CHAPTER 1. INTRODUCTION

1.1 The mouse as a genetic model animal

1.1.1 Evolution divergence of the mouse and the human

The human and mouse lineages diverged roughly 60–100 million years ago according to phylogenetic research based on molecular analysis of certain loci sequences (Eizirik *et al.* 2001). That was at about the same time as the extinction of the dinosaurs (Hedges *et al.* 1996). During that time, genomes have experienced many alternations, including nucleotide substitutions, deletions, insertions, duplications, inversions and translocations, which underscore the differences between the species.

Nonetheless, by comparing human and mouse DNA sequences, similar gene functions were found when gene sequences were conserved (highly similar). Many new genes' functions were identified when their conserved counterparts' functions were known in the other species (Dehal *et al.* 2001; Loots *et al.* 2000; O'Brien *et al.* 1999; Oeltjen *et al.* 1997; Pennacchio *et al.* 2001).

In both anatomy and physiology, mice are similar to humans. These characteristics, in addition to other facts such as their smaller size (25-40 grams for an adult), short generation time (10 weeks from birth to reproducing the next generation), high proliferation ability and the availability of inbred strains, make mice a major animal model for biomedical research (Silver 1995).

1.1.2 The history of the laboratory mouse

Early mouse genetics research started with William Ernest Castle. Castle had been researching mouse genetics and genetic variation using the fancy mouse from 1902 (Silver 1995). Castle and his colleague Clarence Little realized that they needed to use inbred mouse strains (genetically homogeneous) to avoid the complexities of genetic background. Little started the first mating to produce an inbred mouse strain in 1909. Little generated the DBA strains, which carried coat colour loci of dilute (*d*), brown (*b*) and non-agouti (*a*). After moving to the Cold Spring Harbor Laboratory, Little developed other widely used mouse inbred strains B6, B10 and BALB/c (Silver 1995). These inbred mouse lines provided an early standard experimental material for mouse research. The research outcomes using the same strain became comparable even when they were carried out in different laboratories throughout the world.

Genes were slowly mapped onto mouse chromosomes long before the age of molecular biology. In 1915, Haldane first mapped the linkage between the pink-eye dilution (*p*) and albino (*c*) loci, a linkage group which was eventually assigned to mouse chromosome 7 (Haldane 1915). There were only 3 loci mapped on chromosome 7 by 1935 and 8 by 1954 (Fox 2007).

However, to study allele linkage relationships, a physical genome map was needed. With the development of recombinant DNA technology and DNA sequence-based polymorphism analysis technology, many mouse genes were mapped on their genome. These technologies included the discovery and application of Restriction Fragment Length Polymorphisms (RFLPs), which generated different-sized DNA fragments by digesting a DNA sample with a variety of endonucleases separately or digesting a set of DNA samples with one endonuclease. RFLPs were first used to construct physical maps of small genomes (Kiko *et al.* 1979). The endonucleases' digestion sites were mapped on the genome by analysing the pattern of fragments. In addition, RFLPs facilitated gene linkage research and helped to map DNA clones into individual chromosomes (de Martinville *et al.* 1982; Wieacker *et al.* 1983). This could be achieved by hybridizing DNA fragments with endonuclease-digested genomic DNA and analysing the resulting fragment pattern.

1.1.3 Mouse genomics

During the 1990s, large genome sequencing projects were launched to decipher the order of the billions of nucleotides in human DNA (Lander *et al.* 2001), and then *Mus musculus* (Waterston *et al.* 2002). This opened a genomic era in which genes were mapped, gene structures were annotated and genetic variations could be compared between different genetic backgrounds of the same species or between different species.

Compared with the human genome, the mouse has a similar genome size and gene numbers. Its 3.4×10^9 bp genome (NCBI m37.1, July 2007) is similar to the human genome's 3.2×10^9 bp (NCBI 36.2, Sept 2006), while both have 22,000 known genes (www.ensembl.org). The mouse has 19 pairs of autosomes while the human has 22 pairs of autosomes, in addition to the X and Y sex chromosomes. The shared lineage has resulted in an excess of 90% of the mouse genome having synteny (highly similar sequences, which may be inherited from the same ancestor) with the human (Waterston *et al.* 2002). The conserved synteny between the human and the mouse are illustrated in Figure 1-1 and Figure 1-2. The nucleotide sequence of genomes is not able to predict physiological

functions. Scientists set about deciphering the physiological function of every gene. The isolation of mouse embryonic stem (ES) cells (Evans *et al.* 1981) and the germ line transmission capacity of genetically engineered ES cells greatly facilitated this progress.

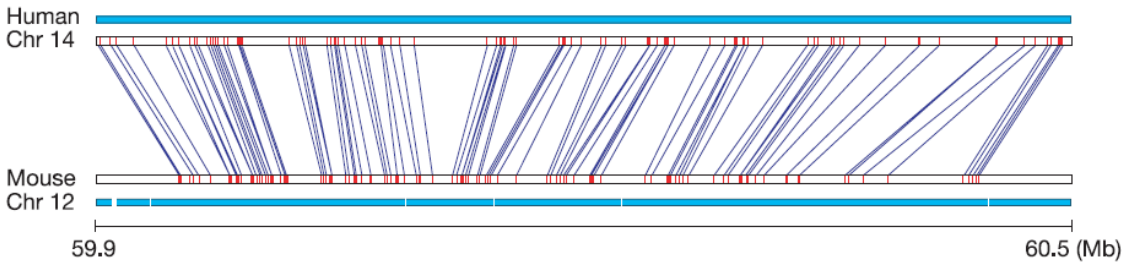


Figure 1-1 Conservation of synteny between human chromosome 14 and mouse chromosome 12 (part)

An example of 600 kb of highly conserved sequence. Reciprocally unique landmarks in human chromosome 14 are detected along a 510 kb segment of mouse chromosome 12. These landmarks are illustrated by blue lines. The cyan bars indicate the conserved synergy regions on both the human and mouse chromosomes. This figure is copied from Waterston *et al*. (Waterston *et al.* 2002)

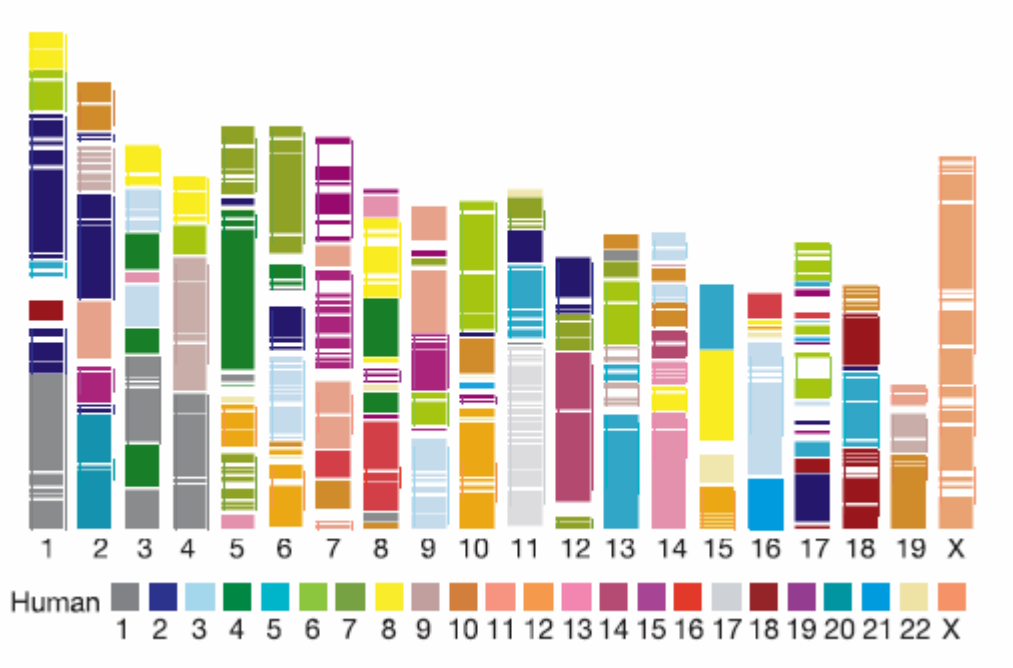


Figure 1-2 Conserved synteny between humans and mice

DNA fragments >300 kb in size with conserved synteny in the human are superimposed on the mouse genome (chromosomes 1–19 and X). Each colour corresponds to a particular human chromosome. The figure is copied form Waterston *et al*. (2002).

1.2 Mouse embryonic stem cells as genetic research tools

Since mouse ES cells were first isolated (Evans *et al.* 1981), they have become widely used as tools in genetic research. In 1984, Bradley cultured ES cells and, following their incorporation into chimeric mice following blastocyst injection, demonstrated that they were able to form functional germ cells (Bradley *et al.* 1984). Furthermore, Evans' laboratory demonstrated that genetic modification of cultured ES cells could be performed and the altered cells were capable of contributing to the mouse germ line (Robertson *et al.* 1986). In 1987, Evans' laboratory generated the first ES cell-derived mice with an altered phenotype (Kuehn *et al.* 1987). In parallel, homologous recombination was developed in mammalian cells (Folger *et al.* 1982; Upcroft *et al.* 1980; Upcroft *et al.* 1980), which was subsequently applied to ES cells (Doetschman *et al.* 1987; Thomas *et al.* 1987). These techniques were primarily used as a genetic loss functional mutagenic approach (knock out technique) (Capecchi 1989). Moreover, several gene-trap mutagenesis systems were developed around the same period (Allen *et al.* 1988; Gossler *et al.* 1989; Kothary *et al.* 1988; Weber *et al.* 1984). All of these findings started a new field of manipulative mouse molecular genetics.

The combination of gene knockout and gene-trap techniques of ES cells has individually mutated ~3,000 genes in the mouse genome (personal communication). This is a spectacular achievement in 20 years, but the progress was a little slow compared with the generation of the genomic sequence and the structural annotation of the genome. Fortunately, the global scientific community has launched collaborative programmes (KOMP – Knockout Mouse Project, EUCOMM – European Conditional Mouse Mutagenesis and NorCOMM – North American Conditional Mouse Mutagenesis), sponsored by the US National Institutes of Health, the European Commission and Genome Canada respectively, to mutate all of the mouse genes within several years. In addition, to maximize the utility of these resources, all collaborating partners have agreed principles, which will allow the resources made by these efforts to be freely available to scientific communities.

Considering the balance of the cost, time and phenotype produced, slightly different knockout strategies were designed. The EUCOMM and NorCOMM projects will focus on generating conditional knockouts. The alleles contain *loxP* sites. Loss of gene function can be achieved in tissue culture or by crossing this mouse with another mouse carrying the germ line or tissue-specific expressed Cre. Although knockouts have many advantages in genetic research, they were performed one by one. Therefore knockouts cannot screen the genome for specific functions.

1.3 Genetic screens

A genetic screen is an approach to identify gene functions by observing and analysing phenotypes of mutants and mapping the chromosomal location of the mutations. Unlike the "gene driven" research strategies (reverse genetics starting from specific genes), genetic screens do not require any previous knowledge of the genes, and are thus regarded as "phenotype-driven" strategies or forward genetics. One advantage of this strategy is that it can examine a set of genes, or even the whole genome for a specific function. Thus, it was used as a high-throughput genetic tool.

In 1980, a large-scale genetic screen was carried out on *Drosophila melanogaster* (Nusslein-Volhard *et al.* 1980). This Nobel Prize-winning work was the first attempt to identify a set of mutant genes in a multicellular organism, which generated mutant phenotypes in a particular biological process (embryonic patterning in this case). In this research, male flies were fed and mutated by Ethyl Methane Sulphonate (EMS) and crossed with wild type virgin female flies carrying a balancer chromosome (a chromosome with one or more inverted segments that suppress recombination). Point mutations generated by EMS in male flies can be inherited by F1 male flies. Single F1 males that carry a mutagenized chromosome in trans to the balancer were then backcrossed to balancer stock to generate F2 males and females that carry the same mutagenized chromosome. Then, F2 male and female flies were crossed to generate homozygous mutants for analysis. Balancer chromosomes not only suppress recombination, but also allow lethal mutations to be maintained in heterozygous mutants without selection.

Genetic screens have been designed and applied to a variety of model organisms, including *E. coli*, yeast, *Caenorhabditis elegans*, filamentous fungi, zebrafish, *Arabidopsis thaliana* and mice, from tissue culture to animals. These screens are reviewed in the following literature (Casselton *et al.* 2002; Forsburg 2001; Grimm 2004; Jorgensen *et al.* 2002; Kile *et al.* 2005; Page *et al.* 2002; Patton *et al.* 2001; Shuman *et al.* 2003; St Johnston 2002).

Genetic screens can be implemented in several alternative directions: gain, loss or modification of function. One application of this strategy is expression cloning, in which individual cells in culture acquire a new characteristic caused by transient or permanent expression of a transformed recombinant cDNA/DNA fragment from a library. For example, using expression cloning, a recombinant cDNA clone encoding human pancreatic growth hormone-releasing factor (GRF) within a larger precursor protein was identified (Mayo *et al.* 1983). In this experiment, RNA from human pancreatic tumour samples was reverse-

transcribed before ligating its complementary DNA into a bacteria expression vector. Independent colonies were hybridized with oligonucleotides corresponding to amino acids 24–27 of GRF. Clones were obtained that were positive in both the hybridization experiment and immunological assay (anti-GRF serum binding). Thus, the clones were confirmed to encode human pancreatic growth hormone-releasing factor. Similar approaches also helped to isolate the associated genes of the human autosomal recessive disorder *xeroderma pigmentosum* (XP). XP is a human hereditary disease with the cancer predisposition on skin exposed to sunshine (Friedberg *et al.* 2005). XP cells are defective in DNA repair, and complementation of this defect has been used to identify eight genetic groups (A–G and variant). Genetic screens were used in the discovery of some XP-associated DNA repair genes, such as the human nucleotide excision repair (NER) gene *ERCC4* (Thompson *et al.* 1994) and *XPC* gene (Legerski *et al.* 1992).

On the other hand, genetic screens can be carried out on mutant libraries, in which each clone has lost gene function caused by the mutagen. For instance, recessive mutations leading to abnormal circadian rhythmicity were identified in N-ethyl-N-nitrosourea (ENU) mutated mice (Siepka *et al.* 2007). Moreover, a randomly modified cell population may gain new functions, thus distinguishing themselves from other cells. For example, through using mouse ES cell gene-trap mutagenesis, early developmentally regulated genes in chimeric embryos were studied based on the mutants' spatial and temporal expression patterns generated by X-gal staining in order to detect the expression of a *lacZ* reporter gene (Wurst *et al.* 1995). This study revealed that 13% of 279 gene-trap clones resulted in restricted expression in chimeric embryos at 8.5 days post coitus (dpc). Fifty-five per cent (155) of chimeras do not have detectable expression, while one third of these 155 chimeras showed detectable reporter gene expression at 12.5 dpc. This research demonstrated that genes were differentially expressed spatially and temporally, and it was possible to study the embryonic development process by analysing a set of mutants.

However, this type of genetic screen has limitations: A. The mutant cells used to generate chimeric mice were heterozygous mutants. This meant that the phenotype of mutant cells was observable only when mutations were dominant. B. Chimeric mouse embryos had wild type cells, which may interfere with the mutant phenotype. C. To obtain pure heterozygous and homozygous mutant mice, much more labour, time and resources would be needed in the breeding of mice.

It would be a significant improvement to overcome the above limitations if homozygosity of mutant cells could be achieved. In principle, if it is feasible to generate homozygous

mutants in ES cells, any gene in the functional pathways within ES cells can be screened. The current approaches to achieve homozygous mutant ES cells will be discussed below.

1.4 Strategies to make homozygous mutations in ES cells

1.4.1 Sequential gene targeting

Gene-targeting technology is widely used to study gene function by destroying a specific gene's function using homology recombination. Homozygous targeted mutations in ES cells can be achieved in two ways, as shown in Figure 1-3: (I). Two targeting vectors with different drug resistance markers are used sequentially to generate mutations in both alleles of a given gene by two steps of targeting (te Riele *et al.* 1990); (II). Alternatively one drug resistance marker gene flanked by two *loxP* sites can be used. After targeting, Cre-mediated recombination can be used to remove the selection cassette from the locus and a second targeting event using the same targeting vector can be selected. This method has been termed vector recycling (Abuin *et al.* 1996).

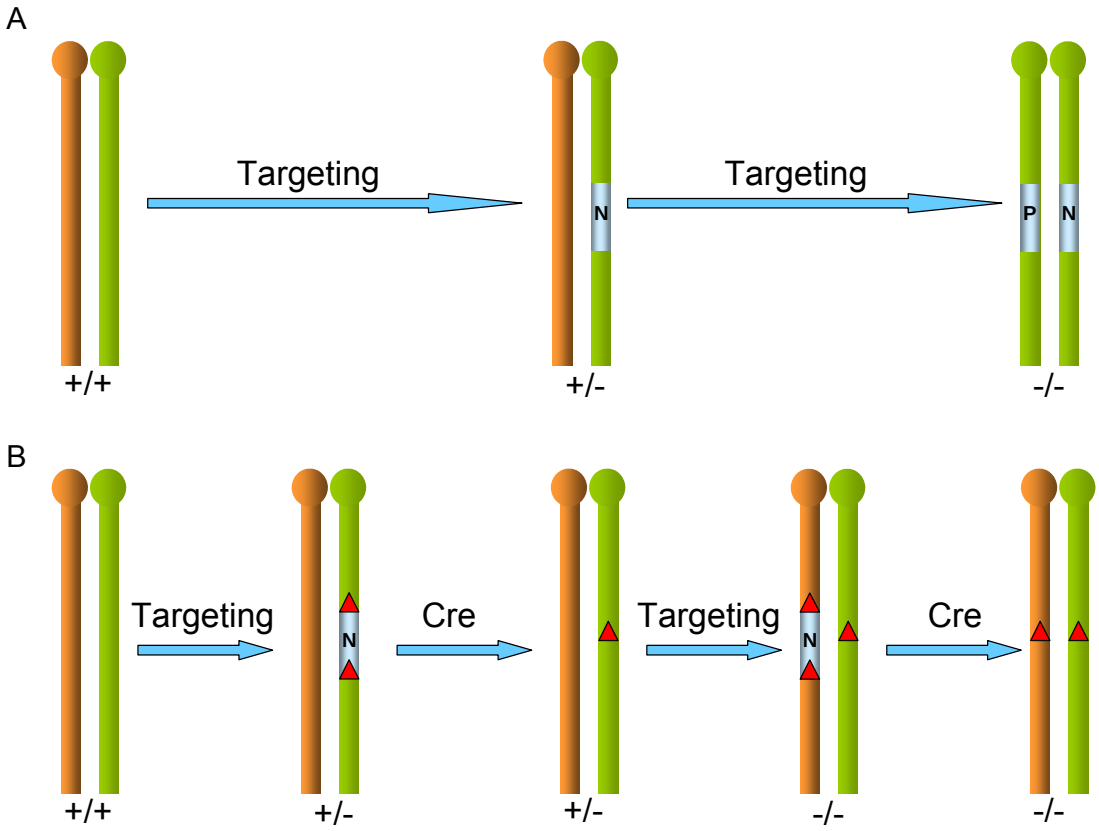


Figure 1-3 Sequential gene targeting

A. Two alleles of any given locus can be targeted sequentially with two selection markers, one with *Neo* (N) and a second with puromycin (P). **B**. Targeting using a *loxP* (red triangle)-flanked *Neo* cassette (N). Cre-mediated recombination can be used to remove the selection marker, which recycles the cassette and additionally prevents bystander effects.

1.4.2 High concentration of G418 selection

Several mechanisms of loss of heterozygosity (LOH) can be used to generate homozygous mutations in ES cells. They are whole chromosome loss followed by chromosome duplication, gene conversion and mitotic recombination. It has been shown that heterozygous targeted ES cells (with a Neomycin resistance gene) can segregate mutations in the same daughter cells in mitosis and these can be specifically selected in the media containing a high concentration of G418 (Mortensen *et al.* 1992).

This high G418 concentration approach is relatively easy to implement and it has been successfully used to recover homozygous mutations in four genes on chromosomes 2, 5, 10 and 17 in the R1 ES cells; this approach is shown in Figure 1-4 (Lefebvre *et al.* 2001). The R1 ES cell was established from a hybrid embryo generated from a cross between mouse inbred strains 129X1 and 129S3 (Festing *et al.* 1999; Nagy *et al.* 1993). As genetic polymorphisms can be identified by simple sequence length polymorphisms (SSLP) between 129 mouse substrains (Simpson *et al.* 1997; Threadgill *et al.* 1997), SSLP analysis was carried out to determine the cause of LOH in these high G418 recovered clones. In homozygous mutant cell lines, PCR analysis of SSLP markers polymorphic between the parental 129S3 and 129X1 chromosomes in R1 ES cells indicated that all distant linked markers (16–66 cM from *Neo* insertion) were homozygous. This confirmed that LOH in the above homozygous targeted cell lines was caused by two mechanisms: (1). chromosome loss and re-duplication; (2). mitotic recombination proximal of the *Neo*-targeted locus.

