Genome-wide recessive screens for DNA mismatch repair genes in mouse ES cells

This dissertation is submitted for the degree of Doctor of Philosophy.

By Zikai Xiong

The Wellcome Trust Sanger Institute
Hughes Hall, University of Cambridge

DECLARATION

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

None of the material presented herein has been submitted previously for the purpose of obtaining another degree.

Zikai Xiong

ACKNOWLEDGEMENT

I would like to take this opportunity to formally thank all the people who supported me during the course of my PhD project.

My deepest thank is to my PhD supervisor Dr Allan Bradley, for his enormous support and excellent advice through the whole project. It was great pleasure to work with Allan and achieve challenging tasks in his laboratory. I thank his patience as well as encouragement during the hard time of this project.

I have special thanks to all Bradley laboratory members for their communication and kind help. I am particularly grateful to my wife Dr Wei Wang for her huge support in both my life and career. We have had numerous scientific discussions and collaborations during these experiments. I learned recombinant retroviral techniques from her. Many thanks are also due to Oliver Dovey for his detailed guidance and contribution to array CGH experiments. I also want to thank Peter Ellis for his help on expression arrays and I am grateful to Gregory Lefebvre, Keith James and Robert Andrews for their help in bioinformatics analysis. I thank Dr Ge Guo and Dr Wei Wang (Male) for their instructions in ES cells' culture and molecular manipulations and valuable discussions. I also thank Dr Yue Huang for discussions at the beginning of the project. I thank Haydn Prosser for providing vectors and advice, Qi Liang for comfirming the recombination medicated cassette exchange ability of my cell line in her experiments. I am grateful to Frances Law and Alastair Beasley for their help in cells' culture and routine laboratory issues. I thank Shaun Cowley, Juan Cadinanos, Roland Rad, George Vassiliou, Antony Rodriguez, Nathalie Conte for valuable scientific discussions.

The advice and feedback from my PhD committee were precious. The members were Pentao Liu, Jessica Downs, Andrew Fraser, Derek Stemple and Allan Bradley, who shared their critical thinking and provided excellent advice for my project.

I must say "Thank you!" to my parents Yan Xiong and Huiqin Zhu. Their endless support never leaves me.

My 4-year full studentship is sponsored by the Wellcome Trust.
Registered charity number 210183

ABSTRACT

Genome-wide genetic screens on libraries containing homozygous mutant mouse embryonic stem (ES) cells make it feasible to examine all mouse genes expressed in ES cells for their role in any biological process which is active in ES cells. Bloom syndrome protein (Blm) deficient ES cells have high mitotic recombination and loss of heterozygosity rates, which allow cells with homozygous mutations to be generated in populations of mutated *Blm*-deficient ES cells. A genome-wide mutation library was generated in this genetic background using gamma radiation. Analysis of isolated clones from this library by high-resolution Comparative Genomic Hybridization (CGH) arrays revealed that each carried several duplications or deletions ranging in size from 0.1 to 50Mb. This mutation library provided good coverage of the mouse genome and it is a new genetic resource for conducting loss-of-function genetic screens in mammalian cells.

In humans, mutations in components of the DNA mismatch repair (MMR) system cause hereditary non-polyposis colorectal carcinoma (HNPCC, Lynch syndrome). Mutations in the *MSH2* and *MLH1* genes account for the majority of the cases. To extend our understanding of the MMR system a recessive screen was implemented in the mutation library to identify new components by selection in 6-thioguanine (6TG). This nucleotide analogue is incorporated into DNA and is recognized by the DNA MMR complex, leading to cell death in wild type cells while MMR-deficient cells are viable. In this screen, several independent mutants with homozygous deletions covering the MMR genes *Msh2* and *Msh6*, and one mutant with a heterozygous deletion at the *Dnmt1* gene were isolated. Mutants were also isolated in which a number of unknown genes were deleted. Joint analysis using genomic and transcriptional arrays discovered that a set of genes, including *Msh2* and *Msh6*, were deleted and silenced in the $6TG^R$ mutants. The most frequently silenced genes are highly likely to be involved in the MMR pathway.

