogenesis based on the hypothesis that such genes will exhibit temporally or spatially restricted expression patterns. They mutated ES cells with a promoter trap vector containing the *lacZ* reporter gene. 279 gene trap clones were assessed in chimeric embryos, and by X-gal staining the expression patterns of the mutated genes were examined. Approximately, one third of genes expressed in ES cells are either temporally or spatially regulated during embryogenesis (Wurst et al., 1995). This work demonstrated the feasibility of the use of the promoter gene trap in an expression screen. However, the generation of a large quantity of chimeras or mice requires significant time, labor and animal resources and is not practical for many laboratories.

ES cells are pluripotent cells. They can not only contribute to all tissues in mice but also differentiate into many cell lineages *in vitro*, therefore allowing prescreening of genes which function in specific types of cells. The gene-trap mutagenesis combined with various *in vitro* ES cell differentiation conditions has

been applied to screen for genes expressed in chondrocytes, cardiomyocytes, skeletel muscle cells, haematopoietic cells, endothelial cells and neurons (Baker et al., 1997, Stanford et al., 1998, Hirashima et al., 2004, Muth et al., 1998, Hidaka et al., 2000, Shirai et al., 1996, Thorey et al., 1998, Stuhlmann, 2003).

Chromatin or chromosomal proteins normally show restricted cellular localization within nuclear compartments or sub-nuclear compartments, thus the genes encoding these proteins may be identified by examining the localization of the gene trap reporter protein in cells (Tate et al., 1998). Genes with altered expression levels in response to many physiological stimuli or signals, such as retinoic acid and gamma-irradiation have also been screened in culture using the promoter gene trap approach (Forrester et al., 1996, Vallis et al., 2002, Mainguy et al., 2000).

Gene trap vectors can also be specially designed to suit individual screens. A secretory trap vector was designed to identify secreted and transmembrane proteins. In this screen, a transmembrane signal sequence was placed adjacent to the *βgeo* gene trap reporter. The transmembrane signal will place the βgeo protein inside the endoplasmic reticulum (ER) so that it doesn't function. To allow the detection of the *βgeo* expression, an additional N-terminal signal sequence from the trapped gene is required to place it outside the ER (Skarnes, 2000). Hundreds of secreted and transmembrane proteins have been identified by this approach (Mitchell et al., 2001). Recently, this secretary trap was modified to identify genes controlling neuronal axon guidance. In this design, an axonal marker is co-expressed with the *LacZ* gene trap reporter to label the neuronal axons. By staining the expression of the *LacZ* reporter and the axon marker in mice, genes with restricted expression patterns in neuronal axons were identified (Leighton et al., 2001).

The gene trap approach is not restricted to ES cells, it can be applied to other cultured cell types to study genes with unique features in these cell lines. For

example, a gene trap screen has been conducted in hematopoietic cells which were induced to undergo apoptosis by growth factor deprivation. Genes with potential survival functions in hematopoietic lineages could be identified based on their induced expression following growth factor deprivation (Wempe et al., 2001). To better understand the complex signaling networks involved in germ cell maturation, gene trap screens have been conducted in Sertoli cells, the somatic cells supporting and controlling male germ cell development (Vidal et al., 2001). Differentiating germ cells were then added to the mutated Sertoli cells to screen for cells showing changes in the expression of the trapped genes. Gene trap strategies have been used in NIH3T3 fibroblasts to identify inhibitors of oncogenic transformation, in cultured B-cells to identify lipopolysaccharide (LPS) responsive genes and in human lung carcinoma cells to identify TGF-beta-responsive genes (Kerr et al., 1996, Andreu et al., 1998, Akiyama et al., 2000). Taken together, these experiments show that gene trap approach is a powerful mutagenesis method with versatile and broad applications in genetic screens.

1.6.3.2.2 Gene trap in phenotype-driven screens

Phenotype-driven screens in diploid genome require a strategy to obtain homozygous mutations. Chinese hamster ovary (CHO) cells contain a partial hemizygous genome. Therefore, recessive mutations within the hemizygous regions can be phenotypically accessed. Screens in CHO cells have been successfully applied to isolate recessive mutations that are sensitive to UV radiation, for example the base excision repair (BER) protein *ERCC1* (Westerveld et al., 1984). A Gene trap screen has also been conducted in CHO cells to identify mutations in glycosylation. Cells with defects in glycosylation are resistant to wheat germ agglutinin. Four individual mutants were isolated in this experiment. By Southern-blot analysis, four gene trap insertion sites were mapped to different positions in a 796 base pair region (Hubbard et al., 1994). The localization bias of the gene trap mutations identified in this screen may reflect the limited hemizygous genome of CHO cells or gene trap vector insertion

“hot spots”. Unfortunately, information about the efficiency of the screen was not provided, nor the identify of the mutated gene was isolated.

1.6.3.3 Methods for introducing gene trap vectors into cells

1.6.3.3.1 Electroporation

The simplest way to perform gene trap mutagenesis is to electroporate the linearized gene trap vector into cells as “naked” DNA. The gene trap vector can integrate into genome randomly, which is normally accompanied by DNA concatermerization. Thus many copies of electroporated linear DNA molecules form head to tail arrays and integrate into host genome together through a process mediated by a DNA repair process known as non-homologous end joining DNA repair (NHEJ) (Brinster et al., 1985, Skarnes, 2000). This method has several limitations. First, multiple copies of the gene trap vector in one locus complicates the identification of the gene trap mutations. Second, the gene trap vector can be truncated during electroporation.

1.6.3.3.2 Retroviral based gene transfer

1.6.3.3.2.1 Retroviral life cycle

The typical retrovirus genome consists of two copies of a single-stranded RNA molecule of about 8-12 kb, depending upon the retroviral species. The genome encodes three major proteins, Gag, Pol and Env. Gag is processed to make the core proteins. Pol has the reverse transcriptase, Rnase H and integrase activities. Env is the viral envelope protein that resides in the lipid layer and mediates the viral-host cell interaction during viral infection. The viral particle consists mostly of gag-derived proteins, genomic RNA, and the reverse transcriptase protein as the virus nucleoprotein core, which are enclosed by the outer lipid-protein shell of the viral envelope. Viral particles infect host cells by

binding to cell surface receptors, a process determined largely by the envelope proteins of the retrovirus. Infection leads to injection of the virus nucleoprotein core. Once inside the cell, a double-stranded DNA is generated from the viral genomic RNA by the reverse transcriptase. Catalyzed by the viral integrase, the double strand viral DNA integrates stably into the host chromosome. The integrated viral DNA is known as proviral DNA. At this stage, the virus is now prepared to initiate a new round of replication. Full-length genomic mRNA is transcribed from proviral DNA by the host cell RNA polymerase, initiated at the beginning of the R region of the 5' LTR (Long Terminal Repeat) and terminating at the end of the R region at the 3'LTR. Full length genomic RNA can be spliced and provides messenger RNAs, from which the viral proteins are synthesized. The full length genomic RNA and the viral nucleoproteins are packed into new viral particles and released from the host cell by budding from the plasma membrane (Coffin J M and E, 1996).

1.6.3.3.2.2 Recombinant retroviral, viral packaging cell lines

Recombinant retroviral vectors have been developed and been wildly used to transfer genes into eukaryote cells because of the capability of retroviral integration, allowing constitutive expression of exogenous genes carried by the retrovirus. Recombinant retroviral vectors have been constructed. The major components of recombinant retroviral vectors include the 5' long terminal repeat, the 3' long terminal repeat and cis -elements essential for viral RNA packaging, such as viral packaging sequence Ψ. The viral proteins can be produced in trans and are thus deleted from the viral genome to accommodate exogenous DNA. Deletion of the trans-elements in a recombinant retrovirus leads to replication

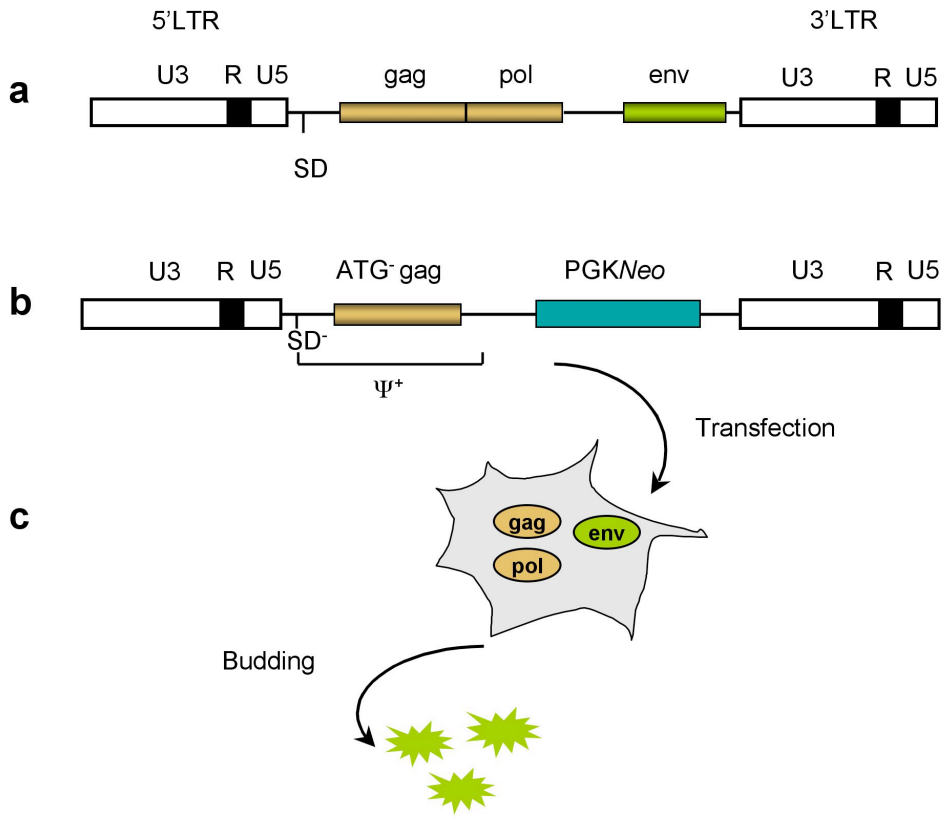


Figure 1-6. Recombinant retroviral vectors and viral production.

a. The structure of wild type retrovirus with a simple genome, illustrated as a provirus containing the long terminal repeat (LTR), genes that encode viral protein core (*gag*), reverse transcriptase (*pol*) and the envelope protein (*env*). SD, viral splice donor. **b.** Schematic of a representative recombinant retroviral vector based on Moloney murine leukemia virus (Mo MuLV). DNA sequences including *pol* and *env* genes are deleted. The splice donor and *gag* sequence remain to facilitate viral packaging, which is indicated as Ψ^+, representing an extended viral packaging signal. To avoid the interference of internal gene expression by viral mRNA splicing and protein translation, the viral splice donor is mutated (SD^-) and the initiation codon of the *gag* gene is deleted (ATG^-gag). Figure is adapted from retroviral vector pBabe (Morgenstern *et al.*, 1990). **c.** To produce infectious virus, proteins that are required for viral reproduction, Gal/Pol and Env, are expressed in a mammalian cell line (viral packaging cell line). The recombinant retroviral vector DNA is transfected into the viral packaging cell line and infectious viral particles are packaged and released from the cells.

deficiency. To produce infectious virus, proteins that are required for viral reproduction, Gal/Pol and Env, are expressed in a mammalian cell line, so called viral packaging cell lines. Once the recombinant retroviral vector DNA is transfected into the viral packaging cell line, infectious viral particles can be produced and released (Fig. 1-6)(Somia, 2004).

1.6.3.3.2.3 Self-inactivating (SIN) retroviral vector

A more recent development is the self-inactivating (SIN) retrovirus that lacks the enhancer or both enhancer and promoter sequences in the integrated provirus. The viral U3 regions of the LTRs possess strong enhancer and promoter activity, which can interfere with the expression of exogenous genes from the internal promoter. Viral enhancers in integrated provirus can activate surrounding cellular genes, such as oncogenes. In some cell lines, these enhancers are targets for epigenetic silencing. A SIN retroviral vector will produce an integrated provirus lacking the viral enhancer and/or promoter. In a typical SIN vector, the enhancer sequence in the U3 region in the viral 3'LTR is removed, while the enhancer in the 5'LTR remains intact. Thus, a full length genomic RNA can be generated by the functional 5'LTR and an infectious viral particle can be produced. However, as a consequence of reverse transcription and second strand synthesis the U3 region of the 5'LTR is copied from the U3 region of the 3'LTR. Thus the integrated provirus will contain the deleted U3 region in both of the 5' and 3'LTRs, leading to an inactivated provirus lacking the enhancer (Fig. 1-7) (Yu et al., 1986, Yee et al., 1987, Soriano et al., 1991).

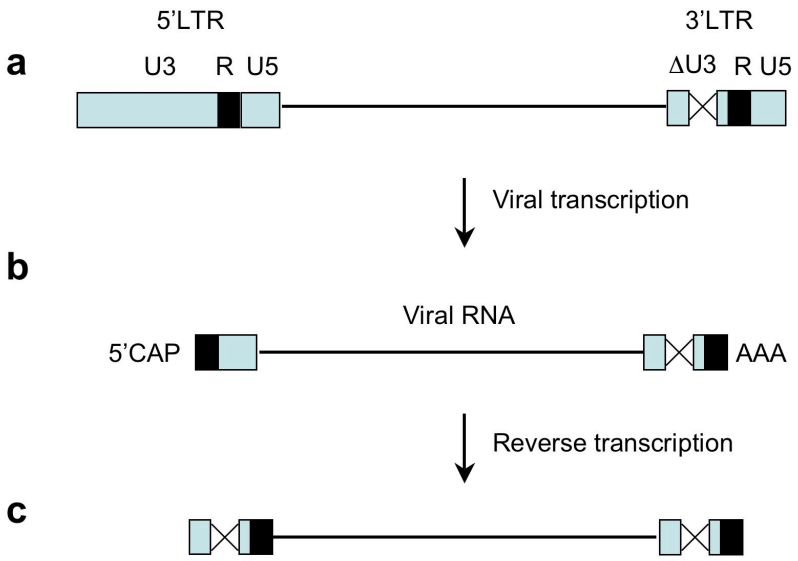


Figure 1-7. Principle of self inactivating (SIN) retroviral vectors.

a. The structure of a recombinant retroviral vector, in which the enhancer fragment in the U3 region of viral 3'LTR is deleted. **b**. Viral RNA is transcribed, initiating from the R region of the 5' LTR and ending at the end of R region of the 3'LTR. **c.** Reverse transcription of viral RNA in host cells generates a linear DNA duplex. The U3 region of the 5'LTR is copied from the U3 region of the 3'LTR in viral genome RNA during this process. The resultant virus will contains a non-functional 5'LTR harboring the deletion, thus it is inactivated.

1.6.3.3.2.4 Retroviral based gene traps

Von Melchner and Ruley developed the first retroviral gene trap vector (von Melchner and Ruley, 1989). In this design, the gene trap cassette is inserted in the U3 region of 3'LTR and replaces the viral enhancer. After viral replication and integration, the provirus carries a duplicated gene trap cassette in both of the 5' and 3'LTRs (von Melchner et al., 1992). Friedrich and Soriano (Friedrich and Soriano, 1991) constructed another version of retroviral gene trap vector, ROSA (reverse orientation splice acceptor) gene trap vector. In this ROSA vector, the gene trap cassette was placed between viral LTRs of a SIN vector in the opposite orientation relative to viral transcription. This reverse orientation was essential in order to avoid removal of the viral packageing sequence Ψ from the full length genomic RNA by splicing from the upstream viral splice donor sequence to the splice acceptor in the gene trap cassette.

Retroviral gene trap vectors can also be made revertible by inserting a *loxP* site into viral U3 region in the 3' LTR. The *loxP* site will be duplicated to the 5'LTR in the integrated provirus, resulting in a provirus flanked by *loxP* sites. By Cre-*loxP* mediated recombination the, *loxP*-flanked provirus can be removed, leaving only a single LTR with a *loxP* site in the genome (Ishida and Leder, 1999).

Gene trap mutagenesis using a retroviral vector has advantages and limitations First, in contrast to electroporation, only a single copy of retrovirus integrates into one genomic locus. Second, by controlling the viral multiplicity, most of cells will contain a single gene trap mutation. Another advantage of this method is that once a stable virus producing cell line is made, large amount of gene trap virus can be produced easily, which significantly improves the throughput of the gene trap mutagenesis method. The major limitation of the retroviral based gene trap method is gene trap "hot spots" caused by non-random retroviral integrations (discussed in the following section).

1.6.3.4 Gene trap "hot spots"

Although gene trap mutagenesis was originally designed as a random method, it has been noticed that some genes appear to be mutated more frequently by a gene trap vector than others (Skarnes, 2000). Recently, the German Gene Trap Consortium (GGTC) has reported a systematic analysis of gene trap "hot spots" by collecting over 10,000 gene-trapped ES clones using four different gene trap vectors, including both electroporation-based and retroviral-based vectors (Hansen et al., 2003). They found that the gene trap insertion sites were dispersed throughout the genome and occurred more frequently in chromosomes with high gene density, which suggests that there is no obvious bias to a single chromosome. 75% of the gene trap mutations appeared only once in the gene trap database, while 25% were "hit" multiple times, suggesting that most genes are accessible to gene trap mutagenesis and "hot spots" (25%) are relatively minor targets. By comparing gene trap "hot spots" arising from different gene trap vectors, they found that some of the "hot spots" (nearly 50%) are common for all vectors, suggesting that the gene trap efficiency could be affected by locus-specific factors. Notably, more than 50% of the hot spots are vector-specific, suggesting that each gene trap vector design will have limited genome coverage. Therefore, it is recommendable to utilize multiple vectors in order to obtain broader genome coverage in gene trap mutagenesis.

The factors that cause gene trap hot spots have not been clearly demonstrated, especially for those specific for individual vectors. Some general factors have been recognized, for example, chromatin structure is expected to affect the gene trap efficiency. Open euchromatic regions that contain transcriptionally active gene are believed to be more permissible to the integration of gene trap vectors. The recovery of cells with gene trap mutations requires that the reporter gene is stably expressed. Therefore, factors that affect the expression or the stability of the gene trap reporter could contribute to a bias of the gene-trap vector, for example, the gene structure and the reading frame. If a fused gene trap

transcript has the endogenous protein translation initiation codon (ATG) before the ATG codon of the gene trap reporter, the translation machinery will prefer the first one and translate a fused protein. In this case, a truncated endogenous protein will be produced if the endogenous ATG and the reporter gene are not in the same reading frame. Consequently, this gene trap mutation will not be recovered. To solve this problem, an IRES (internal ribosome entry site) fragment may be placed between the splicing acceptor and the gene trap reporter. The IRES sequence allows the CAP-independent translation of the reporter from an internal ATG site. Therefore, translation initiation of the reporter will be independent of the reading frame of the trapped gene. In addition, the IRES sequence is able to enhance protein translation, allowing detection of the genes expressed at low levels (Bonaldo et al., 1998).

1.7 Thesis project

The primary goal of the project was to explore the possibility of generating homozygous mutations in *Blm*-deficient mouse ES cells and to investigate the application of a recessive genetic screen for genes involved in MMR surveillance. In this introduction, the function and phenotypic consequences caused by *Blm*-deficiency has been discussed. In mouse ES cells it has been shown that a single allele mutation on an autosomal chromosome can be lost frequently via LOH, generating a bi-allelic (homozygous) mutation. This feature of *Blm*-deficient cells was explored and used as a genetic tool to generate homozygous mutations. Another aim of the study was to identify new MMR components. Although the key players of the MMR system have been identified and their role in repairing DNA replication errors have been studied in detail, knowledge of the MMR system is incomplete, for example, how does the eukaryotic MMR system distinguish the nascent DNA strand in replicating DNA? What is the molecular basis of MMR surveillance? Finally, the knowledge of the MMR system was largely obtained from studies in bacteria and yeast. Although the MMR system seems highly conserved, in higher eukaryotes the MMR system has evolved

more specific functions. Some of these functions have been elucidated (for example in meiosis) while others have not yet been fully defined, for example, the function of MED1/MBD4 in MMR mediated DNA damage surveillance. MED1/MBD4 is a methyl-CpG binding protein. It is notable that this aspect is only found in mammals, but not in yeast, worms or fruit flies since their genomes are deficient in DNA methylation. Therefore, it is important to identify mammalian specific MMR genes. In this regard, performing a genetic screen in a mammalian system for MMR genes is essential to identify elements of this system that can't be identified based on evolutionary conservation.

2.1 Vectors

2.1.1 The Slippage construct, P-Slip

The PGK promoter fragment, including an ATG translation initiation site, was PCR amplified from a PGK-*puro* cassette (YTC 49, a gift from Youzhong Chen) using primers, PGK-5'-*Sal* I and PGK-3'-*Bam*HI, which add *Bam*HI and *Sal*I sites to the ends of the amplified frangment. The *PurobpA* fragment was prepared by PCR from the same PGK-*puro* cassette using PCR primers Puro-5'-*Spe*I and bPA-3'-*Not*I. The PCR fragment was gel purified and digested with *Spe*I and *Not*I to generate a *PurobpA* fragment with *Spe*I and *Sal*I ends. The $(CA)_{17}$ repeat sequence was constructed from oligos with *Bam*HI or *Spe*I ends, then ligated with the PGK promoter fragment and the *PurobpA* fragment into a pBluescript (pBS) plasmid (Stratagene) to create the P-Slip (Puro Slippage) cassette. Using the same strategy, the repeat sequence $(CA)_{16}$ was used to generate the P-Slip-ON plasmid.

Oligonucleotide for $(CA)_{17}$: *Spe*I-$(CA)_{17}$, 5'-CTA GTG TAT C$(TG)_{17}$ TTG and *Bam*HI-$(CA)_{17}$, 5'-GAT CCA A$(CA)_{17}$ GAT ACA.

Oligonucleotide for $(CA)_{16}$: *Spe*I-$(CA)_{16}$, 5'-CTA GTG TAT C$(TG)_{16}$ TTG and *Bam*HI-$(CA)_{16}$, 5'-GAT CCA A$(CA)_{16}$ GAT ACA.

PCR primers for PGK promoter fragment:

PGK-5'-*Sal*I,

5'-ACG CGT CGA CAG GTC GTC GAA ATT CTA CCG GGT AGG GGA GGC GCT TTT

PGK-3'-*Bam*HI,

5'-CGC GGA TCC GTA CTC GGT CCC CAT GGT GGC GTT GGC

PCR primers for *PurobpA* fragment:

Puro-5'-*Spe*I,

5'-GGA CTA GTA AGC CCA CGG TGC GCC TCG CCA CCC G

bPA-3'-*Not*I,

5'-ata aga atg cgg CCg cAG CTG GTT CTT TCC GCC TCA GAA gc

2.1.2 Gene-targeting vectors

Gdf-9-TV1: Gene-targeting construct for *Gdf-9* (growth and differentiation factor 9) locus. This contains a PGK-*Hprt* minigene as the drug selection marker for gene-targeting (Dong et al., 1996).

Gdf-9-TV2: Derived from *Gdf-9-TV1*, in which a *loxP*-flanked PGK-*neo* cassette was inserted into a *Cla*I site in front of the PGK-*Hprt* cassette.

ROSA26/Slip: Gene-targeting construct for the *ROSA26* locus containing the P-Slip cassette. To create the *ROSA26/Slip* gene-targeting vector, a 1.4 kb *Sal*I/*Xho*I fragment from pL313 (a gift from Dr. Pentao Liu) containing a PGK-*BSD* cassette was inserted into the *Sal*I site of the P-Slip vector. A 2.9 kb *Sal*I/*Not*I fragment containing the *PGK/BSD/pSlip* cassette was ligated into the multiple cloning site of a modified pBS vector, which flanks the inserted *PGK/BSD/Slip* cassette with *Nhe*I and *Bgl*II restriction sites. The *PGK/BSD/Slip* cassette was then released by digesting with *Bgl* II and *Nhe* I, and ligated into pROSA26-1 plasmid containing the genomic fragment of *ROSA26* gene and digested with *Nhe*I and *Bgl*II (Zambrowicz et al., 1997).

Dnmt1-V1: A replacement gene-targeting vector for the *Dnmt1* locus assembled using *E.coli* recombination. In *Dnmt1-V1,* a *loxP*- flanked *Neo/Kan* cassette replaces a 5.5 Kb genomic fragment, resulting in deletion of *Dnmt1* exons 2, 3 and 4. This construct was generated by a colleague, Wei Wang.

129Rb-puro: Targeting vector containing a 129Ola-derived retinoblastoma (*Rb*) genomic fragment.

B/cRb-puro: Targeting vector constructed with a BALB/c-derived *Rb* genomic fragment.

2.1.3 Retroviral vectors

pBabeEGFP: A pBabe derivative, containing a *SV40-EGFP* cassette between the pBabe LTRs (a gift from Dr. Xiaozhong Wang).

pBabeOligo: A pBabe derivative, with minimal cis-elements for viral packaging and a multiple cloning sites between pBabeLTRs (a gift from Dr. Xiaozhong Wang).

pLTRloxP: A 1.5 kb *Eco*RI / *Kpn*I fragment containing the pBabe 3'LTR and *SV40-EGFP* cassette was subcloned into pLitmus (Clontech). A *loxP* site flanked by *Xba*I and *Nhe*I sites was synthesized and cloned into the *Xba*I and *Nhe*I sites in the subcloned 3'LTR to create pLTRLoxP.

pBaER: pBabeOligo was linearized using *Eco*RI and then partially digested with *Kpn*I to obtain a 3.5 kb DNA fragment. pLTRloxP was double-digested with *Eco*RI and *Kpn*I to generate the 1.5 kb DNA fragment containing a *SV40-EGFP* cassette and the LTRloxP fragment, which was then ligated with the 3.5 kb DNA fragment from pBabeOligo to create pBaER.

pBaOR: pBabeOligo was linearized by *Bgl*II and then partially digested with *Kpn*I to generate a 3.7 kb DNA fragment. pLTRloxP was digested with *Bgl*II and *Kpn*I to generate a 350 bp LTRLoxP fragment, which was ligated to the 3.7 kb *Kpn*I / *Bgl*II fragment from pBabeOligo to create pBaOR. The SV40 origin was PCR amplified from the pBabepuro vector (Morgenstern and Land, 1990) and cloned into the *Not*I restriction site of pBaOR. PCR primers: *Not*I restriction site underlined
5'Primer_*Not*I: AGA ATG CGG CCG CTT TTT GCA AAA GCC TAG
3'Promer_*Not*I: AGA ATG CGG CCG CGA CCC TGT GGA ATG TGT G;
PCR cycling conditions: 94 °C 30 seconds, 58 °C 30 seconds, 72 °C 1 minute, for 30 cycles.

pCbOR: To generate a retroviral vector with a CMV enhancer and promoter, a 1.2 kb *Xba*I/*Cla*I fragment containing the 5'LTR from pBaOR was first subcloned into pBS. The 110 bp promoter region was replaced using an oligonuclotide containing *Xba*I and *Hind*III sites. This modified 5'LTR was digested with *Xba*I and *Cla*I, and cloned back into pBaOR to replace the original 5'LTR. A 1.6 kb fragment spanning the viral 5' LTR to the end of the 3' LTR was amplified by PCR. The PCR product was digested with *Hind*III and *Apa*I and cloned into *Hind*III and *Apa*I digested pcDNA3-EGFP vector (a gift from Dr. Xiaozhong Wang) to place the viral backbone following CMV promoter.
PCR primers: 5'primer: CGG TCC AGC CCT CAG CAG;
3'Primer_*Apa*I: CGG GGC CCT GAT ACA TGC TGC ATG TG
PCR condition: 94 °C 1.5 minutes, 57 °C 3 seconds and 68 °C 2 minutes for 25 cycles using the Roche Expand Long PCR System (Roche).

pBaERneo: A *Xho*I and *Bam*HI–flanked DNA fragment containing a PGK-*neo* cassette was cloned into pBaER digested with *Sal*I and *Bam*HI.

pBaORneo, pCBaORneo: A *Xho*I and *Bam*HI–flanked DNA fragment containing a PGK-*neo* cassette was cloned into pBaOR and pCBaOR (respectively) digested with *Xho*I and *Bam*HI.

pBeGTV, pCbGTV (RGTV-1): Gene trap retroviral vectors containing *SAβgeo* gene trap cassette. To introduce the *SAβgeo* cassette into the retroviral backbone so that the transcription of *βgeo* is reversed in relation to the viral transcription, a *Sna*BI restriction site was introduced into pSAβgeo (Friedrich and Soriano, 1991) between the *Sal*I and *Kpn*I restriction sites which follow the polyA signal sequence, and the *Xho*I site between *Sal*I and *Kpn*I was deleted. The *SAβgeo* cassette was then obtained by *Xho*I and *Sna*BI double digestion and cloned into *Eco*RV/*Sal*I digested pBaER or *Eco*RV/*Xho*I digested pCbOR to create pBeGTV and pCbGTV, respectively.

2.2 Cell culture

2.2.1 ES cell culture

ES cell culture has been described in detail (Ramirez-Solis et al., 1993). Briefly, ES cells were maintained on mitotically inactivated feeder cell layers (SNL76/7) in standard M15 medium (Knockout Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 15% foetal bovine serum (FBS), 2 mM L-Glutamine, 50 units/ml Penicillin, 40 μg/ml Streptomycin and 100 μM β-Mercaptoethanol (β-ME) (Invitrogen)). Cells were cultured at 37 °C with 5% CO_2. If not specified, ES cell medium was changed daily.

2.2.2 Chemicals used for selection in ES cells

Blasticidin, Blasticidin S HCl (Invitrogen), 1000x stock (2.5 mM) was made in Phosphate Buffered Saline (PBS).
FIAU, 1-(2'-deoxy-2'-fluoro-β-D-arabinofuranosyl)-5-iodouracil, 1000X stock (200 μM) was made in PBS and 5 M NaOH was added dropwise until it dissolved.
G418, Geneticin (Invitrogen), was bought as liquid containing 50 mg active ingredient per milliliter.
Puromycin, ($C_{22}H_{29}N_7O_5$.2HCL, Sigma) 1000X stock (3 mg/ml) was made in MiliQ water.
50X HAT supplement (Hypoxanthine-aminopterin-thymidine) (Invitrogen).
100X HT supplement (Hypozanthine-thymidine) (Invitrogen).
6TG: 2-amino-6-mercaptopurine (Sigma) 10 mM stock was made in PBS and 5 M NaOH was added dropwise until it dissolved.
.

2.2.3 Transfection of DNA into ES cells by electroporation

DNA to be used for electroporation was either prepared by the standard alkaline lysis method followed by the CsCl banding purificaion or was prepared using QIAgen Plasmid Purification Kits (QIAGEN). Before electroporation, DNA was purified by ethanol precipitation and air-dried in a tissue culture (TC) hood. If DNA linearization was required, for example for gene-targeting, plasmid DNA was first digested by an appropriate restriction enzyme before ethanol precipitation. The air-dried DNA was dissolved in TC hood in Tris-Cl (1 mM, pH 8.2) to a final concentration of 0.5-1 μg/ μl and 20 μl was used for each electroporation.

ES cell electroporation has been described in detail (Ramirez-Solis et al., 1993). Briefly, $1x10^7$ ES cells were eletroporated in a 0.4 cm gap cuvette with 10 to 20 μg DNA using Biorad Gene Pulser at 230 V, 500 μF. After electroporation, ES cells were plated onto a 90 mm feeder plate and followed by an appropriate drug selection as needed. If not specified, the drug selection procedure followed the description in (Ramirez-Solis et al., 1993).

2.2.4 Rb-targeting using Isogenic and non-isogenic gene targeting vectors

The Rb-targeting vectors, Rb129Rb-puro and B/cRb-puro were linearized with *Hind*III and electroporated into ES cells and selected with Puromycin (3 μM) for 8 days. 96 Puromycin resistant clones from each cell line were picked and genomic DNA was extracted and digested with *Eco*RI for Southern analysis using a *Rb* probe for gene-targeting events (expected sizes of detected bands were 9.7 kb for wild type and 4.7 kb for targeted). Targeting efficiency was determined as the number of targeted clones versus the number of samples exhibiting the 9.7 kb wild type band on the Southern-blot.

2.2.5 Gene-targeting of *ROSA26/Slip-TV1* construct

10 μg of *ROSA26/Slip-TV1* DNA was linearized with *Kpn*I restriction enzyme and electroporated into ES cells. Blasticidin selection was initiated 48 hours post electroporation and continued for 4 days. At this stage feeder cells and most of the ES cells were dying and detaching from the bottom of the plate. Fresh M15 medium with $2x10^6$ mitoticaly inactivated SNL76/7 feeder cells were added to the plates. The surviving ES cells were allowed to grow in fresh M15 medium for 6 days to form ES cell colonies. 48 ES cell clones from each targeted cell line were picked into 96 well tissue culture plates and expanded for further analysis. For Southern analysis, genomic DNA was extracted and digested with *Eco*RI and hybridized with a *ROSA26* probe.

2.2.6 Cre-mediated recombination

20 μg of Cre-expressing plasmids pOG231 (CMV-Cre from Steve O'Gorman) or pCAAG-Cre (Araki et al., 1995) was electroporated into 0.5 -$1x10^7$ ES cells. After electroporation, cells were diluted in M15 and about 1,500 ES cells were plated onto a 90 mm feeder plate and cultured for 10 days to allow the formation of single ES cell colonies. 96 ES cell clones were picked into a 96 well tissue culture plate. To identify Cre-mediated recombination events, cells cultured in 96 well plates were duplicated into several 96-well tissue culture plates and sib-selection was performed to identify ES clones that lost the drug selection marker flanked by *loxP* sites. The revertant clones were expanded and loss of the drug selection marker was confirmed by Southern analysis.

2.2.7 Clonal survival assay

Cells were seeded at clonal density (200 to 250 cells per well) in one well of a 6-well tissue culture plate, with or without drug selection. 10 days later, the ES

clones were stained with 2% methylene blue in 70% ethanol for 5 minutes and the number was counted.

2.2.8 EMS mutagenesis in ES cells

EMS (Ethyl methanesulphonate) was purchased from Sigma (1.17 g/ml). A stock (20 mg/ml) was made in PBS and diluted to its final concentration immediately before treatment of ES cells. ES cells at approximately 50% confluence were fed with EMS supplemented M15 medium for 16 hours, then cells were washed three times with PBS, and re-fed with fresh M15. Three hours later, cells were trypsinized and counted. A small portion of cells were diluted to low density and plated onto one well of a 6 well tissue culture plate to determine the survival rate. The survival rate was determined by comparing the plating efficiency of EMS treated cells with non-EMS treated cells of the same genotype. The remaining cells were passaged onto fresh feeder plates.

2.2.9 Retroviral approaches

2.2.9.1 Producing retrovirus by transient transfection

Cells for transfection: The Phoenix ecotropic retroviral packaging cell line, a derivative of human embryonic kidney 293T line expressing retroviral gal, pol and env proteins, was obtained from the American Tissue Culture Collection (ATCC, Manassas, Virginia, USA). In general, Phoenix cells were cultured in M10 medium (DMEM supplemented with10% FBS, 2 mM L-Glutamine, 50 units/ml Penicillin, 40 μg/ml Streptomycin) at 37 °C with 5% CO_2. 18 hours prior to transfection, Phoenix cells were plated at a density of 2.1 million cells per 90 mm plate in M10. Three hours before transfection, cells were fed with 9 ml fresh M10 medium (at this time the cells were about 60% confluent).

DNA preparation: DNA for transfection was prepared by QIAgen Plasmid Purification Kit (QIAGEN). 20-25 μg DNA was used for one transfection of cells cultured on each 90 mm plate. DNA was precipitated with ethanol and air-dried in a TC hood then dissolved in 20 μl TE.

Transient transfection: 500 μl of $CaCl_2$ (0.25 M) was added to DNA and mixed. 500 μl HEBS (0.28 M NaCl, 0.05 M HEPES, 1.5 mM Na_2HPO_4, pH 7.10) was added to the DNA, mixed quickly by bubbling vigorously with an automatic pipettor for 5 seconds. The DNA mixture was kept at room temperature for 5 minutes, and then added to cells cultured in 9 ml of M10 medium. 24 hours later, cell medium was removed and 5 ml of 1% DMSO in PBS was added. After 2 minutes at room temperature, cells were washed twice with PBS and 10 ml of viral producer medium (M10 with heat-inactivated FBS) was layered on each plate. Viral supernatant was harvested 48, 60 and 72 hours after transfection.

2.2.9.2 Viral Infection

ES cells were plated at a density of $3X10^6$ per 90 mm feeder plate about 18 hours before infection. The viral supernatant was filtered through 0.45 μm filter, and polybrene (Hexadimethrine Bromide, Sigma, H9268) was added to the viral supernatant to a final concentration of 4 μg/ml, then added to ES cells.

2.2.9.3 Determination of the transfection efficiency

48 hours post transfection, the viral producer cells were trypsinized and collected in M10. Cells were pelleted and resuspended in PBS at a final concentration of $1x10^6$ cells per ml for flow cytometric analysis of EGFP expression.

2.2.9.4 Titration of the retrovirus

ES cells were plated in 6-well tissue culture plates at a density of $0.5x10^6$ per well in 3 ml M15 medium. 24 hours later, viral supernatant was applied. For the virus carrying a *Neo* cassette, G418 selection (180 μg/ml) was initiated 24 hours after viral infection and continued for 8 days. The drug-resistant ES colonies

were stained with 2% methylene blue in 70% ethanol and counted. The titer of the retrovirus is the number of drug resistant ES cell colonies per milliliter of viral supernatant used to infect the cells.

2.2.10 Gene trap mutagenesis and 6TG resistance screen

2.2.10.1 Gene trap mutants by RGTV-1 retrovirus on NGG5-3 cells.

RGTV-1 retrovirus was produced by transient transfection of Phoenix viral packaging cells. 400 ml of viral supernatant was harvested and filtered through 0.45 μm filter. NGG5-3 ES cells were plated on seventeen of 90 mm feeder plates at a density of $2.5x10^6$ cells per plate. 24 hours later, cells were infected with 5 ml of viral supernatant for at least 5 hours. Viral infection was repeated 5 times. G418 selection (180 μg/ml) was initiated 48 hours after first infection and continued for 8 days. One plate was stained by 2% methylene blue in 70% ethanol to determine the number of gene trap clones obtained. The G418 resistant ES cells from the other 16 infected plates were collected by trypsinization, and cells from two plates were combined and plated to 90 mm feeder plates, generating a total of 8 pools. These cells were cultured for 4 days and frozen down for the subsequent selection in 6TG.