Other researchers have also recovered homozygous mutants in ES cells using this strategy (Carmeliet *et al.* 1996; Dufort *et al.* 1998; Reaume *et al.* 1995). However, to recover a homozygous mutant ES cell line, a specific gene-targeting needs to be performed. This is not a suitable method for generating a genome-wide homozygous mutation library in ES cells.

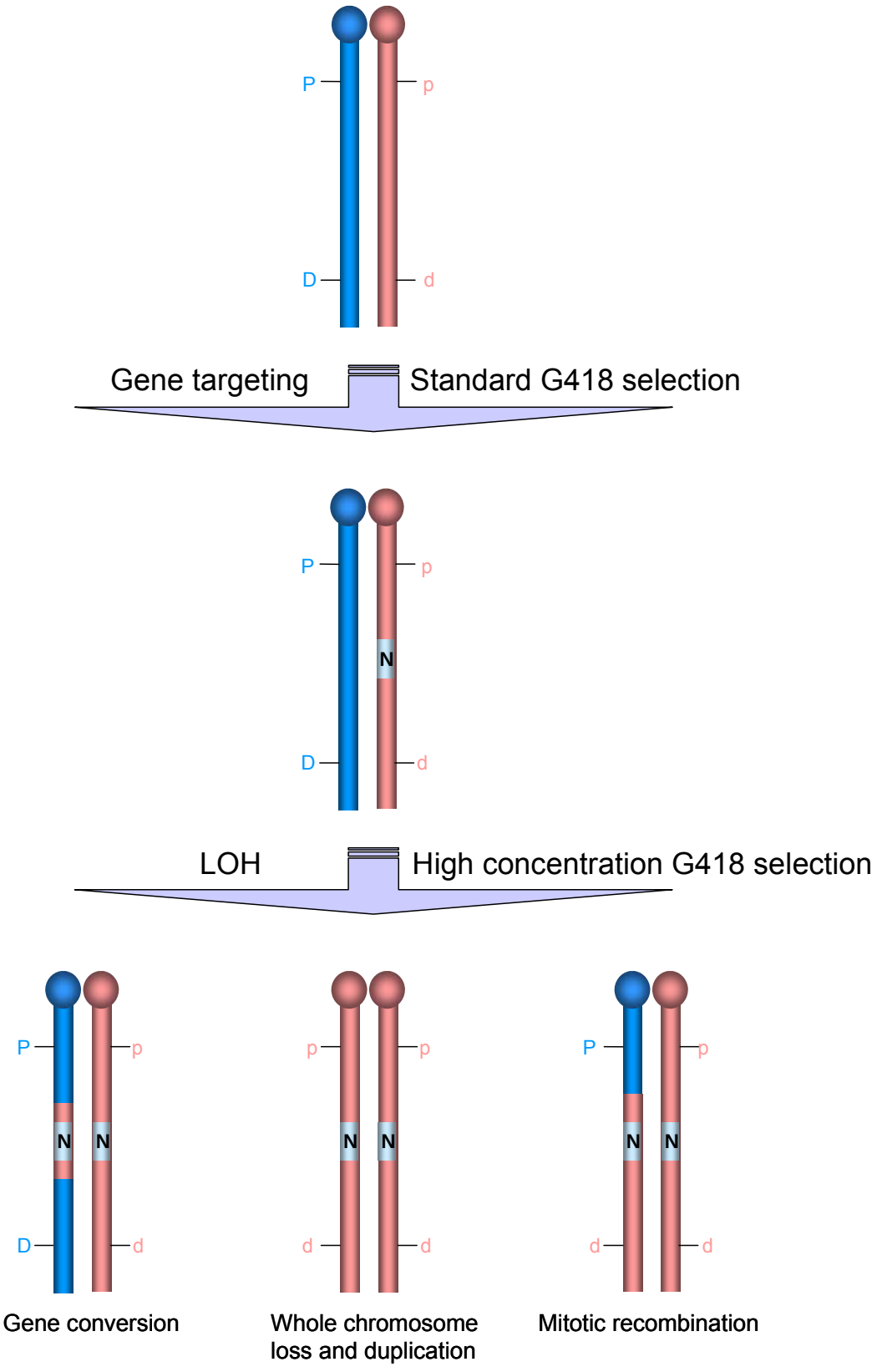


Figure 1-4 **High concentration G418 selection facilitates recovery of homozygous mutants**

Figure 1-4 High concentration G418 selection facilitates recovery of homozygous mutants.

A heterozygous targeting event can be recovered by selection in a standard concentration of G418 for the presence of the *neomycin* cassette (N). A pair of chromosomes of R1 hybrid ES cell are illustrated in blue and red to indicate the origins of the different 129 substrains. The following schematic illustration assumes the targeting event was on the red chromosome. Homozygous mutations can result from gene conversion, whole chromosome loss plus reduplication and mitotic recombination. By analysing SSLP DNA markers, which are located proximal (P/p) or distal (D/d) to the *Neo* insertion, one can distinguish which LOH mechanism was responsible for these homozygous mutants. If LOH was caused by gene conversion, SSLP analysis would show the pattern of P/p and D/d; if LOH was caused by whole chromosome loss and duplication, SSLP analysis would show a pattern of p/p and d/d; if LOH was caused by mitotic recombination, SSLP analysis would show P/p and d/d. Because distal linked markers were homozygous in the four analysed gene-targeting events, whole chromosome loss/duplication and mitotic recombination mechanisms resulted in these homozygous mutants (Lefebvre *et al.* 2001).

1.4.3 Induced mitotic recombination

Mitotic recombination takes place when a crossover occurs between the two homologous non-sister chromatids, which can be segregated in two ways (Figure 1-5): (i). Recombinant chromatids segregate to opposite poles of the cell, so that they separate into two daughter cells (X-segregation); (ii). Recombinant chromatids segregate to the same pole of the cell and thus are retained in the same daughter cell (Z-segregation).

In *Drosophila*, mitotic recombination induced by the Flpe/*FRT* system was developed as a genetic tool to generate and study homozygous mutant cells of the *Drosophila* EGF receptor homologue gene (*Egfr*) in mosaic animals (Xu *et al.* 1993). In this research, recombination occurred between *FRT* sites on chromosomes after a heat shock signal induced Flpe recombinase transcription and expression. The Flpe*/FRT* induced a mitotic recombination system generated in homozygous *Egfr* mutant cells in mosaic flies. This research showed that it was feasible to identify homozygous mutations induced by mitotic recombination in mosaic animals, therefore providing a system to perform genetic screens for mutations affecting many biological processes.

A similar system was developed in mouse and mouse ES cells using Cre recombinase-induced mitotic recombination (Liu *et al.* 2002; Wong *et al.* 2000). In Liu's experiments, two complementary cassettes (Δ3' and Δ5') were constructed, each containing non-functional halves of the human gene *HPRT1* (hypoxanthine guanine phosphoribosyl transferase 1). *LoxP* sites were inserted downstream of the 5' *HPRT* half and upstream of the 3' *HPRT* half. Only when recombination occurred between *loxP* sites, an intact *HPRT* gene was generated and the cells can be resistant to HAT (hypoxanthine/aminopterin/thymine) media (Adams *et al.* 2004; Zheng, Mills *et al.* 1999). In Liu's experiments, recombination was mediated by transient expression of Cre by transforming a Cre expression vector into these ES cells. The frequency of induced mitotic recombination was measured in five loci (three in Ch7 and two in Ch11). Recombination frequencies ranged from 4.2×10^{-5} to 7.0×10^{-3}, which is high enough to make it feasible to generate homozygous mutants from a heterozygous mutated cell library. This study demonstrated that induced mitotic recombination using the Cre/*loxP* system is feasible and products can be directly selected in mouse ES cells.

Using induced mitotic recombination, biallelic inactivation of specific genes has been analysed in mice (Muzumdar *et al.* 2007; Wang, Warren *et al.* 2007). Muzumdar used a system called mosaic analysis with double markers (MADM) to visualize loss of

homozygosity events of *p27kip1*, a cyclindependent kinase inhibitor gene in individual cells. This system contained two chimeric non-functional fluorescent markers upstream of the gene *p27kip1.* The N- and C-terminals of two chimeric markers were separated by an intron containing a single *loxP* site. The gene *p27kip1* was heterozygously mutated. In the presence of Cre, mitotic recombination can be induced. Fluorescent genes can then be activated and produce either red or green fluorescent proteins depending on what kind of mitotic recombination occurred. As a result, cells with any of the wild type alleles, heterozygous alleles or homozygously mutated alleles at the *p27kip1* gene were able to be distinguished by the fluorescence (Muzumdar *et al.* 2007). Similarly, Wang. in our laboratory, examined how mosaic mice with *Trp53* mutations were generated in the presence of Cre or Flpe (Wang, Warren *et al.* 2007). *Lox* sites and *FRT* sites were targeted into the D11Mit71 locus, which was 60 Mb upstream of the $Trp53^{+/-}$ gene. With induced mitotic recombination in mice, homozygous mutations of the *Trp53* gene were generated and this can be observed with tumourigenesis.

Despite its wide applications, induced mitotic recombination still has some characteristics which limit its application to generate genome-wide mutations. First, the different recombination frequencies between loci and chromosomes (Liu *et al.* 2002) matter when creating a genome-wide mutation library. Loci with low mitotic recombination frequencies may not generate a homozygous mutation when recombination is induced and has generated homozygous mutants at high-frequency loci. Second, genetic imprinting specifying allelic expression *in vivo* and *in vitro* may have an adverse impact (Liu *et al.* 2002). A number of regions on different mouse chromosomes are known to be imprinted, and mono-parental inheritance of these regions can cause early embryonic lethality and post-natal developmental defects (Cattanach *et al.* 1994). Third, to induce mitotic recombination in the whole genome, each chromosome needs to be targeted separately to insert *loxP* sites. This requires a great deal of effort and resources.

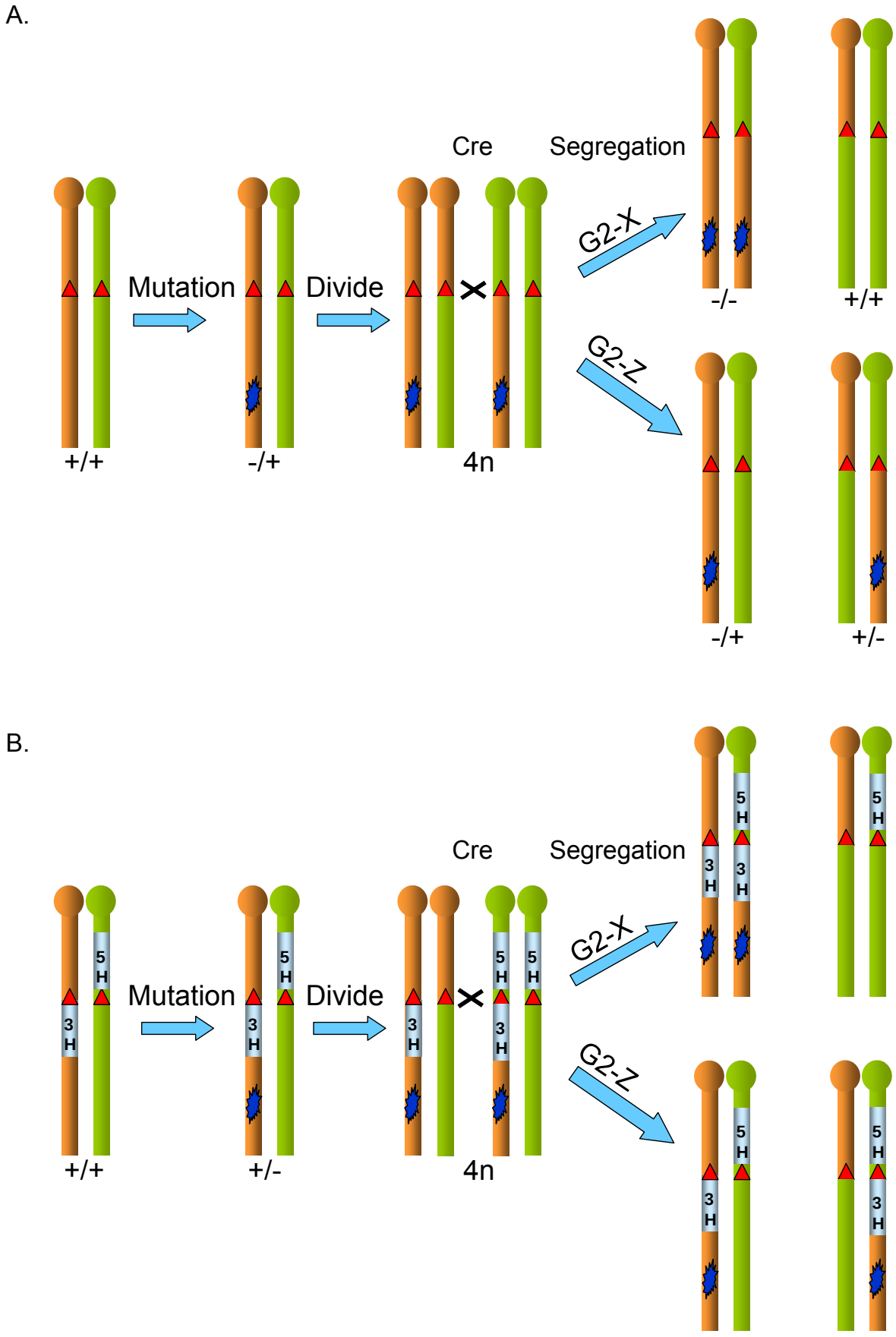


Figure 1-5 **Induced mitotic recombination**

Figure 1-5 Induced mitotic recombination

A. Cre/Flpe-mediated recombination between *loxP/FRT* sites (red triangle) facilitates recombination between non-sister homologous chromatids. X-segregation results in homozygous mutant cells and wild type cells. A blue star indicates a mutation. **B**. Mitotic recombination can be selected. Two halves of a human *HPRT1* gene (each with a *loxP/FRT* site) were targeted separately in trans in a pair of chromosomes. After mutagenesis, Cre/Flpe-mediated recombination generates non-sister homologous chromatid exchange. One chromosome with a mutation and an intact *HPRT1* gene will be generated. Then, daughter cells with this chromosome will be selectable by HAT media.

1.4.4 *Blm* gene deficiency elevates the rate of loss of heterozygosity

1.4.4.1 Bloom syndrome

Bloom syndrome (BS), first discovered by and named after Dr David Bloom in 1954 (Bloom 1954), is a rare inherited autosomal recessive disorder characterized by proportionate pre- and post-natal growth deficiency; sun-sensitivity, hypo- and hyper-pigmented skin; predisposition to malignancy and a high frequency of chromosomal breaks and rearrangements in the affected individuals. There are only a few hundred reported BS cases; many are of Ashkenazi Jewish origin. A predominant mutation containing a 6-bp deletion and 7-bp insertion at nucleotide position 2281 in the *Blm* cDNA was found in >95% of Ashkenazi Jewish persons (60 samples with BS) and only in <5% of unrelated non-Ashkenazic persons (91 samples with BS) (Ellis *et al.* 1998; German 1995; German 1997; Walpita *et al.* 1999). Seventy-one of the 168 registered patients were reported to have developed a variety of cancers (German 1997; German *et al.* 1990).

BS cells demonstrated features of chromosome instability and an increased rate of spontaneous sister chromatid exchanges (SCEs), 10- to 15-fold more frequently than SCEs seen in normal cells, observed through bromodeoxyuridine (BrdU) labelling (German 1969; Kuhn *et al.* 1986; McDaniel *et al.* 1992). In addition, BS cells have shown the loss or the suppression of homologous recombination events (Sonoda *et al.* 1999).

1.4.4.2 Cloning of the *BLM* gene in humans

The Bloom syndrome gene (BLM), assigned to human chromosome 15 (McDaniel *et al.* 1992), was first mapped by genetic linkage analysis to human chromosome 15q26.1 (Ellis *et al.* 1994; German *et al.* 1994). Ellis then mapped the BLM gene into a 250 kb interval by a strategy called somatic crossover point (SCP) mapping (Ellis *et al.* 1995). The SCP mapping approach was used because some cell lines derived from BS patients' lymphocytes exhibit low SCE rates, although all BS patient cells had high SCE rates. In some cell lines from low SCE lymphocytes, polymorphic loci distal to *BLM* on 15q were transheterozygous (both alleles were different mutated versions of the wild type allele), whereas polymorphic loci proximal to *BLM* remained heterozygous in all low SCE lymphocytes. This phenomenon indicated that recombination might happen in the *BLM* gene. By examining proximal and distal polymorphic markers for heterozygosity close to the *BLM* gene, the *BLM* gene was narrowed within a 250 kb genomic region. The *BLM* cDNA was isolated by hybridizing this 250 kb region with cDNA libraries. Upon analysing the *BLM*

cDNA sequence, significant homologies were identified with three known peptides in the RecQ subfamily of DExH box-containing helicases. The BLM protein's amino acid sequence was similar to proteins of human RECQL (44%), *Saccharomyces cerevisiae* SGS1 (43%), and *Escherichia coli* recQ (42%).

1.4.4.3 DNA helicase activity of BLM

The homology of human BLM suggested similar functions to recQ in *E.coli*, SGS1 in yeast and RECQL in humans (Ellis *et al.* 1995). The *E. coli recQ* gene product is a DNA-dependent ATPase and has a helicase activity (Umezu *et al.* 1990) with the ability to unwind the DNA helix (Umezu *et al.* 1993). Two yeast two-hybrid screens showed that the SGS1 protein interacts with Top3p topoisomerase (Gangloff *et al.* 1994) and with Top2p topoisomerase *in vivo* (Watt *et al.* 1995). RECQL is a human gene initially isolated from HeLa cells (Puranam *et al.* 1994). The product of RECQL has DNA-dependent ATPase, DNA helicase, and 3'–5' single-stranded DNA translocation activities (Puranam *et al.* 1994; Seki *et al.* 1994). It was then confirmed that the BLM protein was a RecQ 3'–5' DNA helicase (Karow *et al.* 1997). It unwinds the complementary strands of nucleic acid duplexes using the energy derived from ATP hydrolysis (Soultanas *et al.* 2001).

1.4.4.4 Proteins interacting with BLM

BLM protein exhibits a number of interactions within a large protein complex known as BASC (BRCA1-associated genome surveillance complex), which includes DNA damage repair proteins: MSH2, MSH6, MLH1, ATM, RAD50-MRE11-NBS1, RAD51, RAD51L3, p53 and BRCA1. Many of these proteins are tumour suppressors (Wang *et al.* 2000; Wu *et al.* 2001; Yang *et al.* 2004; Yang *et al.* 2002).

BLM directly interacts with MLH1, MSH2 and MSH6, which are essential players in DNA mismatch repair (Pedrazzi *et al.* 2001; Yang *et al.* 2004). Mutations in these genes are associated with non-polyposis colorectal cancer. Yeast two-hybrid experiments showed that hBLM protein interacts with hMLH, hMSH2 and hMSH6 *in vitro* and that co-immunoprecipitation experiments demonstrated hBLM, hMLH, hMSH2 and hMSH6 existed as a complex *in vivo*. By observing immunofluorescence, it was confirmed that hBLM, hMLH, hMSH2 and hMSH6 co-localize in the nucleus (Pedrazzi *et al.* 2001; Yang *et al.* 2004). However, BLM-defective cell lines were DNA mismatch repair proficient, demonstrated by measuring the repair efficiency on a substrate containing a single G-T mismatch and a strand discrimination signal (a nick) upstream (5') from the mispair (Pedrazzi *et al.* 2001).

In addition to its presence in BASC, p53 has been reported to bind physically to BLM helicase, identified by yeast and mammalian two-hybrid systems (Garkavtsev *et al.* 2001). Residues 237–272 aa (amino acid) of BLM and 285–340 aa of p53 were discovered as interaction sites. It was also shown that recombinant p53 binds to BLM and WRN helicases and attenuates their ability to unwind the synthetic holiday junction *in vitro*. Mutant p53 reduces its ability to bind to holiday junctions and inhibit DNA helicase activity (Yang *et al.* 2002). When treated by hydroxyurea (HU), a ribonucleatide reductase inhibiting chemical which can interfere with DNA replication fork progression (Koc *et al.* 2003), Bloom syndrome fibroblasts (BSF) have a higher percentage of apoptotic cells (~50%) than treated normal fibroblasts, p53-deficient fibroblasts and p53-deficient BSF, in which the apoptotic cells are ~10% (Davalos *et al.* 2003). This result indicates that apoptosis of BSF is p53-dependent.

Phosphorylated histone H2AX (γH2AX) foci recruit repair factors, such as NBS1, BRCA1, Rad50 and Rad51 for repair functions (Celeste *et al.* 2002; Paull *et al.* 2000). H2AX is one of the subfamily proteins of histone H2A and is a non-allelic isoform that replaces major histones within the nucleosome. The phosphorylated form of histone H2AX (γH2AX) is phosphorylated at serine 139 within the conserved COOH-terminal region (Rogakou *et al.* 1998). γH2AX foci are indicators of double-strand breaks (forming 1–3 minutes after double-strand breaks) as a result of stalled replication forks (Paull *et al.* 2000; Ward *et al.* 2001; Xu *et al.* 2003). The p53 function is downstream of BLM in recruitment of Nijmegen breakage syndrome protein 1 (NBS1) and breast cancer-associated protein 1 (BRCA1) as the presence of p53 delays assembling of NBS1 and BRCA1 after replication stress caused by HU (Davalos *et al.* 2003). Normal p53 response and an intact G1/S cell cycle checkpoint were observed when *Blm*-deficient cells were exposed to ionizing radiation, which also indicates that BLM does not have function in the p53 pathway (Ababou *et al.* 2000).