This screening system confirmed that irradiation is efficient in generating genome-wide loss-of-function mutation libraries in the *Blm*-deficient cells, and combined with DNA- and RNA-based array analysis platforms, provides a new approach for phenotype-driven recessive genetic screens in mouse ES cells.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

6TG	6-Thioguanine
BAC	Bacterial Artificial Chromosome
Blm	Bloom syndrome homolog (human) gene
BME	β-Mercaptoethanol
bp	base pairs
Bsd	Blasticidin resistance gene
Bcyd	Blasticidin S HCl
CGH	Comparative Genomic Hybridization
DMSO	Dimethyl Sulfoxide
Dnmt1	DNA methyltransferase (cytosine-5) 1 gene
DSB	Double-Strand Break
ENU	N-ethyl-N-nitrosourea
FBS	Foetal Bovine Serum
FIAU	1-(2'-deoxy-2'-fluoro-β-D-arabinofuranosyl)-5-iodouracil
GPS	Glutamine-Penicillin-Streptomycin
Hprt	Hypoxanthine guanine phosphoribosyl transferase 1
HSVtk	Herpes Simplex Virus type 1 thymidine kinase
IR	Ionizing Radiation
LET	Linear-Energy-Transfer
LOH	Loss of Heterozygosity
Mlh1	mutL homolog 1 (*E. coli*)
Mlh3	mutL homolog 3 (*E coli*)
MMS	Methyl Methane Sulfonate
MMuLV	Moloney Murine Leukemia Virus
MNNG	N-methyl-N'-nitro-N-nitrosoguanidine
Msh2	mutS homolog 2 (*E. coli*)
Msh3	mutS homolog 3 (*E. coli*)
Msh4	mutS homolog 4 (*E. coli*)
Msh5	mutS homolog 5 (*E. coli*)
Msh6	mutS homolog 6 (*E. coli*)
NCBI	National Center for Biotechnology Information
NMD	nonsense-mediated decay
OMIM	Online Mendelian Inheritance in Man
PBS	Phosphate Buffered Saline
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PGK	phophoglycerate kinase
Pms1	postmeiotic segregation increased 1 (*S. cerevisiae*)
Pms2	postmeiotic segregation increased 2 (*S. cerevisiae*)
PTC	premature termination codon
Puro	puromycin N-acetyltransferase gene
RFLP	Restriction Fragment Length Polymorphism
RMCE	Recombination Mediated Cassette Exchange
SDS	Sodium dodecyl sulphate
SIGTR	Sanger Institute Gene Trap Resource
SSB	Single-Strand Break
SSLP	Simple Sequence Length Polymorphism

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.

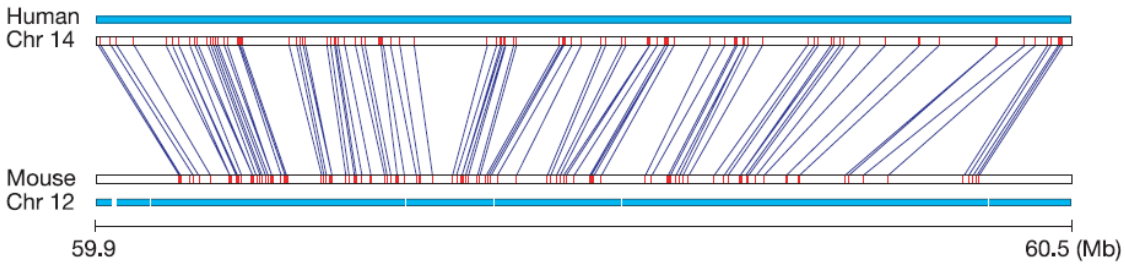


Figure 1-1 Conservation of synteny between human chromosome 14 and mouse chromosome 12 (part)

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)