2.2.10.2 Screen for 6TG resistant gene trap mutants

Gene trap mutants which had been expanded for more than 14 population doublings were plated on 90 mm tissue culture plates at a density of 0.7-1 $x10^7$ cells per plate. For high stringency 6TG selection, 6TG selection (2 μM) was initiated 16 hours later and the 6TG-supplemented M15 medium was changed every day for 8 days. After culturing 4 days in fresh M15 medium, the 6TG resistant colonies were picked. For low stringency 6TG selection, cells were plated directly in 6TG (0.5 μM) -supplemented M15 medium which was changed

every other day for 10 days. After culturing in fresh M15 for 2 days, surviving ES cell clones were picked.

2.2.11 Fluctuation analysis of the MSI rate of P-Slip.

ES cell lines with the targeted P-Slip cassette were plated at single cell density on 90 mm tissue culture plates and cultured for 10 days to allow formation of single ES cell colonies. ES cell clones were picked to a 96 well tissue culture plate and independently expanded to the desired number of cells. The expanded clones were trypsinized and the number of cells in each clone were determined using a Beckman-Coulter blood cell counter, then selected in M15 medium containing Puromycin (3 μM) for 8 days. Puromycin-resistant ES cells were allowed to grow and visible colonies were stained with 2% methylene blue in 70% ethanol. The colony number was counted and the MSI rate of P-Slip was calculated by the Luria-Delbruck method of means with equation: $r = aN \ln(NCa)$, where "a" represents the mutation rate "r" represents the mean number of variation per culture (puromycin resistant clones); "C" represents the number of parallel cultures (luria, 1943).

2.3 DNA methods

2.3.1 Probes

2.3.1.1 General probes

***LacZ* probe**: A probe for gene trap viruses containing the *SAβgeo* gene trap cassette, consisting of a 1.4 kb *Cla*I fragment from pSAβgeo, a plasmid containing the *SAβgeo* cassette in pBS (from Dr. Philippe Soriano).
***Neo* probe**: A probe for gene trap viruses containing the SAβgeo gene trap cassette and consisting of a 700 bp *Pst*I /*Xba*I fragment from the PGK-*Neo* cassette.

***Rb* probe**: 450bp *Pst*I/*Pvu*II fragment from pPHA153 (a gift from Dr. Hein te Riele).

***Gdf-9* probe**: 650bp *Bam*HI/*Sal*I fragment from pGDF9-212D (a gift from Dr. Martin Matzuk).

γ*SAT* probe: Probe for paracentromeric gamma satellite repeats. 1.9kb *Not*I and *Sal*I fragment from pγSAT plasmid (a gift from Niall Dillon).

2.3.1.2 PCR amplified genomic DNA probe

Genomic DNA probes were PCR amplified from AB2.2 mouse genomic DNA and used for Southern-blot analysis.

Msh6 exon2 (F), 5'-GCAACAGTTCTTGTGACTTCTCACCA
(R), 5'-CCTCTTACCTGTATATGGCTTTAACAT, 180 bp

Dnmt1 (F), 5'-GCAGTTTGTTTAAATAGAAGTGTGCATAGT
(R), 5'-GTCCCCTAACACATACCTTCGTGTAT, 685 bp

Tgif (F), 5'-CGCCAGCGCGCTCCGACTTCTTAACT
(R), 5'-GAGCAGCGACGTCACCGCCGGTG, 1.1kb

Rbpsuh (F) 5'-GAATTCCCTTATCTCTAAAAGGAGCATAT
(R) 5'-GACTCCACATTAACACAGAGATGTTAAG, 721 bp

mMRG9 (F), 5'-CGACTGTGGGCCGAAGGTTCGAGGCTGT
(R), 5'-CCGCCTGTCCTTGTACATCGATTAATTAAAACGT, 900 bp

ROSA26 (F) 5'- CTGGATCCTCCCCAATCAAAAGTATAGG
(R) 5'- CTCCCTGTGGCGTATGCCCCAGTATCC, 660 bp

STA1.2 (F), 5'-GAGAGGTCACCATTATTTCTAGAATGGCCTA
(R), 5'-CTGAAGAAATACAGCCTGGATATCCACAGCT, 1.2 kb

AldpS (F), 5'-GCTTCCCAAGTGCTGGGATTAAAGGTATGTGT
(R), 5'-CAGGGTACTGCAGCAAAGGAGCCCAGGT

AldpL (F), 5'-CCATTCAGGACAACCACAGAGTACTGGATCA
(R), 5'-CTCATGTGAGTATATGGACATGTAAGTTGGGTAT

2.3.1.3 cDNA probe

cDNA probes were PCR amplified from AB2.2 cDNA and used for Northern-blot analysis or Southern-blot analysis.

Msh6 exon1 (F), 5'-CGTCAGCCTTCATCATGTCCCGACAA
exon4 (R), 5'-GGTGCCTAGGTCCTCACTATC, 876 bp
Msh2 exon1 (F), 5'-GCAGCCTAAGGAGACGCTGCAGTTG
exon18 (R), 5'-CCGTGAAATGATCTCGTTTACGAAGCTG, 2.8 kb
Mlh1 exon1 (F), 5'-GGCGTTTGTAGCAGGAGTTATTCGG
exon19 (R), 5'-AGACTTTGTATAGATCTGGCAGGTTGGC, 2.3 kb
Dnmt1 exon1 (F), 5'-GCTCCAGCCCGAGTGCCTGCGCTTG
exon6 (R), 5'-CTCTGTGTCTACAACTCTGCGTTTc, 543 bp

2.3.2 Southern blotting and hybridization

1). Southern blotting: Genomic DNA was digested with an appropriate restriction enzyme and the digested fragments were separated by electrophoresis on 0.8% agarose gel in 1XTAE buffer. The gel was soaked in Depurination Buffer (0.25 M HCl) for 10 minutes with gentle agitation, and then transferred into Denaturation Buffer (0.5 M NaOH, 1.5 M NaCl) for 30 minutes with gentle agitation. A capillary blot was set up according to standard methods, using Hybond-N+ membrane (Amersham) using Denaturation Buffer as the transfer buffer. Following overnight transfer, the blot was neutralized in Rinse Buffer (0.2 M Tris-Cl (pH7.4), 2X SSC) for 5 minutes, and baked at 80 °C for 1 hour.

2). Probe preparation: 5-20 ng of probe DNA was labeled using Rediprime^{TM} II Random Prime Labeling System (Amersham) according to the manufactor's protocol and purified with a G-50 column. The probe was denatured at 100 °C for 5 minutes, and chilled on ice for 5 minutes before use.

3). Hybridizaiton: The blot was pre-hybridized at 65 °C for 2 hours in Hybridization Buffer (1.5X SSCP, 1X Denhardts solution, 0.5% SDS, 10% Dextran Sulfate). The denatured probe was added into the hybridization tube and incubated at 65°C for at least 8 hours. The blot was first rinsed briefly in low stringency wash buffer (1X SSC, 0.1% SDS) at room temperature and then washed twice in high stringency wash buffer (0.5X SSC and 0.1% SDS) at 65 °C for 15 minutes. The blot was then exposed to X-ray film (Fuji).

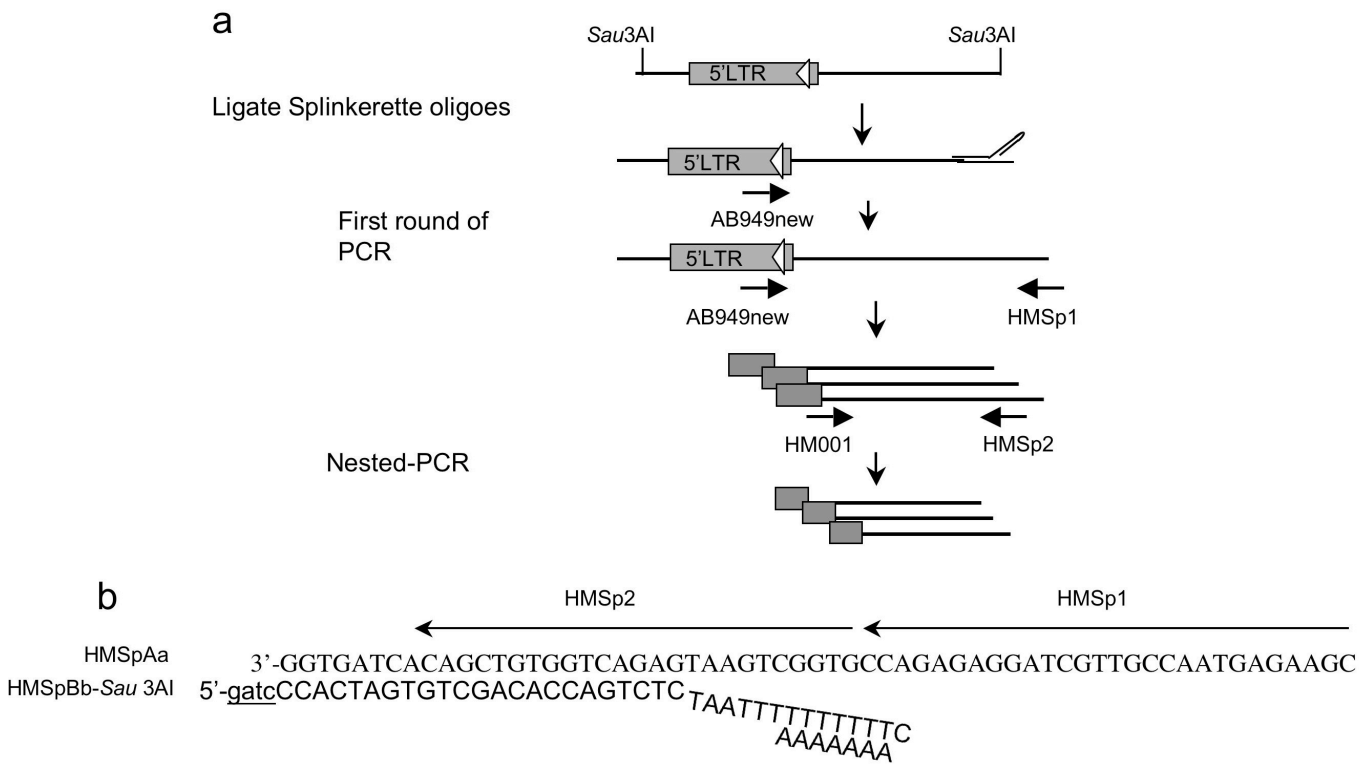


Figure 2-1. Splinkerette PCR. **a.** Schematic representation of Splinkerette PCR procedure**.** The *Sau*3AI digested DNA is ligated with an annealed Splinkerette oligoes (Splinkerette adaptor). The first round pf PCR is carried out with viral primer AB949new and Splinkerette primer HMSp1. The nested PCR is carried out with viral primer HMOO1 and Splikerette primer HMSp2. **b.** The structure of an annealed Splinkerette Oligos, showing the *Sau* 3AI site (under lined font), the loop structure and the position of the HMSp1 and HMSp2 primers.

2.3.3 Isolation of proviral/host junction by Splinkerette PCR

3 μg of genomic DNA was digested with *Sau*3AI in a 30 μl volume at 37 °C for 3 hours. The reaction was stopped by heating at 65 °C for 20 minutes. The Splinkerette Oligos were annealed in a reaction mix containing 150 pmol of HMSpAa-*Sau3* AI, 150 pmol of HMSpBb, 5 μl NEBuffer 2 (New England Biolabs) in a total of 100 μl. After a short incubation at 95 °C for 3 minutes, the mixture was cooled slowly to room temperature. 3 μl of the annealed Splinkerette oligos were used for each ligation reaction in 20 μl volume containing 5 μl digested genomic DNA, 2 μl 10X Ligation Buffer and 5 units T4 DNA Ligase (New England Biolabs). The ligation reaction was carried out at 16 °C overnight. The T4 DNA ligase was then heat-inactivated at 65 °C for 15 minutes. 20 μl of *Cla* I digestion mix containing 10 units *Cla*I (New England Biolabs), 4 μl 10X *Cla*I Buffer (New England Biolabs), 14 μl H_2O was added to the ligation reation and incubated at 37 °C for 2 hours. Unligated oligos were removed by gel filtration using Sephacryl^{TM}S-300 (Amersham) as described in section 2.4.4.

First round of PCR: The 5'LTR proviral junction was amplified with the LTR specific primer, AB949new, and the Splinkerette primer, HMSp1 in 50 μl PCR system containing 20 μl purified genomic DNA, 2 μl AB949new (10 μM), 2 μl HMSp1 (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen), ddH_2O 18 μl. The PCR cycling was performed at 94 °C 2 minutes, cycles of 94 °C 1 minutes, 68 °C 30 seconds, 72 °C 1.5 minutes and followed by 30 cycles of 94 °C 30 seconds, 65 °C 30 seconds, 72 °C 2 minutes and finished by 72 °C incubation for 10 minutes. The first round of PCR product was diluted at a 1 to 200 ratio and 5 μl of the diluted product was used as the template for the nested PCR with primers, HM001 and HMSp2, in 50 μl PCR system containing 1 μl HM001 (10 μM), 1 μl HMSp2 (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen), ddH_2O 35 μl. PCR cycling was performed at

94 °C 1.5 minutes, 30 cycles of 94 °C 30 seconds, 60 °C 30 seconds, 72 °C 1.5 minutes and finished by 72 °C incubation for 7 minutes. The nested PCR products were separated on a 1% agarose gel. The specific PCR fragments were gel purified using QIAquick Gel Extraction Kit (QIAGEN) according to the manufacture's instructions. 5-20 ng of the purified DNA was used for sequencing using two sequencing primers.

To obtain longer provial/host juction fragments, genomic DNA was digested with restriction enzyme *Eco*RI*, Hind*III, *Xba*I*, Spe*I*, Nhe*I*, Bam*HI *or Nco*I. The Splinkerette adapters were made by annealing the related HMSpAa and HMSpBb oligos and ligated to the digested genomic DNA (Fig. 2-1b). The same Splinkerette PCR and LTR primers used for amplifying *Sau*3AI digested genomic DNA were used to amplify the proviral/host junction following the above protocol, with 3 minutes of PCR elongation time.

Splinkerette Oligos:
HMSpAa: 5'-CGA AGA GTA ACC GTT GCT AGG AGA GAC CGT GGC TGA ATG AGA CTG GTG TCG ACA CTA GTG G
HMSpBb-*Sau*3AI
5'-gat cCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$
HMSpBb-*Hind*III
5'-agc tCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$
HMSpBb -*Nco*I
5'-cat gCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$
HMSpBb –*Xba*I
5'-cta gCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$
HMSpBb -*Eco*RI
5'-aat tCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$
HMSpBb -*Bam*HI
5'-gat cCC ACT AGT GTC GAC ACC AGT CTC TAA $(T)_{10}C(A)_7$

PCR primers:
AB949new: 5'-GCT AGC TTG CCA AAC CTA CAG GTG G
HM001: 5'- GCC AAA CCT ACA GGT GGG GTC TTT
HMSp1: 5'-CGA AGA GTA ACC GTT GCT AGG AGA GAC C
HMSp2: 5'-GTG GCT GAA TGA GAC TGG TGT CGA C
Primers for sequencing:
HM002: 5'-ACA GGT GGG GTC TTT CA; HMSp3: 5'-GGT GTC GAC ACT AGT GG

2.3.4 Cre-*loxP* mediated reversal assay of integrated retrovirus

Cre-mediated recombination was performed as described in Section 2.2.6. To identify clones which had undergone Cre-mediated excision, PCR was performed on 200 ng of genomic DNA to amplify the SAβ*geo* gene trap cassette using the primers lacZ(F) and LacZ(R), which amplify a 335 basepair *LacZ* fragment. A pair of PCR primers, Ctbp2 (F) and Ctbp2 (R), were included in the PCR reaction as a control, which amplified a 490 basepair fragment from *CtBP2* (C-terminal binding protein 2) (Fig. 5-11a). Clones which did not show amplification of the *LacZ* fragment were expanded and the clonal survival were checked in 6TG (See Section 2.2.6). The primer sequences:
LacZ (F), 5'- CGA ATA CGC CCA CGC GAT GGG TAA CA
LacZ (R), 5'- CGC TAT GAC GGA ACA GGT ATT CGC TGG T
Ctbp2 (F), 5'-CTC GCC AGC AGC CTT GAT GTC CAC GTT GT
Ctbp2 (R), 5'-CCT GGT GGC ACT GCT GGA TGG CAG AGA CT.
PCR cycling was performed at 94 °C for 2 minutes; 30 cycles of 94 °C 30 seconds, 58 °C 30 seconds, 72 °C 1 minute and finished by incubation at 72 °C for 7 minutes.

2.3.5 Quantitative Southern analysis (QTSouthern)

2 to 4 μg of genomic DNA was digested using 20 units each of *Eco*RV and *Hind*III (New England Biolabs) in 40 μl, and separated on a 0.8% agarose gel. Southern blotting and hybridization were performed as described in section 2.3.2. 20 ng of each probe, *AldpS, AldpL* and *LacZ* was used for each hybridization. *AldpS* and *AldpL* probes were PCR amplified from AB2.2 genomic DNA for X-linked gene Adrenoleukodystrophy Protein Homolog using PCR primers AldpL(F), AldpL(R), AldpS(F), AldpS(R). *AldpS* recognizes a 2.5 kb *Hind*III fragment. *AldpL* recognizes a 3.7 kb *Eco*RV/*Hind*III fragment. The *lacZ* probe recognizes a 3.0 kb *Eco*RV/*Hind*III fragment from the gene trap cassette.

2.3.6 Sequencing analysis of the recovered P-Slip cassette

Individual Puromycin resistant ES colony was picked into a 96 well tissue culture plate, expanded, and genomic DNA was extracted. 50 μl 0.1xTE Buffer (0.1 mM EDTA, 1 mM Tris-Cl (pH 8.0)) was added to each well and incubated at 55 °C overnight to dissolve the genomic DNA. 5 μl dissolved genomic DNA was used for PCR reaction to amplify the 1.1 kb DNA fragment containing the $(CA)_{17}$ repeat sequences. The PCR reaction was performed with primers, PGK (F) and Rosa3'arm, in a total of 50 μl using the Expand Long Template PCR System (Roche) following the manufacturer's instructions. 15 μl of PCR product was treated with 2 units of Exonuclease I (*Exo*I, NEB) and 2 units of Shrimp Alkaline Phosphatase (SAP, Amersham) at 37 °C for one hour to degrade the single-stranded PCR primers and destroy the dNTPs. The enzymes were inactivated at 95 °C for 15 minutes, and 5 μl was used for sequencing with the PCR primers. Underlined is the *Eco* RI restriction site.

PGK (F) 5'-<u>cgg aat tc</u> G GGC AGC GGC CAA TAG CAG CTT TGC T

Rosa3'arm: 5'-<u>cgg aat tc</u>G ATA GAA CTT GAT GTG TAG ACC AGG CTG G

2.3.7 Determination of the CpG methylation pattern

2 μg genomic DNA was digested with 20 unit of *Msp*I or *Hpa*II (New England Biolabs) at 37 °C overnight. The digested DNA was separated on a 1% agarose gel and blotted onto Hybond-N+ membrane (2.3.2). Southern hybridization was carried out with the gamma satellite probe γSAT (2.3.1.1) in Rapid-Hyb Buffer (Amersham Bioscience) according to the manufacturer's instructions.

2.4 RNA methods

2.4.1 RT-PCR

Total RNA was prepared using an RNAqueous™ Kit (Ambion). 5 μg of total RNA was treated with 1 μl amplification grade DNase I (1 unit/μl, Invitrogen) in 10 μl volume for 10 minutes at room temperature (RT) to degrade the residual DNA. 1 μ1 of EDTA (25 mM) was added to the reaction, which was incubated at 65 °C for 15 minutes to inactivate the DNase1. 1 μl of Oligo-dT primer (10 μM) was added to the reaction, followed by incubation at 65°C for 10 minutes. The tube was placed on ice. First strand reaction mix (0.5 μl dNTPs (25 mM), 5 μl 5x first-strand buffer, 2.5 μl DTT (0.1 M), 1 μl Superscript™ II (5 units/μl)) was made and aliquoted into each sample, followed by an incubatation at 50 °C for 60 minutes. 1 μl of Ribonuclease H (2 U/μl) was added to each reaction and incubated at 37 °C for 30 minutes to remove RNA. The resultant cDNA was diluted at a ratio of 1:5 with ddH_2O and 2 – 5 μl was used for each PCR reaction. The PCR cycling conditions were: 94 °C for 2 minutes; 30 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds, 72 °C for 1 minute; 72 °C for 7 minutes. dNTPs, Superscript™ II, Ribonuclease H and Platium Taq were purchased from Invitrogen.
Oligo-dT primer: GGC CAC GCG TCG ACT AGT AC $(T)_{17}$

2.4.2 RT-PCR primers

Dnmt1 **exon1 (F), 5'-GCT CCA GCC CGA GTG CCT GCG CTT G**
exon6 (R), 5'-CTC TGT GTC TAC AAC TCT GCG TTT C

Tgif

Tgif-α, exon1 (F) 5'-GCC ACT CCA CGG CTG CTG GCT CCT

Tgif-β, exon1 (F), 5'-GAG CTG AGG GAT GGA GAT GGT GCT CT

Tgif-γ, exon1 (F), 5'-CTG CCT CGA AAA GAT TTA TGC GAG CAG A
exon3 (R), 5'-TCT CAG CAT GTC AGG AAG GAG CCT G

Parp-2, exon1 (F), 5'-GCA GAG ATC AGG CTC TGG AAG GCG A
exon5 (R), 5'-GTG CTG GCA GCA TAG TCC ATC TGT A

Rbpsuh exon1 (F), 5'-CTC AGT CTC CAC GTA CGT CCC CGA G
exon4 (R), 5'-CAG AAC ATC CAT CTC GTT CCA TTT GCT CT

mMRG9 exon1 (F), 5'-GCG TCT GAC GCT GAG TTG GGT
exon6 (R), 5'-CCT CTC ATC TTG CCC TCT GCA

LacZ-Gsp2 (R): 5'- atg tgc tgc aag gcg att aag

2.4.3 Northern blotting and hybridation

RNA gel electrophoresis: Total RNA was prepared using the RNAqueous™ Kit (Ambion). 10-20 μg total RNA was mixed with 10 μl RNA loading dye, denatured by heating at 70 °C for 15 minutes and chilled on ice for 5 minutes. RNA was then separated by electrophoresis on a formaldehyde-agarose gel (1% agarose, 3% formadehyde, 1X MOPS solution in DEPC-treated H_2O) in RNA gel running buffer (1X MOPS in DEPC-treated H_2O).

Northern blotting: The gel was soaked in DEPC-treated H_2O for 30 minutes and then in 10X SSC for 30 minutes. RNA was blotted to Hybond-N+ membrane (Amersham) using standard capillary transfer method in 10X SSC for 12 hours. The blot was rinsed in 6X SSC for 5 minutes and baked at 90 °C for 1 hour. 20 ng of probe DNA was labeled using Rediprime™ II Random Prime Labeling System (Amersham) according to the manufacturer's protocol and purified using

a G-50 column. The labeled probe was denatured by heating to 100 °C for 5 minutes and then chilled on ice for 5 minutes before use. Hybridization was carried out in Rapid-hyb Buffer (Amersham) according to the manufacturer's instructions for 4 hours. The blot was washed in the low stringency wash buffer (1X SSC and 0.1% SDS) at room temperature for 5 minutes, and then washed once in high stringency wash buffer (0.5X SSC, 0.1% SDS) at 65 °C for 10 minutes. The blot was then exposed to X-ray film (Fuji).

2.4.4 Isolation of the 5' end of gene trap transcripts by 5'RACE

1). First strand cDNA synthesis: First strand cDNA synthesis was performed as described in Section 2.4.1 using a *lacZ* primer, LacZ-GSP1. cDNA was purified using a QIAquick PCR Purification Kit (QIAGEN), and eluted in 30 μl Tris-Cl (10 mM, pH8.2). (Note: If the cDNA synthesis was performed on samples in a 96-well plate, the resultant cDNA was purified using Sephacryl^{TM}S-300 following the protocol below).

2). Sample purification and size selection: Sephacryl^{TM}S-300 Media (Amersham Bioscience) was mixed at a 1:1 ratio with MilliQ water. 200 μl of this mixture was added to each well of a 0.2 μm PVDF filtration plate (Corning Inc.). The filtration plate was placed onto a collection plate and spun for 2 minutes at 600 g, and the flow-through was discarded. This step was repeated once. 200 μl of H_2O was added to each well of the filtration plate, spun for 2 minutes at 600 g, and then the flow-through was discarded. This step was repeated once. cDNA samples were loaded on to the filtration plate. This was placed onto a fresh collection plate and the samples were recovered by spinning at 800 g for 2 minutes.

3). PolyC tailing: On ice, 30 μl purified-cDNA was mixed with 10 μl of the tailing-mixture, containing 8 μl 5xTdT buffer, 2μl dCTP (4 mM), and 1 μl TdT enzyme The samples were kept at 37 °C 10 minutes.

4). First round of PCR: On ice, 10 μl of the tailed-cDNA was mixed with the PCR reaction mixture containing 5 μl 10X PCR buffer, 1.5 μl of $MgCl_2$ (50 mM),

0.5 μl dNTP (25 mM), 1μl LacZ-GSP2 (10 μM), 1 μl AAP (10 μM), 0.5 μl of PlatinumTaq (5 units/μl, Invitrogen) 30.5 μl MilliQ H_20. The PCR reaction was performed using the following conditions: 94 °C for 1.5 minutes; 35 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 72 °C 1.5 minutes; 72 °C for 10 minutes.

5**). Nested-PCR:** First-round PCR products were diluted at a ratio of 1:100. 5 μl of the diluted PCR product was added to a PCR reaction mixture containing 5 μl 10X PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 0.5 μl dNTP (25 mM), 1 μl LacZ-GSP3 (10 μM), 1 μl AUAP (10 μM), 0.5 μl Platinum Taq (5 units/μl, Invitrogen), 35.5 μl MiliQ H_20. PCR was performed using the following conditions:
94 °C for 1.5 minutes; 35 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 72 °C 1.5 minutes; 72 °C for 10 minutes. 10 μl of the nested PCR products were loaded on a 1.0 % agarose gel. Samples with multiple PCR bands were gel purified and each band was cloned using TOPO TA Cloning kits (Invitrogen).

6). Sequencing the nested-PCR product: If the nested-PCR was performed on a small scale, the nested-PCR product was purified using a QIAquick PCR Purification Kit following the manufacturer's instructions and eluted in 30 μl of Tris-Cl (10 mM, pH 8.0). 5 μl was used for each sequencing reaction in a 10 μl volume, containing 4 μl of ABI PRISM^{TM} Big DyeTerminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems), 1 μl of SA-seq (5 μM)).
Sequencing reaction was performed using the following conditions: 94 °C for 1.5 minutes; 40 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 60 °C 4 minutes. If the nested-PCR was performed in a 96 well plate, 10 μl of the nested-PCR product was treated with 1U each of Exonuclease I (Exo I, NEB) and Shrimp Alkaline Phosphatase (SAP, Amersham) as described in Section 2.3.6. 5 μl of the treated PCR products was used for sequencing reaction using SA-seq primers with the same sequencing reaction conditions as above.

7): Purifying the sequencing product: 10 μl of MilliQ water was added to each sample to bring up the volume to 20 μl. To each sample, 50 μl of Precipitation Mix (100 ml 96% ethanol, 2 ml Na_2OAC (3 M, pH 5.2), 4 ml EDTA (0.1 mM, pH 8.0)) was added. The sequencing product in the 96 well PCR reaction plate was

collected by centrifugation at 4000 rpm at 4 °C for 25 minutes. The plate was placed upside-down on tissues to drain the liquid. The samples was washed twice using 100 µl of chilled 70% ethanol followed by a spin at 4000 rpm at 4 °C for 5 minutes. The residual liquid was drained by a final quick spin (200 rpm for 1 minute) with the sequencing plate upside-down on a tissue. The samples were dried at 65°C for 2 minutes. The sequencing reactions were run on a ABI PRISM^{TM} 377 DNA sequencer (Perkin Elmer).

5' RACE-PCR primers: LacZ-GSP1, 5'-GGG CCT CTT CGC TAT TAC GC; LacZ-GSP2, 5'-ATG TGC TGC AAG GCG ATT AAG; SA-GSP3, 5'-GTT GTA AAA CGA CGG GAT CCG CCAT; SA-seq, 5'-TGTCAC AGA TCA TCA AGC TTA TC, AAP and AUAP were purchased from Invitrogen.

3.1 Introduction

3.1.1 Screen for 6TG resistance mutants

The cytotoxity of 6TG and the simple methylation drug MNNG requires a functional MMR system (see Chapter1). The cytotoxic mechanism of 6TG is initiated in cells by Hprt (hypoxanthine-guanine phosphoribosyltransferase), which converts 6TG to 2'-deoxy-6-thioguanosine triphosphate. 2'-Deoxy-6-thioguanosine triphosphate can be incorporated into DNA as a guanine analogue during DNA synthesis. 6TG in DNA is methylated by cellular SAM (S-adenosylmethionine) to form S^6-methylthioguanine (S^6-mG). S^6-mG pairs with thymidine during DNA replication, forming a S^6-mG/T mismatched basepair (Swann et al., 1996). The binding of the MutSα to S^6-mG/T initiates a futile mismatch repair process and signals a G2/M cell cycle arrest and apoptosis process (Waters and Swann, 1997, Karran and Bignami, 1994)(Fig.1-4). Deficiency in four mismatch repair proteins, Msh2, Msh6, Mlh1 and Pms2, lead to a 6TG resistance phenotype in cultured cells because the mismatched S^6-mG/T is not recognized by the MMR machinery (Branch et al., 1993, de Wind et al., 1995, Abuin et al., 2000, Buermeyer et al., 1999). Based on mechanism of 6TG cytotoxity, it is conceivable that cells could mutate to 6TG resistance by alterations in 6TG transport, metabolism, mismatch repair recognition, cell cycle arrest and apoptosis. Therefore, genes involved in these pathways or processes could be phenotypically identified in a genetic screen for 6TG resistance.

3.1.2 MMR system and 6TG tolerance in ES cells

Mouse ES cells express all the known mismatch repair genes and mouse models with mutations in the mismatch repair recognition proteins have been generated through gene-targeting (Wei et al., 2002). Abuin et al (2000) generated *Msh2* deficient, *Msh3* deficient and *Msh2/Msh3* double null ES cells by

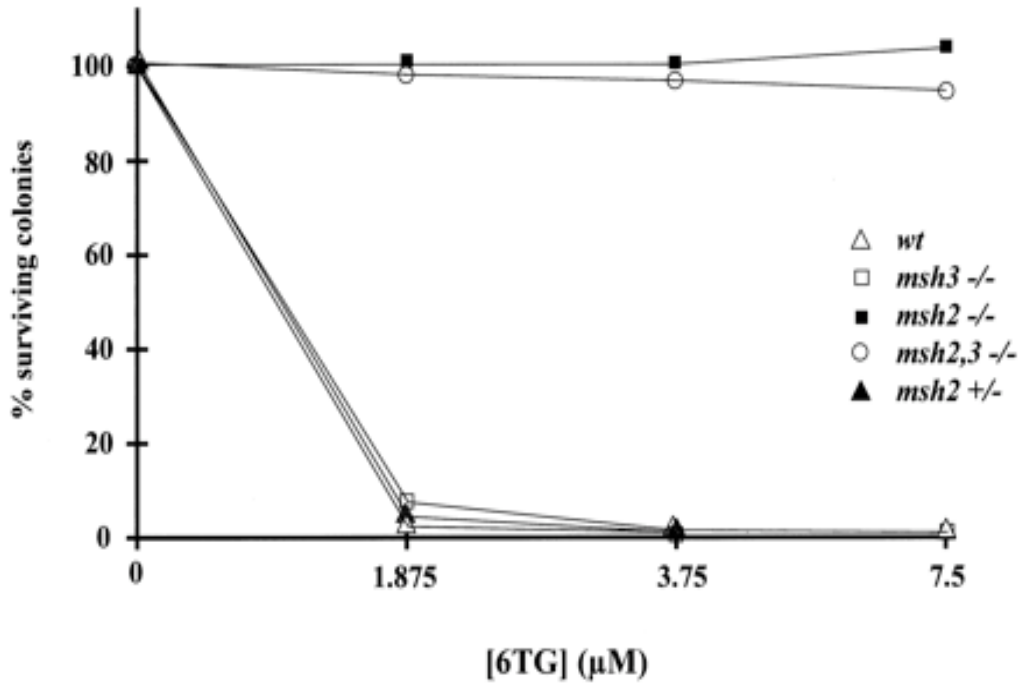


Figure 3-1. Cytotoxicity of 6TG in mouse ES cell lines.

Clonal survival of wild-type and MMR mutant ES cells in the presence of various concentrations of 6TG. Figure is taken from (Abuin et al., 2000).

sequential gene-targeting, and examined the survival of these ES cells in 6TG at various concentrations. A kill curve of 6TG on ES cells was first established (Fig. 3-1). This study revealed that *Msh2* deficient and *Msh2/Msh3* double null ES cells are able to grow in 6TG at concentrations that will kill the wild type ES cells. Another finding of this study was that ES cells with *Msh3* deficiency were not resistant to 6TG. This result is consistent with the role of Msh3 in MMR, in which Msh3 forms a protein complex with Msh2 and binds preferentially 2-4 base pair insertion/deletion lesions (Modrich, 1991). The single nucleotide mismatch (S^6-mG/T) formed by 6TG is, however, recognized by Mutα complex composed of both Msh2 and Msh6 (Waters and Swann, 1997). The effect of Msh6 and the MutL homologues, Mlh1and Pms2 on 6TG cytotoxity has not been examined directly in ES cells. Since the MMR system is highly conserved, it would be expected that mutations in these genes would confer a 6TG resistance phenotype in mouse ES cells since they do in human cells (Karran and Bignami, 1994)

3.1.3 6TG resistance caused by deficiency in *Hprt* gene

Hprt plays a central role in the mechanism of 6TG cytotoxity. It is well established that cells which have lost the function of *Hprt* gene can survive selection in 6TG. The endogenous *Hprt* gene is a single copy gene located on X-chromosome and is thus present in one copy (XY cells) or expressed from just one allele (XX cells). The mono-allelic nature of *Hprt* expression is convenient for measuring mutation rates and types since the recessive *Hprt* mutants could be selected out by 6TG (Chen et al., 2000, Munroe et al., 2000).

Despite the overall similarity of the 6TG resistant phenotype, MMR mutants and *Hprt* mutants can be distinguished in several ways. First, MMR mutants are resistant to HAT. HAT is a mixture of sodium hypoxanthine, aminopterin and thymidine. Aminopterin is a potent folic acid antagonist, which inhibits

dihydrofolate reductase blocking de novo nucleoside synthesis. Cells can only survive in HAT if the purine and pyramidine savage pathways are active. Hypoxanthine is the substrate for purine savage pathway. Thus, *Hprt* mutants are unable to utilize the purine salvage pathway and are sensitive to HAT selection, whereas MMR mutants, containing a functional *Hprt* gene, are resistant to both HAT and 6TG. Secondly, the 6TG resistance phenotype of *Hprt* mutants is affected by the genotype of the neighboring cells. It has been established that *Hprt* mutants will be killed by toxic metabolic intermediates produced by the neighboring *Hprt* positive cells, which is known as cross killing or metabolic cooperation effect. Therefore, to select *Hprt* mutants in 6TG, cells have to be plated at very low density to avoid cross killing by wild type cells (Hooper and Slack, 1977). However, MMR mutants can be selected in 6TG at high cell densities, in which *Hprt* mutants would not survive. Thirdly, It must be emphasized that the 6TG resistance phenotype exhibited by MMR deficient cells is a result of lack of recognition of the S^6-mG/T mismatch. Thus, MMR mutants will possess many mismatched S^6-mG/T nucleotides following 6TG treatment, which is likely to be mutagenic. In contrast to MMR mutants, cells without Hprt activity do not metabolize 6TG and it is not incorporated into the genome. Thus, *Hprt* deficient ES cells can grow normally in 6TG at very high concentrations. 10 μM of 6TG is routinely used to select for *Hprt* mutants. However, MMR mutants exhibit dose dependent selection (Fig. 3-1). Characteristically, *Msh2* mutants grow slower and the plating efficiency is decreased in 6TG with a concentration as low as 2 μM.

3.1.4 EMS mutagenesis in ES cells

EMS (ethyl methanosulfonate) is a monofunctional alkylation agent that ethylates DNA principally at nitrogen positions (mostly the N^7 position of guanine) as well as oxygens such as the O^6 of guanine. EMS is mutagenic in a wide variety of genetic systems from viruses to mammals (Sega, 1984). The mutagenic aspects of EMS in mouse ES cells have been investigated on the endogenous X-linked

Hprt gene. The loss of function mutations in the *Hprt* gene induced by EMS can be selected in 6TG. EMS induces predominantly point mutations, mostly G to A transitions. C to T transitions also occurs but in fewer cases. EMS is a very efficient mutagen, causing null mutations at frequencies as high as 1 mutation per locus per 1,200 treated ES cells in the 129/SvJae background with an EMS dosage of 500 μg/ml (Munroe et al., 2000).

3.1.5 *Blm* deficient ES cells in 6TG resistance screen

The rational for using *Blm* deficient ES cells to generate homozygous mutations has been discussed in Chapter 1. *Blm* deficient ES cells exhibit a high SCE and an elevated a LOH rate (Fig. 3-2 a & b). In the *Blm* deficient ES cells, *Blm* (m1/m3), the LOH rate is about 4.2×10^{-4}(cell/ generation/ locus)(Luo et al., 2000). In practical terms, a single *Blm* deficient ES cell with an autosomal single allele mutation will have segregated several daughter cells with bi-allelic mutations by the time a single cell has expanded to an ES cell colony containing 2,000 to 5,000 cells, which requires about 14 cell population times. Therefore, It is possible to mutate *Blm* deficient ES cells, generating thousands of independent single allele mutations in different genes, and from these cells derive daughters with homozygous mutations.