BLM protein interacts with topoisomerase III alpha physically and directly according to co-immunoprecipitation experiments (Dutertre *et al.* 2002; Wu *et al.* 2000). This observation is consistent with a previous study that showed that *S.cerevisiae* Sgs1p, the homologue of human BLM, interacts with Top3p physically (Gangloff *et al.* 1994), suggesting that BLM may recruit hTOPO III for helicase activity. There are two isoforms of DNA topoisomerase III, α and β, with a weak topoisomerase activity on negatively supercoiled (subtractive helical twisting) DNA (Goulaouic *et al.* 1999; Seki *et al.* 1998). The hTOPO IIIα binding site on BLM was identified between residues 1–212 and 1267–1417 (Wu *et al.* 2000). It was

demonstrated that BLM can stimulate the activity of hTOPO IIIα to unwind negatively supercoiled Phi-X174 DNA, while an hTOPO IIIα binding-defective mutant BLM protein (which still retains helicase activity) does not stimulate hTOPO IIIα activity (Wu *et al.* 2000; Wu *et al.* 2002).

1.4.4.5 Mouse models of Bloom syndrome

RecQ family DNA helicases are defined as proteins sharing a homologous region with *Escherichia coli* RecQ and are regarded as enzymes involved in recombination (Nakayama 2002). Among the five proteins of the human RecQ family, defects in three of them – WRN, RECQ4 and BLM – give rise to autosomal recessive debilitating disorders (Werner syndrome, Rothmund–Thomson syndrome and Bloom syndrome respectively), which are characterized by increased genome instability and a predisposition to develop cancer (Kitao *et al.* 1999; Lindor *et al.* 2000; Yu *et al.* 1996). The mouse BLM protein (1,416 residues) is located on chromosome 7 with 22 exons spanning 80 kb (NCBI m37, Figure 1-6). The mouse BLM and the human BLM proteins share 75% identity of their amino acid sequences (Figure 1-7, Figure 1-8, Figure 1-9). The first exon and the first four nucleotides in exon 2 are non-coding.

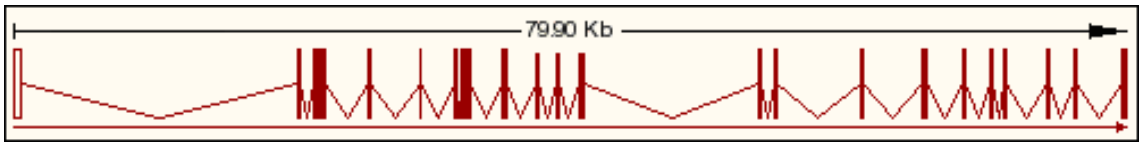


Figure 1-6 Mouse Blm gene transcript

The *Blm* transcript is drawn in scale, which starts from right to left. Twenty-two exons (horizontal bars) are connected by thin lines.

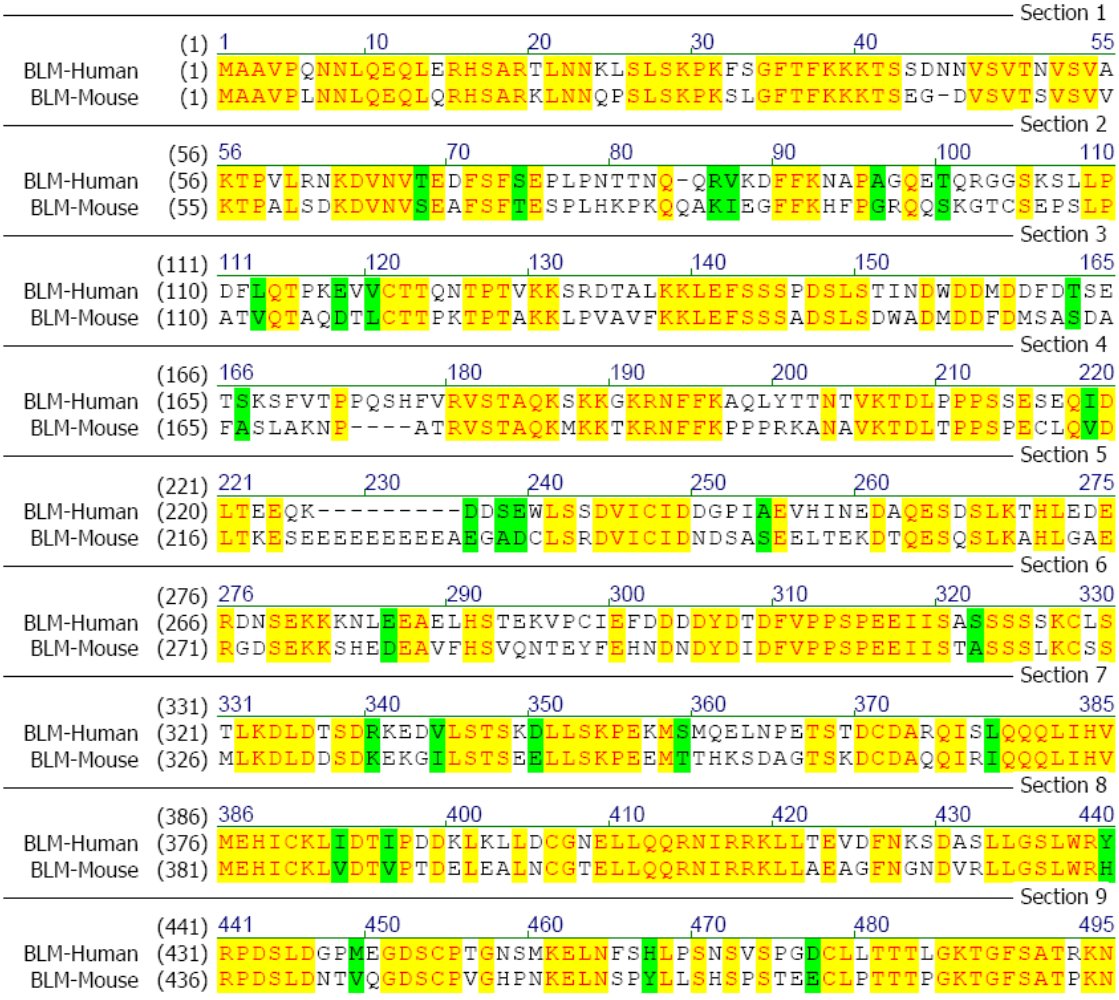


Figure 1-7 The alignment of the BLM protein sequence in humans and mice – 1/3

Human and mouse BLM protein sequences were retrieved from the Ensembl genome viewer and aligned in AlignX software (Invitrogen). The identity position between them is 75%. Red letters with a yellow background indicate identical residues. Black letters with a green background indicate similar residues. Black letters with a white background indicate non-similar residues. This figure shows the first third of the whole alignment.

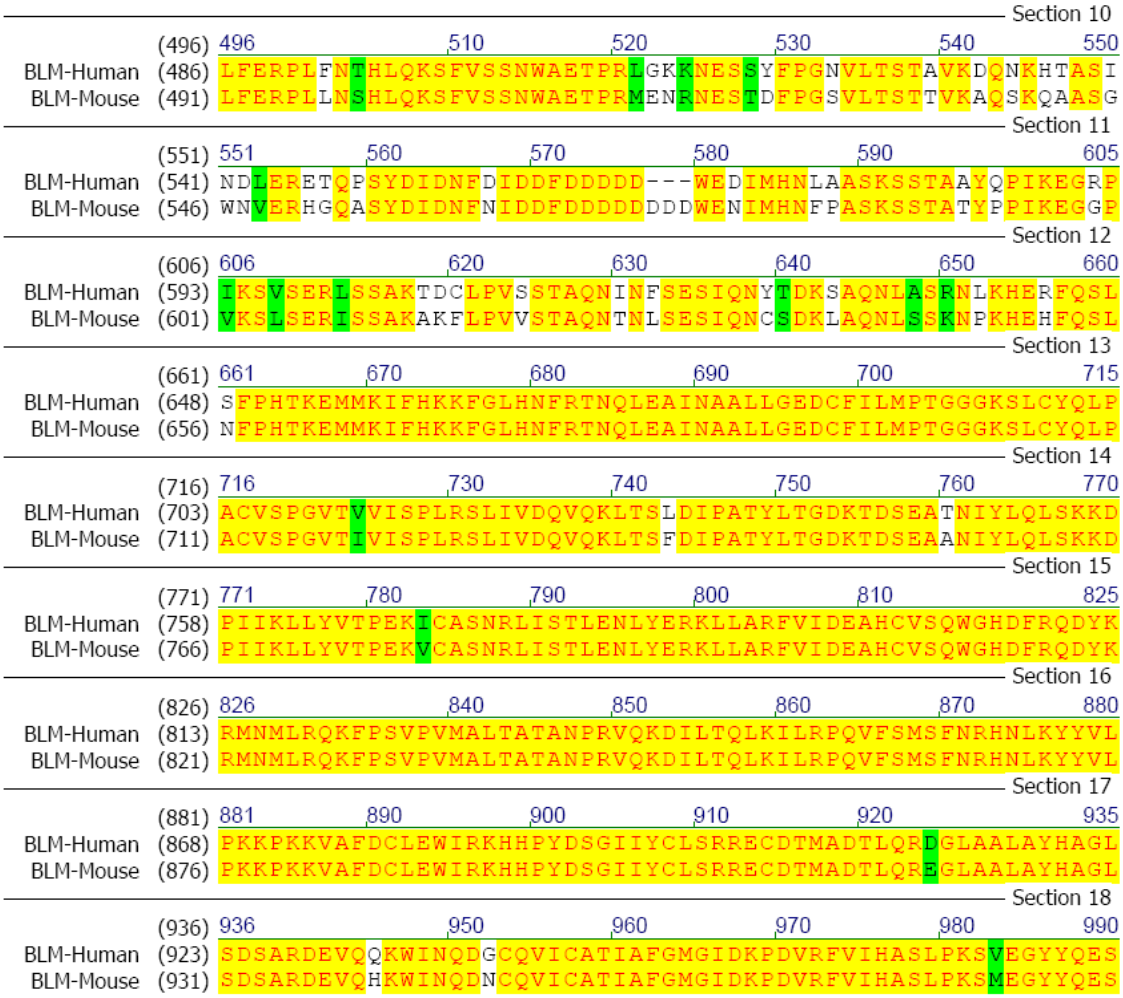


Figure 1-8 The alignment of the BLM protein sequence in humans and mice – 2/3

Human and mouse BLM protein sequences were retrieved from the Ensembl genome viewer and aligned. The identity position between them is 75%. Red letters with a yellow background indicate identical residues. Black letters with a green background indicate similar residues. Black letters with a white background indicate non-similar residues. This figure shows the second third of the whole alignment.

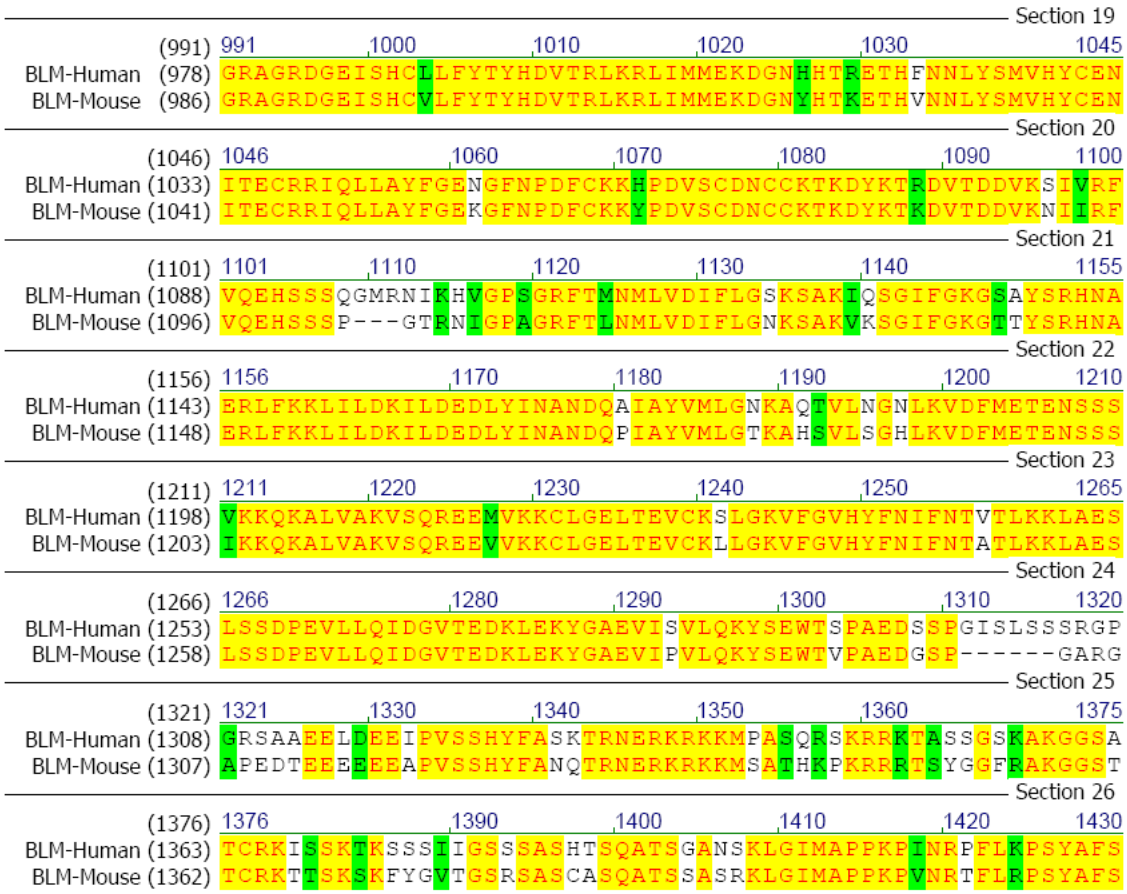


Figure 1-9 The alignment of the BLM protein sequence in humans and mice – 3/3

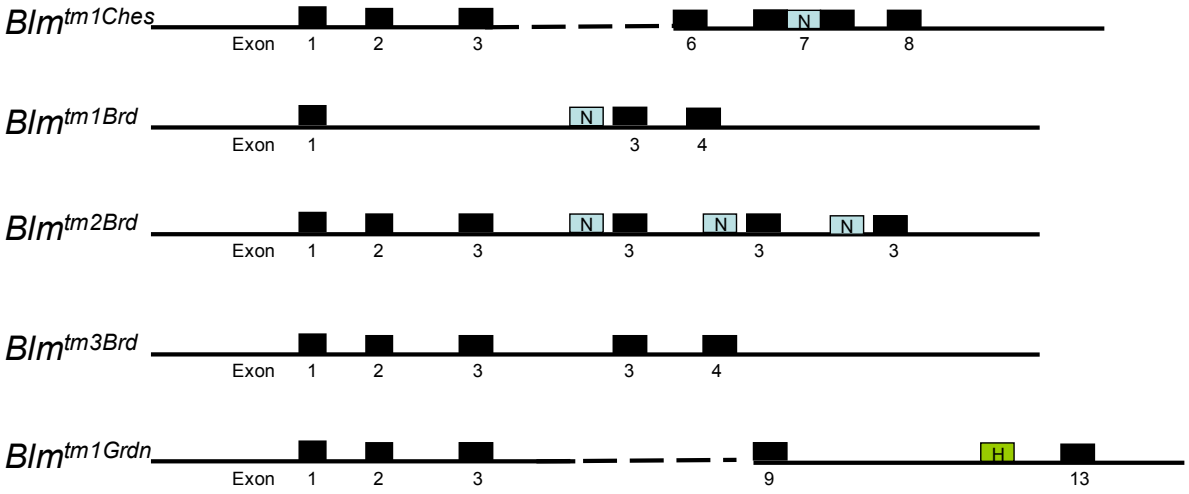
Human and mouse BLM protein sequences were retrieved from the Ensembl genome viewer and aligned. The identity position between them is 75%. Red letters with a yellow background indicate identical residues. Black letters with a green background indicate similar residues. Black letters with a white background indicate non-similar residues. This figure shows the final third of the whole alignment.

Eight targeted alleles of the *Blm* gene have been reported in mouse and/or mouse ES cells: $Blm^{tm1/Ches}$, $Blm^{tm3/Ches}$, $Blm^{tm4/Ches}$, $Blm^{tm1/Grdn}$, $Blm^{tm1/Brd}$, $Blm^{tm2/Brd}$, $Blm^{tm3/Brd}$ and $Blm^{tm1/Khor}$ (Chester *et al.* 1998; Goss *et al.* 2002; Luo *et al.* 2000; McDaniel *et al.* 2003; Yusa, Horie *et al.* 2004). Four of these ($Blm^{tm1/Ches}$, $Blm^{tm3/Ches}$, $Blm^{tm1/Grdn}$ and $Blm^{tm1/Brd}$) were generated by replacement targeting strategies, which produced null alleles by deleting exons (Figure 1-10). Mice with a homozygous targeted *Blm* gene mutation show delayed embryo development and die by embryonic day 13.5 (Chester *et al.* 1998). A high number of sister chromatid exchanges were observed in cultured $Blm^{-/-}$ fibroblasts (Chester *et al.* 1998). Moreover, it was found that the heterozygous mutated *Blm* gene in mice accelerated tumour formation (Goss *et al.* 2002).

$Blm^{tm2/Brd}$ and $Blm^{tm3/Brd}$, produced by insertional targeting resulted in the duplication of exon 3 of the *Blm* gene. They appear to be null alleles according to Western blot analysis using a polyclonal antibody specific to N-terminal 430 amino acids of the human BLM protein. The mice heterozygous for $Blm^{tm2/Brd}Blm^{tm3/Brd}$ are viable and fertile. However, the cross ($Blm^{tm2/Brd,tm3/Brd}$ × $Blm^{tm2/Brd,tm3/Brd}$) only generated two out of the three expected genotypes: $Blm^{tm2/Brd,tm3/Brd}$ and $Blm^{tm3/Brd,tm3/Brd}$, suggesting that $Blm^{tm2/Brd}$ is homozygously lethal and $Blm^{tm3/Brd}$ is viable. In addition, $Blm^{tm2/Brd,tm3/Brd}$ mice demonstrated genome instability and a cancer-prone phenotype, and were therefore regarded as a better model for human Bloom syndrome (Luo *et al.* 2000). Later, it was demonstrated that the homozygous embryonic lethal allele of $Blm^{tm1/Ches}$ can be rescued by the allele of $Blm^{tm3/Brd}$ (McDaniel *et al.* 2003).

As genome instability of *Blm*-deficient cells can potentially lead to the accumulation of mutations in ES cell culture, a system with tetracycline-controlled trans-elements was developed to disrupt the BLM protein temporally. This system has been used to identify genes in genetic screens (Hayakawa *et al.* 2006; Yusa, Horie *et al.* 2004).

A



B

Reference of Blm mutant mice	Major phenotypes
Chester, 1998	Early embryonic lethality; high SCE
Luo, 2000	Viability & lethality; high SCE; tumour; high non-sister exchange rate
Goss, 2002	Heterozygous mice develop tumours

Figure 1-10 Mutant *Blm* alleles and mouse phenotypes

A. Schematic view of the mutant *Blm* gene structure in several *Blm* knockout mice. B*lm*$^{tm1/Ches}$ replaces a 180 bp region in exon 7 with the neomycin cassette (N). *Blm*$^{tm1/Brd}$ deleted exon 2. *Blm*$^{tm2/Brd}$ have copies of exon 3, resulting in premature termination codons in all three possible open reading frames. The in-frame transcript of *Blm*^{tm2Brd} was predicted to produce a truncated peptide of 296 amino acids. *Blm*$^{tm3/Brd}$ has one extra copy of exon 3, resulting in a frame shift and truncated peptide, the same as *Blm*^{tm2Brd}. *Blm*$^{tm1/Grdn}$ replaced exons 10–12 with an *hHprt* cassette (H). **B**. Summary of phenotypes in knockout mice and/or cells generated in the mentioned literature.

1.4.4.6 Increased rate of loss of heterozygosity in *Blm*-deficient mouse ES cells and its applications

Analysis of metaphase spreads from almost all of the BS patients' cells reveals a high frequency of sister chromatid exchange compared with normal cells, which consequently causes increased somatic recombination, illustrated in Figure 1-11 and Figure 1-12 (Weksberg *et al.* 1988). BLM helicase-deficient mouse ES cells ($Blm^{tm1Brd/tm3Brd}$) were reported to have a 40–50 times higher rate of mitotic recombination, resulting in a higher loss of heterozygosity (LOH) rate: 4.2×10^{-4} events per locus per cell per generation, calculated by Luria–Delbruck fluctuation analysis (Luo *et al.* 2000). In principle, this rate makes it possible to segregate homozygous mutant cells from a heterozygous mutant in *Blm*-deficient ES cells within 12 cell doublings: $\text{Log}_2\ (1 / 4.2 / 10^{-4}) = 11.2 \approx 12$.