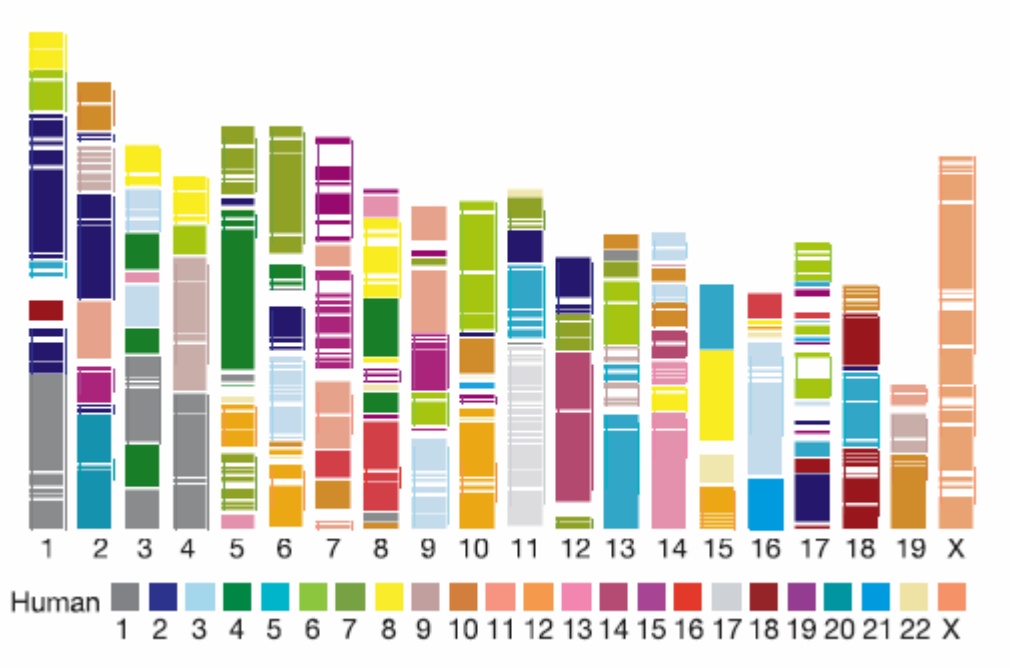


Figure 1-2 Conserved synteny between humans and mice

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).

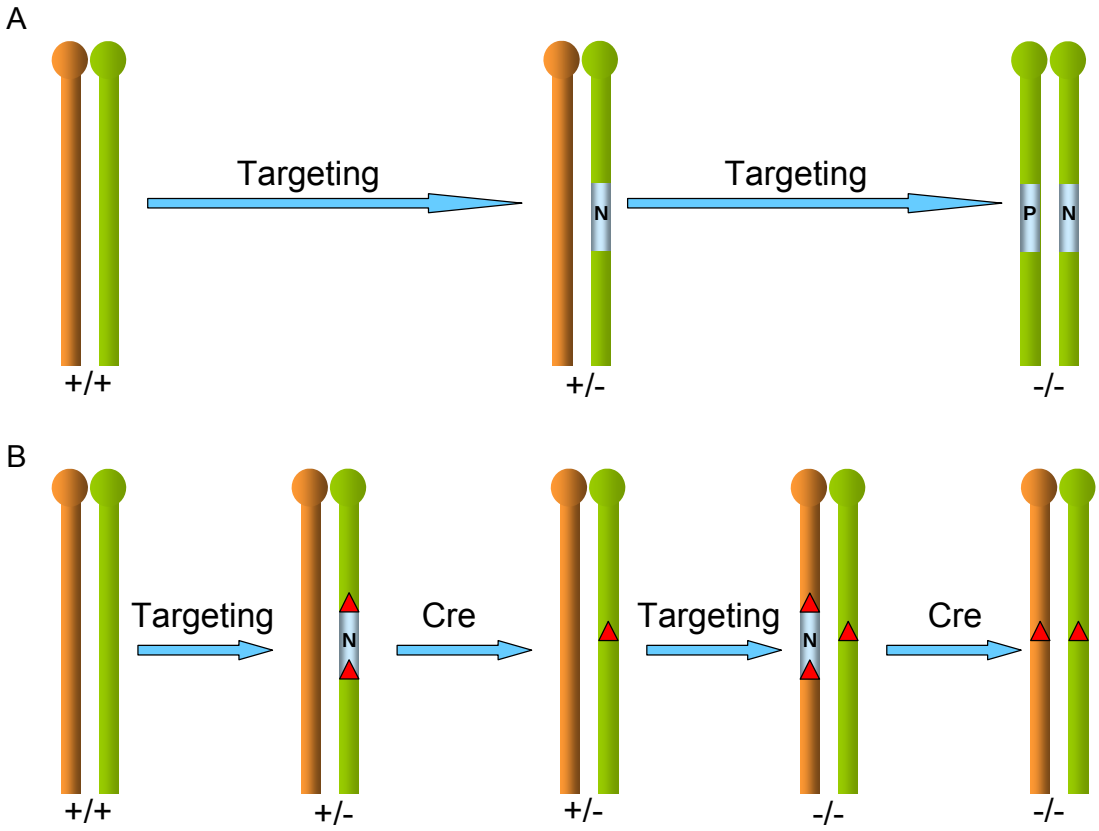


Figure 1-3 Sequential gene targeting

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.