The Blm deficient ES cell line used in this screen is derived from the double targeted *Blm*(m1/m3) cells that contain the Blm^{tm1Brd} and Blm^{tm3Brd} alleles (Luo et al., 2000). The Blm^{tm1Brd} allele was generated by replacement gene-targeting, in which the first coding exon, exon 2, was replace by the *loxP*-flanked PGK-*neo* cassette, resulting in a truncated message RNA lacking the initiation ATG codon (Fig. 3-3a). Blm^{tm3Brd} was derived from a complex insertional targeting event that leads to the production of an aberrant transcript with an extra copy of exon3, which causes a frameshift mutation (Fig. 3-3 a) (Luo et al., 2000). *Blm*(m1/m3) cells are *Hprt* deficient because they were established in AB2.2 ES cell, an *Hprt* deficient cell line that carries the $Hprt^{mb2}$ allele (Kuehn et al., 1987). Therefore,

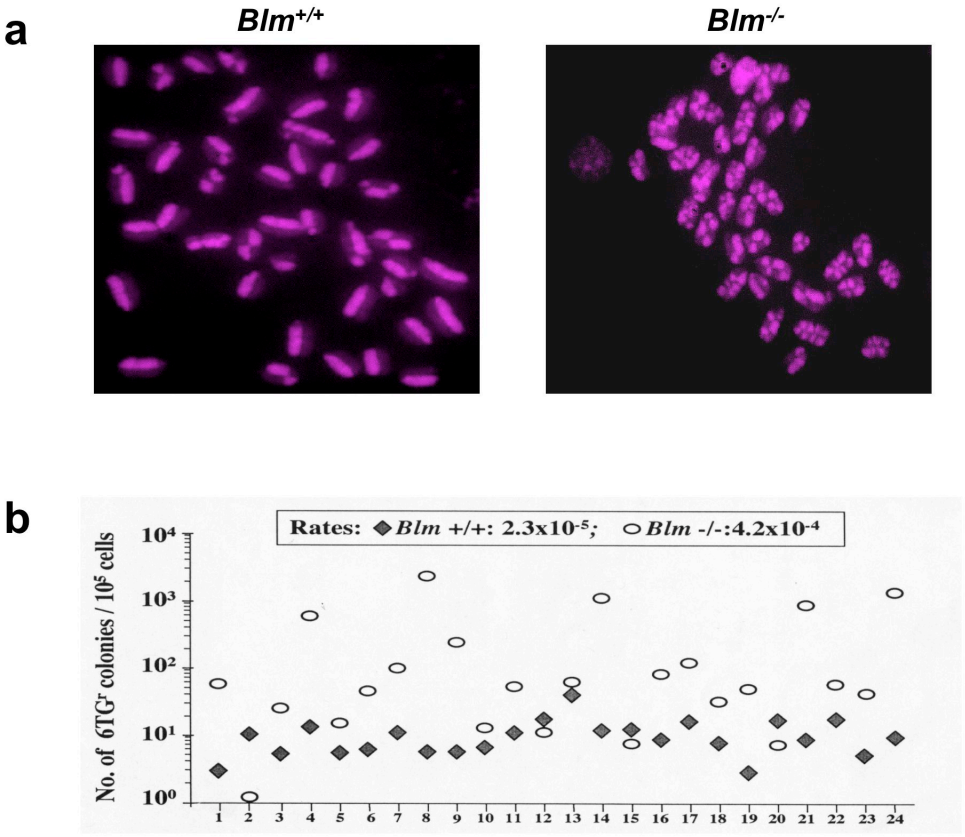


Figure 3-2. Genomic instability in *Blm* deficiency ES cells.

a. Elevated sister chromatin exchange in *Blm* deficiency ES cells illustrated by BrdU staining**. b.** Increased LOH rate in *Blm* deficiency ES cells tested by examing the loss of the single allele gene targeted *PGK-hprt* minigene. The horizontal axis shows 24 individual clones examined. The vertical axis is a log scale for the number of 6TG resistant clones per 10^5 cell. The LOH rate was calculated by Luria-Delbruck fluctuation analysis, method of means.

Figure adapted from (Luo et al., 2000)

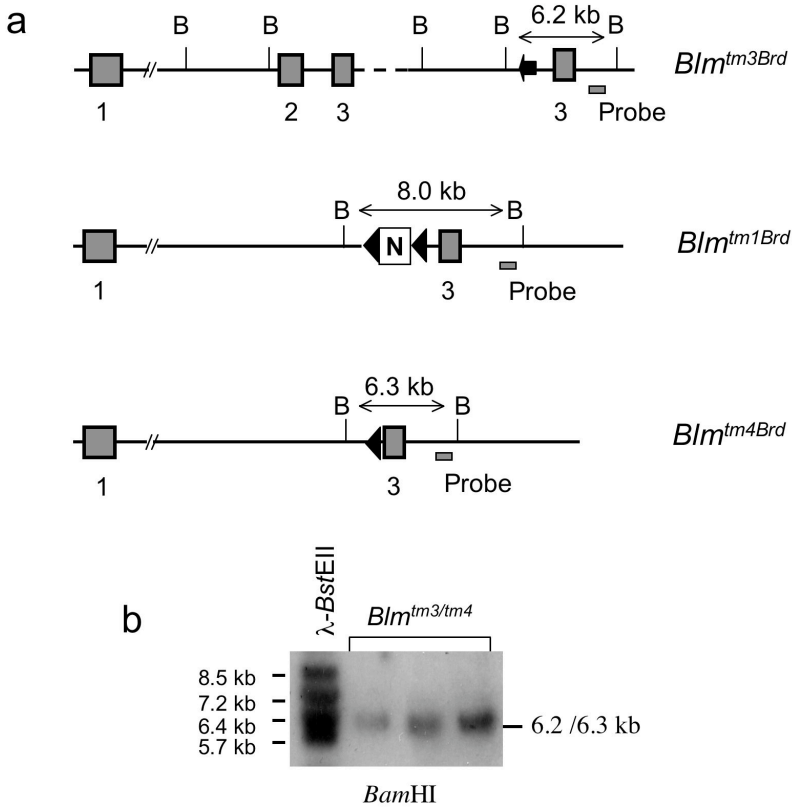


Figure 3-3. Gene-targeted *Blm* alleles

a. Schematic illustration of gene-targeted *Blm* alleles. *Blm*^{tm1Brd3}, possessing an extra exon 3, was derived from an insertional gene-targeting event. *Blm*^{tm1Brd} was generated by replacement gene-targeting, resulting in deletion of exon 2 (Luo et al., 2000). *Blm*^{tm4Brd} allele was derived from *Blm*^{tm1Brd}, in which the *loxP*-flanked *PGK-neo* cassette was removed. **b.** Southern-blot analysis of *Blm*$^{tm3/tm4}$ cells using a *Blm* probe, showing three *Blm*$^{tm3/tm4}$ cell lines that possess *Blm*^{tm3Brd} and *Blm*^{tm4Brd} alleles as indicated by 6.2 kb/6.3 kb *Bam*HI fragments. B, *Bam*HI; N, *PGK-neo* cassette; Black arrow, *loxP* site. Genomic DNA for Southern-blot analysis was digested with *Bam*HI restriction enzyme.

Blm(m1/m3) cells are 6TG resistant. To perform a screen for mutations that cause 6TG resistance in *Blm*(m1/m3) based ES cells, Hprt activity has to be provided. In this chapter, a gene-targeting strategy was adopted to introduce *Hprt* minigenes into both alleles of an autosomal gene in *Blm* deficient cells. Targeting two copies of the *Hprt* minigene also helps to reduce the background of 6TG resistant clones caused by loss of function of the targeted *Hprt* genes by spontaneous mutation or LOH events. EMS mutagenesis was performed in these *Hprt* positive and *Blm* deficient cells to verify that a 6TG-resistant screen would be successful.

3.2 Results

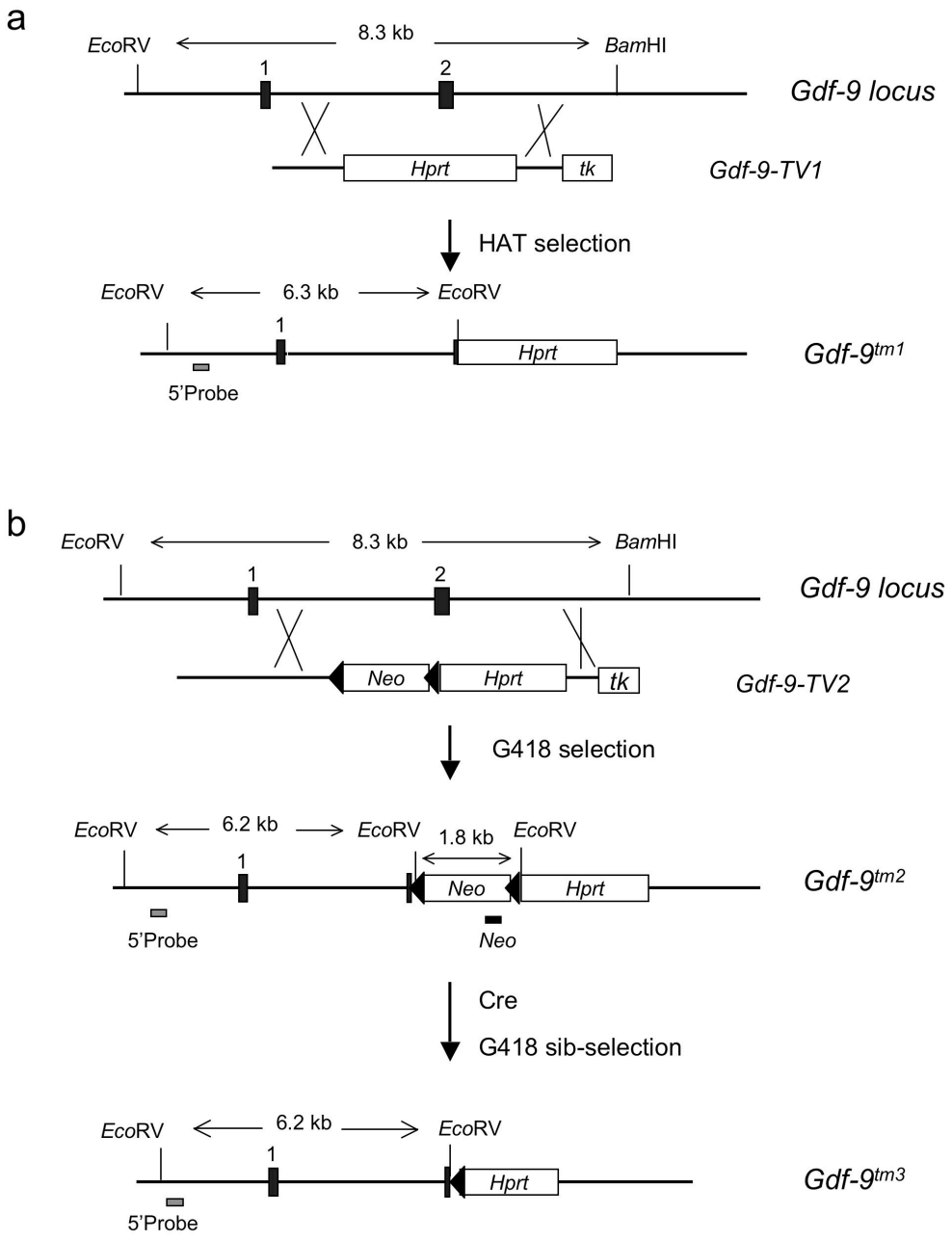
3.2.1 Construction of the *Blm* deficient ES cells carrying two copies of an *Hprt* minigene

Blm (m1/m3) ES cells contain Blm^{tm1Brd} and Blm^{tm3Brd} alleles (Fig. 3-3 a). The Blm^{tm1Brd} allele contains the *loxP*-flanked PGK-*neo* selection marker (Luo et al., 2000). To enable recycling of the *neo* selection marker (Abuin and Bradley, 1996) in subsequent aspects of our experiments, the PGK-*neo* cassette was removed by Cre-*loxP* mediated recombination from *Blm* (m1/m3) cells to generate the marker-free *Blm* deficient cells, *Blm* m3/m4) cell line. Southern-blot analysis using a 3' *Blm* external probe revealed the predicted 6.2 and 6.3 kb *Bam*HI fragments in *Blm* (m3/m4) cells (Fig. 3-3 a & b). Note that the 6.2 kb and 6.3 kb fragments can not be distinguished in the Southern-blot because of the similar size, therefore, they appear as one wild band (Fig. 3-3 a & b).

To provide the Hprt activity, PGK-*Hprt* minigenes were introduced sequentially into both autosomal *Gdf*-9 (*growth differentiation factor 9*) loci by gene-targeting technology. The *Gdf-9* locus was chosen because a *Gdf*-9 gene-targeting vector was available that contains the PGK-*Hprt* minigene and this exhibited high gene-targeting efficiency. Importantly, *Gdf*-9 is only required for sex development in

female mice and *Gdf*-9 deficiency doesn't have adverse effects on ES cell growth (Dong et al., 1996). The first *PGK-Hpr*t minigene was introduced using a previously described *Gdf*-9 targeting vector (Dong et al., 1996). This *Gdf*-9 targeting vector contains a PGK-*Hprt* cassette as the selection marker for gene-targeting and a MCI-*TK* marker for negative selection. The *Gdf*-9 targeting vector was linearized by *Pvu*I and electroporated into *Blm*(m3/m4*)* cells. HAT and FIAU double selection was applied. ES cells clones with targeted *Gdf*-9 alleles were identified by genomic Southern-blot using a 5' external *Gdf*-9 probe. For Southern-blot analysis the genomic DNA from HAT and FIAU resistant ES cells was digested with *Bam*HI and *Eco*RV. The 5' external probe recognizes a 8.0 kb *Bam*HI*/Eco*RV wild type fragment and a 6.3 kb *Eco*RV fragment for the targeted allele (*Gdf9*^{tm1}) (Fig. 3-4 a). The targeting efficiency was 50%. The targeted cell line, *Blm*$^{m3/m4}$/ *Gdf9*$^{tm1/+}$*,* was expanded from a 96 well tissue culture plate to 90 mm tissue culture plate for targeting of the second *Gdf*-9 allele.

The high LOH rate in *Blm* deficient cells increases the probability of losing the single targeted *Hprt* minigene via mitotic recombination. Thus, it is important to generate cells with the PGK-*Hprt* minigene targeted to both alleles. To introduce the second PGK-*Hprt* minigene into *Gdf-9* locus, the *Gdf9-TV2* gene-targeting vector was generated, in which a *loxP*-flanked PGK-*neo* cassette was inserted in front of the PGK-*Hprt* minigene as the selection marker for gene-targeting. *Gdf*9-*TV2* was linearized with *Pvu*I and electroporated into *Blm*$^{tm3/tm4}$ / *Gdf*-9$^{tm1/+}$ cells, followed by G418 selection. The second targeted allele (*Gdf*-9^{tm2}) was identified as a 6.2 kb *Eco*RV fragment by genomic Southern-blot analysis using the 5'*Gdf*-9 probe (Fig. 3-4 b & d). The PGK-*neo* selection marker was then removed by Cre-*loxP* mediated recombination from the *Gdf*-9^{tm2} allele to generate the *Gdf*-9^{tm3} allele (Fig. 3-4 b & d). The resultant *Blm*-deficient, *Hprt*-postive, and *neo*-negative cells were named as NGG, which have *Blm*$^{tm3/tm4}$ and *Gdf*-9$^{tm1/tm3}$ alleles (Fig. 3-4 c). NGG cells were identified by sib-selection for G418 sensitive clones. Genomic Southern-blot analysis using a *neo* probe confirmed the removal of *PGKneo* cassette. Cells with the targeted PGK-*neo* cassette displayed a 1.8 kb


a
EcoRV
8.3 kb
BamHI
1
2
Gdf-9 locus
Hprt
tk
Gdf-9-TV1
HAT selection
EcoRV
6.3 kb
EcoRV
1
Hprt
Gdf-9tm1
5'Probe
b
EcoRV
8.3 kb
BamHI
1
2
Gdf-9 locus
Neo
Hprt
tk
Gdf-9-TV2
G418 selection
EcoRV
6.2 kb
EcoRV
1.8 kb
EcoRV
1
Neo
Hprt
Gdf-9tm2
5'Probe
Neo
Cre
G418 sib-selection
EcoRV
6.2 kb
EcoRV
1
Hprt
Gdf-9tm3
5'Probe

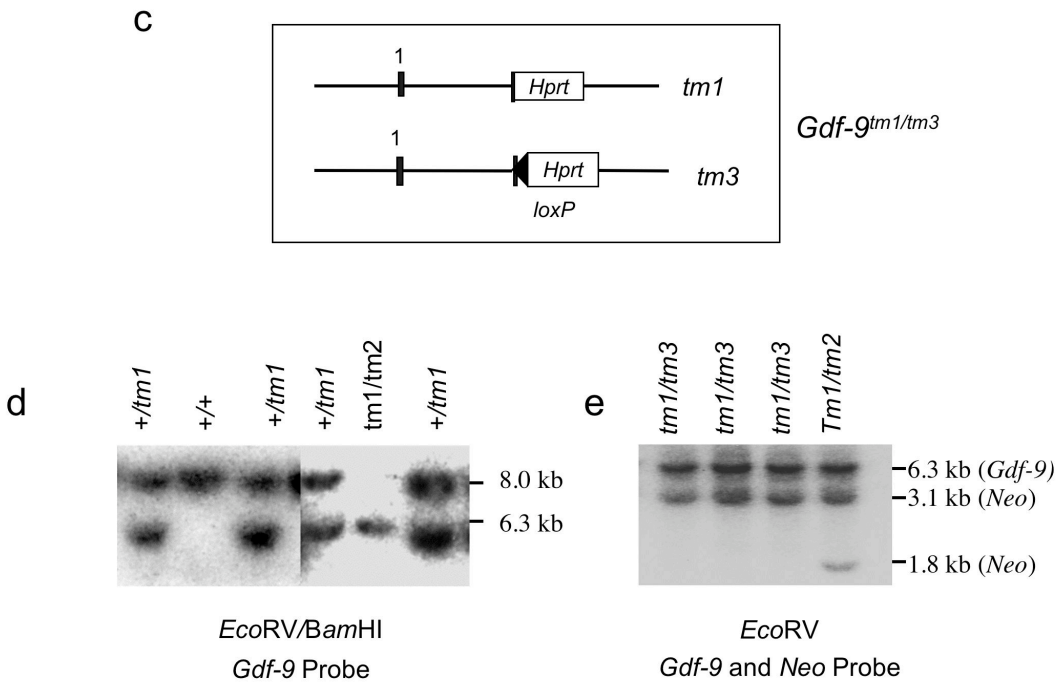


Figure 3-4. Generation of *Hprt*$^{+/+}$ *Blm* deficient ES cells.

a. Generating the *Gdf-9*^{tm1} allele by gene-targeting of wild type *Gdf-9* locus. **b.** Targeting the second *Gdf-9* allele, *Gdf-9*^{tm2} *Gdf-9*^{tm3} alleles. **c.** Schematic of the *Blm*(m3/m4) cell line that harbours the gene-targeted *Gdf-9*^{tm1} and *Gdf-9*^{tm3} alleles (the *Blm* locus is not illustrated). **d.** Southern-blot using the *Gdf-9* probe demonstrates ES cells that are single allele targeted (*Gdf-9*^{tm1}) or double targeted (with *Gdf-9*^{tm1} and *Gdf-9*^{tm2} alleles) **d.** Southern-blot showing the double gene-targeted *Gdf-9*$^{a/b}$ and *Gdf-9*$^{a/c}$ cells. The PGK-*neo* cassette(*Neo*) in the *Gdf-9*^{tm2} allele is represented as an1.8 kb *Eco*RV fragment. The Cre-*loxP* mediated recombination results in the loss of the 1.8 kb fragment (*Neo* cassette) in the *Gdf-9*$^{tm1/tm3}$ cells. The 3.1 kb fragment derived from a non-functional *Neo* gene in AB2.2 ES cells.

*Eco*RV band from the PGK-*neo* cassette and a 3.1 kb band from the X-linked non-functional *neo cassette* in the AB2.2 genetic background, while NGG cells showed only the 3.1 kb band (Fig. 3-4 e). After confirmation of the removal of PGK-*neo* cassette from NGG, one of the NGG cell lines, NGG5, was seeded at low density and allowed to form single ES cell colonies. The single cell colonies were picked and expanded. This single cell recloning process was performed to eliminate cross contamination of *Neo* positive cells during gene-targeting of the second *Gdf*-9 allele. One single cell clone, NGG5-3 was expanded and the genotype was confirmed again by Southern-blot analysis. NGG5-3 cells were then plated in G418 and confirmed to be G418 sensitive.

3.2.2 NGG cells are 6TG sensitive

Before starting a genetic screening for 6TG resistant mutants, it was important to check the PGK-*Hprt* transgenes express at a level sufficient to give 6TG toxicity. NGG cells and two other *Hprt* positive cell lines, AB1 and NG12-D were plated, separately in 5μM of 6TG. AB1cells have a normal endogenous *Hprt* gene, and are therefore sensitive to 6TG. As shown in table 3-1, AB1 cells were fully killed by 6TG. No 6TG resistant cells were recovered from plated NGG cells either, confirming that the targeted PGK-*hprt* minigenes in the *Gdf-9* loci are functional and stable. The NG-12D cell line is one of the parental *Blm*-deficient cell lines with a single targeted *Gdf*-9^{tm1} allele. Approximately 5% of these cells survived 6TG selection, which is consistent with the high LOH rate in *Blm*-deficient cells. The effect of 6TG killing was also examined in *Hprt* -deficient AB2.2 cells. AB2.2 cells were fully resistant to 5 μM 6TG with a clonal survival rate of 100%. An *Msh2* deficient cell line generated by sequential gene-targeting method (Abuin et al., 2000) was also tolerant to 5 μM 6TG, but the clonal survival rate was slightly lower (80%). The lower survival rate of *Msh2* deficient cells in 6TG suggests that 6TG may cause an adverse effect on the growth of MMR deficient cells (Table 3-1).

Cell line	Hprt♥	Survival rate in 6TG✳	Cells plated
AB1	$+^{e}$	0	$2.5x10^{4}$
NG-12D	$+^{t}$	5%	$2.5x10^{4}$
NGG	$+^{t}/+^{t}$	0	$2.5x10^{4}$
Msh2$^{-/-}$	$+^{e}$	83%	$2.5x10^{2}$
AB2.2	-	103%	$2.5x10^{2}$

Table 3-1. Survival rates of cells in 6-TG

NG12-D is a *Blm*-deficient ES cell line with single allele of *Gdf-9* targeted with a PGK-*hprt* minigene. NGG is a *Blm* deficient ES cell line with both alleles of Gdf-9 targeted with PGK-*Hprt* minigenes. *Msh2*$^{-/-}$ is a *Msh2* deficient ES cell line, generated by gene-targeting in AB1 ES cells (Abuin *et al.,*2000).

"$+^{e}$" represents the endogenous X-linked *Hprt* gene; "$+^{t}$" represents the targeted *PGK-Hprt* minigene and *Hprt* negative is illustrated as "-".

✳ Cells were plated in M15 medium with or without 6TG and allowed to grow for 10 days to form visible ES cell colonies. The Survival rate was determined by comparing the number of colonies formed with and without 6TG selection.

3.2.3 Screening for 6TG resistant mutants by EMS mutagenesis

3.2.3.1 EMS treatment of *Blm* deficent NGG cells

$2x10^6$ NGG ES cells were plated in 6-well feeder plates and treated with 600 μg/ml EMS for 15 hours. The cell plating efficiency was determined to be 31% by plating cells at low cell density. After EMS treatment, the survival cells were harvested and the number was counted. About 8,000 (0.4%) cells survived (Material and Method 2.2.8). Thus, the survival rate of EMS treatment was1.3%, which was determined as the survival rate of EMS treated ES cells versus cell plating efficiency (Table 3-2 a). Based on the established EMS mutation frequency in ES cells (about 1 mutant per locus per 1,200 cells surviving EMS treatment (Munroe et al., 2000)), it was estimated that the pool of 8,000 ES cells contains about 6 mutants for each locus. In other words, the EMS mutated cells contain 6 fold genome coverage. To allow for the segregation of homozygous mutant cells, the pool of EMS treated cells were expanded continuously for 10 days, and then a small portion of the pool was expanded for 4 more days to give time for decay of mRNA and protein in the presumptive mutants before being plated for 6TG selection (Fig. 3-5).

3.2.3.2 Screen for 6TG resistant mutants at low cell density

After 14 days passaging following EMS treatment, about 4 $x10^6$ cells were plated on four 150 mm tissue culture plates at a density of $1x10^6$ ES cells per 150 mm tissue culture plate. 6TG (5 μM) selection was applied and continued for eight days. The surviving 6TG resistant clones were cultured in 6TG-free medium for four days to allow the healthy growth of the “sick” mutant cells before the colonies were picked and expanded. Eleven 6TG resistant ES cell clones were recovered in this experiment. No 6TG resistant clones were recovered in two control plates of NGG cells without EMS treatment, suggesting that the 6TG selection was efficient and the double targeted *PGK-Hprt* minigene does not

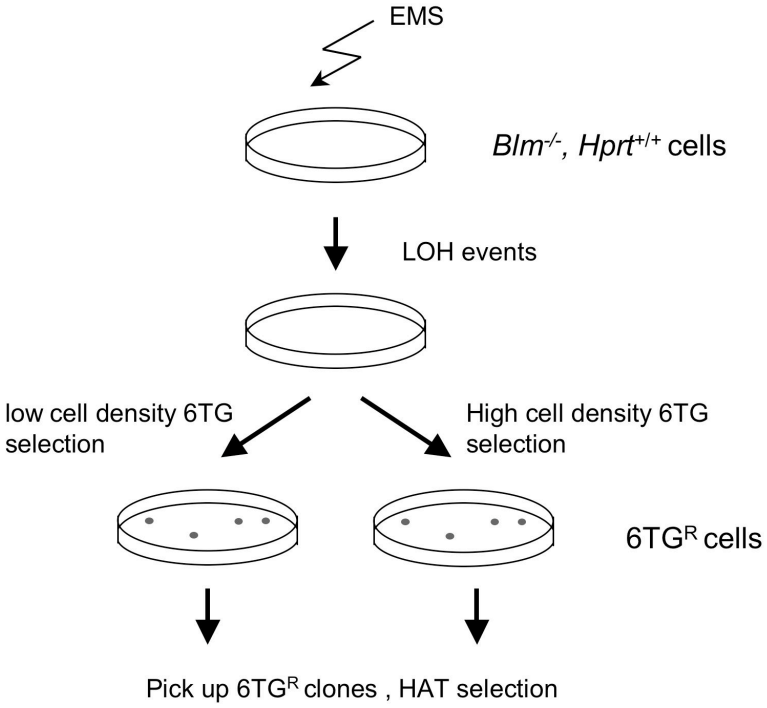


Figure 3-5 EMS mutagenesis and 6TG screens

Step1: EMS mutates NGG (*Blm*$^{tm3/tm4}$, *Gdf-9* $^{tm1/tm3}$) ES cells

Step2: Expand cells for 14 days to obtain homozygous mutants

Step3: plate cells at relevant density and apply 6TG selection

Step4: pick 6TG^{R} ES cell clones and apply HAT selection

Table 3-2 a. EMS treatment of NGG cells

[EMS] μg/ml	Plating efficiency	Survival rate in EMS	Mutation rate[#1]	No of mutants per locus[#2]
0	31%	-	-	-
600	0.4%	1.3%	1/1200	6

Table 3-2 b. Selection for 6TG resistant mutants

Selection	No. of cells selected	Cell density	No. of $6TG^R$ clones	No. of HAT^R clones
Low density	$4x10^6$	$1x10^6$/150 mm	11	5[#3]
High density	$2.3x10^7$	$2.3x10^7$/150 mm	24	24

Table 3-2. EMS mutated NGG cells and 6TG screens.

#1: This mutation rate was derived from published data (Munroe *et al*., 2001).

#2: indicates the pool of EMS mutated ES cells possesses about 6 mutants in each gene based on the assumed mutation rate and $8x10^3$ surviving cells.

#3: The number of HAT^R colonies was extrapolated from the ratio of HAT^R versus total $6TG^R$ clones obtained by examining seven of the eleven $6TG^R$ clones from the low cell density selection.

"-": The parameters of the columns were not applied to the non-EMS treated cells.

revert. The viability of the selected clones in 6TG was confirmed by re-testing their colony forming ability in 6TG. Two groups of mutants would be expected to be recovered from the low cell density 6TG selection. One group will be mutated in the 6TG metabolism pathway, presumably the *Hprt* locus itself, while the other group should include mutants in MMR mediated DNA damage surveillance. To inspect the integrity of the *Hprt* gene, seven of the recovered 6TG resistant clones were expanded and plated on 24 well tissue culture plates in HAT. Four out of seven tested cell lines didn't grow in HAT, suggesting that they contain mutations in *Hprt*. Three out of seven 6TG resistant clones were also HAT resistant. These are potential MMR mutants with intact *Hprt* genes (Table 3-2 b).

3.2.3.3 Screen for 6TG resistance cells at high cell density

A high cell density 6TG selection was performed with $2.3x10^7$ EMS treated and expanded NGG5 cells seeded on one 150 mm tissue culture plate. The 6TG selection was applied in the same way as for the low cell density 6TG selection. Twenty-four 6TG resistant clones were recovered from this screen. These 6TG resistant cells were plated in HAT supplemented cell culture medium and all were confirmed to be HAT resistant, suggesting that they are potential MMR mutants (Table 3-2 b).

3.3 Discussion

3.3.1 Construction of NGG (*Blm*$^{tm3/tm4}$, *Gdf-9* $^{tm1/tm3}$) cells

In this chapter, *Blm* deficient cell lines containing two gene-targeted PGK-*Hprt* minigenes were generated in order to establish a genetic screen for mutations in mismatch surveillance by selection in 6TG. The NGG (*Blm*$^{tm3/tm4}$, *Gdf-9*$^{tm1/tm}$) cells exhibited complete 6TG sensitivity even after long period of culture and multiple passages. This result suggests that the targeted *Hprt* minigenes are maintained stably in *Blm* deficient cells. In this regard, targeting two copies of the

Hprt genes to both alleles of one gene provides a very stable situation for a screen compared with cells with a single targeted *Hprt* minigene. For example the NG12-D cells with a single *Gdf-9* allele targeted with the PGK-*Hprt* minigene frequently segregate 6TG resistant clones (Table 3-1). Importantly, NGG and one of its single cell derivatives, NGG5-3 cell line, are devoid of the commonly used drug selection markers, such as *Puro, Neo*, *Bsd*. Thus they are amenable to further targeting based modifications.

3.3.2 Screen for 6TG resistant mutants

EMS mutagenesis was performed to mutate NGG cells in order to screen for 6TG resistant mutants. EMS is a highly efficient chemical mutagen that causes preferentially loss of function point mutations. Therefore, EMS mutagenesis is an ideal method to quickly generate a large quantity of recessive mutations. To examine the effect of cell density on 6TG killing, two screens were performed on EMS mutated NGG cells at different cell plating densities. The low cell density 6TG selection allowed the recovery of both *Hprt* mutants as well as the potential mutants in mismatch surveillance as shown recovery of the HAT resistant clones. In contrast, all clones recovered in the high cell density 6TG screen were also HAT resistant, showing that the *Hprt* mutants were killed by the metabolic cooperation under high cell density 6TG selection as expected. In the low cell density screen, 5 6TG resistant mutants were also HAT resistant. The total number of cells plated for 6TG selection in the high cell density 6TG screen was about 5 times the number of the cells plated in the low cell density screen. Thus, 25 HAT resistant mutants were expected to be recovered in the high cell density 6TG screen. In fact, 24 clones were recovered, which illustrates the effectiveness of the high cell density selection in recovering potential mismatch surveillance mutants.

These results also suggest that the *Blm* deficient ES cell is a useful tool for isolating recessive mutations. After this work was finished, Yusa et al reported a

screen for recessive mutants of glycosylphosphatidylinositol (GPI)-anchor biosynthesis using a conditional *Blm* deficient ES cell line. They mutated *Blm* deficient ES cells using ENU (*N*-ethyl-*N*- nitrosourea) mutagenesis and selected GPI anchor mutants in a drug called aerolysin, which is capable of killing GPI-anchor positive cells. By cDNA rescue, they identified mutations in 12 out of 23 candidate genes, and by sequencing they confirmed that all of these mutants are homozygously mutated (Yusa et al., 2004), demonstrating that *Blm* deficiency is indeed an efficient means to produce homozygous mutations for recessive genetic screens. Although these experiments demonstrated clearly the power of *Blm* deficiency in a recessive screen, little additional information could be determined about the novel mutated genes. The methods for identifying mutations generated by chemical mutagenesis require localization of the novel mutation by linkage analysis, cDNA rescue and sequencing of the candidate genes. These methods are not suitable to identify the molecular basis of mutants from large scale phenotype driven screens, which aim to isolate novel genes and obtain novel information. Because the major aim of my Ph.D. project is to establish a high throughput method for recessive genetic screens which include gene identification, we did not pursue chemical mutagenesis beyond the proof of principle described here. As an alternative, we used gene trap mutagenesis which does allow high throughput identification of mutated genes.

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.

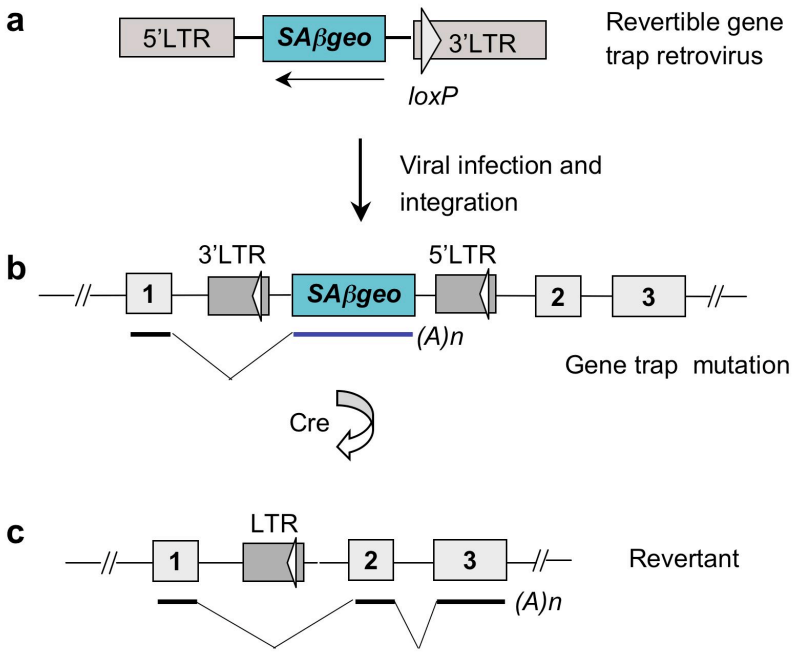


Figure 4-1. Revertible gene trap virus

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

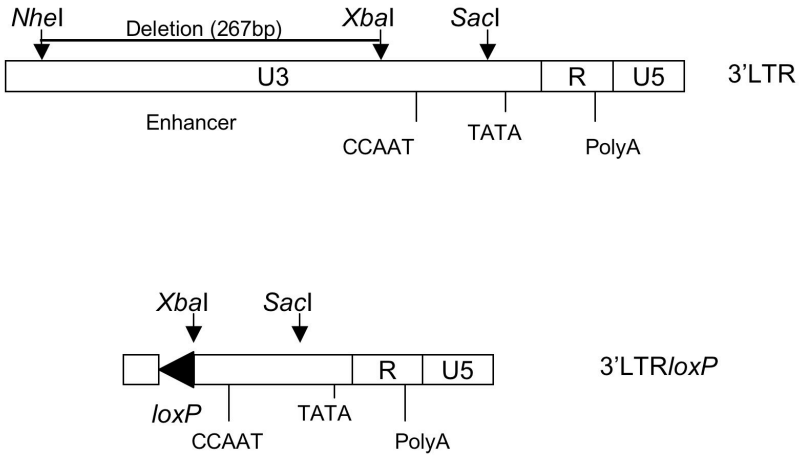


Figure 4-2. Schematic the construction of 3'LTR for revertible retroviral vectors

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.