In 2004, two groups published phenotype-driven, genome-wide recessive genetic screens in *Blm*-deficient ES cells (Guo *et al.* 2004; Yusa, Horie *et al.* 2004). Although they screened for genes in unrelated biological pathways and used different mutagenesis systems (ENU by Yusa *et al*. and a gene-trap retrovirus by Guo *et al*.), both proved the productivity and reviability of the *Blm*-deficient ES cell system. Guo identified a known DNA mismatch repair (MMR) gene, *Msh6*. In addition, she also discovered that a novel gene, DNA methyltransferase (cytosine-5) 1, was involved in the DNA mismatch repair system (Guo *et al.* 2004). Yusa recovered 10 homozygous point mutations in genes involved in glycosyl-phosphatidyl-inositol (GPI) anchor biosynthesis.

However, ENU has a drawback as a mutagen compared to gene trapping: it is time-consuming to identify the location of the mutation, as ENU only produces single nucleotide changes. On the other hand, gene trapping is limited to generate high coverage mutation libraries for the screen. Due to integration site preferences of viruses and their selection systems (Stanford *et al.* 2001), gene traps repeatedly hit some genes. An example is that the gene *Msh6* was homozygously mutated seven times in Guo's work (Guo *et al.* 2004). Also, Guo *et al.* were unable to identify other MMR genes such as *Msh2* and *Mlh1*, deficiencies which have the same phenotype in ES cells as the screen criteria. An advantage of the gene-trap approach is that the embedded Cre/*loxP* system means that the phenotype of mutants can be easily reverted in gene-trap virus mutants, thus confirming the function of mutations.

Using the same insertional mutation *Blm*-deficient ES cell library generated by Guo *et al.*,

Wang *et al.* performed a screen for host factors involved in the Moloney murine leukaemia virus (MMuLV) retroviral infection and they identified the cell surface receptor (*mCat-1*) that was required for MMuLV infection of ES cells (Wang and Bradley 2007). Wang developed a novel negative selection strategy, in which infected cells carry a truncated Herpes Simplex Virus type 1 thymidine kinase produced by the virus. These cells were killed by FIAU (1-(2'-deoxy-2'-fluoro-β-D-arabinofuranosyl)-5-iodouracil), so that uninfected cells were isolated. Due to the intended superinfection, none of the uninfected cells result from low infection efficiency but from mutation of essential genes involved in the infection pathway.

According to the above findings, *Blm*-deficient ES cells have been proven to be an appropriate platform for recessive screens. Mutants of any genes which are expressed in ES cells can be isolated through appropriate screening strategies.

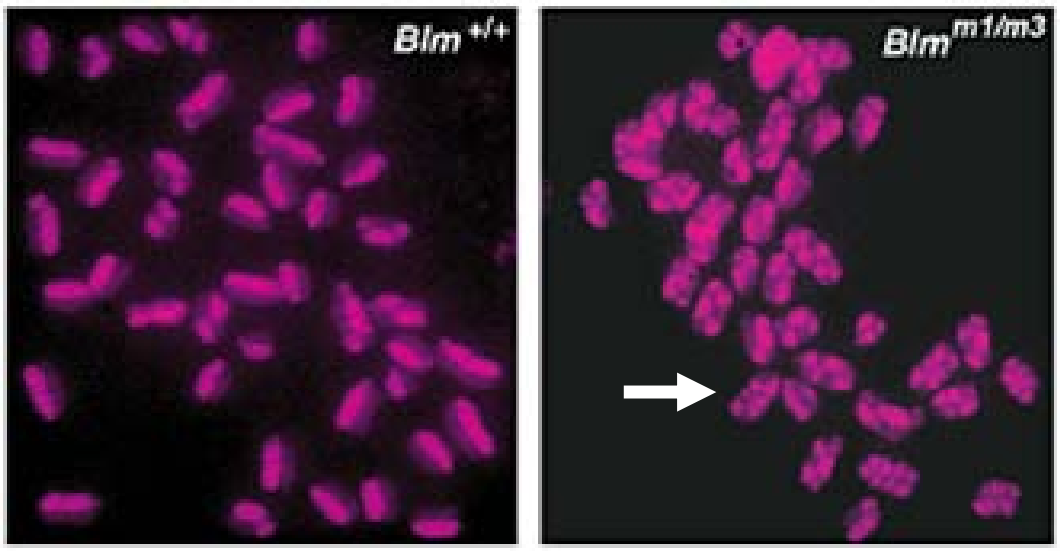


Figure 1-11 ***Blm*** **deficiency affects rates of sister chromatid exchange in ES cells**

Elevated sister chromatid exchange (SCE) in *Blm*$^{m1/m3}$ (*Blm*$^{tm1Brd/tm3Brd}$) cells illustrated by BrdU staining. The sister chromatids were differentially labelled by BrdU, and appear as light and dark intensities, respectively. There are many more SCEs per metaphase spread in the mutant background *Blm*$^{m1/m3}$. The white arrow indicates an SCE. This figure is copied from Luo *et al*. 2000.

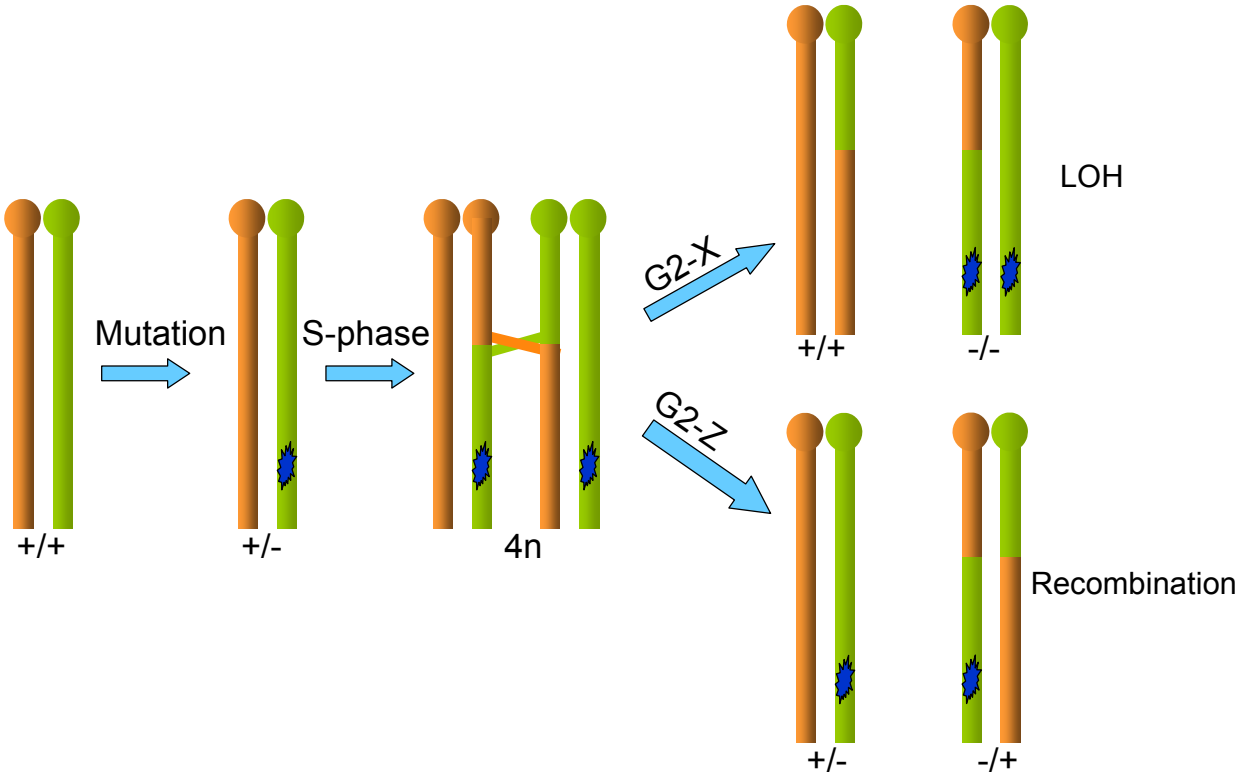


Figure 1-12 Mitotic recombination

A mutation (blue star) occurring before the S phase of the cell cycle can be segregated into either the same cell (LOH) through X segregation in the G2 phase or the two daughter cells through Z segregation in the G2 phase. Non-sister chromatid exchange is shown at the 4n stage after the S phase. The deficiency of BLM protein results in the increased rates of loss of heterozygosity (LOH).

1.5 Experimental mutagens used in mouse genetics

DNA sequences not only change spontaneously at the low frequency (~10^{-8} per nucleotide per generation) (Crow 1995), but can also be mutated by exogenous agents. These include various chemicals, viruses and engineered DNA fragments (gene-trap vectors, transposons and homologous DNA fragments) as well as irradiation. Permanent alterations, for example single-nucleotide substitutions, insertions, duplications, deletions or translocations can be created on the endogenous DNA strands. All mutagens have unique mechanisms to alter DNA, which can be both strengths and weaknesses, as discussed separately later. In this section, some frequently used mutagens and their applications in genetic screens are discussed and summarized in Table 1-1.

1.5.1 N-ethyl-N-nitrosourea

N-ethyl-N-nitrosourea (ENU), in addition to other chemicals, has been used as a mutagen for many years. By the late 1970s, the mouse was confirmed as the mammalian species of choice for genetic studies and was routinely used as a model for human disease. Responding to the need for identifying physiological gene function, a large collection of mutant mice was established by the international research community (Green 1966). Ideally, the best chemical mutagen should comply with some prerequisites: ease of purchase and preparation, easy to handle, not too toxic and, if possible, active on the pre-meiotic germ cells to produce inheritable mutants.

Among over 50 compounds tested as mutagens in mice, ENU is the most potent mutagen so far with an approximate mutation rate of 1 mutant in 1,000 gametes (Chen, Y. *et al.* 2000; Guenet 2004; Russell *et al.* 1979). With this mutation rate, ENU is also five times more efficient than a single exposure of X-ray radiation (6 Gray). The mutagenecity of ENU is due to its capacity to transfer an ethyl group to oxygen or nitrogen radicals in the DNA nucleobases, which causes nucleotide mismatches and ultimately results in base pair substitutions or sometimes base pair losses if not repaired. These single nucleotide mutations include A/T to T/A, A/T to G/C, G/C to A/T, G/C to C/G, A/t to C/G and G/C to T/A (Justice *et al.* 1999). In an ENU genetic screen, usually male mice (G0) are injected with ENU to induce mutations in gametes. Mating the ENU-treated male mice with untreated wild type female mice produces the first generation (G1) offspring. G1 mice may carry a unique mutation and are thus ready for a dominant phenotype screen, for example the discovery process of the *Clock* gene (Vitaterna *et al.* 1994). A mutant G1 mouse exhibited

Table 1-1 A comparison of mutagens to germ line mutations

Mutagen	Mutagenesis frequency	Type of mutation	Primary advantages	Primary disadvantages
None	5×10^{-6} per locus per generation	Spontaneous. Point mutations, small deletions, chromosomal rearrangements and insertions of endogenous retrovirus-like sequences.	Visible phenotypes; only requirement is observant mouse handlers.	Only visible phenotypes detected at very low frequency.
Irradiation	$1–5\times10^{-4}$ per locus per generation	Chromosomal rearrangements: deletions, duplications, inversions and translocations.	Rearrangements act as a molecular landmark for cloning. Easy to saturate genome.	Multiple genes affected, hard to dissect individual gene function.
ENU	1.5×10^{-3} per locus per generation	Primarily generates point mutations, occasionally very small deletions (20–50 bp).	Single-gene mutations, amenable to high throughput.	No molecular landmarks for cloning.
Transgenic/retroviral insertion	5–10% of transgenic animals	Disrupts endogenous gene expression or coding sequence. Sometimes causes chromosomal rearrangements.	Provides a molecular landmark for cloning.	Labour-intensive, not applicable to high-throughput approaches, often associated with complex rearrangements.
Trapping	Almost 100% of transgenic animals *	Disrupts endogenous coding sequence.	Forward-genetic strategy, easy to clone mutated gene, reports endogenous gene-expression pattern.	Unpredictable phenotypes.
Gene targeting	Almost 100% of transgenic animals *	Generates insertions or deletions as designed.	Can design type of mutation as required.	Requires knowledge of gene and its structure, labour-intensive, unpredictable phenotypes.
RNAi (post-transcription modifier)	Almost 100% of delivered cells.	Silencing gene by disrupting its transcript.	Carry molecular landmark if designed. Suitable for high-throughput approaches.	Requires knowledge of gene structure and sequence. Unpredictable phenotypes. Knock-down effects, not 100% of target effects.

* Requires pre-screening of embryonic stem cells *in vitro*. Modified from Stanford *et al*. (Stanford *et al.* 2001).

abolished persistence of rhythmicity. Alternatively, the G1 mice can be backcrossed with wild type mice to produce a mouse line with the same mutation. Intercrossing between this mouse line will generate homozygous mutant mice for recessive screens (Cordes 2005). One example of this type of breeding scheme is a mouse screen for embryonic lethal mutations conducted by K. Anderson (Kasarskis *et al.* 1998)

Only 2% of F1 offspring mice exhibit dominant phenotypes (Hrabe de Angelis *et al.* 2000; Nolan *et al.* 2000). To dissect recessive mutations, complicated and expensive breeding regimes are needed, which have limited a wider use of ENU in mouse recessive screens. To overcome this limitation, mice with large regional deletions or inversions were generated by either Cre-*loxP*-mediated recombination or irradiation (Ohtoshi *et al.* 2006; Su *et al.* 2000; Zheng, Mills *et al.* 1999; Zheng *et al.* 2000). These deletion and inversion mice provide genetic tools to accelerate the efficiency of ENU mutagenesis screens. When an ENU mutation is present in a deletion region, a cross between G1 mutant mice with deletion carrier mice may generate recessive phenotypes (Bergstrom *et al.* 1998).

Large deletions may result in less fitness, infertility and even lethality. Balancer chromosomes, originally developed in *Drosophila* to maintain recessive lethal mutations, were used to circumvent this problem (Zheng, Sage *et al.* 1999). A balancer chromosome carries a large inverted fragment on a chromosome. Meiotic crossover between an inversion chromosome and a normal chromosome are efficiently suppressed. If this occurs, cells with this type of meiotic crossovers are inviable. Balancer chromosomes can be engineered to carry a visible dominant coat marker and a recessive lethal mutation on the chromosomal inversion region. With this arrangement, mice homozygous for the balancer chromosome are not viable. Thus, mutations can be maintained in heterozygous mice and tracked by the visible dominant marker.

A number of genome-wide (Favor *et al.* 1991; Kasarskis *et al.* 1998; Shedlovsky *et al.* 1993; Vitaterna *et al.* 1994) and regional (Justice *et al.* 1986; Rinchik *et al.* 1999; Shedlovsky *et al.* 1988) dominant/recessive genetic screens were carried out by ENU mutagenesis reviewed by Justice *et al*. (Cordes 2005; Justice *et al.* 1999; van der Weyden *et al.* 2002). These mutant mice have provided invaluable insights in defining mouse gene function of molecular, genetic, immunological, reproductive, physiological and pathological similarities to humans. However, some drawbacks limit the utility of ENU in genome-wide screens. ENU is heavily biased towards A/T base pairs (87%) (Justice *et al.* 1999), which means that the ability to mutate genes is dependent on the G/C content of their functional sequences. Moreover, due to the lack of a molecular tag, it is a time-consuming and labour-intensive process to

identify mutations produced by ENU. To locate the mutation of interest within several centimorgens (cM), hundreds of meiotic events are needed. Therefore, better mutagenesis systems are needed to annotate mouse genes through their functions.

1.5.2 Ionizing radiation

1.5.2.1 Effects on DNA

Ionizing radiation (IR) involves energetic particles or electromagnetic waves that have the potential to ionize an atom or molecule through atomic interactions. Examples of ionizing radiation are energetic beta particles, neutrons, alpha particles and energetic photons. Having the shortest wavelengths in the electromagnetic spectrum, gamma (γ) rays and X-rays carry the highest energy and can ionize almost any molecule or atom. Ultraviolet (UV) and visible light are ionizing to very few molecules; microwaves and radio waves are non-ionizing.

In the 1920s, investigation of the effects of ionizing radiation on animals was initiated (Silver 1995). Mutant offsprings were generated from X-ray-irradiated parent mice, without elucidating the connection between irradiation and induced mutation (Little 1924). Then, Muller made this connection and explained the induction of heritable mutations by X-rays (Muller 1927).

Ionizing radiation causes both direct and indirect effects on DNA. Direct effects lead to ionized bases or sugars after the direct absorption of the radiation energy by DNA. Indirect effects are created when DNA reacts with ionized surrounding molecules. For X-rays and Cobalt-60 (^{60}Co) γ irradiation, around 65% of DNA damage is caused by the indirect effects of radicals and roughly 35% by direct ionization (Friedberg *et al.* 2005). In cells, numerous molecules, inorganic ions and water surround DNA. Although potential sources of reactive species that arise as excited molecules, ion radicals or free radicals are eventually converted into chemically stable products by subsequent decay reactions, these reactions create a wide spectrum of products in ionizing radiated DNA. In many reports, ionized molecules created by ionizing radiation were found to damage all cellular components randomly and cause a huge variety of DNA lesions, such as DNA-protein cross-links, base damage, single-strand breaks and double-strand breaks (Frankenberg-Schwager 1990; Goodhead 1989; Hutchinson 1985; Lett 1990; Ward 1988).

Due to the predominance of water in life systems, materials formed after the radiolysis of water are the major sources of indirect damage to DNA (Riley 1994; Sonntag 1987; Ward

1988). Two principle reactions involving water contribute the most to the damage.

i). When the photon energy is high enough, a water molecule is ionized:

$H_2O \rightarrow H_2O^{\cdot+} + e^-$

The $H_2O^{\cdot+}$ immediately loses a proton to release a hydroxyl radical:

$H_2O^{\cdot+} \rightarrow H^+ + {}^{\cdot}OH$

The hydrated electron e^- reacts rapidly with oxygen to produce a superoxide radical:

$e^- + O_2 \rightarrow {}^{\cdot}O_2^-$

Then the superoxide radical can react with protons to form hydrogen peroxide:

$2.O_2^- + 2H^+ \rightarrow O_2 + H_2O_2$

ii). The second primary reaction is excitation of a water molecule followed by homolysis into an H. and a hydroxyl radical:

$H_2O \rightarrow H_2O^* \rightarrow H^{\cdot} + {}^{\cdot}OH$

About 80% of the energy that ionizing radiation delivers to cells results in the first series of reactions and ~20% of the energy causes the second type of reaction (Friedberg *et al.* 2005).

Ionizing radiation damages DNA bases as well as producing strand breaks by attacking DNA sugars. For all of the bases (thymine, cytosine, adenine and guanine), the C5=C6 double bond is the major site of radical attacks by hydroxyl radicals (Bjelland *et al.* 2003; Sonntag 1987; Ward 1988). After γ irradiation, the predominant species are thymidine–tyrosine cross-links, which are found in the presence of oxygen or after the treatment by hydrogen peroxide in the presence of iron or copper ions (Dizdaroglu 1992; Nackerdien *et al.* 1991).

Hydroxyl radicals cause more severe damage in the form of strand breaks, when they attack DNA sugars. 1 Gy dose of X-ray or γ-ray radiation can produce between 600 and 1000 single-strand breaks (SSB) and between 16 and 40 double-strand breaks (DSB) (Ward 1988). An SSB is generated when a hydrogen atom is released from deoxyribose, forming a deoxyribose radical. The mechanism by which this radical formation leads to strand breaks has been studied intensely and there are a number of theories to explain it (Breen *et al.* 1995). Double-strand breaks can be formed in two ways: i). When a single track of radiation creates a cluster of ionizations, two or more ${}^{\cdot}OH$ radicals can form DSB by attacking both strands of DNA simultaneously (Ward 1985; Ward 1990). ii).A DSB can also be produced by a single ${}^{\cdot}OH$ radical attacking one strand of DNA while the other strand

suffers direct damage within a distance of 10 bp (Michael *et al.* 2000). As DSBs can interfere with cell division by stopping DNA replication and as the DSB repair process requires the participation of more proteins, DSB is much harder to fix and it results in a loss of genetic materials. Therefore, double-strand breaks are more lethal to a cell than single-strand breaks (McKinnon *et al.* 2007).

1.5.2.2 Applications of ionizing radiation in mouse genetics

As for the chemical ENU, ionizing radiation was first used in large-scale mouse mutagenesis experiments in the Oak Ridge National Laboratory (USA) and the Medical Research Council Radiobiological Research Unit (UK) (Green 1966). Both programmes initially intended to investigate the effects of various forms of radiation on mice and, by extrapolation, human beings.

Ionizing radiation is simple to use and effective in generating random mutations (Lobrich *et al.* 1996), making it a good mutagen for genetic research. It has been used to generate mice with regional deletions. These mice have been shown to be useful to dissect single gene mutations in the deleted region, such as the head tilt (*het*) gene on chromosome 17 (Bergstrom *et al.* 1998; Chao *et al.* 2003; Goodwin *et al.* 2001; You *et al.* 1997).