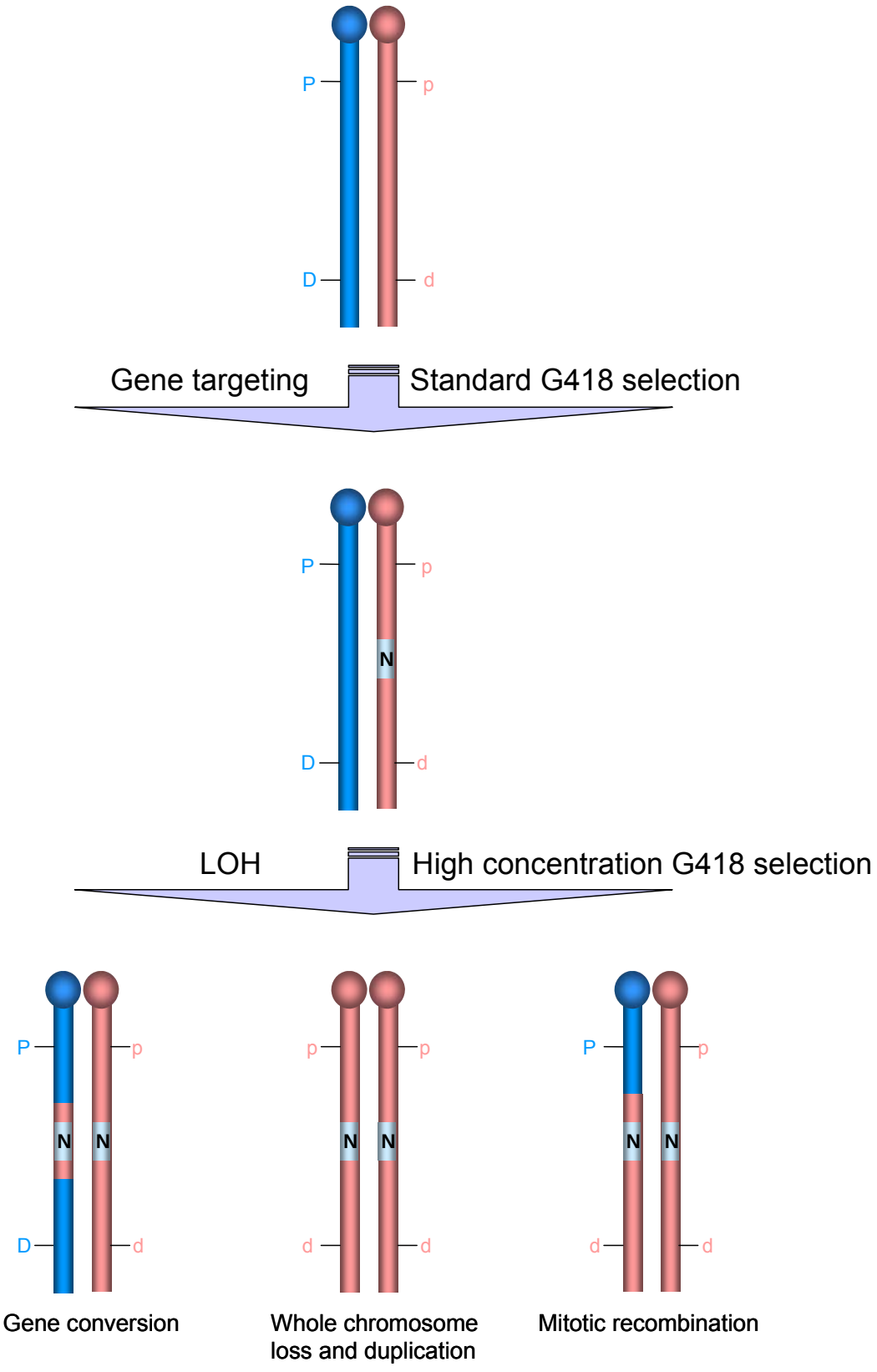


Figure 1-4 **High concentration G418 selection facilitates recovery of homozygous mutants**

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.