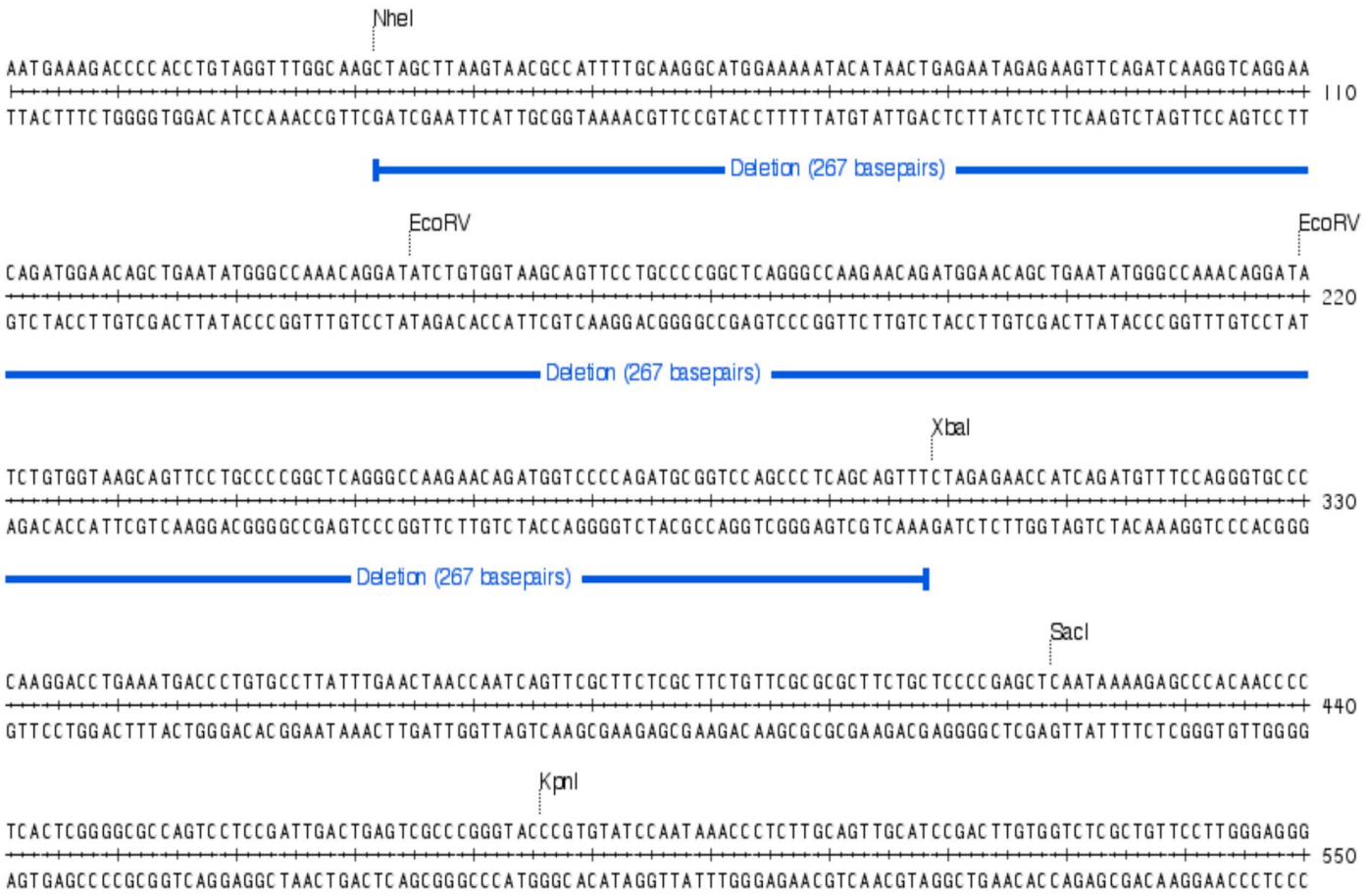


Figure 4-2 b. Sequence of the part of the MoMuIV 3'LTR showing the sequence of the deleted viral enhancer

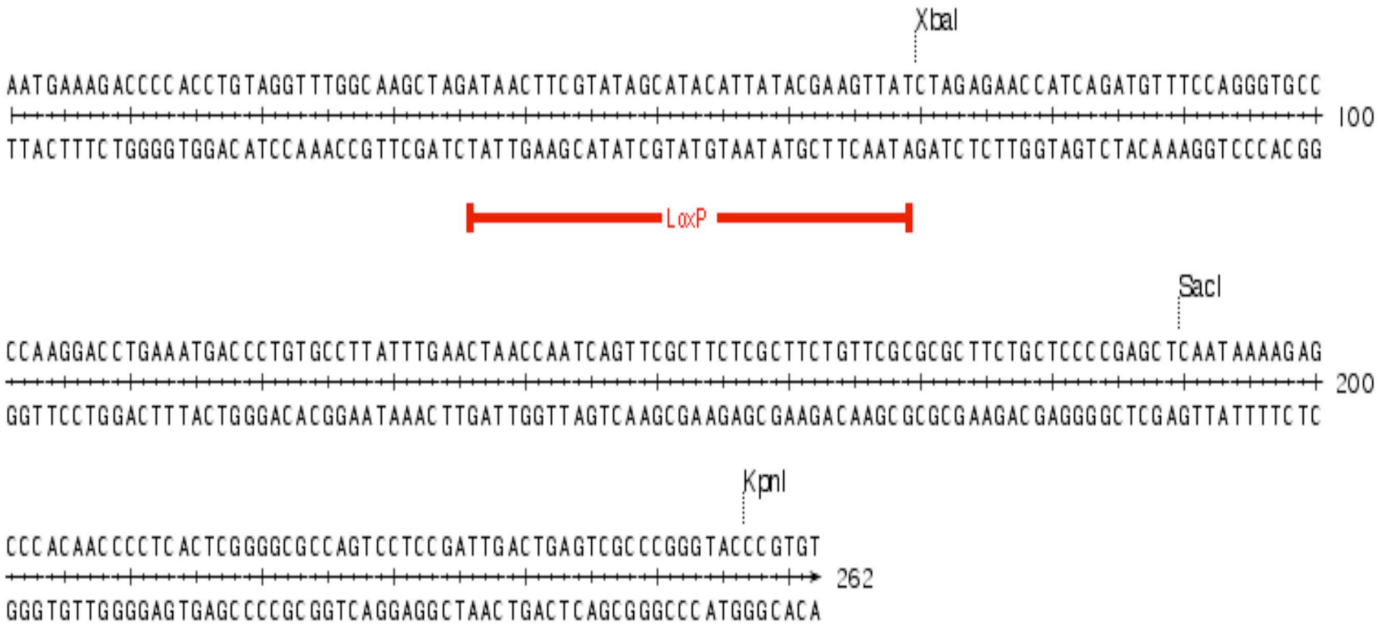


Figure 4-2 c. Sequence of the modified 3'LTR*loxP*, showing the inserted *loxP* site.

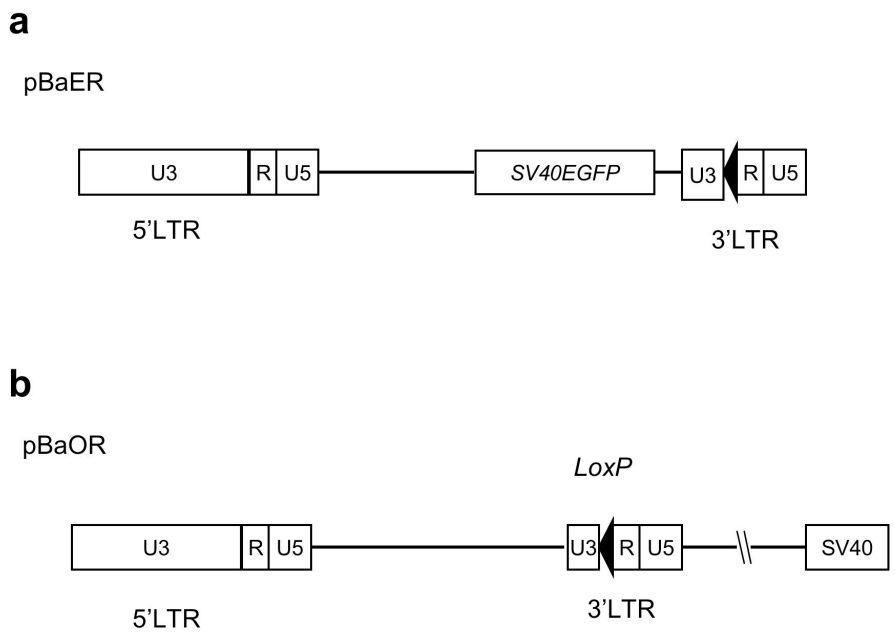


Figure 4-3. Retroviral gene trap vectors. pBaER and pBaOR are MoMulV based retroviral vectors, containing the MoMulV 5'LTR and the modified 3'LTR*loxP*. **a.** pBaER contains a fluorescent marker *EGFP* (enhanced green fluorescent protein)driven by the SV40 early promoter, which can be used to monitor 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.

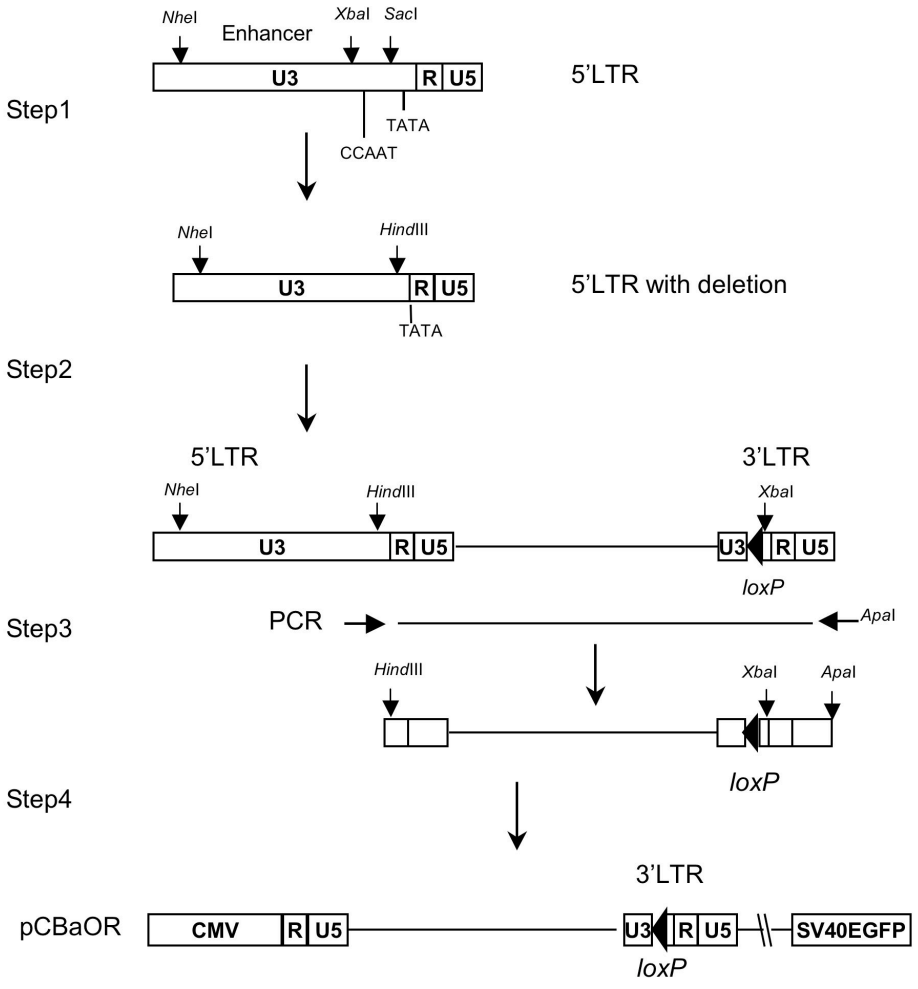


Figure 4-4. Construction of a revertible retroviral vector driven by CMV promoter/enhancer.

Step1: The viral 5'LTR was subcloned and the promoter region between *Xba*I/*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*

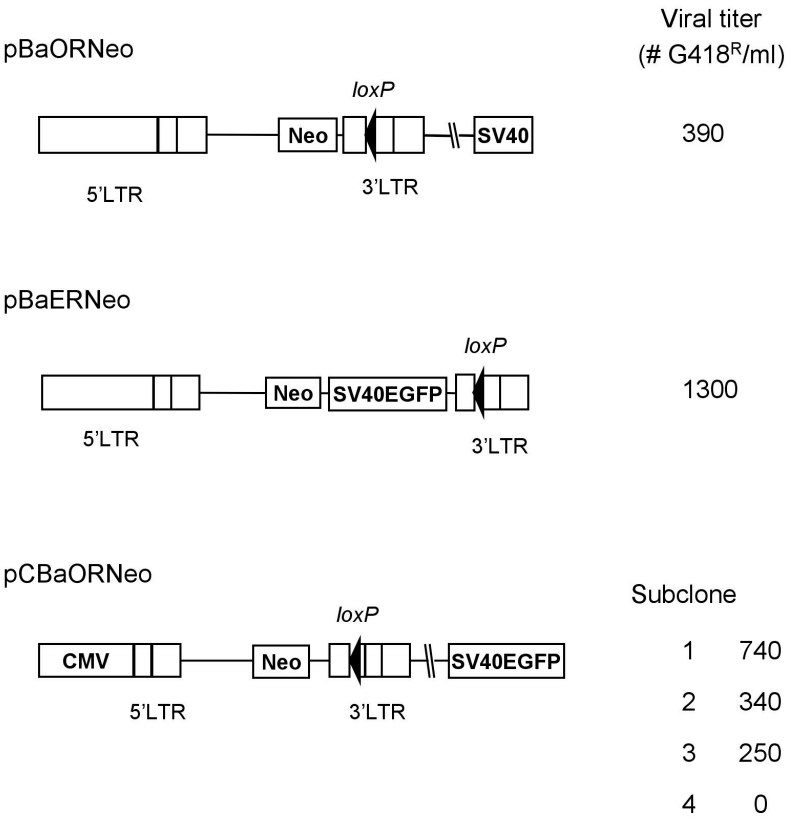
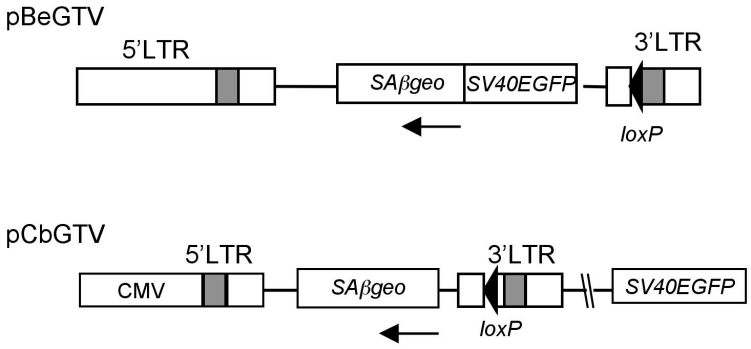


Figure4-5. Viral titer by pBaORNeo, pBaERNeo and four individual pCBaORNeo vectors

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.

a



b

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 *SA*β*geo* 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

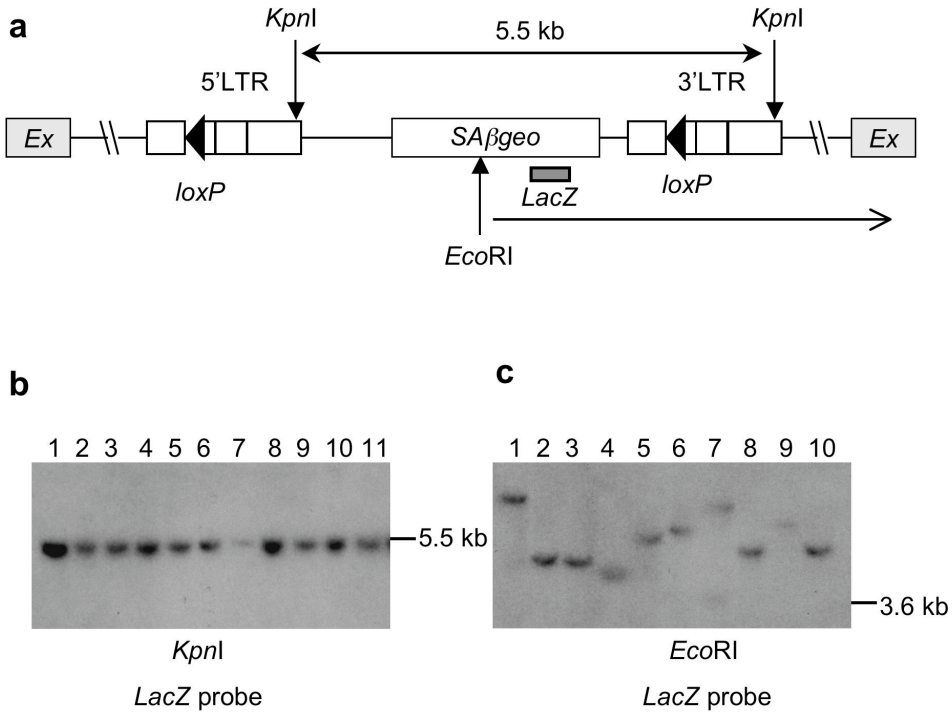


Figure 4-7. Southern-blot analysis of the integrated provirus from RGTV-1

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 *SA*β*geo* 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).

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.

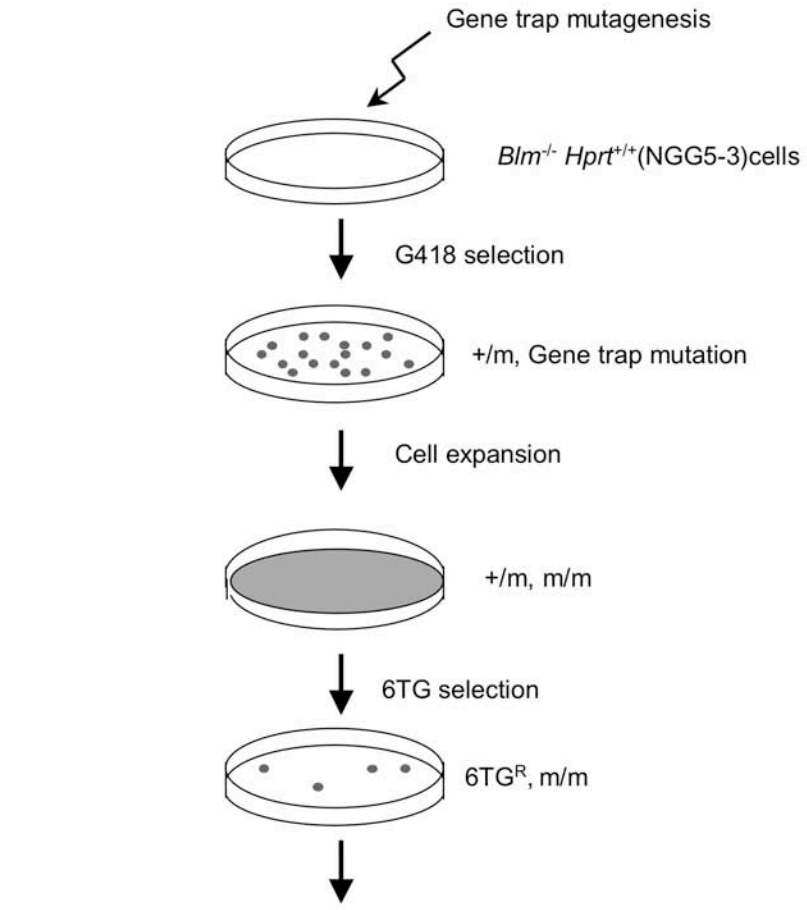


Figure 5-1. Screening strategy for 6TG resistant ($6TG^R$) gene trap mutants

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.

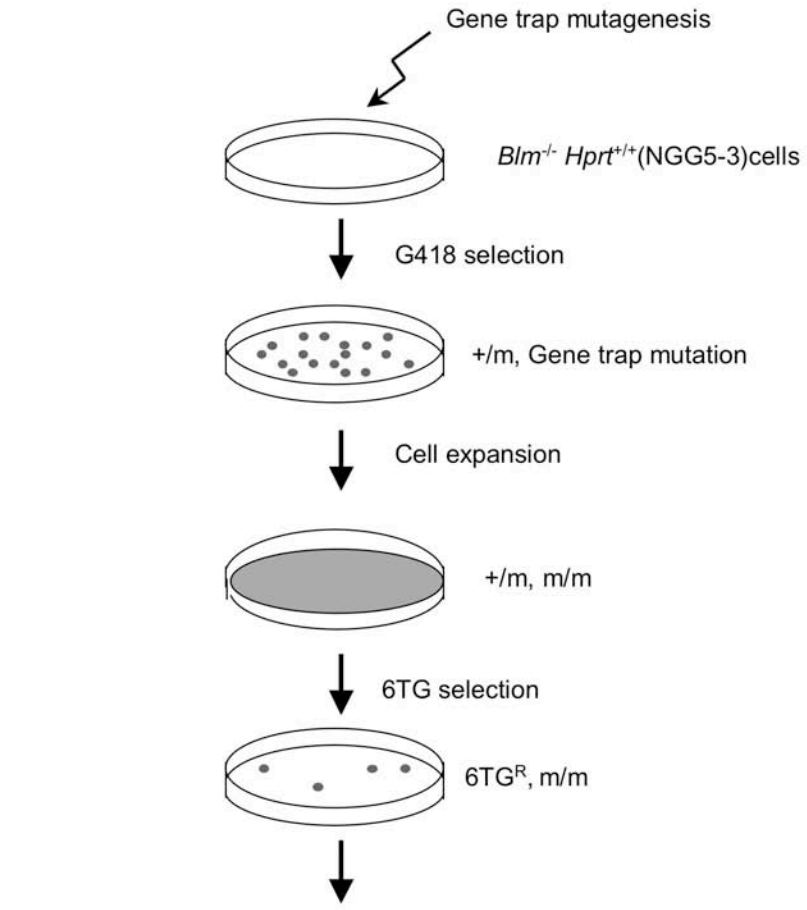


Figure 5-1. Screening strategy for 6TG resistant ($6TG^R$) gene trap mutants

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.

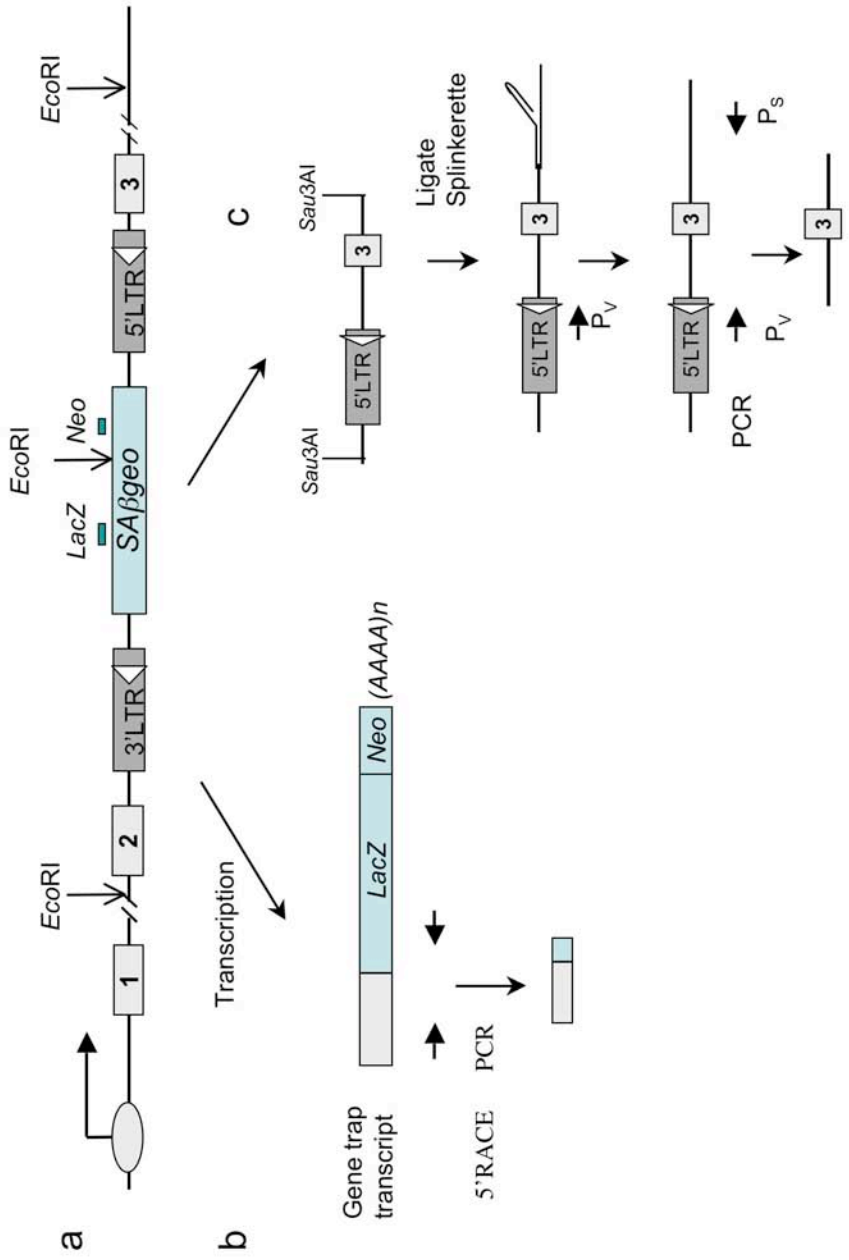

a
EcoRI
1
2
3'LTR
LacZ
Neo
SAβgeo
5'LTR
3
EcoRI
EcoRI
b
Transcription
Gene trap transcript
LacZ
Neo
(AAAA)n
5'RACE
PCR
c
Sau3AI
5'LTR
3
Sau3AI
Ligate Splinkerette
5'LTR
3
PV
5'LTR
3
PS
PV
PCR
3

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

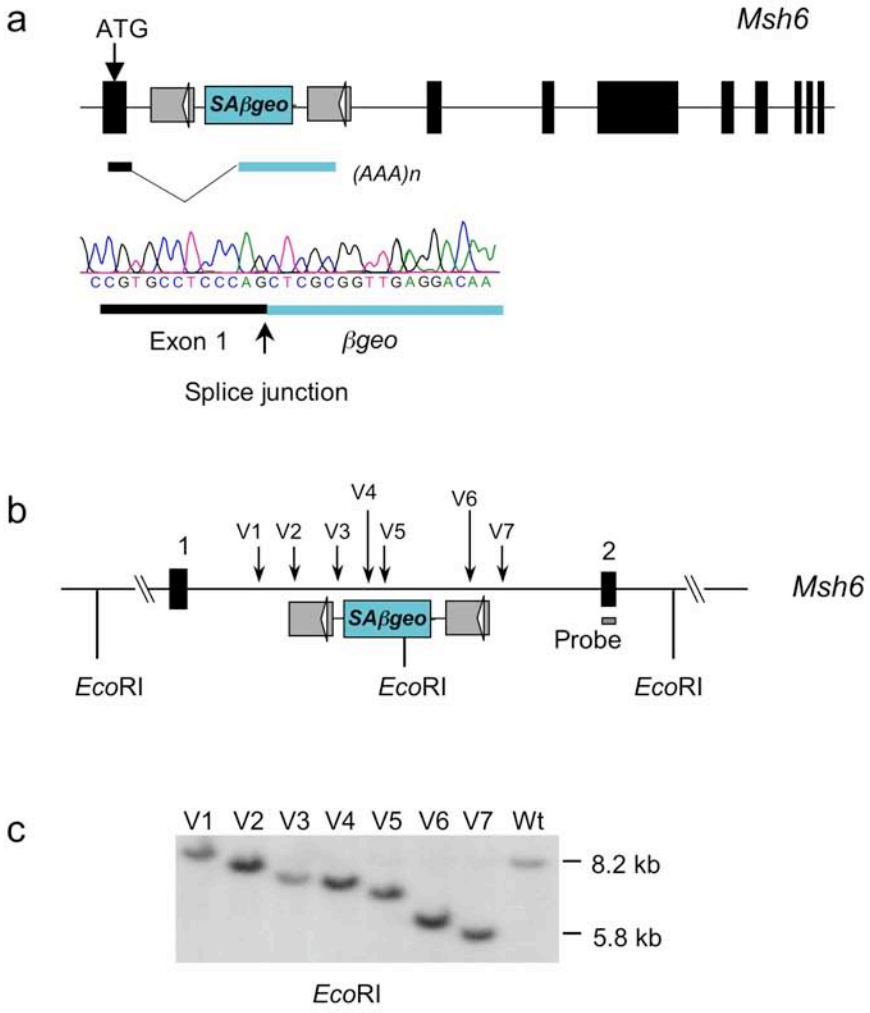


Figure 5-3. Identification of *Msh6* gene trap mutations

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.

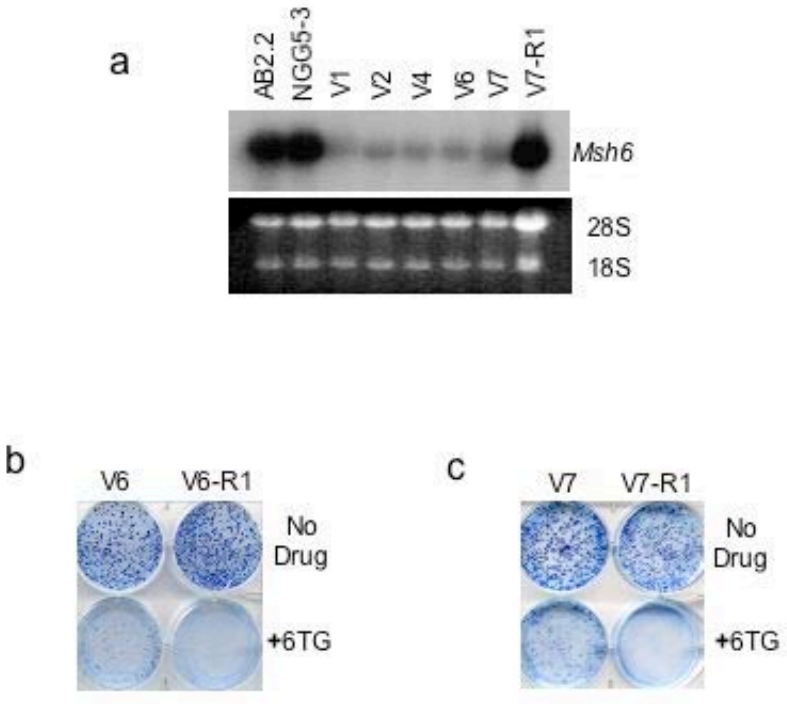


Figure 5-4. Molecular analysis of *Msh6* mutations

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.

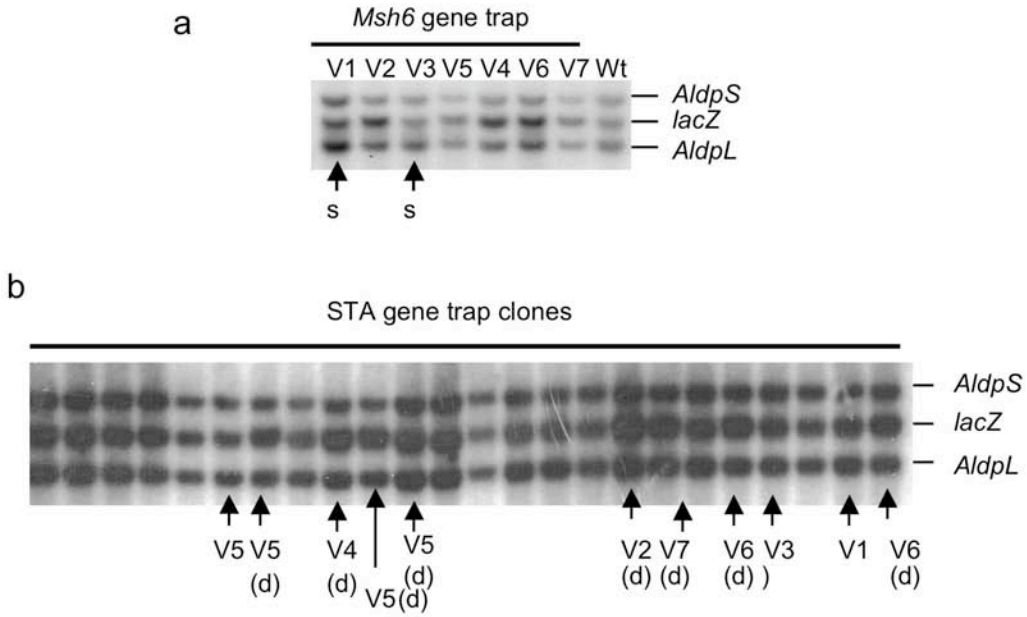


Figure 5-5. Quantitative Southern-blot analysis (QTSouthern) of STA gene trap mutants

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.

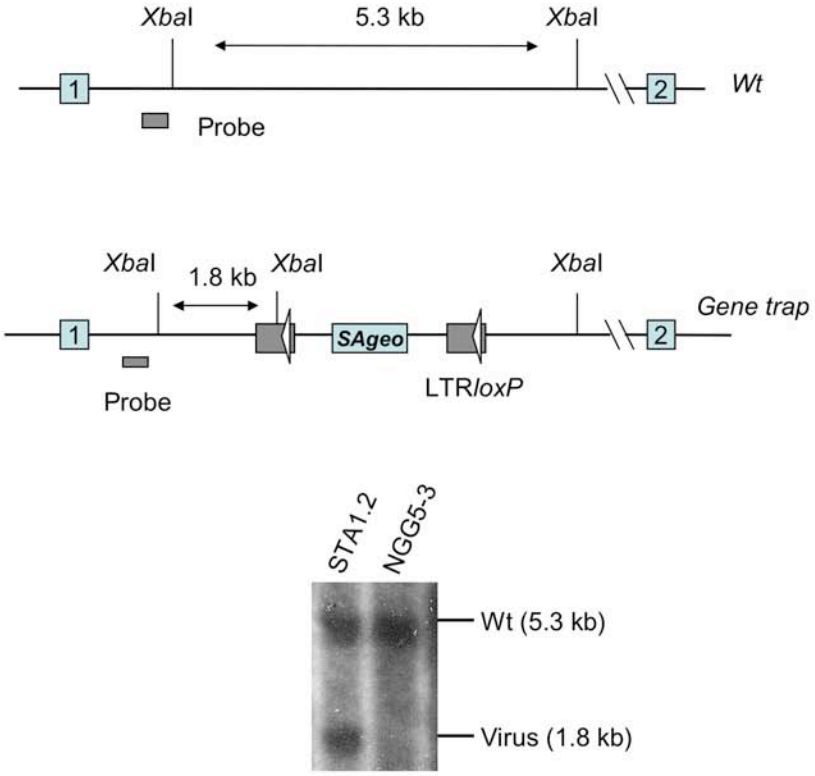


Figure 5-6. Southern-blot analysis of gene trap clone STA1.2,

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).

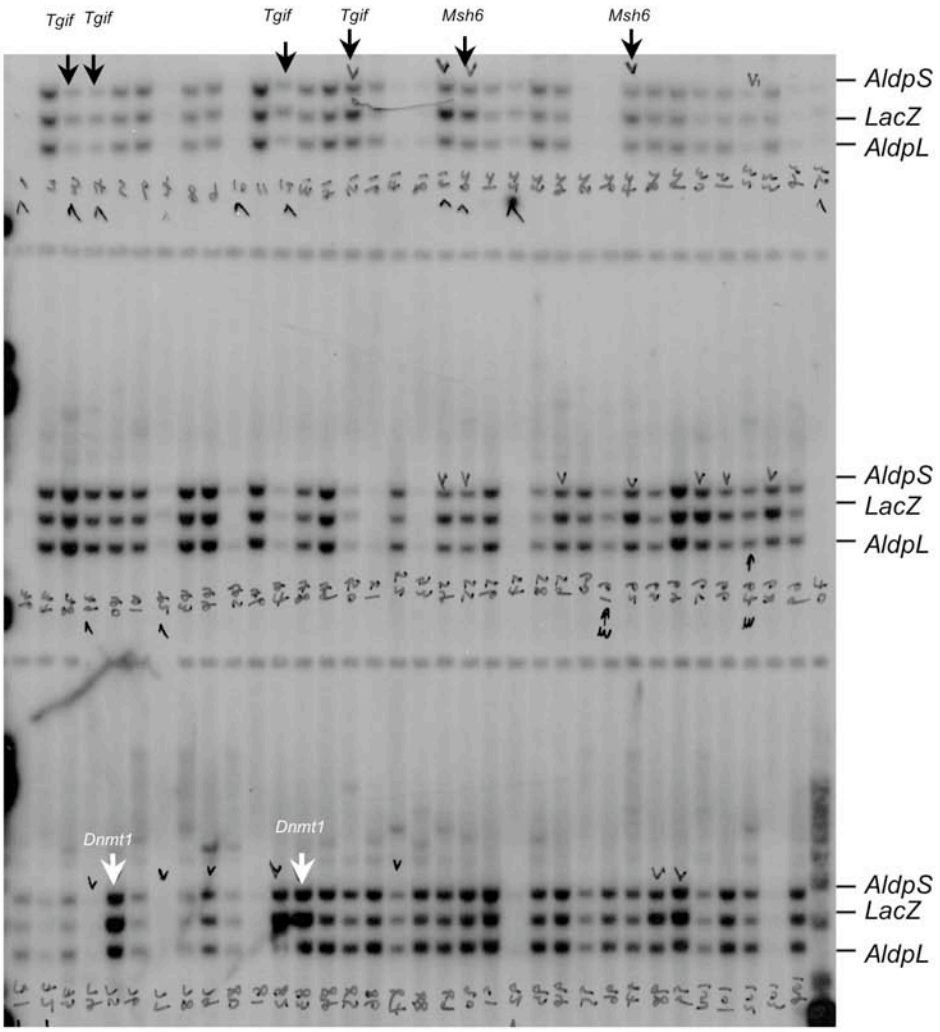


Figure 5-7. QTSouthern analysis of gene trap mutants obtained in the STB screen.

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	–

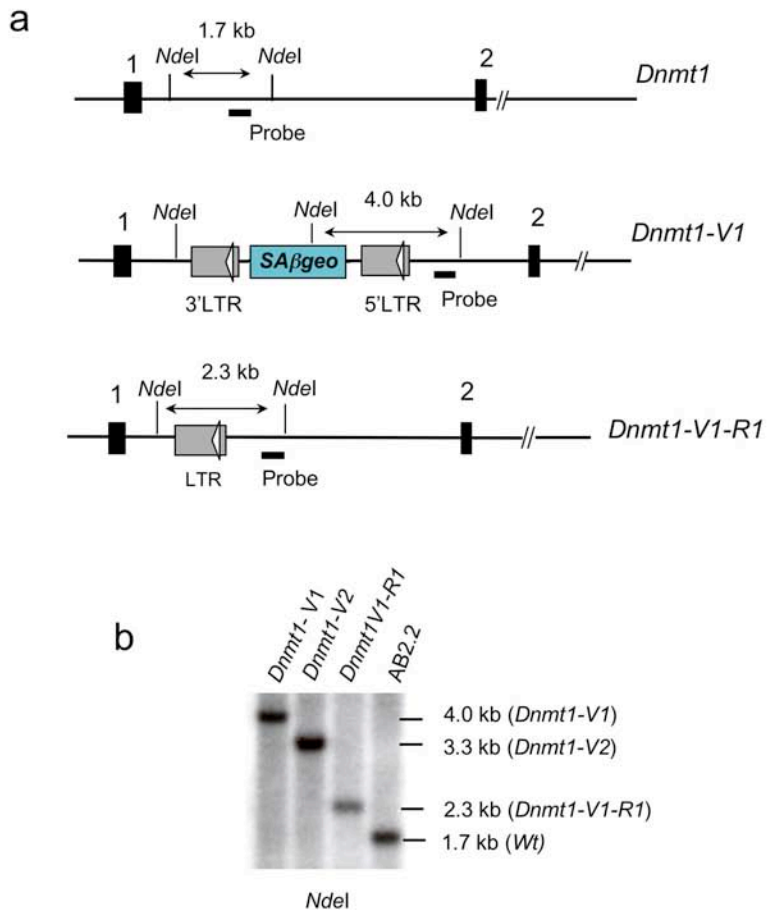


Figure 5-8. Southern-blot analysis of *Dnmt1* gene trap mutations

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.