Although ionizing radiation has been used for decades, the relationship between deletion length and irradiation dosage has not been identified. As DNA is wrapped around nucleosomes and organized in chromatin, irradiation-induced hydroxyl radical clusters can produce double-strand breaks at sites that are several kilobase pairs (kb) or even 700 kb apart (Lobrich *et al.* 1996). Kushi has confirmed that X-rays can be used to make large deletions (200–700 kb, kilobase pairs) around the *Hprt* (hypoxanthine-guanine phosphoribosyltransferase) locus on the X chromosome of the E14 mouse ES cell line (Kushi *et al.* 1998). At the same time, Thomas generated deletions <3 cM around the *Hprt* locus in F1 hybrid ES cells (Thomas *et al.* 1998). Using an artificial selection marker, Schimenti obtained deletions up to approximate 70 Mb induced by irradiation. He targeted the Herpes Simplex Virus *thymidine kinase* (*HSV-tk*) gene with a neomycin cassette at 3 loci of chromosome 5 of F1 hybrid mouse ES cells. The resulting irradiation cells with deletion around these targeted loci were isolated by negative selection for loss of the *HSV-tk* gene (Schimenti *et al.* 2000). Similar work has also been done on the distal part of chromosome 15, and various-sized deletions were isolated (Chick *et al.* 2005). These approaches all have limitations in that deletion sizes might be biased because there may be genes crucial for cell viability around these loci. Homozygous deletions covering any of

these essential genes will result in cell lethality and thus cannot be observed.

It is likely that ionizing radiation will affect several/many genes simultaneously, which can make identification of the functions of individual genes difficult. However, its powerful mutagenicity and the emerging technologies that allow examination of the whole genome at high resolution using comparative genome hybridization (CGH) and gene expression arrays make it easier to elucidate the connections between phenotype and genotype.

1.5.3 Gene targeting

Gene targeting is a technology which can be used to obtain a designed mutation directly and can be applied in reverse genetic studies in mice. Targeted mutations can be achieved by homologous recombination between endogenous genes and a targeting vector (Doetschman *et al.* 1987; Thomas *et al.* 1987). A simple gene-targeting vector is composed of a DNA sequence homologous to the targeted gene and a positive selection marker. Taking advantage of mouse ES cell technology and homologous recombination, targeting events in cultured ES cells can be easily selected in media containing a drug for selection and then identified by Southern blot or PCR. Once the correct targeting event has been identified, this genetically modified mouse ES cell clone can be injected into a host mouse blastocyst. The allele can be established in mice following germ line transmission from chimeras (Bradley *et al.* 1984) to F1 heterozygous mutant mice (Koller *et al.* 1989; Schwartzberg *et al.* 1989; Thompson *et al.* 1989; Zijlstra *et al.* 1989).

Gene-targeting technology provides a powerful means to generate transgenic mice harbouring precise mutations in the gene of interest. The function of the gene can be examined by analysis of the phenotype of gene-targeted mice. Nowadays, gene-targeting vectors can be engineered to target any genes, generating all possible classes of mutations such as loss of function, gain of function, point mutations and knock-in alleles (Billet *et al.* 2007; Skvorak *et al.* 2006; Zheng, Larkin *et al.* 1999). Combined with Cre-*loxP* technology (Zheng *et al.* 2000), an enzyme-mediated site-specific recombination system, the “expression” of a mutation can also be made controllable in a temporally or spatially restricted manner, the so-called conditional knockout (van der Weyden *et al.* 2005). The completion of the mouse genome sequence project made it an appropriate time to carry out gene knockouts genome-wide. The Mouse Genetics Programme at the Wellcome Trust Sanger Institute aims to understand the function of genes and their role in disease by generating large numbers of gene-targeted mutant mice and screening them for characteristic features of diseases. The outcome of this programme will make a significant

impact on the understanding of gene functions in the mouse genome and their homologues in the human genome.

Despite gene-targeting providing unprecedented insights into gene function, this technology is costly, time-consuming and labour-intensive because it can only be applied on a gene-by-gene basis. This approach requires the prior knowledge of genes to design a gene-targeting vector. Therefore, novel phenotypic information about a gene is often missed. Even some null mutations generated by the gene-targeting method do not resemble the types of molecular lesions found in disease. In addition, gene-targeting can only be conducted on an one-by-one basis, and the time required to generate homozygous mice is lengthy. So, despite the success of gene targeting, random mutagenesis screen systems have continued to be explored. Each functional genomics approach has its strengths and weaknesses. By taking advantage of the strength of each approach, the functions of the mammalian genome will be most efficiently understood (Stanford *et al.* 2001).

1.5.4 Insertional mutagenesis systems

1.5.4.1 Gene-trap mutagenesis

The introduction of exogenous retroviral DNA into the mouse germ line was first reported by Jaenisch (Jaenisch 1976). In his work, an insertional mutagenesis through the Moloney leukaemia virus was undertaken. The integration of a retrovirus may produce various mutations, in which the expression of a gene is increased by the viral enhancer element (Jaenisch *et al.* 1981; Lund *et al.* 2002; Mikkers *et al.* 2002). The availability of ES cell technology in the mid 1980s stimulated improved designs of insertional mutagens. The development of gene-trap technology has successfully circumvented the limitation of insertional mutagenesis using wild type retroviruses. Gene-trap mutagenesis produces loss-of-function mutations at random gene targets (Gossler *et al.* 1989). These events can be positively selected through the expression of the marker/reporter genes. The gene-trap vector serves as a molecular tag for cloning mutated candidate genes. Combined with the ES cell technology, gene traps offer a valuable tool for generating loss-of-function mutations on a large scale. Therefore, gene-trap mutagenesis can promote the efficiency of functional genomic studies in mouse ES cells and mice (Stanford *et al.* 2001).

Gene-trap vectors contain a promoterless selection/reporter gene and the expression of the reporter gene requires activity of the cis-elements of a target gene following integration. That means the selection/reporter genes can be activated only when the vectors integrate into the region near an endogenous gene so that the gene-trap vector can utilize its

transcriptional elements. However, the gene-trap events which occur in non-coding regions of the genome won't be selected out/observed due to the selection cassette being non-functional. The basic gene-trap vectors include enhancer-traps, promoter-traps and polyadenylation signal (PolyA) traps (Figure 1-13).

Enhancer-trap vectors contain a minimal promoter sequence (Figure 1-13). This kind of vector was first used to identify and characterize mammalian enhancer sequences from cells (Weber *et al.* 1984). The expression of the selection or reporter cassettes in the vectors can be activated when they integrate into the region next to a cis-acting endogenous enhancer element. This approach has not been widely used in the mice because loss-of-function mutations couldn't be efficiently generated using this approach. This is because the enhancer elements of a gene can exert their function at some distance from the structural gene, so that the insertion of the enhancer-trap vector does not normally disrupt the expression of the endogenous gene. However, the desire to overcome the low mutagenesis efficiency of enhancer-trap vectors led to the development of promoter-trap and other gene-trap vectors.

The key components of promoter-trap vectors are a promoter-less reporter gene upstream of which there is a "strong" splice acceptor (SA) sequence (Figure 1-13). Thus, the expression of the reporter can be driven by the endogenous promoter of a "trapped" gene so that a fused transcript containing a 5' portion of the trapped gene and the coding sequence of the reporter is generated. As a result, the transcription of the endogenous gene is destroyed, creating a loss-of-function mutation in this gene. Moreover, the expression of the mutated gene can be evaluated by detecting the expression of the reporter gene. Due to the nature of promoter-trap vector, this approach can only mutate the genes that are expressed in the cell lines of interest, such as ES cells (Gossler *et al.* 1989). However, promoter-trap vectors are still the most widely used trapping approach. A joint programme of several academic groups has formed the International Gene-Trap Consortium (IGTC, http://www.genetrap.org), in order to generate a public library of murine ES cell lines mutated by gene trapping.

There are, however, many genes that do not express or express at very low levels in undifferentiated ES cells. In order to mutate this category of genes, the promoter gene-trap approach is not suitable because the selection marker of the vector cannot be driven to express from the promoter of the trapped gene. One approach to overcome this deficit in promoter-trap vectors was the development of polyA traps (Zambrowicz *et al.* 1998). A polyA-trap vector utilizes a reporter gene lacking a polyadenylation signal, but possesses a

“strong” splice donor (SD). The reporter gene has its own promoter, so that the transcription of the reporter gene can be initiated independently of the expression status of the endogenous gene. However, the fusion transcript initiated from the promoter of the gene-trap vector is unstable unless the vector inserts into an endogenous gene upstream of a splice acceptor and a polyA signal (Figure 1-13). However, the polyA-trap approach also has some disadvantages. PolyA-trap vectors often trap pseudo splice acceptors and polyA signals in the mouse genome. These pseudo sites are the fossilized legacy from evolutionary events in the genome and are not associated with functional genes. The other problem is the complication due to alternative splicing at the 3’ end of the trapped gene. So the mutagenesis efficiency of this type of vectors is still controversial.

There are several methods that are used to introduce gene-trap vectors into cells including electroporation, retroviral infection and transposable elements. Retroviral based and transposon-based gene transfers are described separately. Electroporation, meaning providing an electric shock to cells which allows the linearized gene-trap vector into cells as “naked” DNA, is the simplest way to perform gene-trap mutagenesis. Although this is an easy approach to introduce the gene-trap vector into cells and the method can be applied on a large scale, there are some obvious disadvantages that limit the application of this method. Firstly, the integration of the gene-trap vector is frequently accompanied by DNA concatemerization, in which many copies of linearized DNA molecules form head to tail arrays and insert into one site of the host genome together. Secondly, the gene-trap vector can be truncated during electroporation. These points increase the complexity of the identification of the gene-trap mutations; however, these have not been a huge problem.

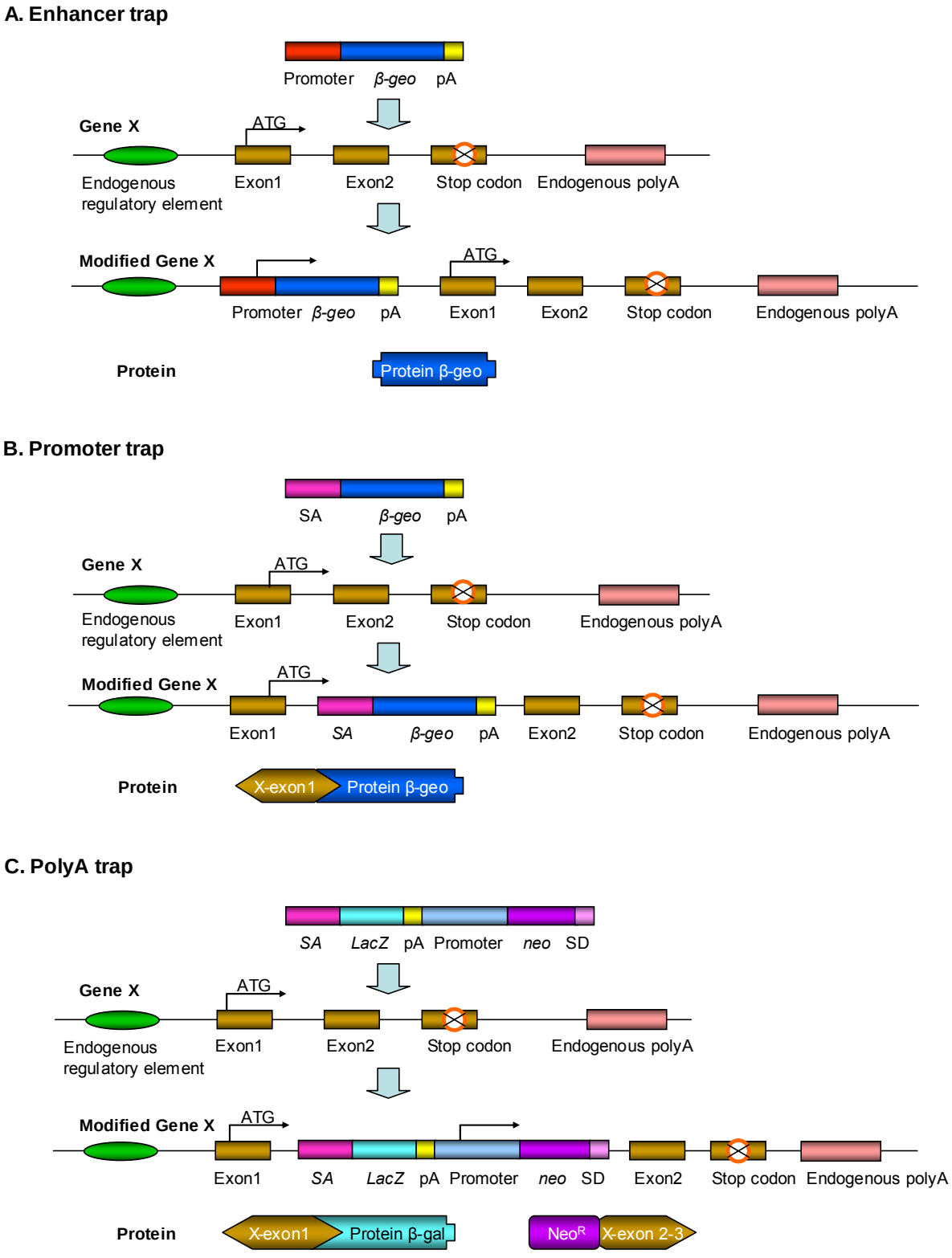


Figure 1-13 Basic gene-trap vectors

Figure 1-13 Basic gene-trap vectors

Enhancer-, promoter- and polyA-trap vectors, which all contain one or two reporter gene(s). Integration of the trap vectors into an endogenous gene “X” of an embryonic stem (ES) cell genome will enable G418 selection (through β-*geo* or *Neo*), whether the insertion has occurred intergenically or intragenically. (*Neo*, Neomycin resistance gene; beta-gal, beta-galactosidase; beta-geo, beta-galactosidase–*Neo* fusion; pA, polyadenylation.)

A. An enhancer-trap vector contains a weak position-dependent promoter immediately upstream of β-geo. Insertion of the enhancer-trap vector within an enhancer’s working region in gene X will lead to the transcription of the β-geo reporter when the enhancer of gene X is activated. As the enhancer elements of a gene are often distant from the gene coding sequences, the insertion of the enhancer-trap vector does not normally disrupt the expression of the host gene. Thus, enhancer-trap vectors usually generate hypomorphic rather than null mutations.

B. A promoter-trap vector contains a splice acceptor (SA) site immediately upstream of a promoterless β-geo gene. Integration in an intron leads to a fusion transcript generated from the upstream exon of gene X and *β-geo* upon transcriptional activation of gene X. Fusion proteins in the wrong frame can result in the abolished function of the trapped gene.

C. A PolyA-trap vector contains a splice acceptor immediately following a *LacZ* gene with a transcriptional terminator, a polyA signal and a *Neo* gene driven by a promoter with a splice donor (SD) but missing a transcription terminator. When inserted into an intron in the correct orientation, a PolyA vector can generate two fusion transcripts and proteins. The first is composed of the upstream exon of gene X and β-gal. The second transcript, which is initiated by the promoter 5’ of *Neo* and terminated by the polyA signal of gene X, contains the *Neo* gene and downstream exons of gene X. Cells containing trapped genes can be isolated by G418 selection.

1.5.4.2 Retroviruses

Retroviruses are a class of enveloped viruses possessing a single-strand RNA molecule as their genome, and replicate via a DNA intermediate (Coffin 1997). Following infection, the viral genome is reverse-transcribed to form double-stranded DNA, which is integrated into the host genome. The viral genomes are usually approximately 10kb in length, mainly containing at least three genes: *gag* (coding for core proteins), *pol* (coding for reverse transcriptase) and *env* (coding for the viral envelope protein). At both ends of the genome are long terminal repeats (LTRs) which contain promoter/enhancer regions and sequences involved in the viral integration process. In addition, there are sequences required for packaging the viral sequence (psi) and RNA splice sites in the env gene (Coffin 1997). Other signals exist. For instance, an 800-nucleotide sequence of the murine leukaemia virus (MLV) genome, starting from downstream of the splice donor and extending into the *gag* gene, is adequate to direct the packaging of a heterologous transcript (Adam *et al.* 1988).

Retroviral vectors can be based upon the Moloney murine leukaemia virus (MMuLV), which is an ectotropic virus. These vectors are capable of infecting both mouse cells, enabling vector development for mouse models, and human cells, for potential use in “gene therapy”. The viral genes (*gag*, *pol* and *env*) are replaced with transgenes of interest and expressed from plasmids in the packaging cell line. Because the non-essential genes in the virus lack the packaging sequence (psi) and are not included in the virion particle, a retroviral vector can be used to transfer exogenous DNA into target cells. Once the proviral double-strand DNA integrates into the genome of host cells, exogenous genes carried by a retrovirus can be efficiently expressed.

Many different recombinant retroviral vectors have been constructed for use in mouse genetics and genomics. A typical recombinant retroviral vector includes the 5’ LTR, the 3’ LTR and viral RNA packaging signal (psi) (Coffin 1997). Because the viral proteins are deleted from the viral genome to accommodate exogenous DNA, the recombinant retrovirus is replication-deficient. To produce an infectious retrovirus, the vector needs to be transfected into and transcribed in a viral packaging cell line, which can express all three proteins that are required for viral reproduction, Gag/Pol and Env. Once the recombinant retroviral vector DNA is transfected into the viral packaging cell line, infectious viral particles can be produced and released (Mann *et al.* 1983). The derived replication deficient retrovirus particle can infect any cells that have the receptor for the virus. However, the

infection event will stop at the integration step in the infected cells and no more virus particles can be produced and released because at this stage, the recombinant viruses lack the Gag/Pol and Env proteins (Figure 1-14).

Von Melchner developed the first retroviral-based gene-trap vector (von Melchner *et al.* 1989). In this design, the gene-trap cassette was inserted into the U3 region of the 3'LTR and replaces the viral enhancer sequence. Thus, after viral infection and integration, the provirus provides a duplicated gene-trap cassette in both 5' LTR and 3'LTR (von Melchner *et al.* 1992). Friedrich and Sorano designed another version of the retroviral gene-trap vector, the ROSA (reverse orientation splice acceptor) gene-trap vector (Friedrich *et al.* 1991). In the ROSA vector, the gene-trap cassette was placed between two viral LTRs but in the opposite orientation relative to the viral transcription direction. This reverse orientation was important to avoid splicing of the viral genome which removed the viral packaging signal (psi) 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 the 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 from a host genome, just leaving a single LTR with a *loxP* site in the genome (Ishida *et al.* 1999).

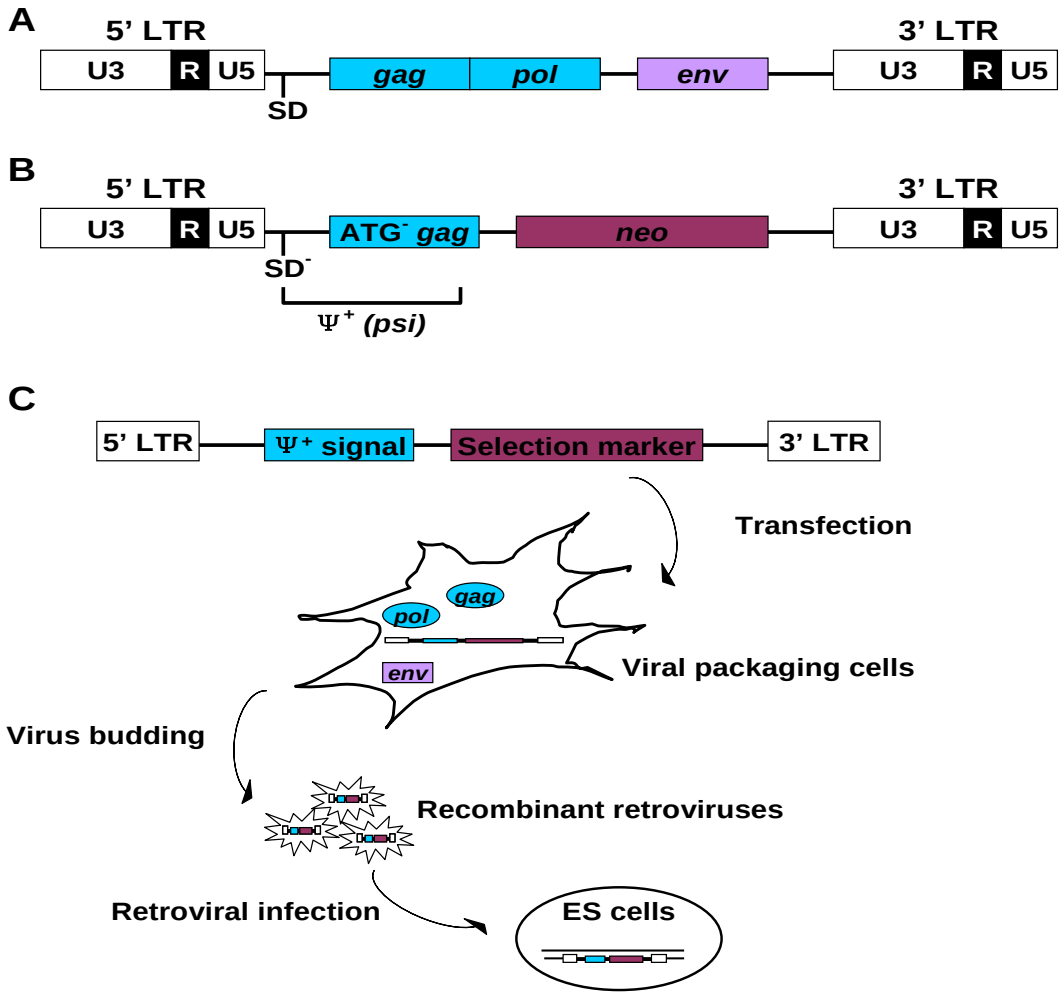


Figure 1-14 Recombinant retroviral vectors and viral production

A. Schematic view of a wild type retrovirus genome. A provirus contains long terminal repeats (LTRs), genes encoding a viral core (*gag*), reverse transcriptase (*pol*) and an envelope protein (*env*). SD, viral splice donor. **B**. Schematic structure of a typical recombinant retroviral Moloney murine leukaemia virus (MMuLV) vector. DNA sequences of the *pol* and *env* genes are deleted. The viral splice donor and *gag* sequence remain to facilitate viral packaging, indicated as Ψ^+. The viral SD is mutated (SD^-) and the initial codon of *gag* is deleted (ATG^-gag) to avoid the interferring effect of gene expression through internal viral mRNA splicing and protein translation. **C**. The essential proteins to produce infectious viruses Gal/Pol and Env are expressed in a mammalian packaging cell line. A recombinant retroviral vector DNA is transfected into the packaging cell line, then the recombinant virus particles are packaged and released.