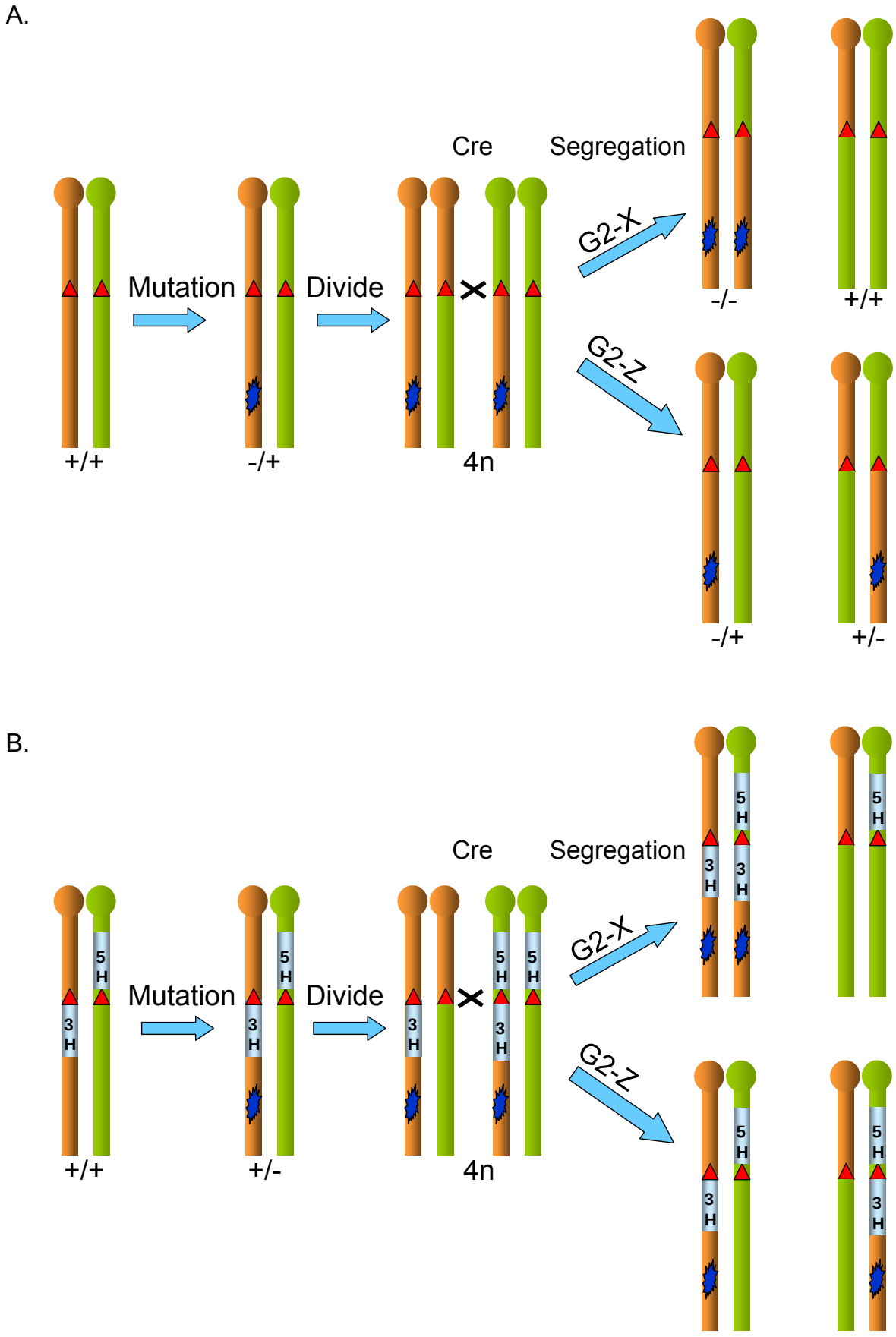


Figure 1-5 **Induced mitotic recombination**

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.