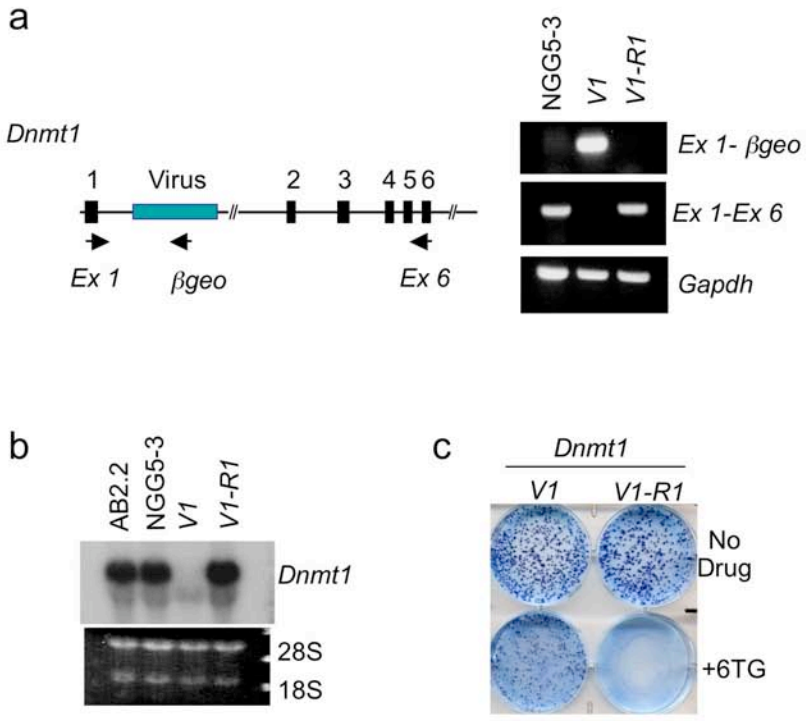


Figure 5-9. Molecular analysis of gene trap *Dnmt1* mutants

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).

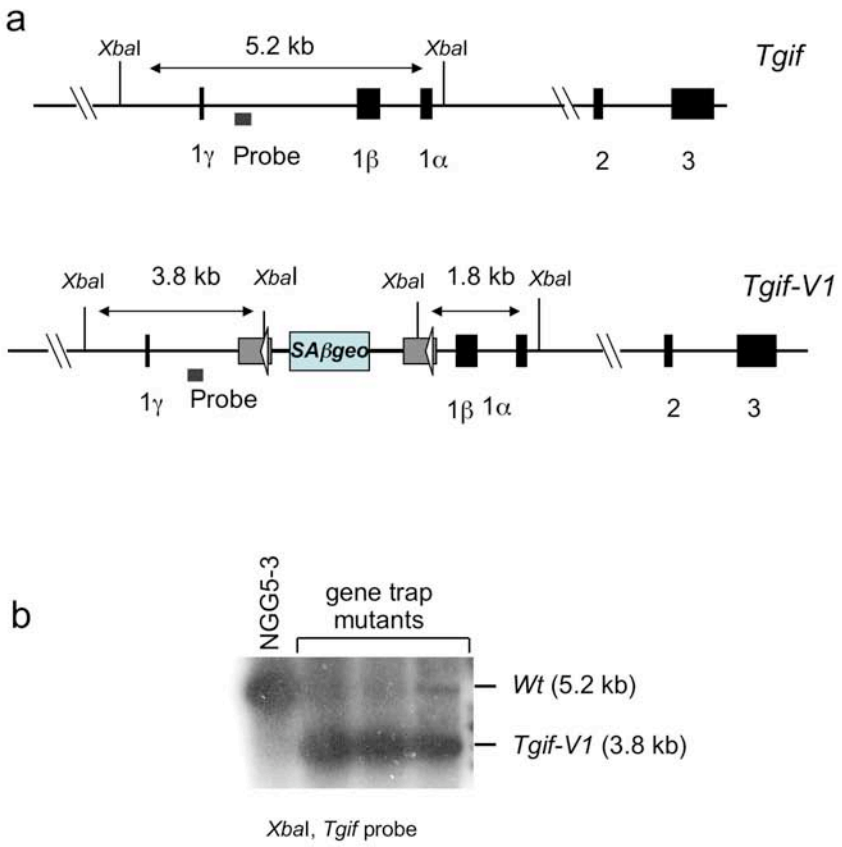


Figure5-10. Southern-blot analysis of *Tgif* gene trap mutants.

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.

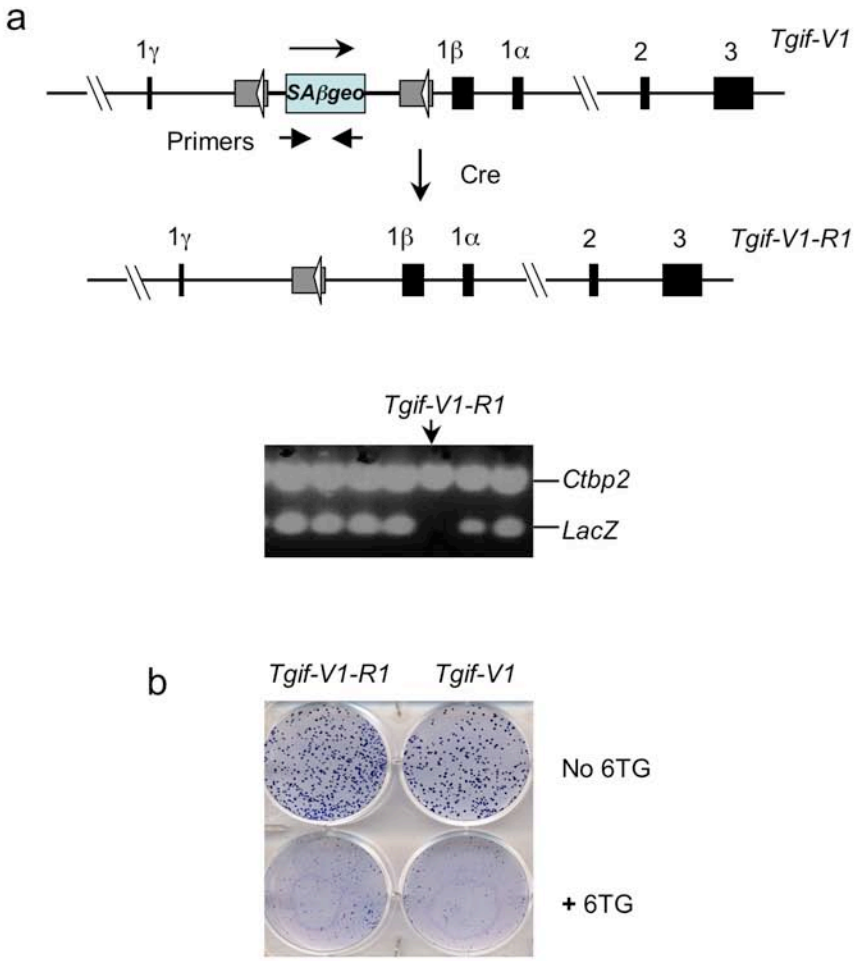


Figure 5-11. Cre reversal assay of *Tgif* gene trap mutation

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.*

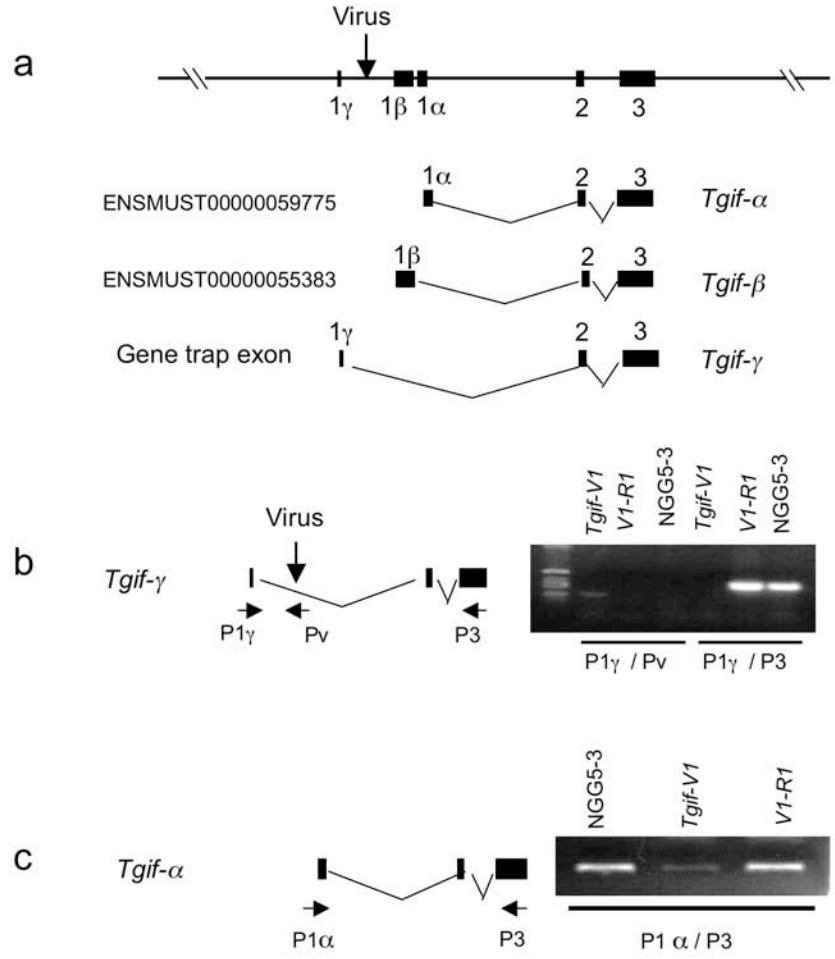


Figure 5-12. RT-PCR analysis of *Tgif* expression

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).

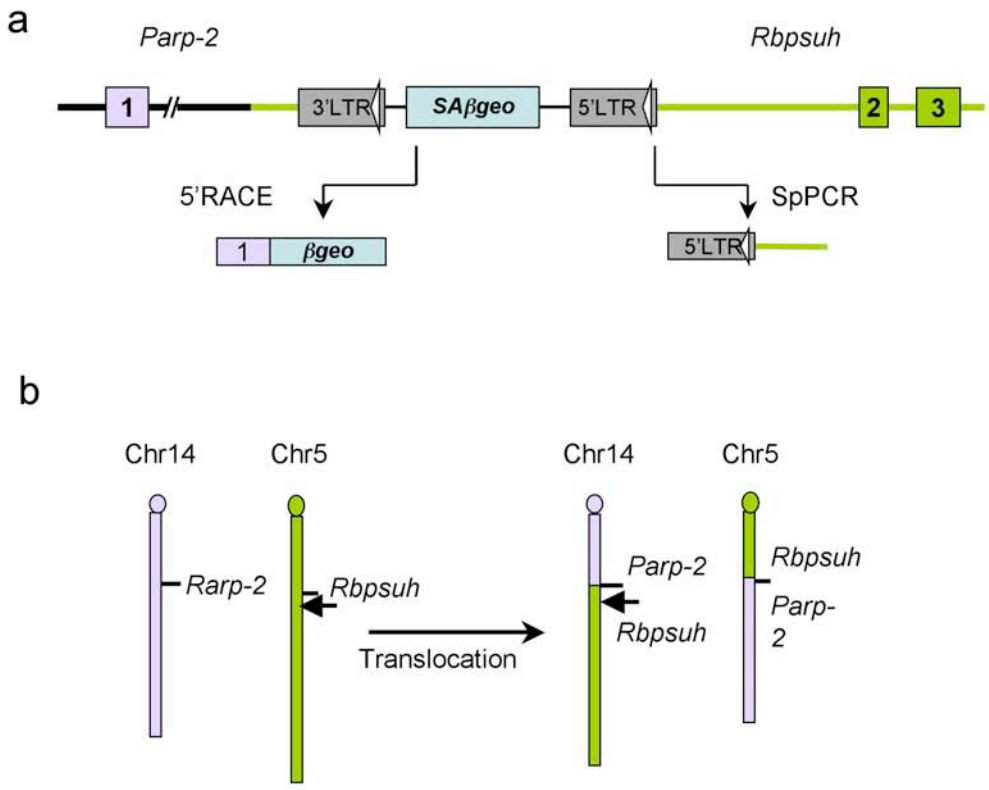
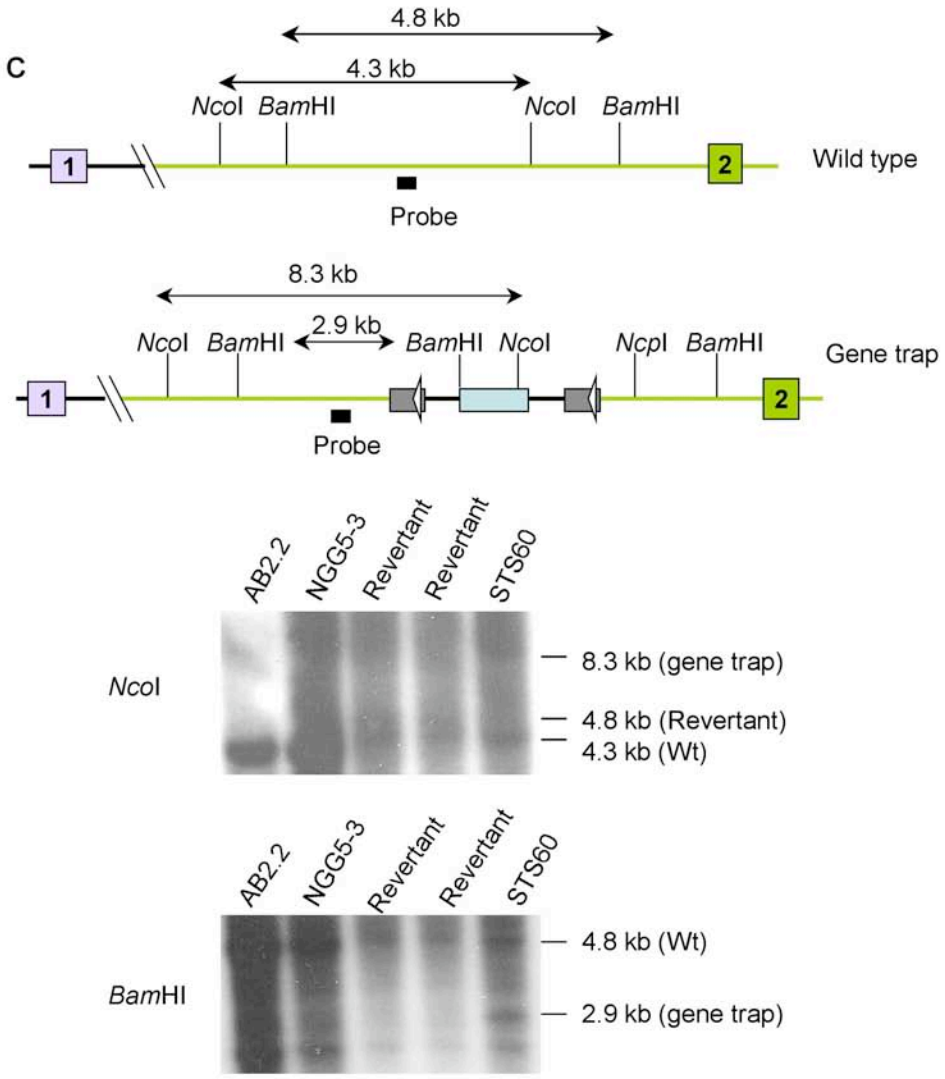


Figure 5-13. Characterization of a translocated *Parp-2/Rbpsuh* locus

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.

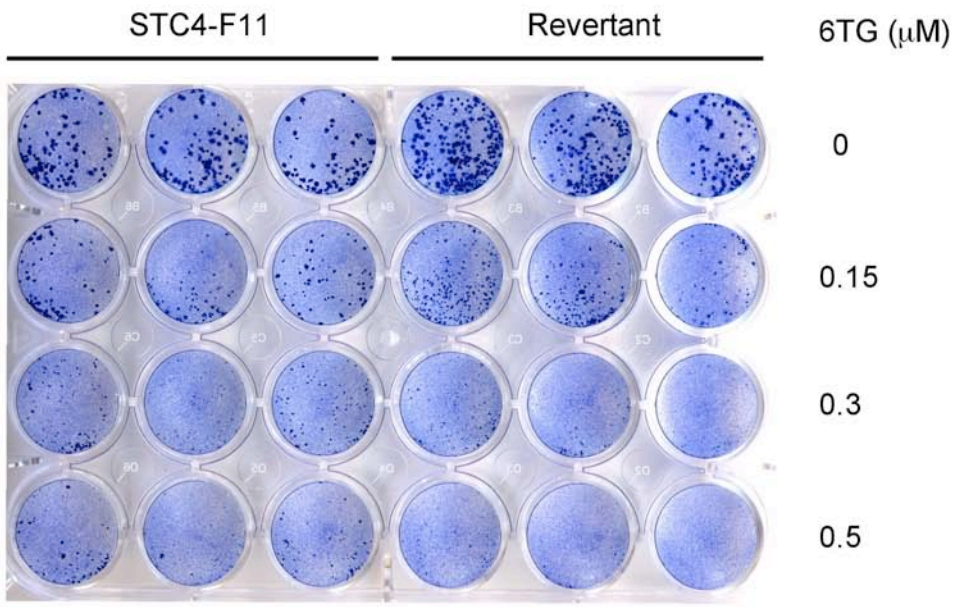


Figure 5-14. Characterization of the gene trap mutation in STC4-F11

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.

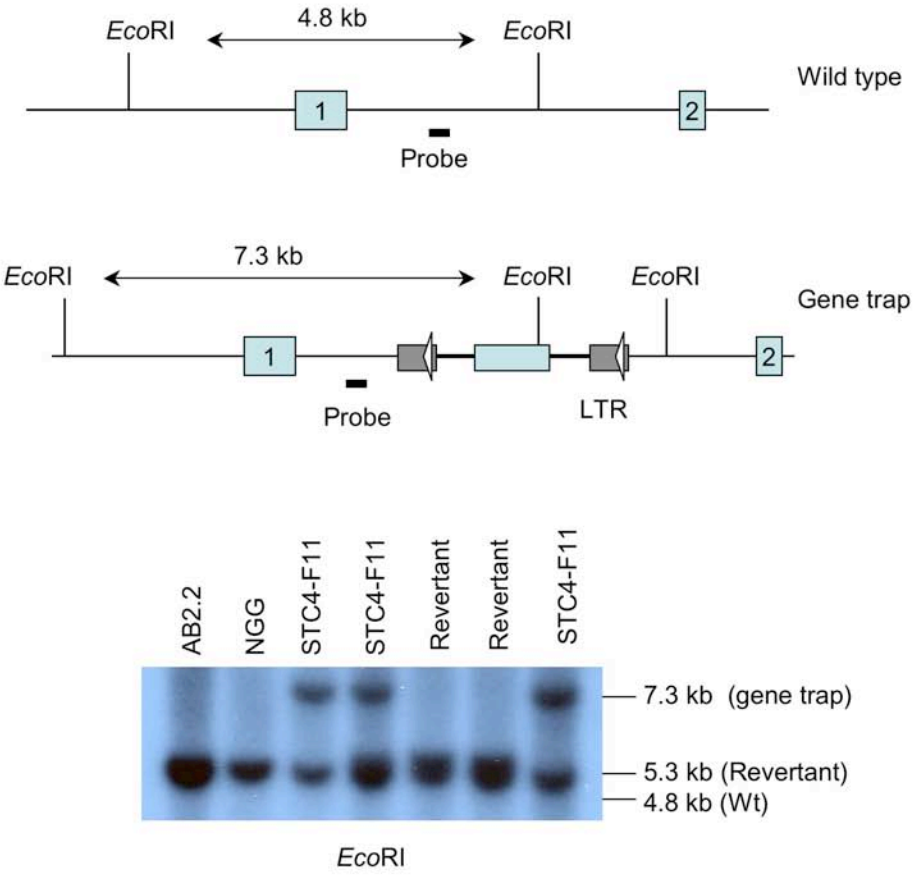


Figure 5-14. Characterization of the gene trap mutation in STC4-F11

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

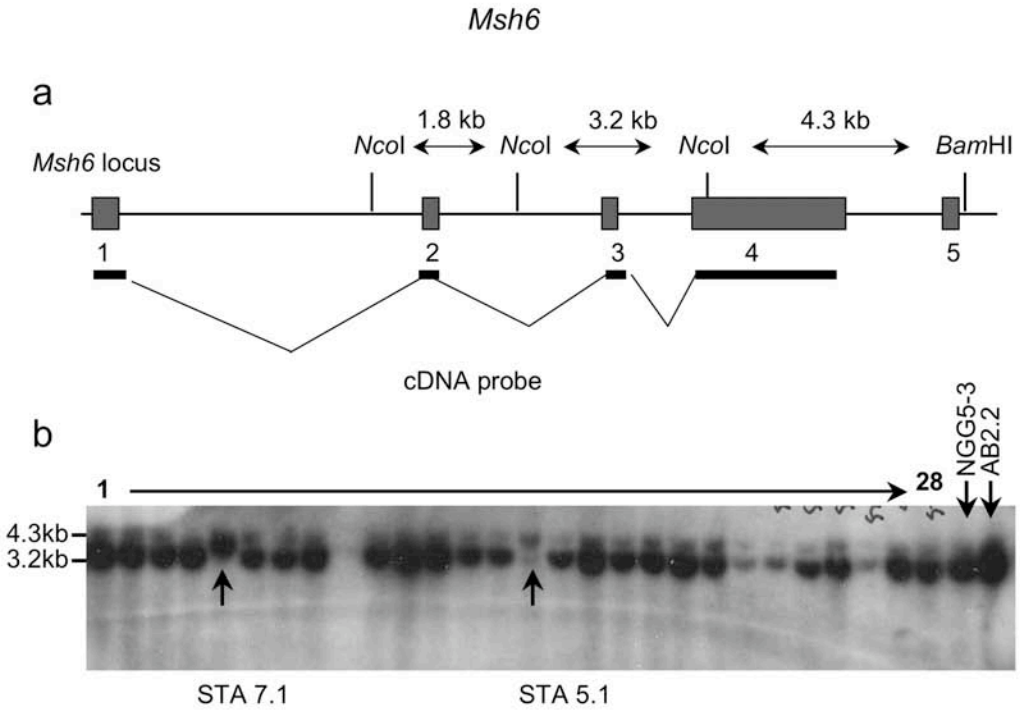


Figure 5-15. Southern analysis of *Msh6* locus in gene trap clones.

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

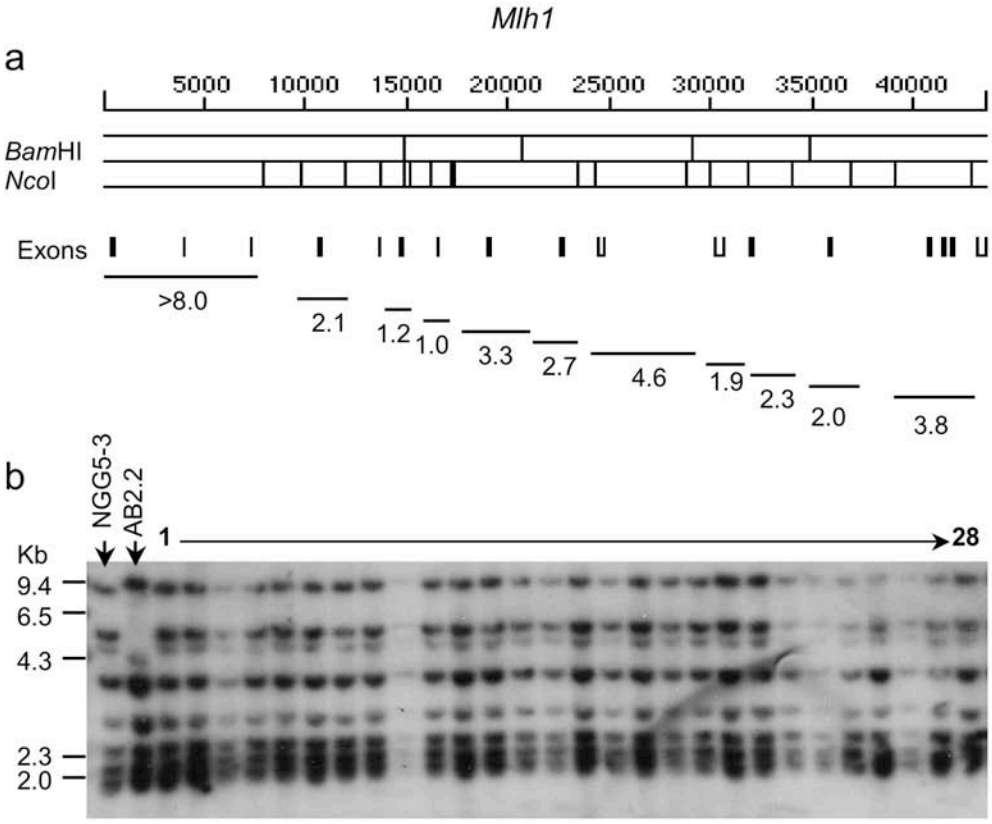


Figure 5-16 Southern analysis of *Mlh1* locus in gene trap clones.

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

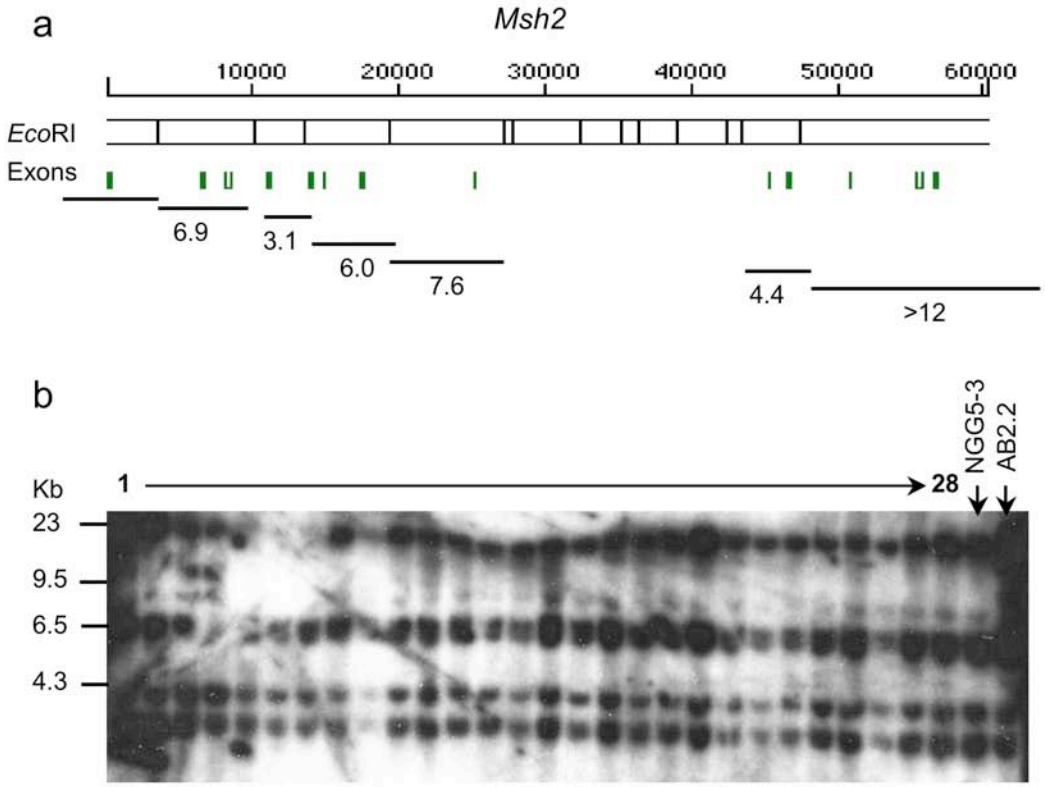


Figure 5-17 Southern analysis of *MSh2* locus in genetrap clones.

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

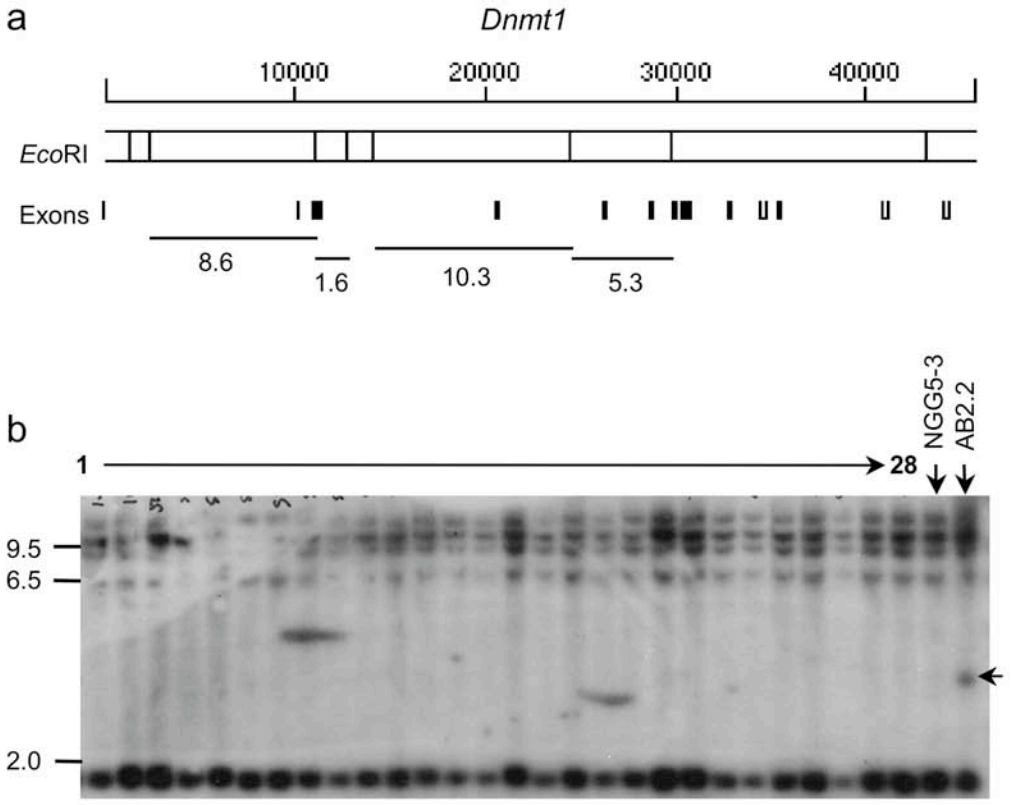


Figure 5-18 Southern analysis of *Dnmt1* locus in gene-trap clones

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.

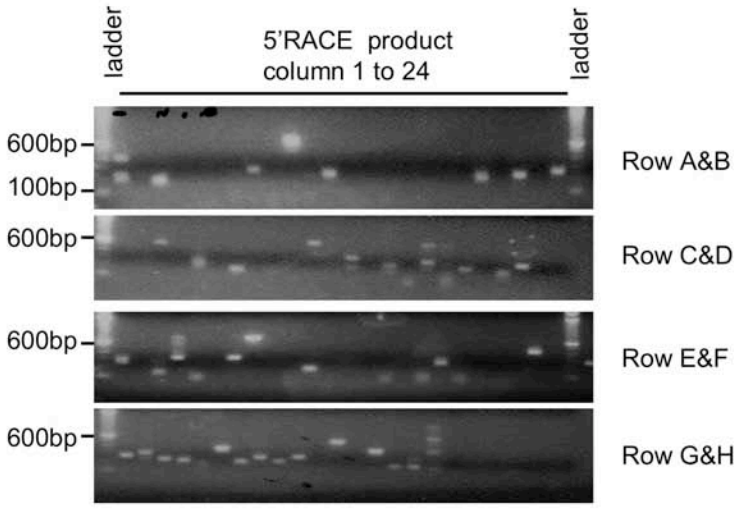


Figure 5-19. 5'RACE amplification of the splicing junction

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.

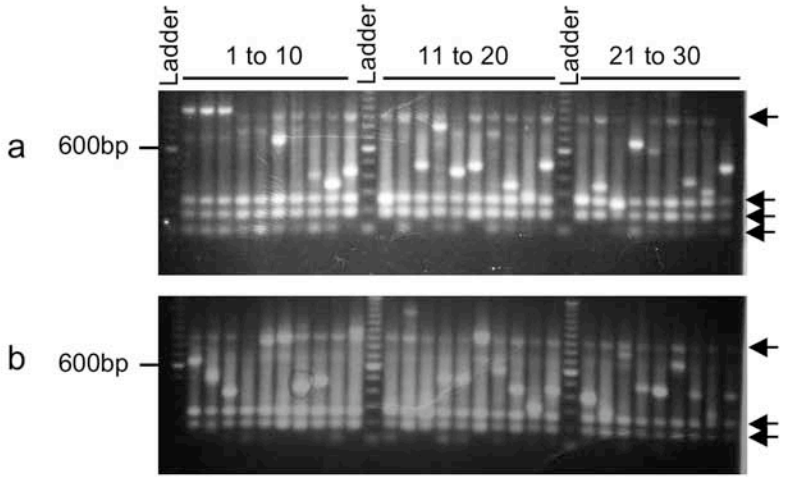


Figure 5-20. Proviral/host junctions identified by SpPCR

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.

6.1 Introduction

6.1.1 Increased spontaneous mutation in MMR deficient cells

Mutations in DNA mismatch repair (MMR) genes are known to affect a number of cellular processes in bacteria, yeast and mammals. The primary role of MMR is to repair mismatched nucleotide pairs generated during DNA replication resulting from DNA polymerase errors. A deficiency in the MMR system will lead to an increased spontaneous mutation rate, known as a "mutator phenotype". Mutator phenotypes reflect deficiencies in several different DNA repair processes, which include nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR) (Charames and Bapat, 2003). The mutator phenotype can be quantified by measuring the mutation rate of reporter genes. For example, ES cells with a mutation in the *Msh2* gene exhibited a 1,000 fold elevated mutation rate in a *HSV-TK* transgene introduced into the genome by gene-targeting (Abuin et al., 2000).

6.1.2 Microsatellite instability as a hallmark of MMR deficiency

Besides an elevated mutation rate, mismatch repair deficiency causes instability in simple sequence repeats (microsatellites) (Levinson and Gutman, 1987). During the replication of a simple sequence repeat, such as mono-,di-, tri- and tetranucleotide repeats, the nascent strand may slip along the template, leading to a bulged mispaired insertion-deletion loop (IDLs), which is a substrate for the DNA mismatch repair system (Parker and Marinus, 1992). Deficiency in the MMR system results in a change in the length of these simple repeats, a phenomena called microsatellite instability (MSI). Simple sequence repeats are abundant in eukaryote genomes, so microsatellite instability has been wildly used as an indicator for MMR deficiency in humans and in mice. In order to check MSI *in vivo*, the endogenous microsatellite repeats are normally recovered by PCR-amplification and sequenced to discover the changes in sequence of the repeats.

However, this method doesn't allow MSI to be quantified and low level of MSI could be easily missed. For example, *Msh6* mutant mice don't exhibit obvious MSI when examined by this method (Edelmann et al., 1997).

An alternative method for examining MSI is by introducing a slippage reporter construct into cells to monitor the MSI. This type of reporter carries a series of short tandem repeats that places the reporter gene out of its reading frame, so that the reporter gene is silent. Slippage mutations that gain or lose repeat units may place the reporter gene in-frame, which can then be identified by selecting for the expression of the reporter gene. This method has been applied to bacteria, yeast and has been used in mammalian cell lines (Farber et al., 1994, Yamada et al., 2003). Abuin et al (2000) constructed a gene-targeting vector carrying a di- nucleotide repeat [poly (CA/GT)] that disrupts the reading frame of a downstream neomycin gene. He targeted this slippage construct into a specific genomic locus and determined the MSI rate by Luria-Delbruck fluctuation analysis and found that this rate was elevated four orders of magnitude in *Msh2*-deficient ES cells and 15 fold in *Msh3*-deficient ES cells (Abuin et al., 2000). Pothof et al (2003) carried out an RNA interference (*RNAi*)–based screen in *C. elegans* for genes that affect the stability of a series of mononucleotides $(A)_{17}$ repeats placed 5' of a fused *gfp/lacZ* (green fluorescence protein/ β-galactosidase) reporter. Many *C.elegans* genes that protect *C.elegans* genome against mutation were identified in this screen, including *C.elegans* homologues of *Msh2, Msh6, Mlh1* and *Pms2*.

6.1.3 Determination of mutation rate by Luria-Delbruck fluctuation analysis

Fluctuation analysis was designed by Luria and Delbruck in 1943 to test whether phage resistant bacteria arose from random mutations or from acquired hereditary immunity. They seeded virus sensitive bacteria at very low density in number of cultures and allowed each to grow to a final culture individually. The number of the phage resistant bacteria in each final culture was then counted.

Based on a random mutation hypothesis, each bacteria has a fixed probability of becoming resistant to phage during each population doubling. This is referred to as the "mutation rate". Thus, the number of the phage resistant bacteria in each culture is the result of a combination of the random mutations that occur at each population doubling and the replication of existing mutants during bacterial growth. The number of mutants will not distribute following Poisson's law, but will vary greatly between each culture, this is called fluctuation. This is actually what Luria and Delbruck observed in their study. By calculating the mean number of mutants, or the proportion of cultures with no mutants, the mutation rate could be obtained by two statistical calculations, which are referred to as Luria and Delbruck method of means, or the P_0 method (luria, 1943). Luria-Delbruck analysis has become one of the most popular methods used to measure the mutation rate in mammalian cells. To determine the mutation rate of cells in a culture by fluctuation analysis, a group of parallel cultures is established with a predetermined small number of cells, ideally at a single cell density. These cells are expanded to a larger number, counted and plated in selective medium to allow the growth of resistant cell colonies. The resistant colonies in each culture are counted and used to calculate the mutation rate according to the mathematical equations provided by Luria-Delbruck.