The retroviral-based gene-trap approach is one of the most commonly used insertional mutagenesis strategies. In contrast to electroporation, by controlling the viral multiplicity, retroviral-mediated mutagenesis can ensure that only a single copy of the entire trap vector integrates into one cell. Also, once a stable virus-producing cell line is made, large amounts of gene-trap viruses can be collected easily, improving the throughput of the gene-trap mutagenesis method. However, analysis of the insertion sites of retroviral vectors derived from human immunodeficiency virus (HIV), avian sarcoma-leucosis virus (ASLV) and murine leukaemia virus (MLV) shows that each viral vector has a unique integration site preference, for example integrating into the 5' end of genes rather than into the 3' end (Mitchell *et al.* 2004). Viral insertion in the 5' untranslated region and the first few introns is more likely to generate null alleles than the 3' insertional sites. Non-random retroviral integration results in trapping "hot spots". This problem also exists in the electroporation-based gene-trap vectors.

Until 2006, the International Gene Trap Consortium, a worldwide collaboration of several gene trapping groups, has reported the generation of ~45,000 independent gene-trapped ES clones using a variety of gene-trap vectors, including electroporation-based vectors and retroviral-based vectors (Nord *et al.* 2006). These insertional events represented ~40% of known mouse genes, which made it possible to systematically analyse gene-trap "hot spots" by both methods. From a study of over 12,000 independent clones generated by one of the collaborating centres, the gene-trap sites were randomly distributed throughout the genome and occurred more frequently on chromosomes with high gene density, which suggests that there is no obvious bias to a single chromosome (Hansen *et al.* 2003). However, among the identified UniGene clusters, 25% of gene-trap mutations were in a limited number of genes, while 75% appeared only once in the database. This result suggests that a proportion of genes are "hot spots" and that the majority of genes are randomly accessible to gene-trap mutagenesis. In addition, they reported that among all the "hot spots", nearly half are common for all vectors, indicating that these hot spots might be caused by locus-specific factors, such as chromatin structure. Open euchromatic regions are more accessible for the insertion of gene-trap vectors. On the other hand, over 50% of the hot spots are vector-specific, indicating that the use of one standard gene-trap vector design can limit genome coverage. To achieve a saturated complexity for insertional mutagenesis with gene-trap vectors, multiple vectors with different features need to be used.

1.5.4.3 Transposon-mediated mutagenesis

Transposable elements or transposons are mobile genetic elements, which have been

identified in many organisms including maize, insects, worms and humans. More than 40% of the human and mouse genomes are composed of transposon-derived sequences (Lander *et al.* 2001; Waterston *et al.* 2002). Transposons were first discovered in maize by Barbara McClintock (McClintock 1948; McClintock 1949; McClintock 1950). She identified the *Ac/Ds* transposons, two members of a family with ~100 transposons. The *Ds*, or dissociation locus, was the first identified mobile locus, but it was incapable of transposition by itself. The second identified locus *Ac*, or activator locus, is autonomous, being able to transpose itself and can also induce the transposition of nonautonomous elements (such as *Ds*). The idea of transposable DNA elements was not fully accepted until the insertion sequences (IS), a transposon-mediated resistance to antibiotics, was discovered in bacteria in 1975 (Hu *et al.* 1975).

The P elements of *Drosophila melanogaster* are widely used transposable elements in fly genetics (Liebl *et al.* 2006). The P elements were cloned in 1982 and around 30–50 copies of P elements were found well dispersed over all the major chromosome arms in the fly genome. In the P-M system of *D.melanogaster*, hybrid dysgenesis, the high rate of mutation in germ line cells, occurs when males of a paternally contributing (P) strain are mated with females of a maternally contributing (M) strain, but it usually does not occur when the reciprocal cross is performed. P strains are distinguished from M strains by multiple genetic elements, the P factors. The P elements were discovered as the genetic causes of hybrid dysgenesis in *D. melanogaster* (Bingham *et al.* 1982; Rubin *et al.* 1982). The full length of these autonomous elements is 2.9 kb with two 31-bp inverted terminal repeats (O'Hare *et al.* 1983; Spradling *et al.* 1982). Due to the alternate splice structure, the P elements transpose only in germ line cells. There are three exons and three introns in the operon of P elements. Introns 1 and 2 are spliced out in somatic cells, resulting in a transposition repressor, which binds to exon 3 to prevent splicing of intron 3. In contrast, all three introns are spliced out in the germ line cells, leading to translation of the P element transposase (Heinz-Albert Becker 2001). With a cut-and-paste mobilizing ability, P elements function as a vehicle for insertional mutagenesis elements and were important tools for advancing *Drosophila* genetics. Like many transposons, P elements are non-functional outside their normal host range, indicating that host factors are involved in transposition (Handler *et al.* 1993).

The *Tc1* transposable element was discovered in 1983 as a repeat sequence in the genome of *Caenorhabditis elegans* (Emmons *et al.* 1983). The homologues of *Tc1* have been found in *Drosophila mauritiana* (Jacobson *et al.* 1986), fungi, plants, fish, frogs and humans (Plasterk *et al.* 1999). *Tc1* and *mariner* elements are members of a large transposon superfamily, the *Tc1/mariner* family (Langin *et al.* 1995). Members of the *Tc1*/*mariner* family

have been used in both mouse and zebrafish (*Danio rerio)* genetics.

By using a comparative sequence reconstruction approach, a *Tc1*-like transposon *Sleeping Beauty* (*SB*) from teleost fish has been synthesized (Ivics *et al.* 1997). The synthetic *SB* has proven to be active in many vertebrate genomes, including fish, mouse (ES) cells and human (ES) cells and it is the first cut-and-paste transposon with activity in mice (Ivics *et al.* 1997; Izsvak *et al.* 2000; Luo *et al.* 1998; Wilber *et al.* 2007). *SB* is a 1.6-kb element flanked by 250-bp inverted terminal repeats and encodes a single protein, the *Sleeping Beauty* transposase, which catalyses the transposition of *SB* from one genomic locus to another. It has been shown that *SB* can be mobile in somatic cells (Yant *et al.* 2000) and germ line cells (Carlson *et al.* 2003; Dupuy *et al.* 2002; Dupuy *et al.* 2001) in mice. The cargo capacity of *SB* can be as long as ~10 kb (Zayed *et al.* 2004). It was found that *SB* tended to insert into AT-rich regions and the sequence of ANNTANNT had a higher frequency of *SB* insertions (Carlson *et al.* 2003). *SB* also showed 100 times more frequent transposition when transposon CpG islands in the transposon are methylated (Yusa, Takeda *et al.* 2004), indicating that heterochromatin formation may play a role in transposition. *SB* has been used as a mutagenesis tool to identify cancer-associated genes in mice. These transposons carry a retroviral long terminal repeat (LTR) and a splice donor, therefore they can activate the expression of proto-oncogenes within a certain distance of their integration sites. Moreover, these transposons have splice acceptors on both strands and polyA signals in both directions. Thus, tumour suppressor transcripts can be disrupted if a transposon inserts into them (Collier *et al.* 2005; Dupuy *et al.* 2005). Although *SB* can transpose to all locations within the genome, there is a strong propensity for “local hopping” events. Three-quarters of insertions are found to be within the same chromosome as the donor loci in mice (Horie *et al.* 2003), which has limited the application of *SB* as a genome-wide mutagenesis system.

Piggybac (*PB*) elements are 2472-bp transposons with two 13-bp inverted terminal repeats and a 594 amino acid transposase (Cary *et al.* 1989; Fraser *et al.* 1995; Fraser *et al.* 1996). This element was discovered as a component of the baculovirus genome Lepidoptera cell lines (Fraser *et al.* 1996). *PB* can carry transgenes up to 9.1 kb without decreasing transposition efficiency. Transgenes up to 14.3 kb were successfully generated (Ding *et al.* 2005), which is bigger than the maximum capacity of retroviral vectors. It has been shown that *PB* transposons insert into the tetranucleotide TTAA site, which is then duplicated after insertion (Fraser *et al.* 1995; Fraser *et al.* 1996). There was no obvious *PB* insertional preference and local hopping has not been observed in any chromosomes. However, 67%

of integration sites were found within the transcriptional region of known or predicted genes, suggesting that *PB* has a unique advantage as a tool for mutagenesis in mice (Ding *et al.* 2005).

1.6 DNA mismatch repair

Both prokaryotic and eukaryotic cells have mutation reduction systems to detect and eliminate various types of DNA alterations/modifications, which reduce the mutation load and limit the accumulation of deleterious DNA changes. The DNA mismatch repair (MMR) system is one of these proofreading systems. Its primary function is to identify and correct errors, such as single nucleotide mismatches and small insertion and deletion (I/D) loops, which mainly arise during DNA replication. MMR is primarily directed at the newly synthesized DNA strand. MMR is highly conserved in prokaryotes and eukaryotes. Loss or damage to the MMR system results in an elevated spontaneous mutation frequency, increased meiotic and mitotic recombination, microsatellite instability, resistance to several cytotoxic DNA-damaging agents and predisposition to cancer in mammals. The MMR system has been extensively reviewed (Buermeyer *et al.* 1999; Hsieh 2001; Kunkel *et al.* 2005; Modrich 1997; Modrich *et al.* 1996). The prokaryotic and eukaryotic MMR systems, the consequence of dysfunctional MMR in humans and mice, and the interaction of MMR with 6-thioguanine will be discussed here.

1.6.1 DNA mismatch repair in prokaryotes

The direct proof that mismatches stimulate their own repair was discovered by Meselson and colleagues who transfected *E. coli* with phage heteroduplex DNA containing one or more mismatched base pairs (Wildenberg *et al.* 1975). Subsequently, a variety of aberrant base pairs caused by chemical and/or physical agents such as O6-methylguanine, UV photo products and cisplatin adducts were identified as being subject to processing by DNA mismatch repair (Duckett *et al.* 1996; Feng *et al.* 1991; Karran *et al.* 1982; Kat *et al.* 1993; Li *et al.* 1996; Ni *et al.* 1999). Essential genes (MutS, MutL and MutH) were identified in bacteria MMR systems as spontaneous mutators, which demonstrated a 100- to 1000-fold increase in their spontaneous mutation frequency (Cox 1976). The functions of these proteins are illustrated in Figure 1-15.

To maintain replication fidelity, the mismatch repair system requires at least five functions: I. Recognition of mispaired nucleotides; II. Discrimination of the parental DNA strand and the newly synthesized DNA strand; III Excision of the incorrectly synthesized nucleotide; IV.

Resynthesis of the correct nucleotide; V. Ligation of the DNA strand.

DNA mismatch repair is initialized when MutS binds to a mismatch. MutS interacts with the β-clamp accessory protein dimer, which is a subunit of the DNA polymerase III holoenzyme (Pol III H.E.). This interaction was observed through the use of protein sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gel-shift assay (Lopez de Saro *et al.* 2001). The function of the β-clamp accessory protein is not known exactly but may help transport MutS to mismatches. MutS binds to ATP and the mismatched DNA (Junop *et al.* 2001) and recruits MutL to activate the latent endonuclease activity of MutH, a member of the type II restriction endonucleases (Ban *et al.* 1998).

The MMR system can discriminate the newly synthesized strand from the original one according to strand methylation status. MutSLH recognizes a hemimethylated GATC site, which is within ~1 kb of the mismatch (Lu *et al.* 1983). A single mismatch can be repaired in wild type *E. coli* but not in a Dam deficient *E. coli* strain, thus confirming that the methylation signal is important for mismatch strand discrimination. By distinguishing the newly synthesized strand from the pre-existing strand, MutH cleaves the new strand (Kunkel *et al.* 2005). The efficiency of repair on the unmethylated chain is nearly 100%; however, no repair is observed if both chains are methylated (Pukkila *et al.* 1983). Newly synthesized DNA is subject to modification by the Dam methylase after a time delay (Lyons *et al.* 1984). As expected, Dam-deficient *E. coli* exhibited a 20 times higher spontaneous mutation frequency (Glickman 1979).

The *E. coli* MMR system recognizes and repairs non-complementary Watson–Crick nucleotide pairs G/T, C/A, G/G, A/A, G/A, A/G, T/T, C/C, A/C, C/T, G/G and T/C. The repair efficiency depends on the sequence context (Jones *et al.* 1987; Kramer *et al.* 1984). Up to four base pairs of insertion/deletion (ID) mismatches are also efficiently processed by the pathway. It has been observed that the heteroduplex molecules containing 5-514 base loops are repaired when a one-base deletion–insertion mismatch is present nearby, and if the mismatch and the loops were spanned 1,448 nucleotides apart, they could still be repaired with high efficiency (95%). Thus, multibase loops in DNA can be resolved as a consequence of co-repair by Dam-directed mismatch repair (Carraway *et al.* 1993; Dohet *et al.* 1986; Learn *et al.* 1989; Parker *et al.* 1992). Based on the above, MMR can correct single nucleotide mismatches and 1–4 nucleotide insertions/deletions. In addition, when larger (0.5 kb) insertion/deletion loops are close (~1 kb) to another one-nucleotide insertion/deletion, both can be repaired.

MutH is able to cleave the unmethylated strand at the hemimethylated GATC site at either side of the mismatches (Lahue *et al.* 1987; Lu 1987). *E. coli* strains deficient for *MutH*, *MutL*, *MutS* or DNA helicase II lack the methyl-directed mismatch repair function; and there are other players in the mismatch repair system, such as exonuclease I (ExoI), exonuclease VII (ExoVII), RecJ exonuclease, exonuclease X (ExoX), single-stranded DNA binding protein (SSB), DNA polymerase III holoenzyme and DNA ligase (Iyer *et al.* 2006). When MutH cleaves the unmethylated strand, a nick is left as an entry point for MutL-dependent recruitment of DNA helicase II and binding of SSB. These two proteins work together to generate single-strand DNA, which is digested with endonucleases. This excision removes the mismatch and DNA polymerase III resynthesizes the new strand until the cleavage point is reached. The new strand is then sealed by DNA ligase.

Thus, the *E. coli* MMR system proceeds in several steps: I). The unmethylated strand (newly synthesized strand) at a hemimethylated GATC site, which is within ~1kb distance of a mismatch, is recognized by MutS and excised by the mismatch repair protein MutH in the presence of MutL; II). Excision of the portion of DNA spanning the incised strand and the mismatch; III). The newly synthesized DNA strand is extended by DNA polymerase III and the gap is sealed by DNA ligase.

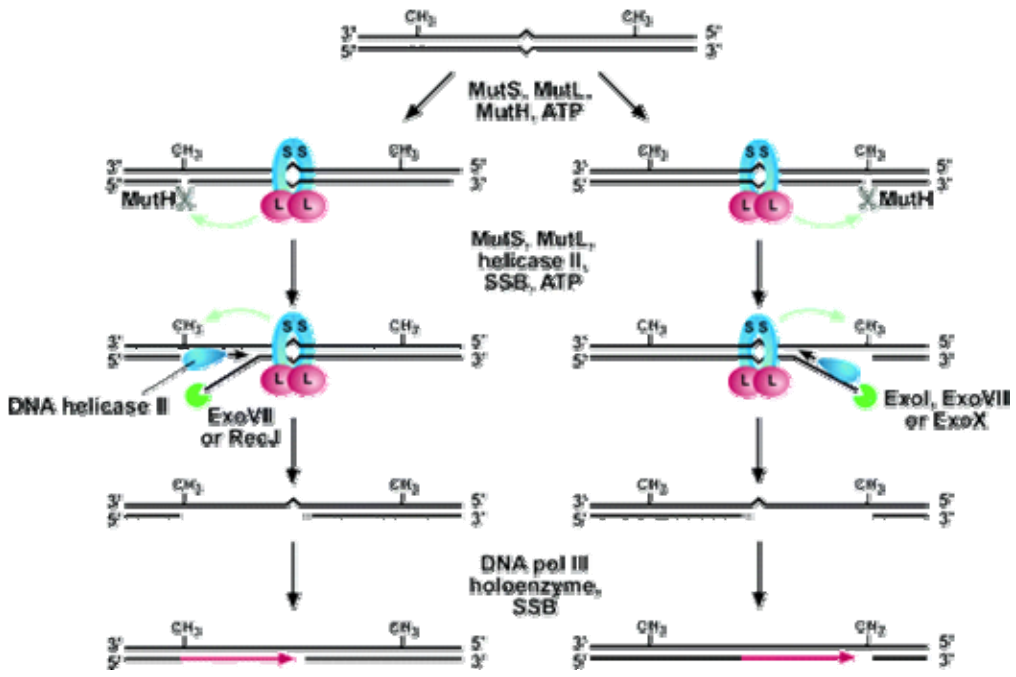


Figure 1-15 ***E. coli*** **mismatch repair system**

Details of the reaction are described in the text. In summary, the hemimethylated GATC sequence may be located either side of a mismatch. The newly synthesized strand is recognized and excised by mismatch repair proteins directed by MutH at the hemimethylated GATC site. DNA polymerase III holoenzyme elongates the excised strand to fill the gap and DNA ligase restores covalent continuity. (Green arrows indicate MutS- and MutL-dependent signalling between the two DNA sites involved in the reaction; blue circle, MutS; red circle, MutL; SSB, single-stranded DNA binding protein; Exo, exonuclease). This figure is copied from Iyer *et al.* (Iyer *et al.* 2006).

1.6.2 DNA mismatch repair in eukaryotes

The DNA mismatch repair system is evolutionarily conserved in prokaryotic and eukaryotic organisms. MMR systems in higher eukaryotes have more specific functions compared with prokaryotic MMR systems. This is reflected by multiple homologues of MutS and MutL in eukaryotes (Table 1-2). Three MutS homologues, MSH2, MSH3 and MSH6, are required to recognize mismatches in yeast and mammals (Wei *et al.* 2002). They form two heterodimeric complexes, MutSα (MSH2 + MSH6) and MutSβ (MSH2 + MSH3) (Genschel *et al.* 1998; Habraken *et al.* 1996; Hughes *et al.* 1992; Kolodner 1996; Marsischky *et al.* 1996).

MutSα and MutSβ have different preferences for recognition sites (Figure 1-16): MutSα binds most base mismatches and one-base insertion/deletion loops (IDLs), whereas MutSβ tends to bind to two- to four- base IDLs (Jiricny 1998; Kolodner *et al.* 1999). The function of MutSα and MutSβ overlap in terms of their recognition of small IDLs, which is consistent with the presence of hMSH2 in both complexes. Microsatellites are repeated in 1–4 nucleotides, which are particularly susceptible to frame shifts. Mutation in hMSH2 causes a higher level of microsatellite instability (MSI), the hallmark of human hereditary non-polyposis colorectal carcinoma (HNPCC), while mutations in hMSH3 and hMSH6 exhibit mild MSI phenotypes (Wei *et al.* 2002). However, evidence has shown that both hMSH3 and hMSH6 can independently participate in the repair of replication errors containing base/base mismatches or one to four extra base insertions, indicating that both proteins may share redundant roles in regulating mutation repairs in human cells (Umar *et al.* 1998). MSH2 and MSH6 exhibit ATP-binding and hydrolysis capacities, providing energy to assist the protein interaction with the MutL homologue (MLH) complexes (Alani *et al.* 1997; Iaccarino *et al.* 1998).