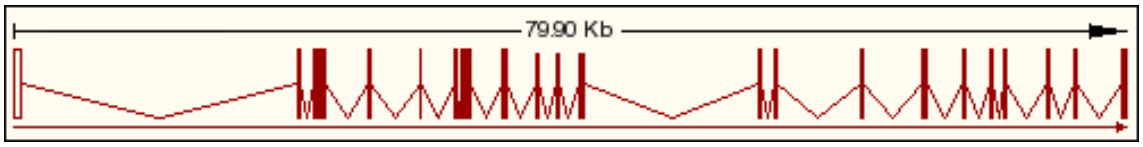


Figure 1-6 Mouse Blm gene transcript

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

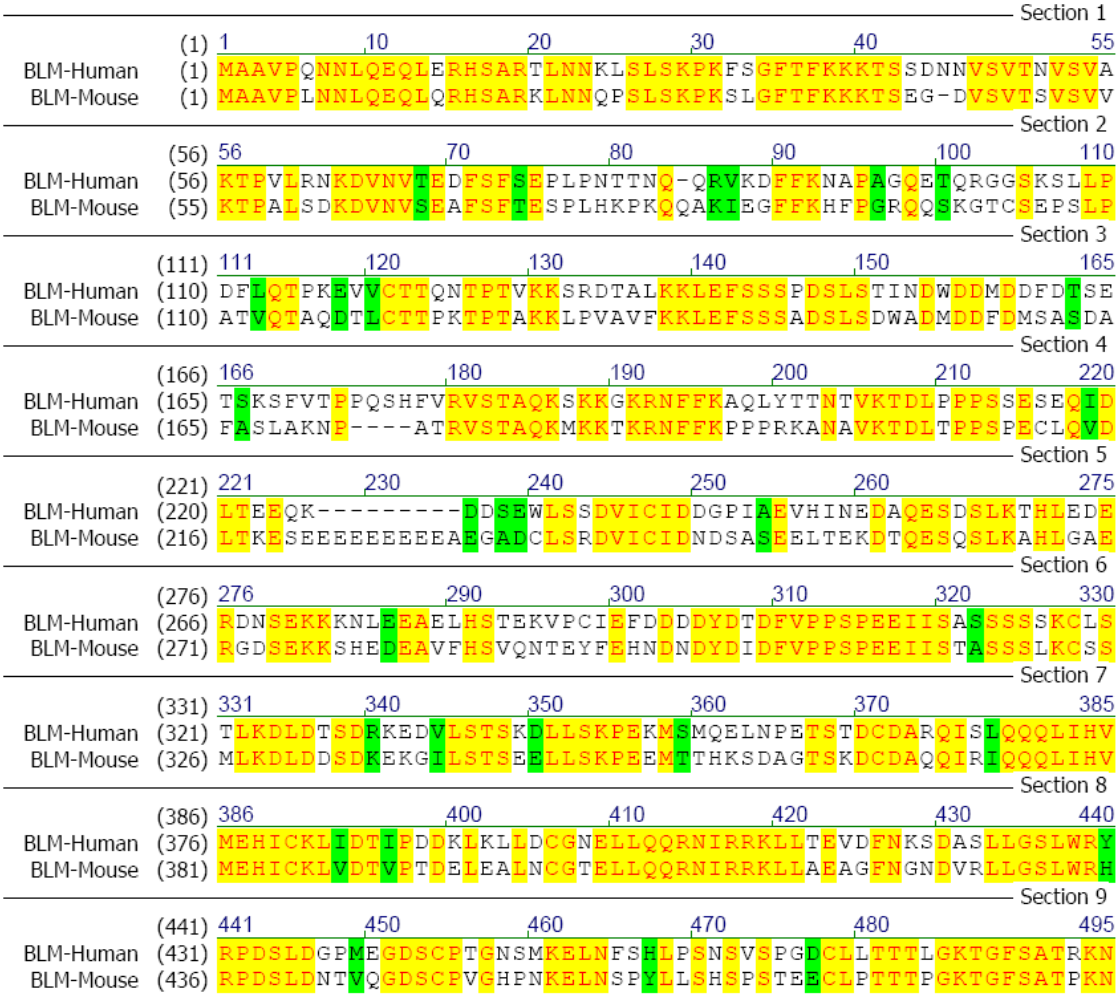


Figure 1-7 The alignment of the BLM protein sequence in humans and mice – 1/3

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.