6.1.4 DNA mismatch repair deficiency leads to increased homologous recombination between diverged sequences.

The DNA mismatch repair machinery is implicated in blocking recombination between diverged sequences (homeologous recombination) in bacteria, yeast and in mice. It has been proposed that the intermediate product of homeogolous recombination contains mismatched nucleotides, which can be recognized by the MMR system (Bailis and Rothstein, 1990, Selva et al., 1995). Evidence that MMR blocks homeologous recombination in the mouse came from gene-targeting experiments performed by homologous recombination. In this experiment, the retinoblastoma (*Rb*) locus in mouse ES cells were targeted using targeting

constructs made with either isogenic genomic DNA or with non-isogenic genomic DNA (te Riele et al., 1992). The non-isogenic targeting construct contained 0.6% base sequence divergence compared to the isogenic targeting construct. In *Msh2*-proficient ES cells, gene-targeting with the non-isogenic targeting constuct was blocked, resulting in a significant decrease in targeting efficiency compared to gene-targeting experiments with the isogenic targeting construct. In an *Msh2*-deficient cell line, homologous recombination with the non-isogenic *Rb* construct was as efficient as with the isogenic constuct (de Wind et al., 1995, Claij and Te Riele, 2002). The effects of mismatch repair proteins Msh2 and Msh3 on preventing homeologous recombination were also assessed by comparing the gene-targeting efficiencies with isogenic and non-isogenic gene-targeting vectors at the *Hprt* locus. In *Msh2*$^{-/-}$ and *Msh2, Msh3* double null cells, the targeting efficiency with non-isogenic vector was comparable to the targeting efficiency with isogenic targeting vectors. However, *Msh3*$^{-/-}$ deficiency itself didn't show a defect in homologous recombination (Abuin et al., 2000).

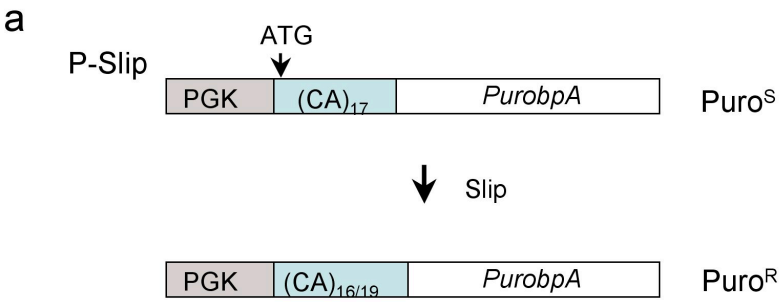
In Chapter 5, potential mismatch repair mutants have been identified in genetic screens for gene trap mutants that are resistant to 6TG, taking advantage of the fact that MMR deficiency causes tolerance to this DNA damaging drug. Many *Msh6* mutants were identified. Other than *Msh6*, novel gene trap mutations were recovered exhibiting tolerance to 6TG. Among these clones is gene trap mutation in *Dnmt1.* The gene trap *Dnmt1* mutant (*Dnmt1-V1*) is a bi-allelic mutant and the expression of *Dnmt1* is fully blocked by the gene trap insertion. The 6TG resistance phenotype was revertible by Cre-mediated deletion of the inserted retrovirus. In this chapter, the function of *Dnmt1* in mismatch repair process is further investigated.

6.2 Results

6.2.1 Construction of the P-Slip slippage cassette

The slippage construct that contains di-nucleotides repeats $(CA)_{17}$ in front of the neomycin reporter gene (*Neo*) has proved to be very useful in examing microsatellite instability in MMR deficiency cells (Farber et al., 1994, Abuin et al., 2000). The gene trap mutants in this study already carry the *βgeo* reporter that renders cells neomycin resistant. To be able to measure MSI activity, a slippage cassette, P-Slip, was designed to contain a series of di-nucleotides repeats following the initiation codon of puromycin phosphotransferase (*puro*) and place it out of it's correct reading frame. Cells harbouring the P-Slip cassette should be sensitive to puromycin due to the frame-shift mutation. If a change in the size of the repeat occurs during the growth of the cells, the reading frame of the P-Slip may be reconstituted, a functional puromycin phosphotransferase protein will be produced and the cell will acquire resistance to puromycin. Therefore, P-Slip allows the determination of the rate of MSI by measuring the number of puromycin resistant colonies.

P-Slip was assembled with a PCR amplified PGK promoter, a *purobpA* fragment lacking the ATG initiation codon, and a strand of synthesized oligonucleotide containing seventeen (CA) repeats ($(CA)_{17}$). The strand of $(CA)_{17}$ repeats was placed just downstream of the ATG start codon, leading to an out of frame mutation in the puromycin cassette (Fig. 6-1). A control vector (P-Slip-ON) was constructed with the same strategy in parallel with P-Slip construct. In contrast to the out of frame P-Slip cassette, P-Slip-ON carries sixteen (CA) repeats, so that the puromycin expression cassette is in frame. P-Slip-ON mimics one type of slippage mutation and thus provides a positive control for the P-Slip cassette. Cells carrying P-Slip-ON cassette should be resistant to puromycin. To check the function of the P-Slip vectors, P-Slip, P-Slip-ON and the parental



b

(CA) 17 repeat

ATGGGGACCGAGTACggatccaacacacacacacacacacacacacaca

cacacacagatacactagtAAGCCCACGGTGCGCCTCGCCACCCGC

GACGA

Figure 6-1. Construction of the slippage vector, P-Slip

a. Schematic representive of P-Slip structure. P-Slip contains a di-nucleotide repeats $(CA)_{17}$ that was inserted in puromycin phosphotransferase(puro) following the ATG , placing the puromycin phosphotransferase out of its normal reading frame. Cells carrying the P-Slip cassette are sensitive to puromycin selection($Puro^S$). Changes in the size of the repeat could place the puromycin phosphotransferase protein in frame and the cell will acquire resistance to Puromycin ($Puro^R$). **b.** The inserted $(CA)_{17}$ repeat (blue colored font) in puromycin coding fragment (black colored font). The $(CA)_{17}$ repeat contains 53 basepairs, which lead to a +2 mutation. Loss of one CA repeat or gain of two repeat units will then restore the reading frame of *Puro*.

PGKpurobpA cassette were linearized and introduced into AB2.2 cells by electroporation. Puromycin resistant ($Puro^R$) cells were selected with 3 μM puromycin for 8 days. $Puro^R$ ES cell colonies were recovered from ES cells with P-Slip-ON and the parental PGKpurobpA, but not from ES cells carrying P-Slip. These results verified that CA repeat in front of the puromycin gene didn't interfere with its activity.

6.2.2 Gene-targeting P-Slip in *ROSA26* locus

In order to provide a precise comparison of the MSI rate between different cell lines, a gene targeting strategy was used to introduce a single copy of P-Slip cassette into a specific genomic locus, *ROSA26*. By using this strategy variation of the MSI rate caused by the copy number of the slippage cassette and the positions effects could be avoided. The *Rosa26* locus was originally identified by retroviral gene trapping in ES cells (Friedrich and Soriano, 1991, Zambrowicz et al., 1997). This locus has been frequently employed for expressing exogenous genes because of the broad spectrum of expression revealed by the gene trap reporter. Also, gene targeting is highly efficient at this locus (Soriano, 1999). Efficient gene-targeting at this locus is important in our study because it allows the P-Slip cassette to be efficiently introduced into multiple cell lines. To make a gene-targeting vector for the *ROSA26* locus (*ROSA26*/Slip-TV), the P-Slip cassette and a PGKBSD selection cassette were inserted between two *ROSA26* genomic arms, of 2.5 kb and 1.9 kb in length. The gene-targeted allele can be identified by Southern analysis using a 5' external probe, which recognizes the targeted allele as a 5.8 kb *Eco*RI fragment and the wild type allele as a 15.5 kb *Eco*RI fragment. Figure 6-2 illustrates the gene-targeting strategy at the *ROSA26* locus and the Southern blot analysis demonstrating three cell lines carrying the targeted allele.

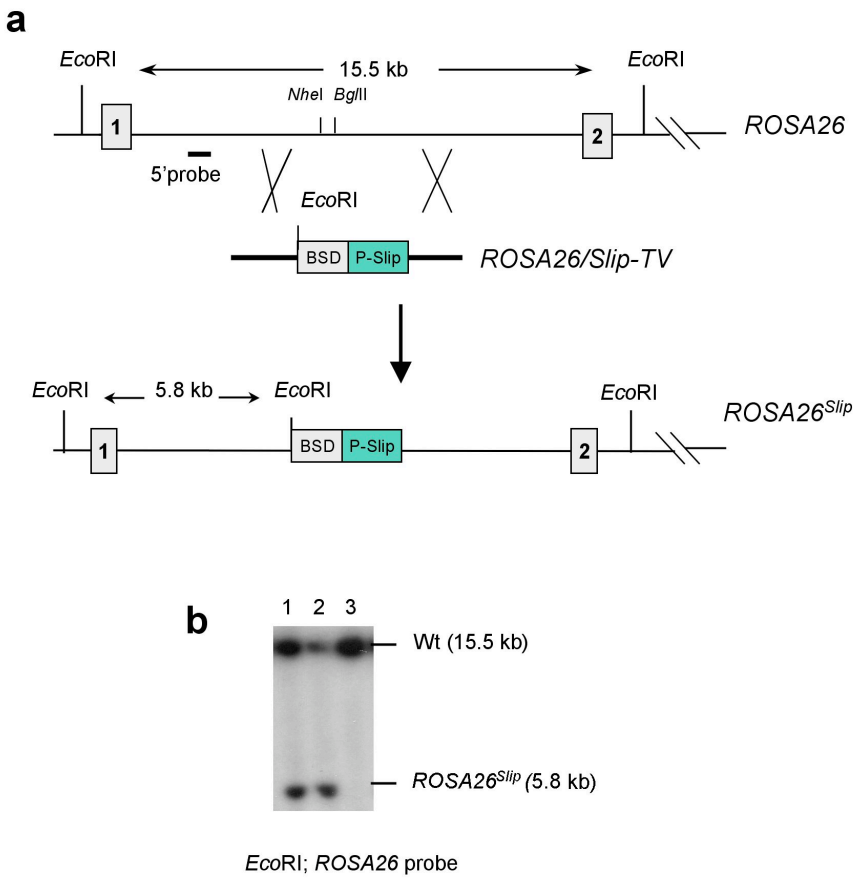


Figure 6-2. Gene-targeting of P-Slip cassette into *ROSA26* locus

a. Schematic of gene-targeting strategy showing the 5' portion of *ROSA26* locus and the gene-targeting vector (*ROSA26/Slip-TV).*
b. Southern-blot showing targeted cells carrying *ROSA26*^{Slip} allele(lane 1 and lane 2).

6.2.3 Determination of MSI rate of the targeted P-Slip cassette in gene trap mutants

The *ROSA26*/Slip gene-targeting vector was targeted into gene-trap mutants that contained the gene-trap cassette in *Msh6, Dnmt1, Tgif,* and *Adprtl2/Rbpsuh* loci. These clones were identified from STB screens (Chapter 5, Table 5-2), and all exhibited resistance to 6TG(2 μM). Both *Msh6* and *Dnmt1* gene-trap mutants are homozygous mutations and the 6TG resistance phenotype is revertible. The *Tgif* gene-trap mutant (*Tgif-V1)* is also a bi-allelic mutant, but the 6TG resistance phenotype is not revertible. The *Parp-2/Rbpsuh* is a complex locus involving two genes, *Parp-2* and *Rbpsuh*. Chromosomal translocation between *Parp-2 and Rbpsuh* loci causes a frame-shift mutation in both genes. The 6TG resistance phenotype is not revertible. The *Blm*-deficient cell line NGG5-3 and a *Blm* wild-type cell line (AB2.2) were used as controls. After confirmation of gene-targeting by Southern analysis using a 5' *ROSA26* external probe, targeted cells from each of the gene-trap cell line were seeded at low density and allowed to form ES cell colones. 12 clones from each cell line were picked into 96 well tissue culture plate and expanded for Luria-Delbruck fluctuation analysis to determine the MSI rate (Methods 2.9). The results are listed in table 6-1. The gene trap *Msh6* mutant exhibited a MSI rate about 10 fold higher than AB2.2 cells. The *Blm*-deficient cells (NGG5-3) exhibited a 2 fold increase in MSI compared to AB2.2 cells. The MSI rate in *Tgif-V1* cells is slightly higher than the *Blm*-deficient cells. However, no obvious change of the MSI rate was observed in the gene trap *Dnmt1-V1*, and *Parp-2/Rbpsuh* cells.

6.2.4 Generation and characterization of Dnmt1-deficient ES cells by gene-targeting

The gene-trap clones were generated on a *Blm*-deficient genetic background. The observation of the microsatellite instability in *Blm*-deficient ES cells raised

Table 6-1. Mutation rate of P-Slip in gene-trap mutants

ES cell lines	Number of selections	Total number of cells	Number of $Puro^R$ clones per select *	Rate of MSI
AB2.2 /#1	12	$7.0x10^7$	16,16,13,5,3,2,0(6)	$2.6 x10^{-7}$
NGG5-3 /#1	12	$4.5 x10^7$	47,11,9.6,5(3),4,3,2(3)	$6.3 x10^{-7}$
Dnmt-V1 /#1	12	$4.6 x10^7$	33,17,15,8,6,4,2,0(5),	$5.6 x10^{-7}$
Tgif-V1/ #1	12	$3.8 x10^7$	72,15,12,11,8,7,6,5(3),2,0	$1.06 x10^{-6}$
STB60/#1	12	$3.5 x10^7$	23(2),5,4,2(4),1(3),0	$6.1 x10^{-7}$
STB20 (*Msh6*) /#1	12	$4.4 x10^7$	152,129,125 100,73,45,44,36,34,30,22,11	$3.6 x10^{-6}$

Table 6-1. Mutation rate of P-Slip in gene-trap mutants

P-Slip was targeted into *ROSA26* locus, the rate of MSI was determined by Luria-Delbruck fluctuation analysis in four gene-trap cell lines, *Dnmt1-V1*, *Tgif*, STB60 (*Parp-l2/Rbpsuh*) and STB20 (a *Msh6*mutant). AB2.2 and NGG5-3 cells were used as controls. " #1" represents the experiment number.

* The order of numbers was arranged arbitrarily from the most to the fewest. The number in parentheses indicates the number of selections giving rise to the corresponding number of drug resistant clones.

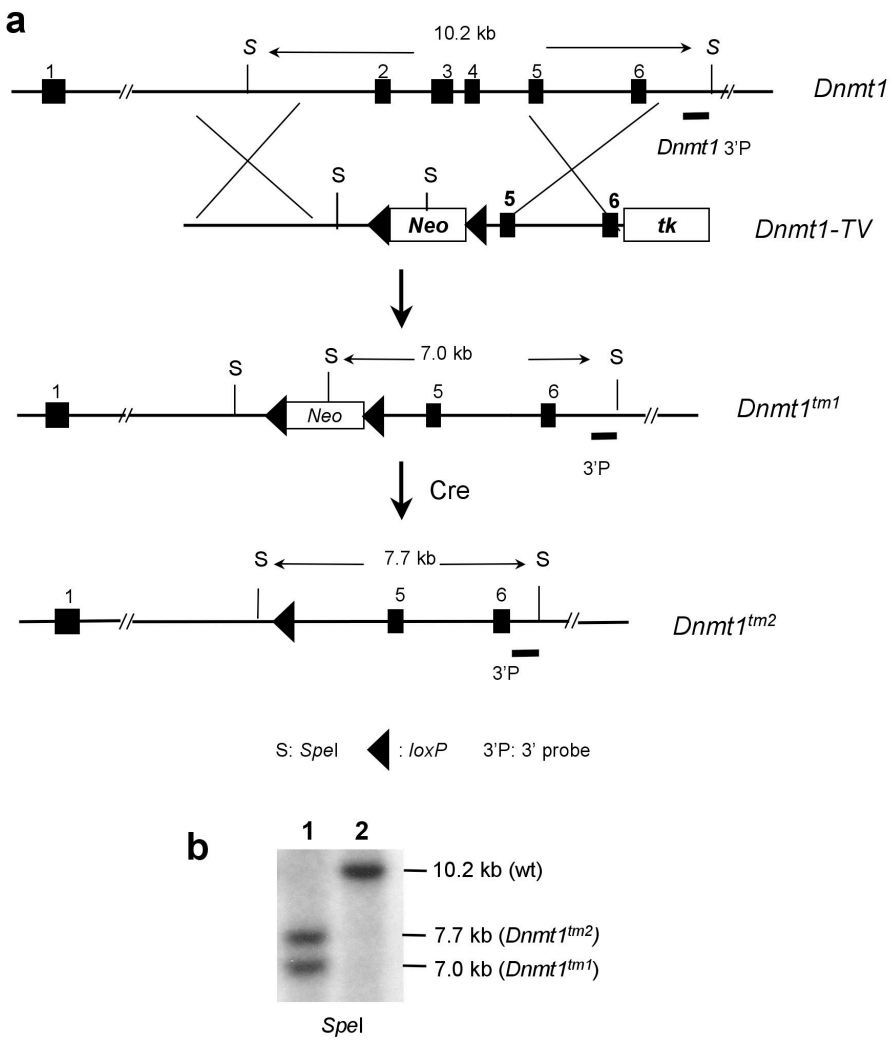


Figure 6-3. Generation *Dnmt1* knockout cells by gene-targeting

a. Schematic of gene-targeting *Dnmt1* locus and the targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles. **b**. Southern blot showing a homozyously gene targeted *Dnmt1-KO* cells carrying *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles. Lane 1, *Dnmt1*$^{tm1/tm2}$; Lane 2, wild type cells

the concern of a genetic interaction between mismatch repair genes and the *Blm*-deficient background, which could interfere with the analysis of MSI in the gene trap clones. To examine the MSI activity without the effect of *Blm*-deficiency, a *Dnmt1*-deficient cell line was generated on the AB2.2 genetic background by gene-targeting. This cell line was named as *Dnmt1-KO* for *Dnmt1* knock out cell line. *Dnmt1-KO* cells contain two targeted alleles, $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$, in which a 5.5 kb DNA fragment including *Dnmt1* exon 2, exon 3 and exon 4 were deleted (Fig. 6-3).

6.2.4.1 Expression analysis of Dnmt1

Dnmt1 encodes a 5.2 kb messenger RNA which derived from 37 exons. Deletion of exon 2, exon 3 and exon 4 in *Dnmt1-KO* cells removes a 359 bp coding region, which results in a frame-shift mutation (Fig. 6-4). Thus, *Dnmt1-KO* cell line is expected to be a null. A pair of PCR primers was designed to amplify a *Dnmt1* cDNA fragment spanning *Dnmt1* exon 1 to exon 6, which is 615 bp in length. In *Dnmt1-KO* cells, a 256 bp DNA fragment will be amplified because of the 359 bp deletion. RT-PCR analysis revealed the predicted 615 bp and 256 bp fragments in heterozygous gene-targeted $Dnmt1^{+/tm1}$ cells and only a 256 bp fragment in the targeted *Dnmt1-KO* ($Dnmt1^{tm1/tm2}$) cells (Fig. 6-5 a). This result confirmed the knockout mutation created by gene-targeting.

To examine the expression of *Dnmt1*, a cDNA probe was PCR amplified spanning *Dnmt1* exon 8 to exon 39. Northern-blot analysis was carried out on total RNA extracted from $Dnmt1^{+/tm1}$, *Dnmt1-KO*, gene-trap *Dnmt1-V1* and AB2.2 cells. This experiment demonstrated that *Dnmt1* is expressed in the heterozygous and homozygous knockout cells at similar level to the wild type cells. Because *Dnmt1* mRNA is over 5 kb, a 359 bp deletion in the gene-targeted alleles is not resolved by Northern-blot analysis. *Dnmt1* expression could not be detected in the gene-trap *Dnmt1-V1* cells (Fig. 6-5 b).

TCGCGCGAAAAAGCCGGGGTCTCGTTCAGAGCTGTTCTGTCGTCTGC
AACCTGCAAGATGCCAGCGCGAACAGCTCCAGCCCGAGTGCCTGCG
CTTGCCTCCCCGGCAGGCTCGCTCCCGGACCATGTCCGCAGGCGgctc
aaagacttggaaagagatggcttaacagaaaaggagtgtgtgagggagaaattaaacttactgcatg
aattcctgcaaacagaaataaaaagccagttgtgtgacttggaaaccaaattacataaagaggaattat
ctgaggaaggctacctggctaaagtcaagtccctcttaaataaggatttgtccttggagaacggaacac
acactctcactcaaaaagccaacggttgtcccgccaacgggagccggccaacctggagagcagaaa
tggcagactcaaatagatccccaagatccaggcccaagcctcggggacccaggagaagcaagtcg
gacagtgacacccttt**CAGTTGAAACTTCACCTAGTTCCGTGGCTACGAGGAG**
AACCACCAGGCAGACCACCATCACGGCTCACTTCACGAAGGGCCCCA
CTAAACGGAAACCCAAGGAAGAGTCGGAAGAGGGGAACTCGGCTGAG
TCGGCTGCAGAGGAGAGAGACCAG

Figure 6-4. 5' portion of *Dnmt1* coding sequence.

The deletion in the knockout cell line, *Dnmt1-KO,* is represented as lowercase fonts and underlined, which includes 356 base pairs. The ATG codon is in red. The deletion results in a frame shift mutation.

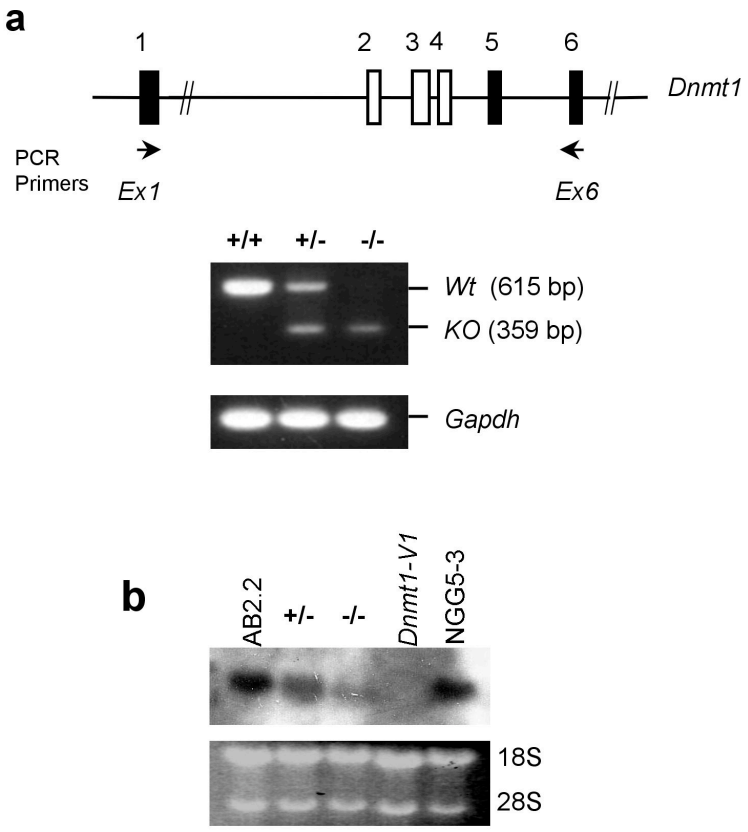


Figure 6-5. Expression analysis of *Dnmt1*

a. RT-PCR amplification of *Dnmt1* cDNA spanning exon 1 to exon 6, demonstrating the deleted transcript (359 bp) in the heterozygous targeted *Dnmt1*$^{+/tm1}$ *(+/-)* and double targeted *Dnmt1*$^{tm1/tm2}$*(-/-)* cells. Note that Exon 2, exon 3 and exon 4 are deleted in the gene-targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles (open boxes). *Gapdh,* RT-PCR control. **b.** Northern analysis showing *Dnmt1* expression in the gene trap *Dnmt1-V1* cells, the heterozygous targeted *Dnmt1*$^{+/tm1}$ (+/-) cells, double targeted *Dnmt1*$^{tm1/tm2}$ cells (*Dnmt1-KO*; +/-), and NGG5-3 (*Blm*-deficient) cells. Note that the expression of *Dnmt1* is absent in *Dnmt1-V1* cells. Expression of *Dnmt1* is reduced in *Dnmt1-KO* cells compared to *Dnmt1* wild type AB2.2 and NGG5-3 cells. 18s and 28s, loading controls.

6.2.4.2 Functional analysis of Dnmt1 activity

Dnmt1 protein encodes a cytosine-5 methyltransferase activity, which transfers methyl groups to cytosine at CpG sites (Li et al., 1992). The methylation level at CpG sites can be examined by comparing the restriction digestion pattern generated by restriction enzymes, *Msp* I and *Hpa* II. *Msp* I and *Hpa* II are isoschizomers that recognize the CCGG sites, except that *Hpa* II is sensitive to methylated cytosine at CCGG sites (Fig. 6-6 a). The CpG site in the CCGG sequence is subjected to methylation by Dnmt1. Therefore, genomic DNA with a higher level of CpG methylation will be more resistant to *Hpa* II digestion (Chapman et al., 1984). To measure Dnmt1 activity, genomic DNA was extracted from heterozygous gene-targeted *Dnmt1*$^{+/tm1}$, homozygous gene-targeted *Dnmt1*$^{tm1/tm2}$, gene-trap *Dnmt1-V1* cells and *Dnmt1* wild type control AB2.2 cells and digested with *Msp* I and *Hpa* II, respectively. The digested genomic DNA was then separated on agarose gel and visualized by staining with EtBr (ethidium bromide). Comparison of the *Msp* I and *Hpa* II digestion of each sample revealed that genomic DNA extracted from *Dnmt1*$^{tm1/tm2}$ and *Dnmt1-V1* cells are globally undermethylated compared to AB2.2 cells (Fig. 6-6 b).

A Southern-blot analysis was conducted to assess the methylation of centromeric repeats. Gamma satellite repeats are the major component of centromere repeats composed of a 234 bp repeat unit (Lundgren et al., 2000). Centromeric repetitive sequences are normally highly methylated. To investigate the methylation level at centromeric repeats, *Msp* I and *Hpa* II digested genomic DNA were blotted from an agarose gel to nylon membranes and probed with a gamma satellite repeats probe (Fig. 6-6 b). Southen analysis revealed that gamma satellite repeats were undermethylated in the gene-trap *Dnmt1-V1* cells since *Hpa* II digested genomic DNA exhibited the same defined hybridization pattern as *Msp* I digestion. However, the methylation of gamma satellite repeats

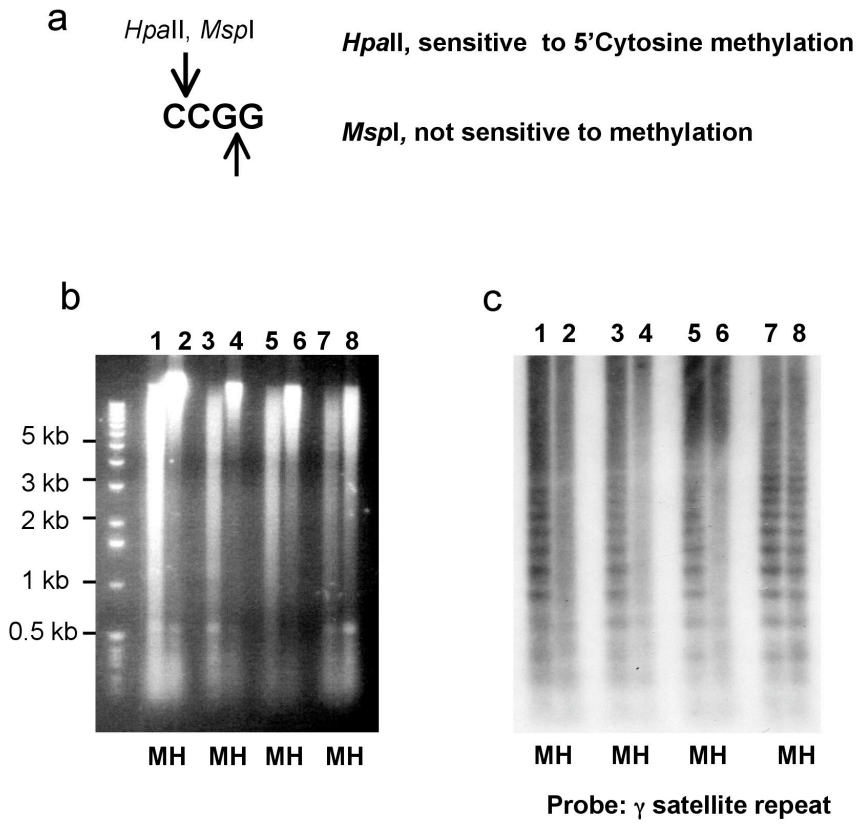


Figure 6-6. Determinatin of DNA methylation.

a. *Hpa*II and *Msp*I are isoschizomers that recognize the same CCGG sites. *Hpa*II is sensitive to the methylated CpG. Methylated CCGG can only be digested with *Msp*I, but not *Hpa*II. **b**. Genomic DNA digested with either *Msp*I or *Hpa*II (left panel) and probed with γ satellite repeat (right panel**).** *Msp*I digested genome: odd numbered lane labed with M; *Hpa*II digested genome: even numbers lanes labled with H. Genomic DNA are from *Dnmt1*-proficient AB2.2 cells (lane 1-2), *Dnmt1*$^{+/tm1}$(lane 3-4), *Dnmt1* $^{tm1/tm2}$ *(Dnmt1-KO*) (lane 5-6*)*, gene-trap *Dnmt1-V1* cells (lane 7-8).

Table 6-2. Mutation rate of P-Slip in *Dnmt1*- deficient cells

ES cell lines	Number of drug selections	Total number of cells	Number of Puro resistant clones per select *	Mutation Rate
AB2.2 /#2	24	$2.8x10^7$	10,3,21,1(3),0(16)	$3.2x10^{-7}$
AB2.2 /#3	24	$3.3x10^7$	29,10,5,2,1,0(19)	$5.0x10^{-7}$
NGG5-3 /#2	24	$1.9x10^7$	7,5,4,3,2(6),1(4),0(10)	$7.0x10^{-7}$
NGG5-3 /#3	19	$1.8x10^7$	4,3,2(5),1(3),0(9)	$5.0x10^{-7}$
Dnmt1-V1 /#2	24	$1.8x10^7$	6,2,(3),1(2),0(18)	$3.7x10^{-7}$
Dnmt1-KO/#2	24	$2.2x10^7$	84,7,7,4,3,2,2,1(5),0(12)	$1.4x10^{-6}$
Dnmt1-KO/#3	23	$3.0x10^7$	128,12(2),5,4(2),3(2),2(2),1(6),0(7)	$1.5x10^{-6}$

Table 6-2. Mutation rate of P-Slip in *Dnmt1*-deficient cells.

P-Slip was targeted into *ROSA26* locus, the rate of MSI was determined by Luria-Delbruck fluctuation analysis in *Dnmt1*-deficient cells. Dnmt1-V1 represents the gene-trap *Dnmt1* mutants. Dnmt1-KO represents the gene-targeted $Dnmt1^{tm1/tm2}$ cells. This experiment was done twice and the data was listed separately and marked as " #2" and " #3 " for the second, and third experiments, respectively. Note "#1" represents the Luria-Delbruck fluctuation analysis experiment listed in table 6-1

* The order of numbers was arranged arbitrarily from the most to the fewest. The number in parentheses indicates the number of selections giving rise to the corresponding number of drug resistant clones.

seemed to be maintained to a similar level in both heterozygous gene-targeted *Dnmt1*$^{+/tm1}$ and homozygous gene-targeted *Dnmt1*$^{tm1/tm2}$ cells as in *Dnmt1* wild type AB2.2 cells, which was demonstrated by the lack of the defined hybridization pattern on *Hpa* II digested genomic DNA (Fig. 6-6 c). These results suggest that the gene-targeted *Dnmt1*^{tm1} and *Dnmt1*^{tm2} alleles generate hypomorphic mutations of *Dnmt1*.

6.2.5 Dnmt1 deficiency in the Dnmt1 knockout cells causes increased MSI

To test the MSI rate in the gene-targeted Dnmt1-deficient cells, the P-Slip cassette was targeted into *ROSA26* locus in *Dnmt1-KO* cells. 20 or 24 single cell clones were obtained by low density plating from the P-Slip targeted *Dnmt1-KO* cells, gene-trap *Dnmt1-V1* cells, *Blm*-deficient cells and the AB2.2 cells. MSI in the targeted P-Slip were determined by Luria-Delbruck fluctuation analysis as described previously. This experiment was repeated once. The results are listed in table 6-2. The MSI activity in *Dnmt1-KO* cells is 3 to 5 fold higher that the MSI rate in AB 2.2 cells and about 2 fold higher than the rate in Blm-deficient cells. The gene-trap *Dnmt1-V1* cell line, which is on the *Blm*-deficient genetic background, however, exhibited a MSI rate lower than *Dnmt1-KO* cells. Because *Dnmt1* expression is fully blocked in *Dnmt1-V1* cells, the difference of MSI rate between *Dnmt1-V1* cells and *Dnmt1-KO* cells are less likely to be derived from residual function in *Dnmt1-V1* cells. Clearly, there is a genetic interaction between *Dnmt*1 and the *Blm gene*. In fact, methylation analysis has shown that *Dnmt1-KO* cells have partial Dnmt1 function. Thus, It would be interesting to know if a null *Dnmt1* mutation in AB2.2 cells would give a higher level of MSI.

6.2.6 The puromycin resistant ES cells carry a single copy of the targeted P-Slip cassette.

Gene-targeted ES cell clones are picked from many ES cells colonies growing on a 90 mm tissue culture plate. Since the gene-targeting efficiency is low, the

majority of the ES cells clones selected in an experiment are those carrying random insertions of the gene-targeting vector. Thus, the gene-targeted ES cell clones may be contaminated with small numbers of cells with randomly inserted transgenes. This contamination is normally ignored because generally less than 10 % of cells represent contaminants. Fluctuation analysis involves plating gene-targeted clones at low density to recover single cell clones, so there is the possibility that a single cell clone may be derived from cells that carry a randomly inserted P-Slip vector. As random insertion is often accompanied by a head-tail concatermerization of the gene-targeting vector, multiple copies of P-Slip cassette will be inserted into genome, logically this clone would have a higher likelihood of being able to restore puromycin expression. To ensure that the puromycin resistant slippage clones do contain the targeted P-Slip cassette at the *ROSA26* locus, 12 puromysin resistant clones were recovered from each cell line and genomic DNA were extracted. Southern analysis using *ROSA26* probe on these cells verified that the P-Slip cassette was targeted into *ROSA26* locus (data not shown).

6.2.7 Mutated CA repeat in Puromycin resistant ES cells

To determine the nature of the mutations in the $(CA)_{17}$ repeats element acquired to produce puromycin resistance, the P-Slip cassette was amplified by PCR from the recovered puromycin resistant ES cell clones and sequenced. Analysis of the sequence data revealed that all clones examined exhibited insertions or deletions in the $(CA)_{17}$ repeat region (Table 6-3). The most frequently identified mutation was the insertion of two extra (CA) repeats, which occurs in all genetic backgrounds, in different cell lines. This observation suggests that a “+2” insertional mutation at dinucleotide repeats may be a common DNA polymerase II error.

Table 6-3. Sequence analysis of P-Slip mutations

ES cells	Mutation
AB2.2	+2
AB2.2	+2
AB2.2	+2
AB2.2	+2
NGG5-3	-7
NGG5-3	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	-10
Dnmt1-KO	-7
Dnmt1-KO	+2
Dnmt1-KO	+2
Dnmt1-KO	-10

Table 6-3 Sequence analysis of P-Slip mutations.

Individual Puromycin resistant ES cell clones were recovered from AB2.2, NGG5-3, and the *Dnmt1-KO* cells that carrying the targeted P-Slip cassette. The (CA) repeat was recovered by PCR amplification from genomic DNA and sequenced to identify the mutations. The sequence revealed that insertion of two repeats units (+2) are the most frequently mutations. Loss of seven or ten repeats was observed.

6.2.8 Dnmt1 does not block homeologous recombination

The efficiency of recombination between homologous DNA stretches is highly dependent on their sequence identity (Rayssiguier et al., 1991, Shen and Huang, 1986, Nassif and Engels, 1993, Waldman and Liskay, 1988, te Riele et al., 1992). The mismatch repair protein Msh2 has a role in blocking homeologous recombination, which has been demonstrated by comparision of gene-targeting efficiency using isogenic and non-isogenic *Rb* targeting vectors (described previously). The isogeneic vector, 129Rb-puro is derived from a 129 genomic DNA library. The non-isogenic vector B/cRb-puro is derived from BALB/c-derived library. Both vectors contain 10.5 kb genomic arms and use the puromycin phophotransferase gene (puro) as the selection marker. The nonisogenic construct B/cRb-puro contains 0.6% base sequence divergence with respect to the isogenic 129Rb-puro construct (te Riele et al., 1992, Claij and Te Riele, 2002). This experiment revealed that the targeting efficiency with the non-isogenic B/cRb-puro vector is about 14 fold lower than with the isogenic 129Rb-puro vector in *Msh2* proficient cells. However, in *Msh2*-deficient cells, 129Rb-puro and B/cRb-puro displayed similar levels of gene-targeting efficiency (Claij and Te Riele, 2002).

In order to examine if *Dnmt1* is involved in blocking of homeologous recombination, the 129Rb-puro and B/cRb-puro were used to target the *Rb* locus in AB2.2 cells (Dnmt1-proficient), *Dnmt1*-deficient cells *Dnmt1-KO* and a *Msh2*-deficient ES cell line (*Msh2*$^{-/-}$) (Abuin et al., 2000). Abuin et al. (2000) has shown that homologous recombination with non-isogenic *Hprt* gene-targeting vector could occur at similar efficiency as with isogenic *Hprt* gene-targeting vector in

Table 6-4. Homologous recombination with Rb targeting vectors

ES cell line	Gene-targeting frequency		Isogenic versus Non-isogenic
	129Rb-puro(%)	B/cRb-puro(%)	
AB2.2	11 of 89 (12%)	2 of 90(2.2%)	5.5X
Msh2$^{-/-}$	1 of 48(2%)	2 of 81(2.4%)	0.8X
Dnmt1-KO	15 of 82(18%)	0 of 93(<1%)	>18X

Table 6-4. Homologous recombination with isogenic and non-isogenic Retinoblastoma (*Rb*) targeting vectors.

Gene-targeting experiments were carried out with isogenic *Rb* targeting vector (129Rb-puro) and non-isogenic (B/cRb-puro) vectors. The gene-targeted clones were identified by Southern analysis. The targeting frequency was represented as numbers of targeted clones versus total number of Puromycin resistant clones that exhibited clear Southern-hybridization signals.

these *Msh2*$^{-/-}$ cells. Thus, this *Msh2*$^{-/-}$ cell line serves as a positive control for our experiments. The gene-targeting efficiency for each targeting vector is listed in table 6-4. The isogenic targeting vector has a targeting efficiency 5.5 fold higher than the targeting efficiency of the non-isogenic targeting vector in AB2.2 ES cells. In contrast, the isogenic and nonisogenic targeting vector exhibited similar targeting efficiency in *Msh2*$^{-/-}$ ES cells, suggesting the homeolgous recombination experiment itself works. However, the targeting efficiency of the non-isogenic targeting vector in *Dnmt1-KO* cells was much lower than the targeting efficiency with the isogenic targeting vector, suggesting that homeologous recombination is still blocked in *Dnmt-1* deficient ES cells (Table 6-4).