Table 1-2 DNA mismatch repair (MMR) homologues

E.coli	*S.cerevisiae*	*Arabidopsis thaliana*	*Mus musculus/Homo sapiens*
MutS	*MSH1*	*MSH1*	-
	MSH2	*MSH2*	*Msh2/MSH2*
	MSH3	*MSH3*	*Msh3/MSH3*
	MSH4	*MSH4*	*Msh4/MSH4*
	MSH5	-	*Msh5/MSH5*
	MSH6	*MSH6*	*Msh6/MSH6*
	-	*MSH7*	-
MutL	*MLH1*	*MLH1*	*Mlh1/MLH1*
	MLH2	-	-
	MLH3	-	*Mlh3/MLH3*
	PMS1	*PMS1*	*Pms1/PMS1*
	PMS2	-	*Pms2/PMS2*
MutH	-	-	-

Homologues of bacterial (*E.coli*) DNA mismatch repair genes in yeast, plants and mammals. Mammalian *MSH2*, *MSH3*, *MSH6*, *MLH1*, *PMS2* and *PMS1* are involved in replication repair. *MSH4*, *MSH5*, *MLH1* and *MLH3* play a role in the meiotic process. *MSH2* functions in homologous recombination. *MSH2*, *MSH3*, *MSH6*, *MLH1* and *PMS2* are involved in DNA damage surveillance. *Msh1* in *S. cerevisiae* is required for normal mitochondria function. *Msh/MSH*: mutS homologue; *Mlh/MLH*: mutL homologue; *Pms/PMS*: postmeiotic segregation increased protein.

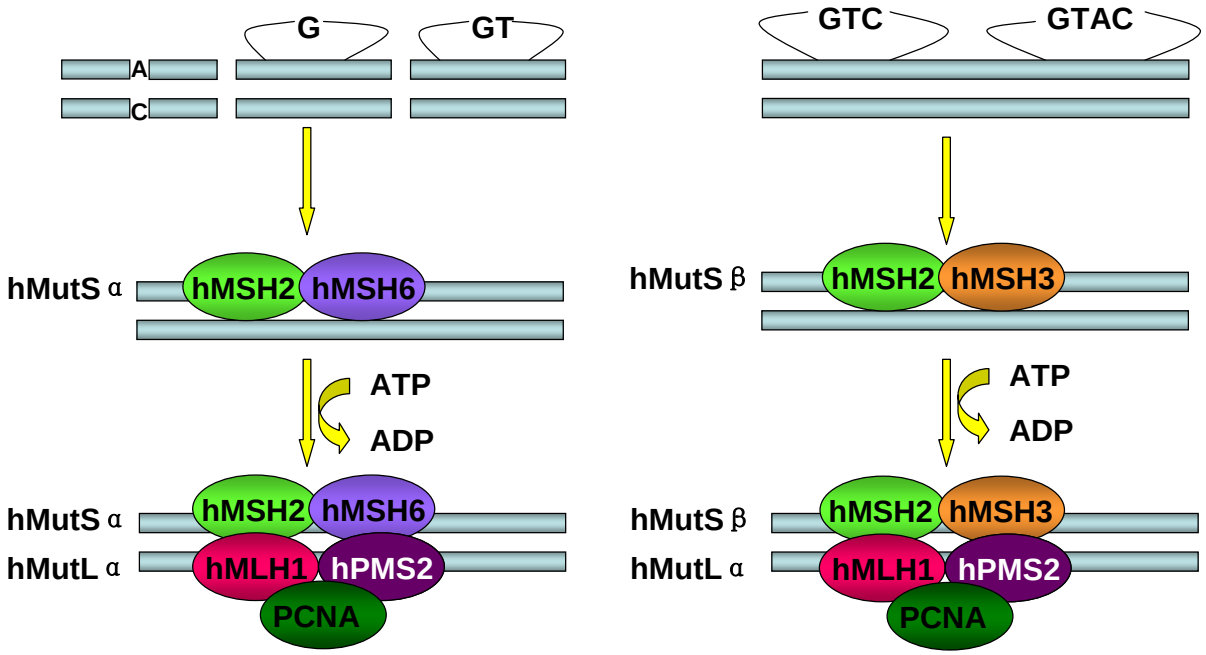


Figure 1-16 Human DNA mismatch recognition preferences

Single nucleotide mismatches and one- to two- nucleotide insertion/deletion loops (IDLs) are detected and bound by the MutSα protein complex (*MSH2* + *MSH6*); three or more nucleotide IDLs are recognized and bound by the MutSβ complex (*MSH2* + *MSH3*). MutLα, including *MLH1* + *PMS2,* and PCNA initiate strand discrimination and excision in an ATP-dependent manner. hMSH: human MutS homologue; hMLH: human MutL homologue; hPMS: human post-meiotic segregation increased protein; PCNA: proliferating cell nuclear antigen.

In humans, two MutL homologues MLH1 and PMS2 (post-meiotic segregation 2), form a heterodimer protein complex MutLα (Flores-Rozas *et al.* 1998). Post-meiotic segregration is a type of segregation generated when a recombinant DNA molecule containing an uncorrected mismatched base pair leads to one mutant progeny and one wild type progeny at the next replication. MutLα binds the DNA/protein complexes of MutSα and/or MutSβ, in the presence of the proliferating cell nuclear antigen (PCNA) (Gu *et al.* 1998; Habraken *et al.* 1998; Habraken *et al.* 1997). These proteins initialize strand discrimination, excision and resynthesis. Like the bacteria MMR system, eukaryotic MMR also needs to discriminate the newly synthesized strand. However, a functional MutH homologue has not been found in yeast and mammals.

MSH2 and MLH1 play a central role in MMR, as mutations in either MSH2 or MLH1 fully abolish the MMR function (de Wind *et al.* 1995; Edelmann *et al.* 1996). Besides PMS2, MLH1 can also form complexes with PMS1 (post-meiotic segregation 1) and MLH3. These protein complexes have minor functions in MMR as it was observed that mutations of PMS1 or MLH3 lead to less severe MSI phenotypes compared to MLH1 mutations (Lipkin *et al.* 2000; Papadopoulos *et al.* 1994). In yeast, Mlh1 and Pms1 are the functional homologues of MutL, which acts with the Msh2-heteroduplex complex containing a G-T mismatch (Prolla, Christie *et al.* 1994; Prolla, Pang *et al.* 1994). Yeast Mlh1 can also complex with Mlh2 and Mlh3 and these two protein complexes have been shown to inhibit mutation of simple sequence repeats (Flores-Rozas *et al.* 1998; Harfe *et al.* 2000). Although the excision process in MMR is not fully understood, it is clear that in eukaryotic cells it requires exonuclease 1 (Exo1), a 5'−3' exonuclease. EXO1 interacts with MSH2 and MLH1 in yeast and mammals (Genschel *et al.* 2002; Tishkoff *et al.* 1997; Tran *et al.* 2001). Mutation of *EXO1* increases the spontaneous mutation rate in yeast (Amin *et al.* 2001). Mice deficient of EXO1 exhibit an elevated spontaneous mutation rate as well as microsatellite instability (Wei *et al.* 2003).

Research in bacteria, yeast and mammals confirmed that MMR proteins suppress homeologous recombination. Homeologous recombination anneals complementary DNA strands, which often contain mismatched nucleotides. Chromosome duplications caused by recombination between homeologous sequences were detected in MutS-deficient and MutL-defective bacteria (Petit *et al.* 1991). In addition, it has been shown that loss of MMR causes increased homologous recombination in bacteria and yeast, leading to the recovery of viable recombinants in interspecies crosses (Chen *et al.* 1999; Datta *et al.* 1997; Rayssiguier *et al.* 1989; Selva *et al.* 1995). Recombination does not usually occur between the genomes of two closely related bacteria, *Escherichia coli and Salmonella typhimurium*,

though their genome sequences are ~80% identical. However, a 50- to 300- fold increase of interspecies recombination is observed in *MutS*, *MutL* and *MutS* mutants (Rayssiguier *et al.* 1989). The role of *Msh2*, *Msh3*, *Msh6*, *Mlh1* and *Pms1* in homologous recombination has been tested in yeast mitotic recombination assays, by using a homologous recombination-activated functional selection marker gene (Nicholson *et al.* 2000; Selva *et al.* 1997; Selva *et al.* 1995). *Msh2* deficiency significantly increases homologous recombination between diverged DNA sequences while the mutants of *Msh3*, *Msh6*, *Mlh1* or *Pms2* only exhibit a minor effect on homologous recombination. Consistent with these observations, *Msh2*-deficient mice have shown a highly elevated homologous recombination between divergent DNA sequences, but Msh3 knockout cells only showed a small increase in homologous recombination (Abuin *et al.* 2000; de Wind *et al.* 1995).

MMR proteins also function in meiosis. Mutations in yeast *Msh2*, *Mlh1* and *Pms1* genes caused post-meiotic segregation due to a loss of ability to repair heteroduplexes (Alani *et al.* 1994; Prolla, Christie *et al.* 1994). Two mammalian *MutS* homologues, *MSH4* and *MSH5*, are expressed specifically in testes and ovaries and play a role during meiosis. MSH4 and MSH5 form a heterodimer protein complex in yeast and humans (Bocker *et al.* 1999). Mutations in either *Msh4* or *Msh5* genes cause reduced meiotic crossovers and increased nondisjunction of homologous chromosomes at meiosis I. However, mutant *Msh4* and *Msh5* do not interfere with the normal mismatch repair function, suggesting that they are not involved in replicative DNA repair (Hollingsworth *et al.* 1995; Ross-Macdonald *et al.* 1994). MSH4- and MSH5-deficient mice exhibit impaired chromosome pairing and synapsis at the meiosis prophase 1: males and females are sterile (Edelmann *et al.* 1999; Kneitz *et al.* 2000). However, mutations of the major mismatch recognition proteins MSH2, MSH3 and MSH6 do not exhibit abnormal meiosis (de Wind *et al.* 1995; Edelmann *et al.* 1997). In contrast, *Mlh1* plays a role in both replicative DNA repair and meiosis. *Mlh1* knockout mice exhibit microsatellite instability and sterility phenotypes in both males and females (Baker *et al.* 1996). *Mlh1* mutant yeast also exhibits a reduction in meiotic crossovers as well as a deficiency in mismatched DNA repair during meiosis (Hunter *et al.* 1997). The MutL homologue *Mlh3* is expressed in mouse meiotic cells and in human testes. Co-immunoprecipitation experiments have shown that the MLH3 protein interacts with meiosis-specific MSH4 in mouse meiotic cell extracts (Santucci-Darmanin *et al.* 2002).

In the normal mammalian genome, methylation occurs only at cytosines 5' to guanosines (Ng *et al.* 1999). CpG dinucleotides in mammals have a high frequency of methylation. However, some small (0.5 to several kb) stretches of DNA (CpG islands) containing the expected frequency of CpGs do not exhibit the expected level of methylation (Ng *et al.*

1999). These CpG islands usually locate in the proximal promoter regions of 40–50% of human genes. As a growing number of cancer genes have been identified that harbour dense methylation in normally unmethylated promoter CpG islands, for example the 5' CpG island of *hMLH1* (Herman *et al.* 1998) in cancers, reviewed by Jones and Herman (Herman *et al.* 2003; Jones *et al.* 1999), and the apparent link between MMR genes and human cancer, it is worth investigating the functional interaction between MMR proteins and DNA methyltransferases. DNA methyltransferase 1 (DNMT1), one of the three functional DNA methyltransferase enzymes, was discovered to be involved in MMR through analysis of *Dnmt1* hypomorphic, *Mlh1*-deficient mice (Trinh *et al.* 2002). Increased lymphomagenesis and decreased intestinal tumourgenesis observed in this compound mutant indicated that *Dnmt1*-mediated methylation status is associated with MMR in the maintenance of an intact genome. *Dnmt1* was identified in a recessive genetic screen for 6-thioguanine tolerance (Guo *et al.* 2004). A role for *Dnmt1* in preventing genome damage was indicated by the observation that inactivation of the *Dnmt1* gene results in elevated 1–2 nucleotide microsatellite instability, as detected by either a slippage assay or PCR assay (Guo *et al.* 2004; Kim *et al.* 2004; Wang *et al.* 2004). Further experiments demonstrated that stable transfection of a sense *DNMT1* construct down-regulates *MLH1* and *MSH2* expression by hypermethylating the promoter regions of these two genes in the human colon cancer cell line SW1116 (Fang *et al.* 2006). The hypermethylation of the promoter region of *hMLH1* was also observed in MSI-positive tumours (Gurin *et al.* 1999). In Fang's research, sense and antisense DNMT1 constructs were introduced into SW1116 cells separately. An inhibitor of DNMT1, 5-aza-2'-deoxycytidine (Robert *et al.* 2003) and the CpG-specific methylase *SssI* were used to treat cells as negative and positive controls of CpG methylation. The Western blot results demonstrated that the DNMT1 protein expression was increased or decreased in transfected cell lines containing sense or antisense DNMT1 constructs, respectively. The methylation status of *hMLH1* and *hMSH2* promoters was determined by the bisulfite conversion (unmethylated cytosine to uracil) based methylation-specific PCR (Herman *et al.* 1996). In SW1116 cells expressing the sense DNMT1 construct, the expression of *hMLH1* and *hMSH2* was down-regulated through hypermethylation of their respective promoters; and the expression of the antisense DNMT1 construct resulted in promoter demethylation and up-regulated transcription of the *hMSH2* and *hMLH1*. These are further indirect evidence that the DNMT1 protein is involved in the MMR system.

1.6.3 MMR in hereditary non-polyposis colon cancer

Hereditary non-polyposis colorectal carcinoma (HNPCC), also called Lynch syndrome, is a cancer predisposition syndrome that is transmitted in an autosomal dominant way (Lynch *et*

al. 1993). HNPCC usually develops after the fourth decade and the cancer syndrome is found in the proximal colon, as well as the endometrium, ovary, stomach and small intestine (Lynch *et al.* 1993). HNPCC comprises about 5% of all colorectal carcinomas, the fourth most common cancer in humans. These tumour samples were found to contain a homozygous mutation in some of the mentioned MMR genes (Buermeyer *et al.* 1999). In 80%–95% of HNPCC cases microsatellite instability, the duplication/deletion of one to four base pair repeats was identified (Lynch *et al.* 1999; Rowley 2005).

The facts that MMR defects in *E. coli* and *S. cerevisiae* (Strand *et al.* 1993) were responsible for miscrosatellite instability and the similar MSI phenotype was observed in some colorectal cancer samples (Parsons *et al.* 1993; Thibodeau *et al.* 1993) stimulated research on the potential relationship between human MMR genes and colorectal cancer. A prediction that the phenotypes of the mutation involved in HNPCC might result from defect(s) of MMR genes was raised (Aaltonen *et al.* 1993; Ionov *et al.* 1993). Moreover, linkage analysis had shown that some human homologues of *E. coli* MMR genes, for instance *hMSH2* and *hMLH1*, were close to HNPCC loci (Lindblom *et al.* 1993; Peltomaki *et al.* 1993). These discoveries led to cloning of MMR genes, for example *hMSH2* (Fishel *et al.* 1993) and the research of their function in HNPCC. Until now, clinical research has revealed that the disease is predisposed by heterozygous mutations or alterations of DNA mismatch repair genes including *MLH1*, *MSH2*, *MSH6*, *MLH3*, *PMS1* and *PMS2*, among which *MSH2* (~60%) and *MLH1* (~30%) account for the majority of the cases, while the other MMR genes are less frequently mutated in HNPCC (OMIM™).

As a consequence of the mutating phenotype, including MSI and nucleotide mismatch in HNPCC, other genes have an elevated probability of being mutated (Gurin *et al.* 1999). MMR-deficient cells obtain a selective growth advantage, which can increase the likelihood of developing cancer. Frame shift mutations have been identified in the genes *APC*, *BLM*, *TGF-betaRII*, *IGF2R*, *BAX*, *BRCA1*, *p53* and the MMR genes *MSH3* and *MSH6* in genetically unstable human colon cancer (cells) or sporadic gastrointestinal tumours (Calin *et al.* 1998; Gurin *et al.* 1999; Huang *et al.* 1996; Markowitz *et al.* 1995; Rampino *et al.* 1997).

1.6.4 Mouse models for human DNA mismatch repair gene defects

To examine MMR gene function in mammals, mouse models with defective MSH2, MSH3, MSH4, MSH5, MSH6, MLH1, MLH3, PMS1 and PMS2 proteins have been established by gene-targeting technology. As reviewed by Beurmeyer and Wei (Buermeyer et al. 1999; Wei

et al. 2002), many of these mouse models exhibit phenotypes similar to HNPCC in human and provide an opportunity to understand the physiological functions of these gene products (summarized in Table 1-3). However, mouse models are mainly homozygous to MMR mutations while human HNPCC samples mainly contain heterozygous MMR mutations.

Genetic inactivation of *Msh2*, *Msh3*, *Msh6*, *Msh3/Msh6*, *Mlh1* or *Pms2* confers cancer predisposition (Baker *et al.* 1995; Baker *et al.* 1996; de Wind *et al.* 1995; de Wind *et al.* 1999; de Wind *et al.* 1998; Edelmann *et al.* 1996; Edelmann *et al.* 2000; Kawate *et al.* 1998; Prolla *et al.* 1998; Reitmair *et al.* 1995; Yao *et al.* 1999). These mice have early onset of lymphoma and other tumours, such as gastrointestinal cancer and skin tumours. These mice show 50% mortality within a year of birth, except for *Msh3*-deficient mice, in which 50% mortality is reached after 18 months. Mice with both *Msh3* and *Msh6* deficiency exhibit a similar phenotype to *Msh2* mice. As both of these genes are essential for MMR, mutations in either of them result in microsatellite instability in mice.

Mlh1 and *Pms2* deficiency is not only tumorigenic in mice, but also leads to sterility in both males and *Mlh1*-deficient females (Baker *et al.* 1995; Baker *et al.* 1996). In *Mlh1*-deficient males, the homologous chromosomes of spermatocytes pair and proceed to pachytene; however, meiosis is arrested at the pachytene stage in which spermatids then fail to develop. In *Pms2*-deficient male mice, sterility is due to abnormal homologous chromosome pairing during pachytene.

Msh4, *Msh5*, *Pms1* and *Mlh3* are not responsible for replicative DNA repair, thus mice deficient for these genes do not develop early onset cancer. Cells of these mice also exhibit low or no microsatellite instability. *Msh4* and *Msh5* are specifically expressed in germ cells and play a central role in meiosis. Inactivation of either of them results in sterile males and females (de Vries *et al.* 1999; Edelmann *et al.* 1999; Kneitz *et al.* 2000). Both male and female *Pms1* mutant mice are fertile. *Pms1* deficiency causes minor MSI at mononucleotide repeats but not dinucleotide repeats (Prolla *et al.* 1998). However, mutation of *Mlh3* leads to sterile males and females. MLH3 protein is needed for MLH1 protein's binding to meiotic chromosomes. Both MLH3 and MLH1 proteins localize to meiotic chromosomes from the mid-pachynema stage of prophase I. The loss of *Mlh3* causes meiotic arrest (Lipkin *et al.* 2002).

Table 1-3 Mouse models of deficient mismatch repair genes

Genotype	50% survival (months)	Insertion/deletion loop repair ability		Microsatellite instability incidence		Tumour		Fertility		Reference
		1 bp	2–4 bp	Mononucleotide	Dinucleotide	Frequency	Spectrum	Male	Female	
$Msh2^{-/-}$	6	no	no	N/A	High	High	L, GI, Sk and other	Fertile	Fertile	1
$Msh3^{-/-}$	18	yes	no	Moderate	High	Low	GI	Fertile	Fertile	2
$Msh4^{-/-}$	>16	yes	yes	N/A	N/A	None	None	Sterile	Sterile	3
$Msh5^{-/-}$	>16	yes	yes	N/A	N/A	None	None	Sterile	Sterile	4
$Msh6^{-/-}$	11	no	yes	None	Low	High	L, GI and other	Fertile	Fertile	5
$Msh3^{-/-}/Msh6^{-/-}$	6	no	no	High	High	High	L, GI, Sk and other	Fertile	Fertile	6
$Mlh1^{-/-}$	6	no	no	High	High	High	L, GI, Sk and other	Sterile	Sterile	7
$Mlh3^{-/-}$	-	no	no	Low	Low	Low/None	-	Sterile	Sterile	8
$Pms1^{-/-}$	>18	N/A	N/A	Low	Low	None	None	Fertile	Fertile	9
$Pms2^{-/-}$	10	no	no	High	High	High	L and Sa	Sterile	Fertile	10

Modified from (Wei *et al.* 2002).
Abbreviations: L: lymphoma; GI, gastrointestinal; Sk, skin cancer; Sa, Sarcoma; N/A, data not available.
References:
1. (de Wind *et al.* 1995; de Wind *et al.* 1998; Reitmair *et al.* 1995; Smits *et al.* 2000)
2. (de Wind *et al.* 1999; Edelmann *et al.* 2000)
3. (Kneitz *et al.* 2000)
4. (de Vries *et al.* 1999; Edelmann *et al.* 1999)
5. (de Wind *et al.* 1999; Edelmann *et al.* 2000; Edelmann *et al.* 1997)
6. (de Wind *et al.* 1999; Edelmann *et al.* 2000)
7. (Baker *et al.* 1996; Edelmann *et al.* 1996; Kawate *et al.* 1998)
8. (Lipkin *et al.* 2002)
9. (Prolla *et al.* 1998)
10. (Baker *et al*. 1995; Prolla *et al*. 1998; Yao *et al*. 1999)

1.6.5 MMR deficiency and tolerance of DNA methylation

The MMR system can recognize some types of DNA damage and initiate repair and G2/M cell cycle arrest and apoptotic cell death. Thus, the MMR system prevents the accumulation of DNA lesions and plays a DNA damage surveillance role. Mutations of *MutS* or *MutL* in Dam-defective bacteria rescue the hypersensitivity to simple methylating agents, including methyl-nitrosourea (MNU) and N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) (Karran *et al.* 1982). It was also found that the human lymphoblastoid MT1 B-cell line is resistant to MNNG and was mismatch repair-deficient (Kat *et al.* 1993). The link between MMR and methylation damage tolerance was established when MNNG tolerance is reduced after transforming normal human chromosome 3 (containing the *MLH1* gene) into the *Mlh1*-deficient cancer cell line HCT116 (Koi *et al.* 1994). Introducing human chromosome 2 (containing *MSH2* and *MSH6*) into colon cancer cell line HEC59 (*MSH2* mutant) and HCT15 (*MSH6* mutant) also confirmed the link between MNNG tolerance and the mutations of genes *Msh2* and *Msh6* (Umar *et al.* 1997).