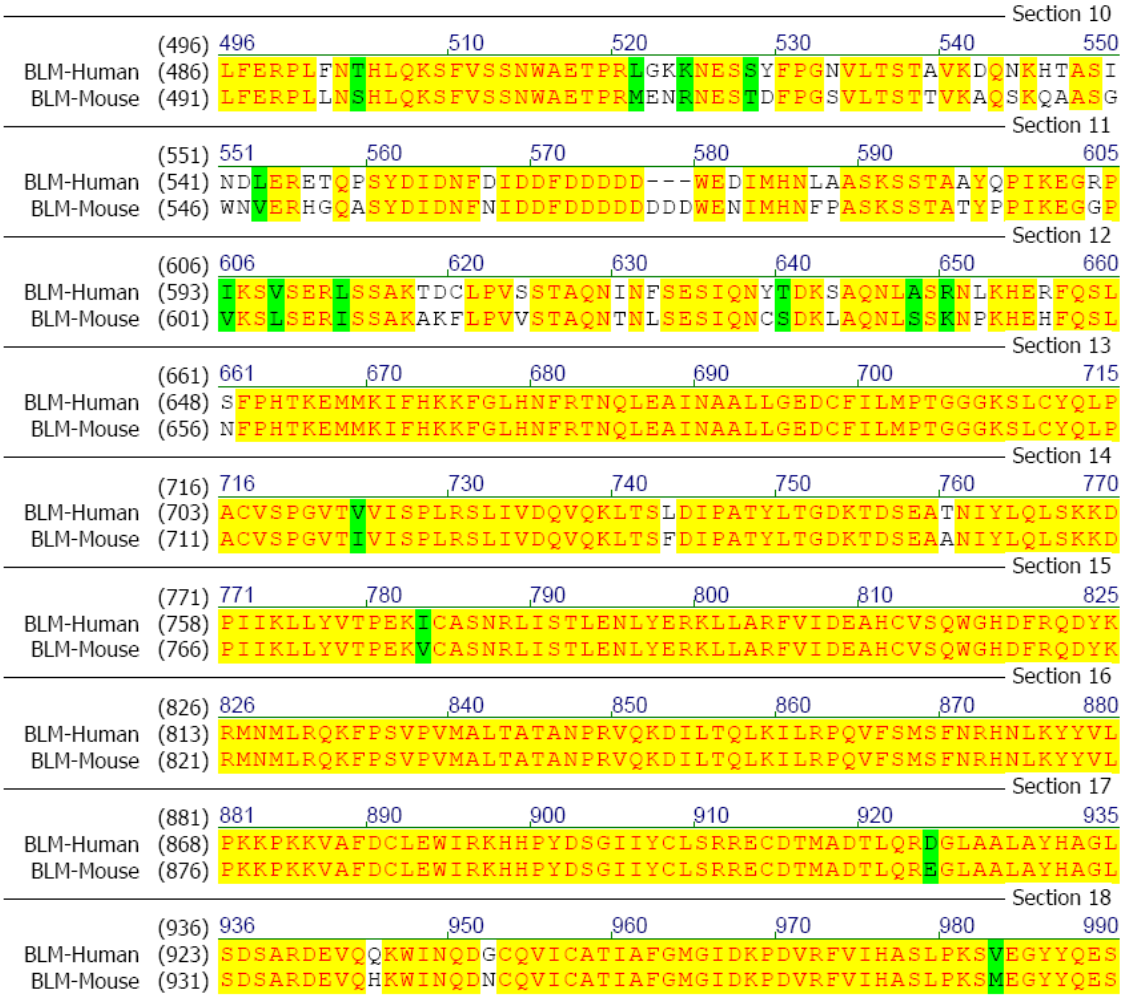


Figure 1-8 The alignment of the BLM protein sequence in humans and mice – 2/3

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.