6.3 Discussion

6.3.1 Summary

Simple sequence repeats (microsatellite sequence) are believed to be hot spots for mutations caused by DNA polymerase slippage. Deficiency in the MMR system leads to a significant increase in microsatellite instability (MSI). In this chapter, a puromycin slippage cassette (P-Slip) was developed to investigate MSI. P-Slip contains a strand of di-nucleotides repeats $(CA)_{17}$ placed out of frame in the puromycin phosphotransferase gene(*puro*), rendering the *puro* non-functional. The rate of MSI in P-Slip can be determined by measuring the number of puromycin resistant cells. P-Slip has been introduced by gene-targeting into *ROSA26* locus of various cell lines. By Luria-Delbruck fluctuation analysis, MSI rates were determined in *Msh6, Dnmt1, Tgif*, and *Parp-2/Rbpsuh* gene-trap mutants. Comparison of MSI rates between gene-trap mutants, AB2.2 ES cells and *Blm*-deficienct control cells revealed an elevated MSI rate in the *Msh6* gene-trap clones. However, a 2.5 fold increase in MSI rate was observed in *Blm*-deficient cells compared to wild-type AB2.2 cells, which complicated interpretation of the MSI rate of gene-trap clones. Gene-trap *Dnmt1, Tgif* and *Parp-2/Rbpsuh* exhibited MSI rates similar to *Blm*-deficient ES cells.

To precisely calculate the MSI rate, it is necessary to separate the gene-trap mutations from Blm-deficiency. A *Dnmt1*-deficient cells line *(Dnmt1-KO*) was generated by gene-targeting in AB2.2 ES cells. *Dnmt1-KO* cells carry $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$ alleles, both of which contain a deletion of exon 2, exon 3 and exon 4. Genomic DNA from *Dnmt1-KO* cells was hypomethylated at CpG sites compared to wild type AB2.2 cells. Comparison of the methylation level of gamma centromeric repeats revealed that the deletion mutation created in the *Dnmt1-KO* cells does not abolish *Dnmt-1* function totally, suggesting that the knockout $Dnmt1^{tm1}$ and $Dnmt1^{tm2}$ alleles are hypomorphic alleles. The MSI rate of the targeted P-Slip cassette in the *Dnmt1-KO* cells was determined by Luria-Delbruck fluctuation analysis, which is nearly 5 fold higher than that in AB2.2 cells. The mutations in the P-Slip cassette were recovered and sequenced to confirm the insertions or deletions in the di- nucleotide repeat. This experiment revealed that “+2” insertions could be a common DNA polymerase II error. Most importantly, these experiments pointed out that *Dnmt1* is involved in maintaining the stability of simple sequence repeats, a function of MMR system.

One mismatch protein, Msh2, has a role in suppressing recombination between diverged DNA sequences. This aspect was also studied in the *Dnmt1-KO* cells in this chapter using gene-targeing vectors that was derived either from isogenic or non-isogenic genomic DNA. No increase in the targeting efficiency was observed with non-isogenic targeting vectors in the *Dnmt1-KO* cells, suggesting that *Dnmt-1* may not be vital in blocking homeologous recombination between diverged DNA sequences.

6.3.2 MSI in *Dnmt-1* deficient cells is not a result of changes in MMR gene expression

Dnmt-1 is required for maintenance of DNA methylation patterns during DNA replication, and functions with other *do novo* methylases, such as *Dnmt3A* and *Dnmt3B*, in establishing the DNA methylation pattern during development (Li et

al., 1992). Gene expression is regulated epigenetically by DNA methylation. For example, promoter methylation leads to transcriptional silencing of a gene's expression. It has been known that mismatch repair gene *Mlh1*is silenced in some human tumor samples resulting from promoter hypermethylation (Veigl et al., 1998). Moreover, a hypomethylated genome is associated with a global increase in gene expression (Jackson-Grusby et al., 2001). MSI observed in *Dnmt1* mutants could be caused indirectly by changes in the expression of mismatch repair genes. To investigate this, the expression of four major mismatch repair genes, *Msh2, Msh6, Mlh1* and *Pms2* were examined by RT-PCR in the gene-trap *Dnmt1-V1* mutant. Compared to the *Dnmt1*-proficient NGG5-3 ES cells, the expression of these MMR genes was slightly increased (data not shown). Although RT-PCR cannot provide precise comparison of gene's expression between each sample, the observation of a general increase in MMR genes' expression in *Dnmt-1* deficient ES cells suggests that the increased MSI in *Dnmt1*-deficient ES cell is not likely to be caused by indirect reduction in the expression of MMR genes.

6.3.3 DNA replication, repair and methylation are coordinated processes

In humans, DNA replication, repair and methylation are coordinated by PCNA (proliferating cell nuclear antigen). PCNA forms a ring-shaped trimeric complex that can encircle double-stranded DNA and slide along it. PCNA itself is the DNA polymerase-processivity factor, which is required for DNA replication (Krishna et al., 1994). Moreover, PCNA provides a sliding platform that can mediate the interaction of proteins with DNA in a non-sequence specific manner. Some DNA repair systems, NER (nucleotide-excision repair) and MMR, are linked to the DNA replication process via interaction with PCNA. PCNA can interact with both eukaryote mismatch repair protein complex MutSα and MutSβ (Umar et al., 1996, Bowers et al., 2001, Lau and Kolodner, 2003). Since the primary function of MMR is to repair mismatched nucleotides generated during DNA replicaiton, the coupling of MMR system with DNA replication is believed to be vital. The

DNA methylation protein, Dnmt1, is recruited to the replicating DNA by binding to PCNA. This interaction allows Dnmt1 to perform its function as a maintenance methyltransferase, and methylate the newly synthesized hemimethylated DNA (Chuang et al., 1997, Vertino et al., 2002).

6.3.4 Evidence of links between DNA methylation and DNA mismatch repair

There is evidence suggesting a physical and functional interaction between DNA methylation and MMR systems. Chen et al (1998) observed that *Dnmt1* mutation causes genomic instability, a mutator phenotype. They measured the mutation rate by fluctuation analysis at both the endogenous X-linked *Hprt* (hypoxanthine phosphoribosyltransferase) gene and an integrated viral thymidine kinase (*tk*) transgene in *Dnmt-1* deficient and proficient ES cells. This experiment revealed a 10 fold increase at *Hprt* gene and 6 fold increase at the *tk* gene in the *Dnmt1*-deficient cells. By PCR analysis and Southern analysis, they found that over 60% of the mutants contain genomic rearrangements, which may be explained by aberrant mitotic recombination. However, a 10 fold increase in the mutation rate of the X-linked *Hprt* gene requires further explanation as this mutation rate cannot be explained by mitotic recombination alone. The mutator phenotype displayed by *Dnmt-1* deficient cells in their study is consistent with the observation of increased microsatellite instability in our study.

Recently, a methyl-CpG binding protein MBD4 was linked to DNA mismatch repair and MMR mediated genomic surveillance. MBD4, originally named as MED1 was identified as a protein interacting with the DNA mismatch repair protein, MLH1, in human cells (Bellacosa et al., 1999). Mammalian MBD4 protein contains glycosylase activity that enzymatically removes thymine (T) from a mismatched T/G base pair at CpG sites (Hendrich et al., 1999). Deamination of 5-methylcytosine (m^5C) to T at CpG sites frequently causes T/G mismatches. MBD4 has been shown to be important in suppressing the mutability of the m^5C.

MBD4 deficiency in mice caused an increase in CpG mutability and tumorigenesis (Millar et al., 2002, Wong et al., 2002). However, the function of MBD4 is not limited to repairing T/G mismatch at CpG sites. MED1 can bind to fully and hemimethylated DNA but not to unmethylated DNA *in vitro*. Moreover, transfection of a dominant negative mutated MED1 into cultured cells leads to microsatellite instability in an episomal slippage construct that contains tandem CA repeats (Bellacosa et al., 1999). These observations made MBD4/MED1 an attractive MMR candidate. Further studies revealed that the frameshift mutations in MBD4/MED1 coding sequence occurred frequently in colon, endometrial, pancreatic and gastric tumors exhibiting high microsatellite instability (Riccio et al., 1999, Bader et al., 1999, Yamada et al., 2002). However, MBD4 deficient mice generated by gene-targeting technology do not exhibit MSI (Millar et al., 2002, Wong et al., 2002). The link between MBD4 and the MMR system is supported by recently studies on cultured embryonic fibroblasts (MEFs) derived from *Mbd4*-deficient mice. In this study the apoptosis response to some DNA damaging drugs were examined and revealed that *Mbd4*-deficiency leads to tolerance of simple methylating agents like MNNG. *Mbd4*-deficient MEF cells also displayed similar tolerance to other DNA damaging drugs, such as platinum drugs, which forms intra and interstrand DNA adducts (Cortellino et al., 2003). It is established that the DNA mismatch repair system plays a role in DNA damage surveillance, in which it recognizes the damaged DNA and induces cell cycle arrest and cell death. MMR mutations allow cells to survive treatment with some DNA damaging drugs. The drug tolerance effects exhibited by *Mbd4*-deficient MEFs is phenotypically similar to the DNA damage tolerance exhibited by MMR deficiency, which is characterized by the accumulation of DNA lesions in cells. The DNA damage tolerance effect was also observed in the small intestine in the *Mbd4*-deficient mice (Sansom et al., 2003). These results suggest that the methyl CpG binding protein MBD4 is a multifunctional protein, which is involved in mismatch repair and its related DNA damage surveillance. Since MBD4 displays differences in the binding affinity to fully methylated DNA and hemimethylated DNA, it logically follows that MBD4 may be the molecular obligator that

transduces the DNA methylation signal laid by Dnmt1 along the replicating DNA to the DNA mismatch repair machinery that involves MLH1.

In *Escherichia coli*, the mismatch repair protein MutH distinguishes the newly synthesized DNA strand from the parental DNA strand using the hemi-methylated adenine at GATC site as a signal. Mutation in bacteria *Dam* DNA methylase, the enzyme methylating the GATC sequence, causes a mutator phenotype. Although DNA mismatch repair machinery is highly conserved in evolution, a functional homologue of strand distinguishing protein MutH has not been identified. *Dnmt1* provides an attractive candidate for such an activity since Dnmt1 protein is located to the DNA replication site and it maintains the CpG methylation post DNA replication. Also, our study demonstrated that *Dnmt1* deficiency caused microsatellite instability. However, some evidence suggests that DNA methylation may not be essential for strand discrimination in eukaryotes, especially organisms that are deficient in genomic methylation for example, *S.cerevisiae*, *Drosophila* and *C.elegans* (Modrich and Lahue, 1996). Alternatively, strand discontinuity occurred during DNA replication is thought to be the signal which directs the MMR system to the newly synthesized strand (Holmes et al., 1990). PCNA provides a molecular linker between mismatch repair proteins and DNA polymerase at the replication fork, so that the nicks in the nascent strand, as a result of DNA synthesis, could direct the MMR process into the newly synthesized DNA (Gu et al., 1998, Johnson et al., 1996, Kokoska et al., 1999, Umar et al., 1996).

The MSI exhibited by *Dnmt1*-deficient (Dnmt1-KO) ES cells is about 5 fold higher than methylation proficient ES cells. In contrast, the MSI exhibited by *Msh2*-deficient ES cells is elevated nearly 4 orders of magnitude (Abuin et al., 2000). This result suggests that *Dnmt1* may not provide a vital strand distinguishing signal for MMR system. However, it cannot be ruled out that Dnmt1 may facilitate the strand distinguishing process in higher eukaryotes. In fact, CpG hemimethylation was shown to synergize with single strand nicks in directing

repair to the unmethylated strand in monkey CV1 cells (Hare and Taylor, 1985). It is worthy to point out that the Dnmt1-KO allele generated in this study appeared to be a hypomorph allele. The moderate microsatellite instability exhibited in the Dnmt1-KO cells compared to that in the MMR deficient cells may reflect the remaining DNA methylation activity. To better address this question, it is necessary to measure the rate of MSI in *Dnmt1* null cells.

6.3.5 Blm affects DNA mismatch repair

It was observed in this study that *Blm* deficiency caused mild microsatellite instability. The *Blm* gene was also identified in the *C.elegans* slippage screen for the reversal of a *lacZ* slippage reporter that has a single nucleotide repeats $(A)_{17}$ (Pothof et al., 2003). In this screen 61 genes were identified showing increased genomic instability. To test if the reversal of the *lacZ* reporter depended on the existence of the $(A)_{17}$ repeats, the reversal of an out of frame *lacZ* reporter lacking the single nucleotide repeat was investigated and this experiments revealed that the majority of the genes identified in the slippage screen didn't require the existence of the $(A)_{17}$ repeats to reverse the expression of the *lacZ* reporter. *Blm* is among these genes. It was proposed that reversal of the *lacZ* reporter expression in these mutants was caused by other mechanisms other than MMR, for example, insertion or deletion mutations caused by illegitimate homologous recombination (Pothof et al., 2003). In our study, the sequence analysis of the slippage puromycin resistance clones confirmed the insertion or deletion mutations occurred in the $(CA)_{17}$ repeats. It doesn't appear to be a result of homologous recombination. Thus, MSI in *Blm* deficient ES cells requires further explanation. The human BLM protein associates with the MMR proteins, MSH2, MSH6 and MLH1 in a protein complex, including Brca1 (Wang et al., 2001). Although a functional role of this interaction is not clear, the existence of a physical interaction between BLM and the MMR proteins may indicate a functional connection. In our study, *Dnmt1* and *Blm* double null ES cells (the gene trap *Dnmt1* mutant) exhibited a MSI rate similar to the MSI rated exhibited

in the *Blm* null ES cells, which suggested a genetic interaction between *Blm* and *Dnmt1* (or *Dnmt1* mediated MMR process). To better understand this interaction, it will be important to compare the MSI rate in the MMR deficient and MMR and *Blm* double deficient cells. DNA replication, methylation and repair are coordinated processes, in which Blm, Dnmt1 and MMR proteins play major roles (El-Osta, 2003). Although an interaction between Blm and MMR hasn't been reported, it was observed that deficiency in the MMR process caused an increase in DNA methylation, which may result from the unleashed Dnmt1 activity from MMR related processes (Ahuja et al., 1997). The genetic interaction between Blm, Dnmt1 (MMR) may be a result of the unbalanced DNA replication and the MMR processes.

Part I: A system for recessive screens

7.1 *Blm*-deficient cells

Five years ago, a viable *Blm*-deficient knockout mouse was generated in our lab. The observation of a high LOH rate associated with a *Blm*-deficiency both in ES cells and in the mouse triggered the possibility of generating homozygous mutations via the high LOH rate. Later, this became the primary goal of my thesis work, which was designed to explore the usefulness of this high LOH rate in a genetic screen for recessive mutations. The *Blm*-deficiency results in an LOH rate of 4.2×10^{-4} per locus /cell/generation. At a practical level this rate means that homozygous daughter cells would be segregated from an ES cell carrying a single allele mutation after that cell has expanded to an ES cell colony consisting of 2,500 cells. In order to conduct a genetic screen in *Blm*-deficient cells, positive selection is required to identify the rare homozygous mutants from a pool of heterozygous mutants. Rooted in the lab's interests in DNA repair and cancer, I decided to identify mutations in the mismatch repair pathway. Mismatch repair deficiency is one of the major molecular lesions involved in non-polyposis colorectal cancer. Four mouse mismatch repair genes, *Msh2, Msh3*, *Pms1* and *Mlh1* have been knocked out in our lab. *Msh2*-deficienct ES cells were shown to be tolerant to high doses of 6TG while MMR-proficient cells were killed completely. This provided an ideal positive selection system. Taking advantage of the highly efficient chemical mutagenesis by EMS, I was able to screen for 6TG resistant mutants on large numbers of EMS-treated *Blm*-deficient ES cells. This experiment verified the use of *Blm*-deficient cell line (NGG) that was generated to carry two *Hprt* minigenes in a screen for 6TG resistance, and established the basic screening procedure for MMR mutants (Chapter3). Importantly, the recovery of potential MMR candidates in this screen encouraged me to develop a gene trap mutagen and to conduct a screen for 6TG resistant mutants (Chapter 4 and Chapter 5).

7.2 Gene trap mutagenesis

Gene trap mutagenesis was first developed in the late 80's, when various gene-trap vectors were developed to mutate as well as tag the expression of endogenous genes. The ability to determine the molecular basis of the mutations makes gene trap one of the favorite methods in cultured cells. New gene trap systems were being developed by a couple of senior Ph.D. students in our lab. Hence, a gene trap approach was chosen to mutate the *Blm*-deficient cells. To provide a quick confirmation of the gene trap mutations, I modified the original *Rosa*β*geo* gene trap vector to make it revertible. Based on the Cre-*loxP* system and the characteristics of the retroviral life cycle, a self-inactivating retroviral gene trap vector (RGTV-1) was constructed carrying a loxP site in its 3'LTR. After insertion into the host genome, the provirus will be flanked by two *loxP* sites. Therefore, the gene trap cassette could be deleted by Cre-mediated recombination. RGTV-1 could be packaged efficiently in viral packaging cells and has been shown to be an efficient mutagen in this study. 10,000 gene trap mutants were using the RGTV-1 retrovirus, from which many 6TG resistant clones were recovered. One DNA mismatch repair gene, *Msh6*, was mutated more than seven times by the gene trap virus and all these mutants were homozygous and revertible. This result confirms the efficiency of the *Blm*-deficient genetic background in producing homozygous mutants and highlights the impact of *Blm*-deficiency on establishing recessive genetic screens in mammalian cells in culture.

7.3 Broad applications of the Blm-deficiency in recessive screens

The use of *Blm*-deficient ES cells as a genetic background for recessive genetic screens has broad applications. Cultured ES cells could provide rapid access to the phenotypes associated with a significant fraction of the genes in the genome. At least 10,000 genes are expressed in un-differentiated ES cells. These genes are required to elaborate the fundamental components for a mammalian cell and

physiological systems for essential functions, for example structural components, metabolism, cell division and DNA repair. ES cells can also be induced to differentiate into a wide variety of cell types *in vitro*, providing access to genes involved in signaling cascades and in cell differentiation programs. Moreover, somatic cell lines could be derived from *Blm*-deficient mice. Thus, a *Blm*-deficiency based screening strategy could also be designed to study tissue or cell type specific gene functions.

7.4 Transponson-mediated mutagenesis

Although gene trap mutagenesis has advantages over chemical mutagenesis because it provides rapid molecular access, it is believed that gene trap mutagenesis has limitations in its genomic coverage. It is not possible to mutate every gene with only one type of gene trap vector (Hansen et al., 2003). To achieve broader genome coverage in large-scale genetic screens, it is necessary to utilize various gene trap vector designs, and to consider recently developed mutagenesis methods for example, the sleeping beauty (SB) transposon system (Ivics et al., 1997).

The sleeping beauty (SB) transposon system provides an alternative method to transfer a gene trap cassette into the genome of a cell. Sleeping beauty (SB) belongs to the Tc1/mariner superfamily of transposons. It was reconstituted from transpositionally inactive transposon sequences in fish by eliminating the inactivating mutations accumulated during evolution (Ivics et al., 1997). The SB system is composed of a SB transposon element and the transposase, which is expressed separately from an expression vector. The SB transposon element contains two terminal inverted repeats (IR). The exogenous DNA is placed between the two IRs. The insertion of the SB transposon element into the host genome occurs by a SB transposase-mediated cut and paste process, during which the transposase binds to the terminal IRs. The insertion of the SB transposon itself could cause an insertional mutation if the expression of host

gene is interrupted. The SB system has been shown to be very efficient at DNA integration in vertebrate cells. Recently it was demonstrated that the SB system transposed efficiently in the germ line of mice. This brought about a new method to establish mutations in mice potentially on a large scale. In addition, the SB element can be used as a vehicle to transfer a gene trap cassette into the host genome. Horie et al (2003) constructed a SB gene trap vector and conducted gene trap mutagenesis in the mouse. They demonstrated that their gene trap vector could insert into endogenous genes at a frequency of 7%. However, several limitations of the SB system must be considered when applying it in cultured cells. One is the local nature of transposition. The application of the SB system in mice requires the construction of a founder mouse that harbors the SB transposon element at a defined genomic locus. Crossing a mouse expressing the transposase with the founder mouse will then induce SB transposition in the offspring. It has been shown that the SB element will preferentially insert into a genomic locus within several megabases of its original integration site. The SB gene trap screen conducted by Horie revealed that three quarters of transpositions sites are actually located in the original chromosome that harbours the SB element. Obviously, this will restrict the use of this strategy in cultured cells if a broader genomic coverage is favoured. Another way to apply the SB gene trap system is to introduce both the SB gene trap element and the transposase into cells by transient transfection, for example by electroporation. The SB gene trap element and the transposase-expressing vector will co-exist episomally in the host cells for a short period of time and transposition will occur from the vector to the genome. Although this episomal method is very efficient in cultured somatic cells, the transposition efficiency in mouse ES cells is very low. Electroporation of 10^7 ES cells will generate less than one hundred ES cells with inserted SB elements. Thus, to perform genome wild mutagenesis in ES cells, it is necessary to improve the efficiency of the SB system in ES cells. DNA methylation has been linked to the activity of SB transposon recently. It has been known that hypomethylation leads to activation of an endogenous transposon in plants and the activation of an endogenous retrotransposon

element in the mouse (Miura et al., 2001, Gaudet et al., 2004). It would be interesting to know if a hypomethylated genome could increase the efficiency of the SB system in ES cells.

7.5 Combination of deletional mutations with *Blm*-deficiency

In addition to insertional mutagenesis, deletions induced by gamma-irradiation (γ-irradiation) could be a powerful mutagenesis method. γ-irradiation causes double strand breaks in DNA. If they are not properly repaired, double strand breaks result in chromosomal abnormalities, such as deletions. γ-irradiation has an advantage of a broad genome coverage, and is highly efficient. The mutation frequency at a specific genomic locus in mouse ES cells can be as high as one per 1000 treated cells with a dosage of 400 rads (You et al., 1997). Deletion mutation can be identified by microarray-based comparative genomic hybridization (CGH). CGH has been shown to be a powerful tool in detecting chromosomal imbalances. A mouse BAC microarray resource has been developed with a 1 Megabase (Mb) interval to identify DNA copy number alterations in cells and tumour samples. With this method, deletions ranging from several hundred kilobases to megabases can be identified. Furthermore, CGH arrays can distinguish double copies of genomic alterations from single copy alterations; thus, homozygous mutations can be distinguished from heterozygous mutations (Chung et al., 2004). For a genetic screen, a small deletion that spans a distance of several kilobases to several hundered kilobases is preferred because a small deletion will allow a rapid localization of candidate genes within the deleted region. Moreover, large deletions will often mutate multiple genes, which may affect cell growth or cause cell lethality. Developing BAC microarrays with a high resolution is essential for identifying small chromosomal deletions with CGH studies.

In summary, the recessive genetic screen system established using *Blm*-deficiency cells in this study provides an opportunity to identify novel genes with

interesting phenotypic consequences. Although this system has limitations, for example, the incomplete genome coverage, with improvements in the mutagenesis strategy this method will provide an important genetic tool for functional genomics.

7.6 RNA interference (RNAi), a new era for mutagenesis

RNA interference was first defined in *C.elegans* as a response to double strand RNA (dsRNA), which causes sequence specific knockdown of a gene's function. About a decade ago, it was a commonly held view that injection of antisense-orientated RNA of a gene into *C.elegans* could inhibit that gene's function. However, Guo et al. (1995) observed that sense-orientated RNA could induce a similar result as that shown with anti-sense RNA. Later Fire et al. (1998) observed that a mixture of both sense and antisense-oriented RNA led to a10 fold increase in the efficiency of inhibiting that gene's function. They named this striking phenomenon RNA interference (RNAi) and proposed that it is the double strand RNA, but not antisense RNA that triggers the gene inactivation process.

Double strand RNA mediated gene inactivation is believed to be a conserved process. Genetic and biochemical studies in plants, *C.elegans* and *Drosophila* have resulted in the identification of some components of the RNAi process and a basic understanding of the initiation of RNAi. The basic model of RNAi includes three major steps: first, double strand RNA is cleaved by a member of RNase III family to form short double strand RNAs of 21-25 nucleotides (nt) in length. Second, a RNA-protein complex (RISC, RNA-induced silencing complex) is assembled containing the 22nt double-stranded RNA and RNA nuclease activity. Third, RISC is directed to the corresponding mRNA and destroys it (Hannon, 2002).

The RNAi phenomena has been quickly employed and developed to be one of the most powerful genetic tools in *C.elegan*s, by which a loss of function mutation

in any gene can be generated simply by introducing corresponding double strand RNA molecules into worms (Hannon, 2002). This process is so efficient that only a few molecules of double strand RNA are required in one cell to initiate the RNAi process; and this process can spread through the whole worm and pass through the germ line. RNAi in *C.elegans* can be performed by soaking worms in water containing RNAi molecules, by injecting RNAi molecules into their gonads or by feeding worms with bacteria that express double strand RNA. A library of bacterial strains have been constructed to express double strand RNAs targeting almost 86% of predicted *C.elegans* genes. This tool has proven to be extremely helpful in high throughput genetic screens (Kamath et al., 2003).

The application of RNAi technology in mammalian cells has lagged behind because of the cytotoxic reaction of mammalian cells to double strand RNA, known as the interferon response, in which double strand RNA induces a non-specific global translation inhibition (Hannon, 2002). Recently Elbashir et al. (2001) demonstrated that chemically synthesized short double strand RNA of 21 to 22 nucleotides in length (siRNA) could strongly induce gene-specific inactivation, while avoiding the non-specific translation inhibition effect. This finding opened the door for the application of RNAi technology in mammalian cells. However, the siRNA mediated RNAi effect in mammalian cells cannot be inherited or spread to adjacent cells in contrast to the RNAi effect in *C.elegans*. Thus, siRNA-mediated RNAi can only be active for a short time. To solve this problem, Brummelkamp et al. (2002) developed a mammalian expression vector to direct the synthesis of short hairpin-structured RNA transcripts (shRNA) using the RNA polymerase III promoter. The shRNA is composed of a siRNA-like double strand RNA stem and a single-stranded loop structure, which can be cleaved in cells by the RNAi machinery and initiate the RNAi process. The stable expression of shRNA in cells allows the persistent suppression of gene expression. Thus, shRNA technology has been quickly adopted as a powerful tool in generating loss of function mutation in mammalian cells. Recently, two groups have reported the application of RNAi in large-scale genetic screens in

cultured cells by the construction of retrovirus-based shRNA expressing libraries (Berns et al., 2004, Paddison et al., 2004). Berns et al. (2004) targeted nearly 8,000 human genes in their shRNA library and obtained on average 70% inhibition of expression for approximately 70% of the targeted genes in the library. With this library they were able to identify new components of the p53-dependent proliferation arrest process. This experiment validated the efficiency of RNAi-mediated genetic screens. Paddison et al. (2004) targeted nearly 10,000 human genes and over 5,000 mouse genes. To explore the efficiency of their shRNA library, they screened for components of the 26S proteosome complex in one quarter of their shRNA clones. Nearly 50% of the shRNA clones that were expected to target proteosomal proteins could be recovered. These experiments have shown that RNAi is providing an exciting opportunity for recessive genetic screens in mammalian cells in culture.

RNAi technology has some limitations though. The most pronounced one is the incomplete inhibition of a given gene's activity. The expression inhibition induced by shRNA leads mostly to partial loss of function mutations. Thus, many weak hypomorphic mutations would be missed in a large-scale genetic screen. In fact, 50% of the expected 26S proteosome components were missed in the screen Paddison et al. conducted (2004). Also, the design of the shRNA construct requires prior knowledge of the expressed sequence of a gene. Therefore, RNAi cannot be used as a random mutagen. Thus, RNAi-mediated genetic screens are complementary to but do not replace genetic screens based on *Blm*-deficient genetic background, which can use various mutagens to generate loss of function mutations.

Part II: DNA methylation and mismatch repair surveillance

7.7 *Dnmt1*, a MMR surveillance gene

I have demonstrated that *Dnmt1* deficiency leads to instability of simple sequence repeats, which implies that *Dnmt1* is a potential MMR protein. One major question left unaddressed in this study is how *Dnmt1* is involved in DNA mismatch repair. Evidence that links *Dnmt1* with MMR has been presented in detail in Chapter 6. Here, I would like to summarize this evidence in brief in order to highlight future investigations into the relationship between DNA methylation and MMR.

Based on the accumulated evidence that links a deficiency in DNA methylation with MMR, *Dnmt1* may act in MMR in several independent but not exclusive mechanisms. First, *Dnmt1* could function in MMR by a physical interaction with MMR related proteins. Evidence supporting this view is that both *Dnmt1* and MMR proteins can interact with PCNA. These interactions are central to both the DNA replication-coupled MMR process and the DNA methylation process. Interestingly, it was observed that when mismatch repair deficient cells were infected with a retrovirus, the retrovirus was transcriptionally silenced by methylation. However, in mismatch repair proficient cells, the infected retrovirus was transcriptionally active (Ahuja et al., 1997). This observation hinted that lack of methylation of the retrovirus in MMR proficient cells might be the result of a lack of *Dnmt1* because *Dnmt1* is recruited to the MMR process. Second, *Dnmt1* could be involved in MMR indirectly by providing a methylation signal to other MMR related proteins, for example, the methyl CpG binding protein, MBD4. MBD4 binds preferentially to the methylated CpG sites. MBD4 is a glycosylase. It can enzymatically remove thymine (T) from a mismatched T:G basepair at CpG sites. It has been shown that an MBD4 mutation caused an increase in mutation rate at CpG sites. Moreover, MBD4 mutation caused an increase in microsatellite instability. Recently, it was demonstrated that MBD4 deficiency

leads to tolerance of DNA damage in a way similar to mutations in MMR surveillence (Cortellino et al., 2003, Sansom et al., 2003). Thus, it would be interesting to know if MBD4 is an interpretor of the hypomethylated genome caused by *Dnmt1* deficiency. Third, it is possible that *Dnmt1* may facilitate the strand distinguishing process in higher eukaryotes. The evidence supporting this view is that CpG hemimethylation was shown to synergize with single strand nicks in directing repair to the unmethylated strand in monkey CV1 cells (Hare and Taylor, 1985). Finally, it cannot be excluded that *Dnmt1* can act in MMR by epigenetically altering the expression of the MMR genes. However, this seems unlikely because *Dnmt1* deficiency doesn't lead to decreased expression of *Msh2*, *Msh6*, *Mlh1* or *Pms2* genes in mouse ES cells (data not shown).

Based on these observations, it would be interesting to investigate if *Dnmt1* and MBD4 double mutants have synergistic effects on microsatellite instability. Because MBD4 deficiency results in tolerance to many forms of DNA damage, including platinum drugs and ionizing radiation, it would be interesting to know if *Dnmt1*-deficiency could cause a similar effect. It has been reported that *Dnmt3a* and *Dnmt3b* facilitate *Dnmt1* in maintaining DNA methylation levels in ES cells, thus it would be interesting to know if a double-knockout of *Dnmt1* and *Dnmt3a* or *3b* would increase the microsatellite instability by profound genomic demethylation. This experiment would help to address the question of whether *Dnmt1* is directly involved in MMR or whether it is the CpG methylation level that is facilitating the MMR process. It is also important to conduct biochemical studies to isolate potential *Dnmt1* binding proteins and to investigate whether the existence of *Dnmt1* could facilitate the binding and processitivity of mismatched nucleotides *in vitro* with and without the hemimethylated CpG signals. Finally, it is important to emphasize that the DNA methylation pattern varies between cells and tissues, which is vital in establishing and maintaining cell and tissue specific gene expression patterns. For example, depletion of *Dnmt1* in MEFs caused massive alteration of DNA expression. This resulted in a p53-related cell cycle arrest and apoptosis (Jackson-Grusby et al., 2001). Whereas the gene trap

Dnmt1 mutants and the knockout *Dnmt1* ES cells exhibited normal cell growth, which is consistent with the previous observation that *Dnmt1* is not required for ES cell growth and survival (Chen et al., 1998). These results suggest that *Dnmt1* is a protein with multiple functions. A better understanding of the relationship between *Dnmt1* and DNA mismatch repair requires the dissection of individual *Dnmt1* functions, which means isolating the downstream effectors of the individual processes *Dnmt1* is involved in and generating subtle mutations in *Dnmt1* that only affect a subset of *Dnmt1* activities.

7.8 Clinical implications

The finding that *Dnmt1*-deficiency leads to 6TG tolerance has some clinical implications. 6TG and mercaptopurine have been used as important drugs in the treatment of acute leukemia (Elion, 1989). As demonstrated in this study *Dnmt1*-deficiency causes a high level of 6TG tolerance. 6TG tolerance will not only eliminate the effect of the treatment, but may induce adverse effects in patients. As DNA damaging drugs, 6TG and mercaptopurine are believed to cause nucleotide mismatches and double strand breaks. 6TG tolerance will lead to DNA damage accumulating in cells, which may cause mutations in tumour suppressor genes, therefore, increasing the risk of cancer. Thus, it may be important to investigate how frequently DNA hypomethylation occurs in patients with acute leukemia and re-evaluate the efficiency of 6TG treatment in *Dnmt1*-deficient animal models. Azathioprine, a drug used in immune suppression in transplant surgery, is converted *in vivo* to mercaptopurine. Thus, it is also relevant to evaluate the effect of *Dnmt1* deficiency in the application of azathiopurine as well.

References

Ababou, M., Dumaire, V., Lecluse, Y. and Amor-Gueret, M. (2002) *Oncogene,* **21,** 2079-88.

Abuin, A. and Bradley, A. (1996) *Mol Cell Biol,* **16,** 1851-6.

Abuin, A., Zhang, H. and Bradley, A. (2000) *Mol Cell Biol,* **20,** 149-57.

Acharya, S., Wilson, T., Gradia, S., Kane, M. F., Guerrette, S., Marsischky, G. T., Kolodner, R. and Fishel, R. (1996) *Proc Natl Acad Sci U S A,* **93,** 13629-34.

Aebi, S., Fink, D., Gordon, R., Kim, H. K., Zheng, H., Fink, J. L. and Howell, S. B. (1997) *Clin Cancer Res,* **3,** 1763-7.

Aebi, S., Kurdi-Haidar, B., Gordon, R., Cenni, B., Zheng, H., Fink, D., Christen, R. D., Boland, C. R., Koi, M., Fishel, R. and Howell, S. B. (1996) *Cancer Res,* **56,** 3087-90.

Ahuja, N., Mohan, A. L., Li, Q., Stolker, J. M., Herman, J. G., Hamilton, S. R., Baylin, S. B. and Issa, J. P. (1997) *Cancer Res,* **57,** 3370-4.

Akiyama, N., Matsuo, Y., Sai, H., Noda, M. and Kizaka-Kondoh, S. (2000) *Mol Cell Biol,* **20,** 3266-73.

Alani, E., Reenan, R. A. and Kolodner, R. D. (1994) *Genetics,* **137,** 19-39.

Allen, M., Heinzmann, A., Noguchi, E., Abecasis, G., Broxholme, J., Ponting, C. P., Bhattacharyya, S., Tinsley, J., Zhang, Y., Holt, R., Jones, E. Y., Lench, N., Carey, A., Jones, H., Dickens, N. J., Dimon, C., Nicholls, R., Baker, C., Xue, L., Townsend, E., Kabesch, M., Weiland, S. K., Carr, D., von Mutius, E., Adcock, I. M., Barnes, P. J., Lathrop, G. M., Edwards, M., Moffatt, M. F. and Cookson, W. O. (2003) *Nat Genet,* **35,** 258-63.

Allen, N. D., Cran, D. G., Barton, S. C., Hettle, S., Reik, W. and Surani, M. A. (1988) *Nature,* **333,** 852-5.

Ame, J. C., Rolli, V., Schreiber, V., Niedergang, C., Apiou, F., Decker, P., Muller, S., Hoger, T., Menissier-de Murcia, J. and de Murcia, G. (1999) *J Biol Chem,* **274,** 17860-8.

Amin, N. S., Nguyen, M. N., Oh, S. and Kolodner, R. D. (2001) *Mol Cell Biol,* **21,** 5142-55.

Andreu, T., Beckers, T., Thoenes, E., Hilgard, P. and von Melchner, H. (1998) *J Biol Chem,* **273,** 13848-54.

Araki, K., Araki, M., Miyazaki, J. and Vassalli, P. (1995) *Proc Natl Acad Sci U S A,* **92,** 160-4.

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M. and Wyllie, A. (1999) *Oncogene,* **18,** 8044-7.

Bailis, A. M. and Rothstein, R. (1990) *Genetics,* **126,** 535-47.

Baker, R. K., Haendel, M. A., Swanson, B. J., Shambaugh, J. C., Micales, B. K. and Lyons, G. E. (1997) *Dev Biol,* **185,** 201-14.

Baker, S. M., Bronner, C. E., Zhang, L., Plug, A. W., Robatzek, M., Warren, G., Elliott, E. A., Yu, J., Ashley, T., Arnheim, N. and et al. (1995) *Cell,* **82,** 309-19.

Balling, R. (2001) *Annu Rev Genomics Hum Genet,* **2,** 463-92.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A. T., Matsumoto, Y., Golemis, E. A., Genuardi, M. and Neri, G. (1999) *Proc Natl Acad Sci U S A,* **96,** 3969-74.

Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, R. K. and Gehring, W. J. (1989) *Genes Dev,* **3,** 1288-300.

Bergstrom, R. A., You, Y., Erway, L. C., Lyon, M. F. and Schimenti, J. C. (1998) *Genetics,* **150,** 815-22.

Berns, A. (1991) *J Cell Biochem,* **47,** 130-5.

Berns, K., Hijmans, E. M., Mullenders, J., Brummelkamp, T. R., Velds, A., Heimerikx, M., Kerkhoven, R. M., Madiredjo, M., Nijkamp, W., Weigelt, B., Agami, R., Ge, W., Cavet, G., Linsley, P. S., Beijersbergen, R. L. and Bernards, R. (2004) *Nature,* **428,** 431-7.