Human colorectal carcinoma cell lines and hamster cell lines, in which mismatch repair is deficient, demonstrate resistance to DNA methylating agents such as N-methyl-N-nitrosourea, MNNG and 6-thioguanine (6TG). Cancer cell lines with *MSH2*, *MSH6* or *MLH1* gene defects are also resistant to 6TG (Aebi *et al.* 1997). Mouse ES cells with disrupted *Msh2* and *Msh6* also demonstrate methylation tolerance caused by MNNG or 6TG (Abuin *et al.* 2000; de Wind *et al.* 1995). DNA cytosine-5-methytransferase 1 (*Dnmt1*) and *Msh6* were identified in a genetic screen for cells defective in MMR. *Dnmt1*-deficient ES cells exhibited methylation tolerance to 6TG and microsatellite instability (Guo *et al.* 2004).

The cytotoxicity of MNNG is due to its ability to methylate guanine to form O^6-methyguanine (O^6-meG). O^6-meG can be repaired by methylguanine methyltransferase (MGMT), which removes the methyl group from O^6-meG. However, O^6-meG is not cleared in MMR-deficient cells when MGMT is proficient. Thus, MNNG tolerance is not due to MGMT activity (Karran *et al.* 1992). O^6-meG can pair with either thymidine (T) or cytosine (C) in the process of DNA replication and form the mismatches of O^6-meG/T or O^6-meG/C. These base mismatches can be recognized and bound by MutSα in MMR-proficient eukaryotic cell extracts but not in MMR-deficient cell extracts (Duckett *et al.* 1996; Griffin *et al.* 1994), suggesting that MMR mediates MNNG cytotoxicity.

1.6.6 6-thioguanine introduces DNA mismatches, cycle arrest and apoptosis

The purine antimetabolite, 6-thioguanine (6TG structure illustrated in Figure 1-17) is widely used as a therapeutic agent for leukaemia and as an immunosuppressant in organ transplant patients. It is utilized by the purine nucleotide synthesis pathway and is incorporated into replicating DNA leading to base mismatch within the DNA duplex. Due to the structural similarity between 6TG and MNNG, it is believed that the mechanisms of their cytotoxicity are similar (Figure 1-18). 6TG is metabolized by hypoxanthine guanine phosphoribosyl transferase (*HPRT)*, thus acquiring a sugar phosphate group to form 2'-deoxy-6-thioguanosine-triphosphate, an active guanine nucleotide analogue in DNA synthesis. The observation that HPRT-deficient cells are fully resistant to 6TG is consistent with the function HPRT in 6TG metabolism. After 6-thioguanine is incorporated into the newly synthesized DNA strand, it undergoes a non-enzymatic methylation by an intracellular methyl group donor, S-adenosylmethionine (SAM) and forms S^6-methylthioguanine [^{SMe}G] (Karran *et al.* 2003; Swann *et al.* 1996). During DNA replication, two types of mismatches are produced in DNA duplexes, [^{SMe}G]/T and [^{SMe}G]/C. Functional MutSα (MSH2 and MSH6) recognizes and binds to DNA containing a mismatched pair of [^{SMe}G]/T or [^{SMe}G]/C, thus leading to cell cycle arrest, usually in the G2/M phase, and apoptotic cell death (Hawn *et al.* 1995; Waters *et al.* 1997).

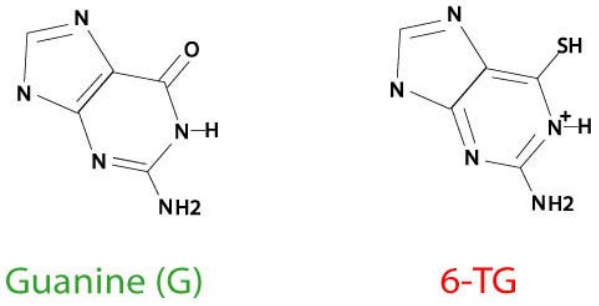


Figure 1-17 The molecules of guanine and its analogue 6-thioguanine (6TG)

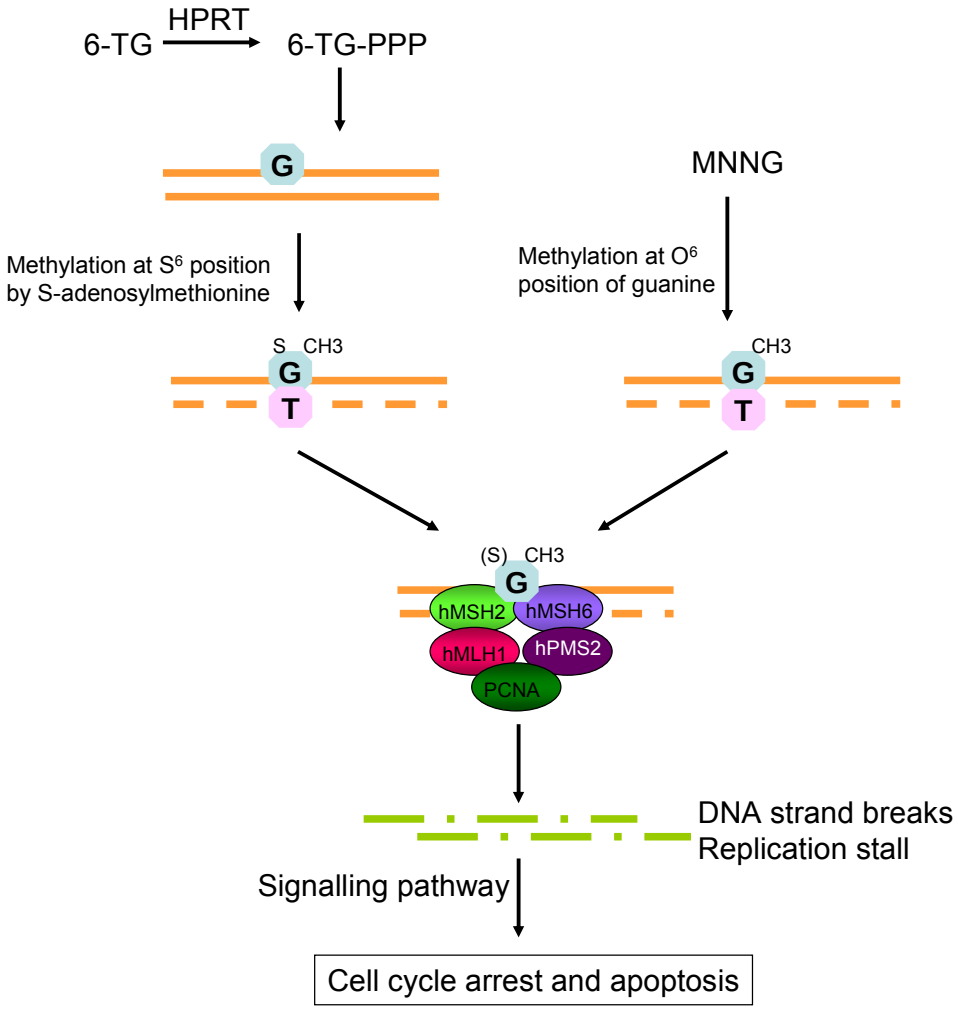


Figure 1-18 Cytotoxicity of MNNG and 6TG

The cytotoxicity of 6TG is initiated in cells with hypoxanthine guanine phosphoribosyl transferase (*HPRT)*, which metabolizes 6TG to form 2'-deoxy-6-thioguanosine-triphosphate (6TG-PPP). 6TG-PPP can be incorporated into DNA as a guanine analogue during DNA replication. S-adenosylmethionine (SAM) offers a methyl group to 6TG at the S^6 position to form an S^6-methylthioguanine [^{SMe}G]. Similarly, the methylation agent MNNG methylates the O^6 position of guanine to form an O6-methylguanine [^{Me}G]. Both [^{SMe}G] and [^{Me}G] pair with thymidine (T) in the next DNA replication cycle. The MutSα complex (MSH2 and MSH6) recognizes this type of mismatch leading to multiple repair interactions with MutLα (MLH1 and PMS2) and proliferating cell nuclear antigen (PCNA), during which single- and double-strand breaks are generated. The density of SSBs is very high and repair patches will overlap with DSBs. This will activate signalling pathways and lead to cell cycle arrest and apoptosis. Figure adapted from Guo *et al*. (Guo 2004)

DNA double-strand breaks (DSBs) can be indirectly measured by pulsed-field gel electrophoresis and the single cell gel electrophoresis assay (also known as comet assay), DNA protein cross-links and a reduction of mRNA synthesis have also been shown in 6TG-treated cells (Ling *et al.* 1992; Pan *et al.* 1992). Interestingly, DSBs are produced at similar levels in both MMR-proficient and MMR-deficient cells after 6TG treatment. Most DSBs appear within one day of 6TG addition (Yan *et al.* 2003).

6TG treatment induces a dose-dependent increase in DNA single-strand breaks (SSBs), which occur during the second and later cell cycles after incorporation of 6TG into DNA (Fairchild *et al.* 1986; Pan *et al.* 1990). SSBs are more frequent and longer lived in MMR-proficient cells, and are thus believed to signal G2/M cell cycle arrest. The duration of SSB formation correlated with the time course of 6TG-induced G2/M arrest (Yan *et al.* 2003).

The p53 pathway may be involved in MMR-mediated DNA damage surveillance. It has been shown that the expression of p53 and the p53-induced protein p21/waf-1 was upregulated in temozolomide (a methylating antitumour compound) treated MMR-proficient lymphoblast cells (D'Atri *et al.* 1998). Similar upregulation in p53 and Fas receptor was observed during MNNG-causing apoptosis. Inhibition of Fas receptor activation by an anti-Fas neutralizing antibody decreased apoptosis in proliferating lymphocytes by 61% (Roos *et al.* 2004). Although the p53 pathway is involved in lymphocyte cell cycle arrest and apoptosis, its function in methylation agents-induced MMR-mediated pathways in other cell lines is not clearly understood. The human kidney fibroblast cell line 293T is deficient for p53 activity, but can still be arrested at the G2/M phase and apoptosis induced by MNNG treatment (Cejka *et al.* 2003; di Pietro *et al.* 2003).

Recently, several experiments have established an association between the ATR (ATM- and Rad3-related)/Chk1 (checkpoint kinase 1) signalling pathway and 6TG cytotoxicity (Stojic *et al.* 2004; Yamane *et al.* 2004). Within 23 stomach tumour samples, microsatellite instability causing frame shift mutations was observed in ATR/CHK1 pathway genes, including CHK1 (9%), ATR (21%), MED1 (43%) and MMR genes hMSH3 (56%), hMSH6 (43%) (Menoyo *et al.* 2001). A co-immunoprecipitation experiment demonstrated that ATR physically binds to MSH2. Phosphorylation of SMC1 (structure maintenance of chromosome 1) at S966 and CHK1 at S317 were abolished in both siRNA-mediated MSH2-deficient and ATR-deficient cells (Wang *et al.* 2003). In HeLa cells, siRNA against ATR or Chk1 kinase reduced the G2/M checkpoint and enhanced the apoptosis following 6TG treatment (Yamane *et al.* 2004).

MED1 protein, also known as methyl-CpG binding domain protein 4 (MBD4), was isolated as an MLH1 binding protein in yeast and the MED1/MLH1 physical interaction was confirmed in human cells by co-immunoprecipitation (Bellacosa *et al.* 1999). MED1 is a methyl-CpG binding protein, which binds to hemimethylated and fully methylated DNA but not to unmethylated DNA. MED1 methyl-CpG-binding domain (MBD) exhibits endonuclease activity on single- and double-stranded DNA. The fact that MED1 binds to MLH1 and MED1 acts as an endonuclease indicates that MED1 may play a role in MMR. A lack of MBD in cells results in microsatellite instability, similar to that observed in MMR-deficient cells (Bellacosa *et al.* 1999). This is consistent with the presence of mutant MED1 in human cancers with MSI (Bader *et al.* 1999; Menoyo *et al.* 2001; Riccio *et al.* 1999). However, MSI is not observed in *Mbd4*-deficient mice (Millar *et al.* 2002; Wong *et al.* 2002). MED1 functions in the maintenance of genome integrity and response to DNA damage (Parsons 2003). *Med1*-deficient mouse embryonic fibroblasts fail to undergo G2/M cell cycle arrest and apoptosis (Cortellino *et al.* 2003), and mice without the MED1 function showed significantly reduced apoptotic responses 6h following treatment with a range of cytotoxic agents including gamma irradiation, cisplatin, temozolomide and 5-fluorouracil (5-FU) (Sansom *et al.* 2003).

Casein kinase 2 also regulates 6TG-induced G2/M cell cycle arrest and apoptosis. Inactivation of CK2 through siRNA significantly reduces G2/M arrest but increases a sub-G1 population and strongly induces apoptosis in 6TG-treated HeLa cells. Upon treatment of CK2-inhibited HeLa cells with a general caspase inhibitor, z-VAD, 6TG-induced apoptosis was inhibited. This demonstrates a role for caspases in 6TG-induced apoptosis (Yamane *et al.* 2005). The regulatory function of CK2 in 6TG-induced cell cycle arrest may be mediated by Chk1. The regulatory β-subunit of CK2 (CK2 β) is important for the formation of tetrameric CK2 complexes. Immunoprecipitation use of Cos-1 cell lysates with a polyclonal anti-CK2β antibody leads to the co-precipitation of CHK1 (Guerra *et al.* 2003). The activation of CHK1 is associated with its phosphorylation by the ATR/ATM family of kinases (Zhou *et al.* 2000).

Recently published results showed that Brac1 contributes to the activation of the G2/M cell cycle arrest following 6TG-induced DNA mismatch damage (Yamane *et al.* 2007). The BRCA1-defective human breast cancer cell line HCC1937 exhibits higher tolerance to 6TG than BRCA1-proficient cells. Considering that BRCA1 plays a central role in the BRCA1-associated genome surveillance complex (BASC) (Wang *et al.* 2000), its regulatory function in MMR and 6TG-induced cell cycle arrest may involve other BASC proteins.

1.6.7 Genetic screen for MMR genes

Mismatch repair is a critical DNA surveillance system to prevent the inheritance of replication errors and cancer development. However, some of our knowledge about MMR systems in mammals is missing, for example the process of strand discrimination and excision. Genetic screens provide an efficient approach to isolate new MMR components. As previously described, cells that are deficient for the different MMR components MSH2, MSH6, MLH1 and PMS2 are resistant to methylating agents, for instance 6TG (Abuin *et al.* 2000; Karran *et al.* 1996). Therefore, 6TG has been used as a selective agent in genetic screens for the isolation of methylation-tolerant mutants as a means of identifying new MMR components.

In 2004, Dr Ge Guo reported the results of a recessive screen conducted on a library of insertional mutated ES cells. The *Dnmt1* gene was identified as a normal component of the DNA MMR system. In this screen, the mutations were induced with a gene-trapping retrovirus. *Dnmt1* mutants were not only resistant to 6TG but also exhibited other features of MMR, for instance an increase in the microsatellite instability rate (Guo *et al.* 2004). However, some parts of the MMR system in humans and mice are still unknown, such as the process of strand excision. Considering the limitations of gene-trap mutagenesis and the limited complexity of Guo's mutation library, I decided to repeat this MMR screen on a new mutation library generated by gamma irradiation to take advantage of the high efficiency of irradiation mutagenesis and relatively unbiased mutation loci.

1.6.8 Project design

My project was designed to identify genes involved in the DNA mismatch repair mechanism (Figure 1-19). One goal was to explore the feasibility of generating homozygous mutations in the *Blm*-deficient ES cells by gamma irradiation. *Blm*-deficient ES cells show an elevated loss of heterozygosity (LOH). As a result, heterozygous mutations (deletions) in *Blm*-deficient ES cells will have a chance to be segregated into the same daughter cell and become homozygous. As discussed earlier, gamma irradiation is a convenient method to create large chromosome aberrations, such as duplications and deletions. In addition, irradiation-induced aberrations are believed to be random, making it a suitable approach for generating genome-wide mutations. Using a series of irradiation doses and whole genome tiling path arrays, comparative genomic hybridization (CGH) experiments were performed to measure the sizes of the irradiation-generated deletions. A mutant *Blm*-deficient ES cell

library was established by gamma irradiation. Cells in the mutant library were cultured for sufficient cell cycles to accumulate LOH-induced homozygous mutations. Using a screening strategy similar to that previously described by Dr Ge Guo (Guo *et al.* 2004), 6TG-resistant mutants were isolated. DNA from these mutants was analysed using a genome-wide CGH tiling path array to detect large genomic alterations. Whole genome expression arrays were used to identify transcriptional variations of mutant clones. The homozygously deleted genes can be identified by these expression arrays when they fail to be detected by array CGH. In addition, expression array can detect the absence of gene expressions when genes are heterozygously deleted. Also, the comparison between mutants can provide evidence for other transcriptional changes. Finally, the expression array analysis is useful as MMR-associated genes can also be identified when the irradiation-induced mutations lead to an alteration in other genes' expression levels. Homozygous deletions and absence of expression of known MMR genes such as *Msh2, Msh6, Mlh1* and *Pms2* are expected.

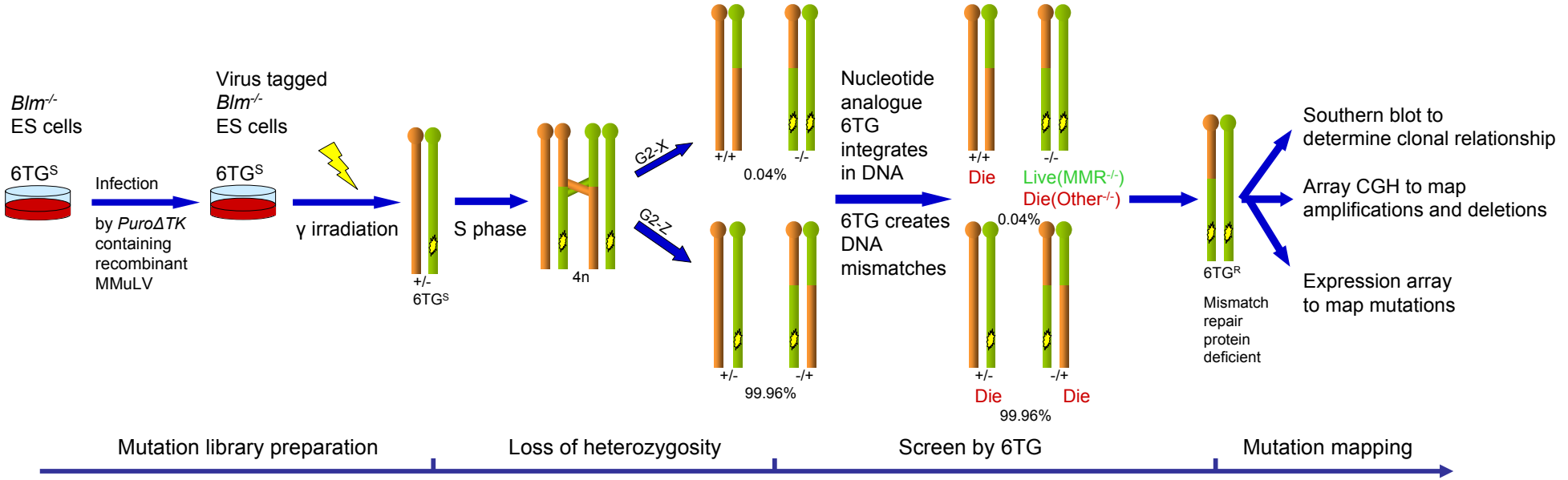


Figure 1-19 A genome-wide recessive screen for DNA mismatch repair genes

Blm-deficient ES cells were infected and tagged by *PuroΔtk* containing recombinant MMuLV before being mutated by gamma irradiation. Irradiated cells carry mainly heterozygous mutations, which may segregate into the same daughter cell with consequent loss of heterozygosity. During the 6TG screening process, DNA MMR gene homozygous mutants should be viable while other wild type and mutant cells should not survive. Mutants isolated in this screen are individually picked and expanded. DNA and RNA samples are prepared. Southern blot hybridizing by a *PuroΔtk* probe can determine the clonal relationship between mutants. Array CGH identifies deletions. Expression arrays are used to identify mutant genes not detected using genomic arrays.