Bertram, M. J., Berube, N. G., Hang-Swanson, X., Ran, Q., Leung, J. K., Bryce, S., Spurgers, K., Bick, R. J., Baldini, A., Ning, Y., Clark, L. J., Parkinson, E. K., Barrett, J. C., Smith, J. R. and Pereira-Smith, O. M. (1999) *Mol Cell Biol,* **19,** 1479-85.

Beumer, K. J., Pimpinelli, S. and Golic, K. G. (1998) *Genetics,* **150,** 173-88.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E. and et al. (1989) *Genes Dev,* **3,** 1273-87.

Bischof, O., Kim, S. H., Irving, J., Beresten, S., Ellis, N. A. and Campisi, J. (2001) *J Cell Biol,* **153,** 367-80.

Blander, G., Kipnis, J., Leal, J. F., Yu, C. E., Schellenberg, G. D. and Oren, M. (1999) *J Biol Chem,* **274,** 29463-9.

Bocker, T., Barusevicius, A., Snowden, T., Rasio, D., Guerrette, S., Robbins, D., Schmidt, C., Burczak, J., Croce, C. M., Copeland, T., Kovatich, A. J. and Fishel, R. (1999) *Cancer Res,* **59,** 816-22.

Bode, V. C. (1984) *Genetics,* **108,** 457-70.

Bonaldo, P., Chowdhury, K., Stoykova, A., Torres, M. and Gruss, P. (1998) *Exp Cell Res,* **244,** 125-36.

Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B. and Schaffner, W. (1985) *Cell,* **41,** 521-30.

Bowers, J., Tran, P. T., Joshi, A., Liskay, R. M. and Alani, E. (2001) *J Mol Biol,* **306,** 957-68.

Boyer, J. C., Umar, A., Risinger, J. I., Lipford, J. R., Kane, M., Yin, S., Barrett, J. C., Kolodner, R. D. and Kunkel, T. A. (1995) *Cancer Res,* **55,** 6063-70.

Bradley, A., Evans, M., Kaufman, M. H. and Robertson, E. (1984) *Nature,* **309,** 255-6.

Branch, P., Aquilina, G., Bignami, M. and Karran, P. (1993) *Nature,* **362,** 652-4.

Brinster, R. L., Chen, H. Y., Trumbauer, M. E., Yagle, M. K. and Palmiter, R. D. (1985) *Proc Natl Acad Sci U S A,* **82,** 4438-42.

Brosh, R. M., Jr., Li, J. L., Kenny, M. K., Karow, J. K., Cooper, M. P., Kureekattil, R. P., Hickson, I. D. and Bohr, V. A. (2000) *J Biol Chem,* **275,** 23500-8.

Brown, S. D. and Balling, R. (2001) *Curr Opin Genet Dev,* **11,** 268-73.

Brummelkamp, T. R., Bernards, R. and Agami, R. (2002) *Science,* **296,** 550-3.

Buermeyer, A. B., Deschenes, S. M., Baker, S. M. and Liskay, R. M. (1999) *Annu Rev Genet,* **33,** 533-64.
Busch, D. B., Cleaver, J. E. and Glaser, D. A. (1980) *Somatic Cell Genet,* **6,** 407-18.
Campbell, C. E., Gravel, R. A. and Worton, R. G. (1981) *Somatic Cell Genet,* **7,** 535-46.
Capecchi, M. R. (1989) *Science,* **244,** 1288-92.
Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., Declercq, C., Pawling, J., Moons, L., Collen, D., Risau, W. and Nagy, A. (1996) *Nature,* **380,** 435-9.
Cejka, P., Stojic, L., Mojas, N., Russell, A. M., Heinimann, K., Cannavo, E., di Pietro, M., Marra, G. and Jiricny, J. (2003) *Embo J,* **22,** 2245-54.
Chao, H. H., Mentzer, S. E., Schimenti, J. C. and You, Y. (2003) *Genesis,* **35,** 133-42.
Chapman, V., Forrester, L., Sanford, J., Hastie, N. and Rossant, J. (1984) *Nature,* **307,** 284-6.
Charames, G. S. and Bapat, B. (2003) *Curr Mol Med,* **3,** 589-96.
Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L. and Jaenisch, R. (1998) *Nature,* **395,** 89-93.
Chen, Y., Yee, D., Dains, K., Chatterjee, A., Cavalcoli, J., Schneider, E., Om, J., Woychik, R. P. and Magnuson, T. (2000) *Nat Genet,* **24,** 314-7.
Chester, N., Kuo, F., Kozak, C., O'Hara, C. D. and Leder, P. (1998) *Genes Dev,* **12,** 3382-93.
Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G. and Li, B. F. (1997) *Science,* **277,** 1996-2000.
Chung, Y. J., Jonkers, J., Kitson, H., Fiegler, H., Humphray, S., Scott, C., Hunt, S., Yu, Y., Nishijima, I., Velds, A., Holstege, H., Carter, N. and Bradley, A. (2004) *Genome Res,* **14,** 188-96.
Claij, N. and Te Riele, H. (2002) *Oncogene,* **21,** 2873-9.
Coffin J M , H. S. H. and E, V. H. (1996) *Retroviruses,* New York Cold Spring Harbor Laboratory.
Copeland, N. G., Jenkins, N. A. and Court, D. L. (2001) *Nat Rev Genet,* **2,** 769-79.
Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., Skalka, A. M., Meropol, N. J., Alberti, C., Larue, L. and Bellacosa, A. (2003) *Proc Natl Acad Sci U S A,* **100,** 15071-6.
D'Atri, S., Tentori, L., Lacal, P. M., Graziani, G., Pagani, E., Benincasa, E., Zambruno, G., Bonmassar, E. and Jiricny, J. (1998) *Mol Pharmacol,* **54,** 334-41.
Davies, S. L., North, P. S., Dart, A., Lakin, N. D. and Hickson, I. D. (2004) *Mol Cell Biol,* **24,** 1279-91.
de Wind, N., Dekker, M., Berns, A., Radman, M. and te Riele, H. (1995) *Cell,* **82,** 321-30.
Deaven, L. L. and Petersen, D. F. (1973) *Chromosoma,* **41,** 129-44.

Devon, R. S., Porteous, D. J. and Brookes, A. J. (1995) *Nucleic Acids Res,* **23,** 1644-5.
di Pietro, M., Marra, G., Cejka, P., Stojic, L., Menigatti, M., Cattaruzza, M. S. and Jiricny, J. (2003) *Cancer Res,* **63,** 8158-66.
Doetschman, T., Gregg, R. G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S. and Smithies, O. (1987) *Nature,* **330,** 576-8.
Dong, J., Albertini, D. F., Nishimori, K., Kumar, T. R., Lu, N. and Matzuk, M. M. (1996) *Nature,* **383,** 531-5.
Drummond, J. T., Li, G. M., Longley, M. J. and Modrich, P. (1995) *Science,* **268,** 1909-12.
Duckett, D. R., Drummond, J. T., Murchie, A. I., Reardon, J. T., Sancar, A., Lilley, D. M. and Modrich, P. (1996) *Proc Natl Acad Sci U S A,* **93,** 6443-7.
Dufort, D., Schwartz, L., Harpal, K. and Rossant, J. (1998) *Development,* **125,** 3015-25.
Dutertre, S., Ababou, M., Onclercq, R., Delic, J., Chatton, B., Jaulin, C. and Amor-Gueret, M. (2000) *Oncogene,* **19,** 2731-8.
Edelmann, W., Cohen, P. E., Kneitz, B., Winand, N., Lia, M., Heyer, J., Kolodner, R., Pollard, J. W. and Kucherlapati, R. (1999) *Nat Genet,* **21,** 123-7.
Edelmann, W., Yang, K., Umar, A., Heyer, J., Lau, K., Fan, K., Liedtke, W., Cohen, P. E., Kane, M. F., Lipford, J. R., Yu, N., Crouse, G. F., Pollard, J. W., Kunkel, T., Lipkin, M., Kolodner, R. and Kucherlapati, R. (1997) *Cell,* **91,** 467-77.
El-Osta, A. (2003) *Bioessays,* **25,** 1071-84.
Elion, G. B. (1989) *Science,* **244,** 41-7.
Ellis, N. A., Lennon, D. J., Proytcheva, M., Alhadeff, B., Henderson, E. E. and German, J. (1995) *Am J Hum Genet,* **57,** 1019-27.
Evans, M. J., Carlton, M. B. and Russ, A. P. (1997) *Trends Genet,* **13,** 370-4.
Evans, M. J. and Kaufman, M. H. (1981) *Nature,* **292,** 154-6.
Farber, R. A., Petes, T. D., Dominska, M., Hudgens, S. S. and Liskay, R. M. (1994) *Hum Mol Genet,* **3,** 253-6.
Fedier, A., Schwarz, V. A., Walt, H., Carpini, R. D., Haller, U. and Fink, D. (2001) *Int J Cancer,* **93,** 571-6.
Fink, D., Aebi, S. and Howell, S. B. (1998) *Clin Cancer Res,* **4,** 1-6.
Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M. and Kolodner, R. (1993) *Cell,* **75,** 1027-38.
Flores-Rozas, H. and Kolodner, R. D. (1998) *Proc Natl Acad Sci U S A,* **95,** 12404-9.
Floyd, J. A., Gold, D. A., Concepcion, D., Poon, T. H., Wang, X., Keithley, E., Chen, D., Ward, E. J., Chinn, S. B., Friedman, R. A., Yu, H. T., Moriwaki, K., Shiroishi, T. and Hamilton, B. A. (2003) *Nat Genet,* **35,** 221-8.
Forrester, L. M., Nagy, A., Sam, M., Watt, A., Stevenson, L., Bernstein, A., Joyner, A. L. and Wurst, W. (1996) *Proc Natl Acad Sci U S A,* **93,** 1677-82.
Franchitto, A. and Pichierri, P. (2002) *J Cell Biol,* **157,** 19-30.
Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M. and Ahringer, J. (2000) *Nature,* **408,** 325-30.

Frei, C. and Gasser, S. M. (2000) *J Cell Sci,* **113 (Pt 15),** 2641-6.
Friedrich, G. and Soriano, P. (1991) *Genes Dev,* **5,** 1513-23.
Gaudet, F., Rideout, W. M., 3rd, Meissner, A., Dausman, J., Leonhardt, H. and Jaenisch, R. (2004) *Mol Cell Biol,* **24,** 1640-8.
Gazzoli, I. and Kolodner, R. D. (2003) *Mol Cell Biol,* **23,** 7992-8007.
Genschel, J. and Modrich, P. (2003) *Mol Cell,* **12,** 1077-86.
German, J., Roe, A. M., Leppert, M. F. and Ellis, N. A. (1994) *Proc Natl Acad Sci U S A,* **91,** 6669-73.
Goodwin, N. C., Ishida, Y., Hartford, S., Wnek, C., Bergstrom, R. A., Leder, P. and Schimenti, J. C. (2001) *Nat Genet,* **28,** 310-1.
Goss, K. H., Risinger, M. A., Kordich, J. J., Sanz, M. M., Straughen, J. E., Slovek, L. E., Capobianco, A. J., German, J., Boivin, G. P. and Groden, J. (2002) *Science,* **297,** 2051-3.
Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C. (1989) *Science,* **244,** 463-5.
Griffin, S., Branch, P., Xu, Y. Z. and Karran, P. (1994) *Biochemistry,* **33,** 4787-93.
Groden, J., Nakamura, Y. and German, J. (1990) *Proc Natl Acad Sci U S A,* **87,** 4315-9.
Gu, L., Hong, Y., McCulloch, S., Watanabe, H. and Li, G. M. (1998) *Nucleic Acids Res,* **26,** 1173-8.
Gupta, R. S. (1980) *Somatic Cell Genet,* **6,** 115-25.
Gupta, R. S., Chan, D. Y. and Siminovitch, L. (1978) *Cell,* **14,** 1007-13.
Hannon, G. J. (2002) *Nature,* **418,** 244-51.
Hansen, J., Floss, T., Van Sloun, P., Fuchtbauer, E. M., Vauti, F., Arnold, H. H., Schnutgen, F., Wurst, W., von Melchner, H. and Ruiz, P. (2003) *Proc Natl Acad Sci U S A,* **100,** 9918-22.
Hare, J. T. and Taylor, J. H. (1985) *Proc Natl Acad Sci U S A,* **82,** 7350-4.
Harfe, B. D. and Jinks-Robertson, S. (2000) *Genetics,* **156,** 571-8.
Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. and Bird, A. (1999) *Nature,* **401,** 301-4.
Hickson, I. D. (2003) *Nat Rev Cancer,* **3,** 169-78.
Hidaka, M., Caruana, G., Stanford, W. L., Sam, M., Correll, P. H. and Bernstein, A. (2000) *Mech Dev,* **90,** 3-15.
Hirashima, M., Bernstein, A., Stanford, W. L. and Rossant, J. (2004) *Blood,* **104,** 711-8.
Hitoshi, Y., Lorens, J., Kitada, S. I., Fisher, J., LaBarge, M., Ring, H. Z., Francke, U., Reed, J. C., Kinoshita, S. and Nolan, G. P. (1998) *Immunity,* **8,** 461-71.
Hitotsumachi, S., Carpenter, D. A. and Russell, W. L. (1985) *Proc Natl Acad Sci U S A,* **82,** 6619-21.
Hollingsworth, N. M., Ponte, L. and Halsey, C. (1995) *Genes Dev,* **9,** 1728-39.
Holmes, J., Jr., Clark, S. and Modrich, P. (1990) *Proc Natl Acad Sci U S A,* **87,** 5837-41.
Hooper, M. L. and Slack, C. (1977) *Dev Biol,* **55,** 271-84.
Hsieh, P. (2001) *Mutat Res,* **486,** 71-87.
Hubbard, S. C., Walls, L., Ruley, H. E. and Muchmore, E. A. (1994) *J Biol Chem,* **269,** 3717-24.

Hunter, N. and Borts, R. H. (1997) *Genes Dev,* **11,** 1573-82.
Ishida, Y. and Leder, P. (1999) *Nucleic Acids Res,* **27,** e35.
Ivics, Z., Hackett, P. B., Plasterk, R. H. and Izsvak, Z. (1997) *Cell,* **91,** 501-10.
Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E. and Jaenisch, R. (2001) *Nat Genet,* **27,** 31-9.
Jaenisch, R. (1988) *Science,* **240,** 1468-74.
Jaenisch, R., Jahner, D., Nobis, P., Simon, I., Lohler, J., Harbers, K. and Grotkopp, D. (1981) *Cell,* **24,** 519-29.
Johansson, M. (1999) *Genomics,* **57,** 442-5.
Johnson, R. E., Kovvali, G. K., Guzder, S. N., Amin, N. S., Holm, C., Habraken, Y., Sung, P., Prakash, L. and Prakash, S. (1996) *J Biol Chem,* **271,** 27987-90.
Justice, M. J., Noveroske, J. K., Weber, J. S., Zheng, B. and Bradley, A. (1999) *Hum Mol Genet,* **8,** 1955-63.
Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P. and Ahringer, J. (2003) *Nature,* **421,** 231-7.
Karow, J. K., Constantinou, A., Li, J. L., West, S. C. and Hickson, I. D. (2000) *Proc Natl Acad Sci U S A,* **97,** 6504-8.
Karran, P. (2001) *Carcinogenesis,* **22,** 1931-7.
Karran, P. and Bignami, M. (1992) *Nucleic Acids Res,* **20,** 2933-40.
Karran, P. and Bignami, M. (1994) *Bioessays,* **16,** 833-9.
Karran, P. and Marinus, M. G. (1982) *Nature,* **296,** 868-9.
Kat, A., Thilly, W. G., Fang, W. H., Longley, M. J., Li, G. M. and Modrich, P. (1993) *Proc Natl Acad Sci U S A,* **90,** 6424-8.
Kerr, W. G., Heller, M. and Herzenberg, L. A. (1996) *Proc Natl Acad Sci U S A,* **93,** 3947-52.
Kile, B. T., Hentges, K. E., Clark, A. T., Nakamura, H., Salinger, A. P., Liu, B., Box, N., Stockton, D. W., Johnson, R. L., Behringer, R. R., Bradley, A. and Justice, M. J. (2003) *Nature,* **425,** 81-6.
King, D. P., Zhao, Y., Sangoram, A. M., Wilsbacher, L. D., Tanaka, M., Antoch, M. P., Steeves, T. D., Vitaterna, M. H., Kornhauser, J. M., Lowrey, P. L., Turek, F. W. and Takahashi, J. S. (1997) *Cell,* **89,** 641-53.
Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H., Jr., Kolodner, R. D., Kucherlapati, R., Pollard, J. W. and Edelmann, W. (2000) *Genes Dev,* **14,** 1085-97.
Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A. and Boland, C. R. (1994) *Cancer Res,* **54,** 4308-12.
Kokoska, R. J., Stefanovic, L., Buermeyer, A. B., Liskay, R. M. and Petes, T. D. (1999) *Genetics,* **151,** 511-9.
Kolodner, R. D. and Marsischky, G. T. (1999) *Curr Opin Genet Dev,* **9,** 89-96.
Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A. and Rossant, J. (1988) *Nature,* **335,** 435-7.
Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M. and Kuriyan, J. (1994) *Cell,* **79,** 1233-43.

Kuehn, M. R., Bradley, A., Robertson, E. J. and Evans, M. J. (1987) *Nature,* **326,** 295-8.
Landel, C. P., Chen, S. Z. and Evans, G. A. (1990) *Annu Rev Physiol,* **52,** 841-51.
Lau, P. J. and Kolodner, R. D. (2003) *J Biol Chem,* **278,** 14-7.
Lee, S., Cavallo, L. and Griffith, J. (1997) *J Biol Chem,* **272,** 7532-9.
Leighton, P. A., Mitchell, K. J., Goodrich, L. V., Lu, X., Pinson, K., Scherz, P., Skarnes, W. C. and Tessier-Lavigne, M. (2001) *Nature,* **410,** 174-9.
Levinson, G. and Gutman, G. A. (1987) *Nucleic Acids Res,* **15,** 5323-38.
Li, E., Bestor, T. H. and Jaenisch, R. (1992) *Cell,* **69,** 915-26.
Li, G. M. and Modrich, P. (1995) *Proc Natl Acad Sci U S A,* **92,** 1950-4.
Lindsay, E. A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H. F., Scambler, P. J., Bradley, A. and Baldini, A. (2001) *Nature,* **410,** 97-101.
Lipkin, S. M., Moens, P. B., Wang, V., Lenzi, M., Shanmugarajah, D., Gilgeous, A., Thomas, J., Cheng, J., Touchman, J. W., Green, E. D., Schwartzberg, P., Collins, F. S. and Cohen, P. E. (2002) *Nat Genet,* **31,** 385-90.
Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., Elliott, R. M. and Collins, F. S. (2000) *Nat Genet,* **24,** 27-35.
Liu, Y., Kao, H. I. and Bambara, R. A. (2004) *Annu Rev Biochem,* **73,** 589-615.
Lonn, U., Lonn, S., Nylen, U., Winblad, G. and German, J. (1990) *Cancer Res,* **50,** 3141-5.
Lu, X. and Lane, D. P. (1993) *Cell,* **75,** 765-78.
Lund, A. H., Turner, G., Trubetskoy, A., Verhoeven, E., Wientjens, E., Hulsman, D., Russell, R., DePinho, R. A., Lenz, J. and van Lohuizen, M. (2002) *Nat Genet,* **32,** 160-5.
Lundgren, M., Chow, C. M., Sabbattini, P., Georgiou, A., Minaee, S. and Dillon, N. (2000) *Cell,* **103,** 733-43.
Luo, G., Santoro, I. M., McDaniel, L. D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R. A. and Bradley, A. (2000) *Nat Genet,* **26,** 424-9.
luria, S. a. D. M. (1943) *Genetics,* **28,** 491-511.
Mainguy, G., Montesinos, M. L., Lesaffre, B., Zevnik, B., Karasawa, M., Kothary, R., Wurst, W., Prochiantz, A. and Volovitch, M. (2000) *Nat Biotechnol,* **18,** 746-9.
McDaniel, L. D., Chester, N., Watson, M., Borowsky, A. D., Leder, P. and Schultz, R. A. (2003) *DNA Repair (Amst),* **2,** 1387-404.
Menissier de Murcia, J., Ricoul, M., Tartier, L., Niedergang, C., Huber, A., Dantzer, F., Schreiber, V., Ame, J. C., Dierich, A., LeMeur, M., Sabatier, L., Chambon, P. and de Murcia, G. (2003) *Embo J,* **22,** 2255-63.
Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., Berns, A. and Romeyn, L. (2002) *Nat Genet,* **32,** 153-9.
Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R. and Bird, A. (2002) *Science,* **297,** 403-5.
Mitchell, K. J., Pinson, K. I., Kelly, O. G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P. A., Goodrich, L. V., Lu, X., Avery, B. J., Tate, P., Dill, K.,

Pangilinan, E., Wakenight, P., Tessier-Lavigne, M. and Skarnes, W. C. (2001) *Nat Genet,* **28,** 241-9.
Miura, A., Yonebayashi, S., Watanabe, K., Toyama, T., Shimada, H. and Kakutani, T. (2001) *Nature,* **411,** 212-4.
Modrich, P. (1991) *Annu Rev Genet,* **25,** 229-53.
Modrich, P. (1997) *J Biol Chem,* **272,** 24727-30.
Modrich, P. and Lahue, R. (1996) *Annu Rev Biochem,* **65,** 101-33.
Morgenstern, J. P. and Land, H. (1990) *Nucleic Acids Res,* **18,** 3587-96.
Munroe, R. J., Bergstrom, R. A., Zheng, Q. Y., Libby, B., Smith, R., John, S. W., Schimenti, K. J., Browning, V. L. and Schimenti, J. C. (2000) *Nat Genet,* **24,** 318-21.
Muth, K., Bruyns, R., Thorey, I. S. and von Melchner, H. (1998) *Dev Dyn,* **212,** 277-83.
Nakayama, H. (2002) *Oncogene,* **21,** 9008-21.
Nassif, N. and Engels, W. (1993) *Proc Natl Acad Sci U S A,* **90,** 1262-6.
Nelson, F. K., Frankel, W. and Rajan, T. V. (1989) *Mol Cell Biol,* **9,** 1284-8.
Nicholson, A., Hendrix, M., Jinks-Robertson, S. and Crouse, G. F. (2000) *Genetics,* **154,** 133-46.
Nolan, G. P. and Shatzman, A. R. (1998) *Curr Opin Biotechnol,* **9,** 447-50.
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., et al. (2002) *Nature,* **420,** 563-73.
Oren, M. and Rotter, V. (1999) *Cell Mol Life Sci,* **55,** 9-11.
Paddison, P. J., Silva, J. M., Conklin, D. S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., Chang, K., Westbrook, T., Cleary, M., Sachidanandam, R., McCombie, W. R., Elledge, S. J. and Hannon, G. J. (2004) *Nature,* **428,** 427-31.
Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J. J. and Jiricny, J. (1995) *Science,* **268,** 1912-4.

Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D. and et al. (1994) *Science,* **263,** 1625-9.

Pardo, P. S., Leung, J. K., Lucchesi, J. C. and Pereira-Smith, O. M. (2002) *J Biol Chem,* **277,** 50860-6.

Parker, B. O. and Marinus, M. G. (1992) *Proc Natl Acad Sci U S A,* **89,** 1730-4.

Peltomaki, P. (2001) *Mutat Res,* **488,** 77-85.

Perrimon, N. (1998) *Int J Dev Biol,* **42,** 243-7.

Pothof, J., van Haaften, G., Thijssen, K., Kamath, R. S., Fraser, A. G., Ahringer, J., Plasterk, R. H. and Tijsterman, M. (2003) *Genes Dev,* **17,** 443-8.

Potter, T. A., Zeff, R. A., Frankel, W. and Rajan, T. V. (1987) *Proc Natl Acad Sci U S A,* **84,** 1634-7.

Prolla, T. A., Baker, S. M., Harris, A. C., Tsao, J. L., Yao, X., Bronner, C. E., Zheng, B., Gordon, M., Reneker, J., Arnheim, N., Shibata, D., Bradley, A. and Liskay, R. M. (1998) *Nat Genet,* **18,** 276-9.

Prolla, T. A., Christie, D. M. and Liskay, R. M. (1994) *Mol Cell Biol,* **14,** 407-15.

Rajan, T. V., Halay, E. D., Potter, T. A., Evans, G. A., Seidman, J. G. and Margulies, D. H. (1983) *Embo J,* **2,** 1537-42.

Ramilo, C., Gu, L., Guo, S., Zhang, X., Patrick, S. M., Turchi, J. J. and Li, G. M. (2002) *Mol Cell Biol,* **22,** 2037-46.

Ramirez-Solis, R., Davis, A. C. and Bradley, A. (1993) *Methods Enzymol,* **225,** 855-78.

Rayssiguier, C., Dohet, C. and Radman, M. (1991) *Biochimie,* **73,** 371-4.

Rayssiguier, C., Thaler, D. S. and Radman, M. (1989) *Nature,* **342,** 396-401.

Reaume, A. G., de Sousa, P. A., Kulkarni, S., Langille, B. L., Zhu, D., Davies, T. C., Juneja, S. C., Kidder, G. M. and Rossant, J. (1995) *Science,* **267,** 1831-4.

Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggeri, B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G. and Bellacosa, A. (1999) *Nat Genet,* **23,** 266-8.

Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986) *Nature,* **323,** 445-8.

Roos, W., Baumgartner, M. and Kaina, B. (2004) *Oncogene,* **23,** 359-67.

Ross-Macdonald, P. and Roeder, G. S. (1994) *Cell,* **79,** 1069-80.

Ruggiero, B. L. and Topal, M. D. (2004) *J Biol Chem,* **279,** 23088-97.

Saintigny, Y. and Lopez, B. S. (2002) *Oncogene,* **21,** 488-92.

Sansom, O. J., Zabkiewicz, J., Bishop, S. M., Guy, J., Bird, A. and Clarke, A. R. (2003) *Oncogene,* **22,** 7130-6.

Schmutte, C. and Fishel, R. (1999) *Anticancer Res,* **19,** 4665-96.

Schroeder, T., Fraser, S. T., Ogawa, M., Nishikawa, S., Oka, C., Bornkamm, G. W., Honjo, T. and Just, U. (2003) *Proc Natl Acad Sci U S A,* **100,** 4018-23.

Sega, G. A. (1984) *Mutat Res,* **134,** 113-42.

Selva, E. M., Maderazo, A. B. and Lahue, R. S. (1997) *Mol Gen Genet,* **257,** 71-82.

Selva, E. M., New, L., Crouse, G. F. and Lahue, R. S. (1995) *Genetics,* **139,** 1175-88.

Sengupta, S., Linke, S. P., Pedeux, R., Yang, Q., Farnsworth, J., Garfield, S. H., Valerie, K., Shay, J. W., Ellis, N. A., Wasylyk, B. and Harris, C. C. (2003) *Embo J,* **22,** 1210-22.

Sharov, A. A., Piao, Y., Matoba, R., Dudekula, D. B., Qian, Y., VanBuren, V., Falco, G., Martin, P. R., Stagg, C. A., Bassey, U. C., Wang, Y., Carter, M. G., Hamatani, T., Aiba, K., Akutsu, H., Sharova, L., Tanaka, T. S., Kimber, W. L., Yoshikawa, T., Jaradat, S. A., Pantano, S., Nagaraja, R., Boheler, K. R., Taub, D., Hodes, R. J., Longo, D. L., Schlessinger, D., Keller, J., Klotz, E., Kelsoe, G., Umezawa, A., Vescovi, A. L., Rossant, J., Kunath, T., Hogan, B. L., Curci, A., D'Urso, M., Kelso, J., Hide, W. and Ko, M. S. (2003) *PLoS Biol,* **1,** E74.

Shen, P. and Huang, H. V. (1986) *Genetics,* **112,** 441-57.

Shen, P. and Huang, H. V. (1989) *Mol Gen Genet,* **218,** 358-60.

Shirai, M., Miyashita, A., Ishii, N., Itoh, Y., Satokata, I., Watanabe, Y. G. and Kuwano, R. (1996) *Zoolog Sci,* **13,** 277-83.

Siminovitch, L. (1976) *Cell,* **7,** 1-11.

Skarnes, W. C. (2000) *Methods Enzymol,* **328,** 592-615.

Slebos, R. J. and Taylor, J. A. (2001) *Biochem Biophys Res Commun,* **281,** 212-9.

Somia, N. (2004) *Methods Mol Biol,* **246,** 463-90.

Sonoda, E., Sasaki, M. S., Morrison, C., Yamaguchi-Iwai, Y., Takata, M. and Takeda, S. (1999) *Mol Cell Biol,* **19,** 5166-9.

Soriano, P. (1999) *Nat Genet,* **21,** 70-1.

Soriano, P., Friedrich, G. and Lawinger, P. (1991) *J Virol,* **65,** 2314-9.

Spillare, E. A., Robles, A. I., Wang, X. W., Shen, J. C., Yu, C. E., Schellenberg, G. D. and Harris, C. C. (1999) *Genes Dev,* **13,** 1355-60.

Stanford, W. L., Caruana, G., Vallis, K. A., Inamdar, M., Hidaka, M., Bautch, V. L. and Bernstein, A. (1998) *Blood,* **92,** 4622-31.

Stanford, W. L., Cohn, J. B. and Cordes, S. P. (2001) *Nat Rev Genet,* **2,** 756-68.

Stern, C. (1936) *Genetics,* **21,** 625-730.

Stewart, E., Chapman, C. R., Al-Khodairy, F., Carr, A. M. and Enoch, T. (1997) *Embo J,* **16,** 2682-92.

Stojic, L., Mojas, N., Cejka, P., Di Pietro, M., Ferrari, S., Marra, G. and Jiricny, J. (2004) *Genes Dev,* **18,** 1331-44.

Stuhlmann, H. (2003) *Methods Enzymol,* **365,** 386-406.

Su, H., Wang, X. and Bradley, A. (2000) *Nat Genet,* **24,** 92-5.

Susse, S., Janz, C., Janus, F., Deppert, W. and Wiesmuller, L. (2000) *Oncogene,* **19,** 4500-12.

Swann, P. F., Waters, T. R., Moulton, D. C., Xu, Y. Z., Zheng, Q., Edwards, M. and Mace, R. (1996) *Science,* **273,** 1109-11.

Swing, D. A. and Sharan, S. K. (2004) *Methods Mol Biol,* **256,** 183-98.

Tate, P., Lee, M., Tweedie, S., Skarnes, W. C. and Bickmore, W. A. (1998) *J Cell Sci,* **111 (Pt 17),** 2575-85.

te Riele, H., Maandag, E. R. and Berns, A. (1992) *Proc Natl Acad Sci U S A,* **89,** 5128-32.

te Riele, H., Maandag, E. R., Clarke, A., Hooper, M. and Berns, A. (1990) *Nature,* **348,** 649-51.
Thomas, K. R. and Capecchi, M. R. (1987) *Cell,* **51,** 503-12.
Thompson, L. H., Busch, D. B., Brookman, K., Mooney, C. L. and Glaser, D. A. (1981) *Proc Natl Acad Sci U S A,* **78,** 3734-7.
Thompson, L. H., Rubin, J. S., Cleaver, J. E., Whitmore, G. F. and Brookman, K. (1980) *Somatic Cell Genet,* **6,** 391-405.
Thorey, I. S., Muth, K., Russ, A. P., Otte, J., Reffelmann, A. and von Melchner, H. (1998) *Mol Cell Biol,* **18,** 3081-8.
Tidd, D. M. and Paterson, A. R. (1974) *Cancer Res,* **34,** 738-46.
Tishkoff, D. X., Boerger, A. L., Bertrand, P., Filosi, N., Gaida, G. M., Kane, M. F. and Kolodner, R. D. (1997a) *Proc Natl Acad Sci U S A,* **94,** 7487-92.
Tishkoff, D. X., Filosi, N., Gaida, G. M. and Kolodner, R. D. (1997b) *Cell,* **88,** 253-63.
Umar, A., Buermeyer, A. B., Simon, J. A., Thomas, D. C., Clark, A. B., Liskay, R. M. and Kunkel, T. A. (1996) *Cell,* **87,** 65-73.
Umar, A., Koi, M., Risinger, J. I., Glaab, W. E., Tindall, K. R., Kolodner, R. D., Boland, C. R., Barrett, J. C. and Kunkel, T. A. (1997) *Cancer Res,* **57,** 3949-55.
Vallis, K. A., Chen, Z., Stanford, W. L., Yu, M., Hill, R. P. and Bernstein, A. (2002) *Radiat Res,* **157,** 8-18.
Veigl, M. L., Kasturi, L., Olechnowicz, J., Ma, A. H., Lutterbaugh, J. D., Periyasamy, S., Li, G. M., Drummond, J., Modrich, P. L., Sedwick, W. D. and Markowitz, S. D. (1998) *Proc Natl Acad Sci U S A,* **95,** 8698-702.
Vertino, P. M., Sekowski, J. A., Coll, J. M., Applegren, N., Han, S., Hickey, R. J. and Malkas, L. H. (2002) *Cell Cycle,* **1,** 416-23.
Vidal, F., Lopez, P., Lopez-Fernandez, L. A., Ranc, F., Scimeca, J. C., Cuzin, F. and Rassoulzadegan, M. (2001) *J Cell Sci,* **114,** 435-43.
von Melchner, H., DeGregori, J. V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H. E. (1992) *Genes Dev,* **6,** 919-27.
von Melchner, H. and Ruley, H. E. (1989) *J Virol,* **63,** 3227-33.
Waldman, A. S. and Liskay, R. M. (1988) *Mol Cell Biol,* **8,** 5350-7.
Wang, X. W., Tseng, A., Ellis, N. A., Spillare, E. A., Linke, S. P., Robles, A. I., Seker, H., Yang, Q., Hu, P., Beresten, S., Bemmels, N. A., Garfield, S. and Harris, C. C. (2001) *J Biol Chem,* **276,** 32948-55.
Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J. and Qin, J. (2000) *Genes Dev,* **14,** 927-39.
Wang, Y. and Qin, J. (2003) *Proc Natl Acad Sci U S A,* **100,** 15387-92.
Wasmuth, J. J. and Vock Hall, L. (1984) *Cell,* **36,** 697-707.
Waters, T. R. and Swann, P. F. (1997) *Biochemistry,* **36,** 2501-6.
Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M.,

Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., et al. (2002) *Nature,* **420,** 520-62.

Wei, K., Clark, A. B., Wong, E., Kane, M. F., Mazur, D. J., Parris, T., Kolas, N. K., Russell, R., Hou, H., Jr., Kneitz, B., Yang, G., Kunkel, T. A., Kolodner, R. D., Cohen, P. E. and Edelmann, W. (2003) *Genes Dev,* **17,** 603-14.

Wei, K., Kucherlapati, R. and Edelmann, W. (2002) *Trends Mol Med,* **8,** 346-53.

Wempe, F., Yang, J. Y., Hammann, J. and von Melchner, H. (2001) *Genome Biol,* **2,** RESEARCH0023.

Westerveld, A., Hoeijmakers, J. H., van Duin, M., de Wit, J., Odijk, H., Pastink, A., Wood, R. D. and Bootsma, D. (1984) *Nature,* **310,** 425-9.

Willers, H., McCarthy, E. E., Alberti, W., Dahm-Daphi, J. and Powell, S. N. (2000) *Int J Radiat Biol,* **76,** 1055-62.

Wobus, A. M. (2001) *Mol Aspects Med,* **22,** 149-64.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A. M., Lipkin, M. and Edelmann, W. (2002) *Proc Natl Acad Sci U S A,* **99,** 14937-42.

Worton, R. G. (1978) *Cytogenet Cell Genet,* **21,** 105-10.

Wu, J., Davis, M. D. and Owens, R. A. (1999) *J Virol,* **73,** 8235-44.

Wu, L., Davies, S. L., Levitt, N. C. and Hickson, I. D. (2001) *J Biol Chem,* **276,** 19375-81.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D. P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A. and et al. (1995) *Genetics,* **139,** 889-99.

Yamada, N. A., Castro, A. and Farber, R. A. (2003) *Mutagenesis,* **18,** 277-82.

Yamada, T., Koyama, T., Ohwada, S., Tago, K., Sakamoto, I., Yoshimura, S., Hamada, K., Takeyoshi, I. and Morishita, Y. (2002) *Cancer Lett,* **181,** 115-20.

Yamane, K., Taylor, K. and Kinsella, T. J. (2004) *Biochem Biophys Res Commun,* **318,** 297-302.

Yang, Q., Zhang, R., Wang, X. W., Spillare, E. A., Linke, S. P., Subramanian, D., Griffith, J. D., Li, J. L., Hickson, I. D., Shen, J. C., Loeb, L. A., Mazur, S. J., Appella, E., Brosh, R. M., Jr., Karmakar, P., Bohr, V. A. and Harris, C. C. (2002) *J Biol Chem,* **277,** 31980-7.

Yee, J. K., Moores, J. C., Jolly, D. J., Wolff, J. A., Respess, J. G. and Friedmann, T. (1987) *Proc Natl Acad Sci U S A,* **84,** 5197-201.

You, Y., Bergstrom, R., Klemm, M., Lederman, B., Nelson, H., Ticknor, C., Jaenisch, R. and Schimenti, J. (1997) *Nat Genet,* **15,** 285-8.
Yu, S. F., von Ruden, T., Kantoff, P. W., Garber, C., Seiberg, M., Ruther, U., Anderson, W. F., Wagner, E. F. and Gilboa, E. (1986) *Proc Natl Acad Sci U S A,* **83,** 3194-8.
Yu, Y. and Bradley, A. (2001) *Nat Rev Genet,* **2,** 780-90.
Yusa, K., Horie, K., Kondoh, G., Kouno, M., Maeda, Y., Kinoshita, T. and Takeda, J. (2004) *Nature,* **429,** 896-9.
Zambrowicz, B. P. and Friedrich, G. A. (1998) *Int J Dev Biol,* **42,** 1025-36.
Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G. and Soriano, P. (1997) *Proc Natl Acad Sci U S A,* **94,** 3789-94.
Zhang, H., Richards, B., Wilson, T., Lloyd, M., Cranston, A., Thorburn, A., Fishel, R. and Meuth, M. (1999) *Cancer Res,* **59,** 3021-7.
Zhang, Y., Proenca, R., Maffei, M., Barone, M., Leopold, L. and Friedman, J. M. (1994) *Nature,* **372,** 425-32.
Zheng, B., Sage, M., Cai, W. W., Thompson, D. M., Tavsanli, B. C., Cheah, Y. C. and Bradley, A. (1999) *Nat Genet,* **22,** 375-8.