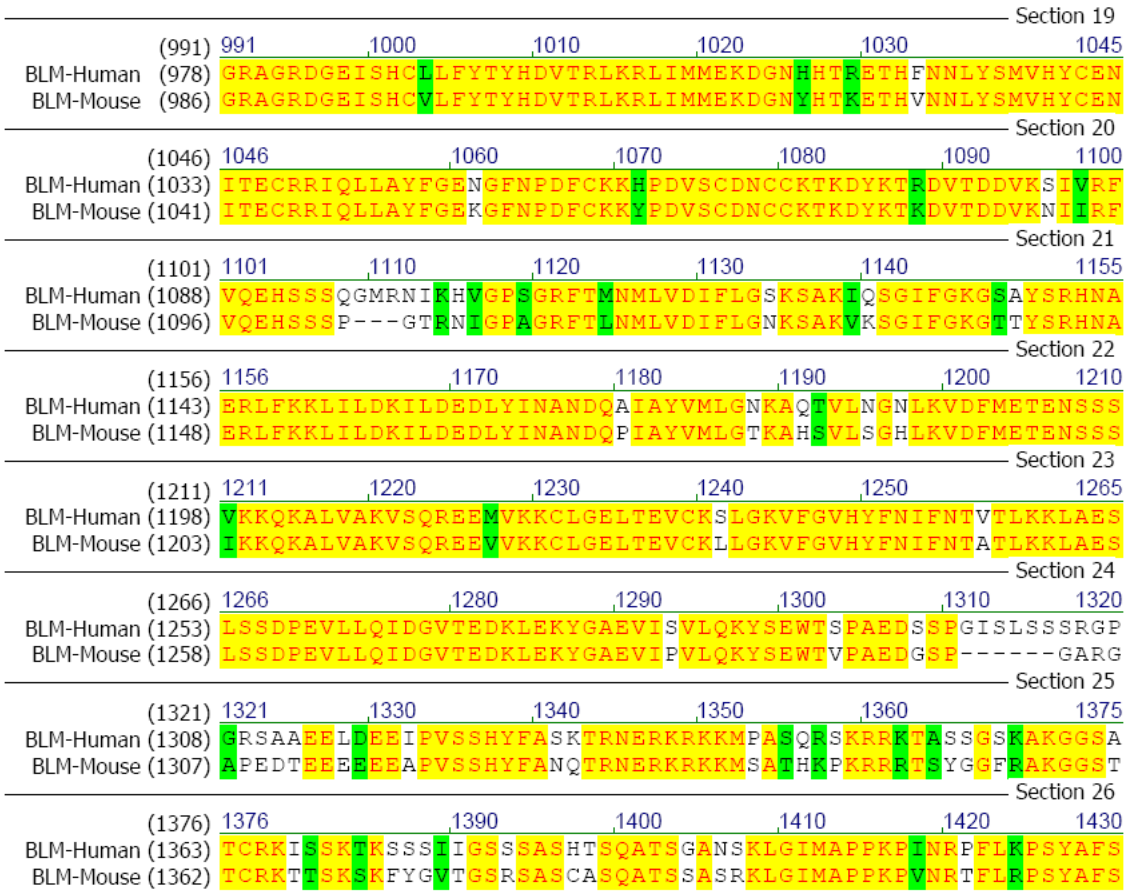


Figure 1-9 The alignment of the BLM protein sequence in humans and mice – 3/3

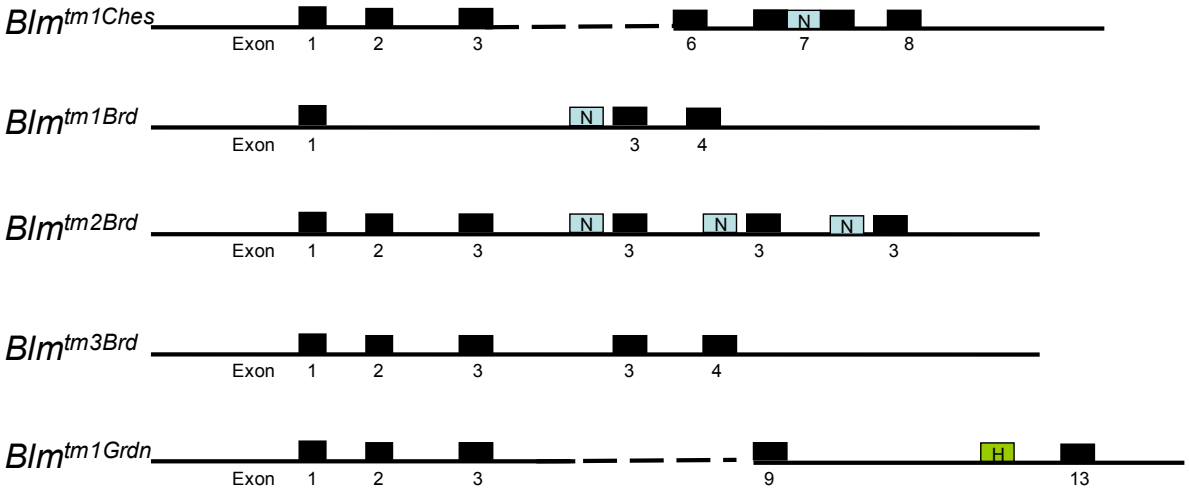
Human and mouse BLM protein sequences were retrieved from the Ensembl genome viewer and aligned. The identity position between them is 75%. Red letters with a yellow background indicate identical residues. Black letters with a green background indicate similar residues. Black letters with a white background 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.

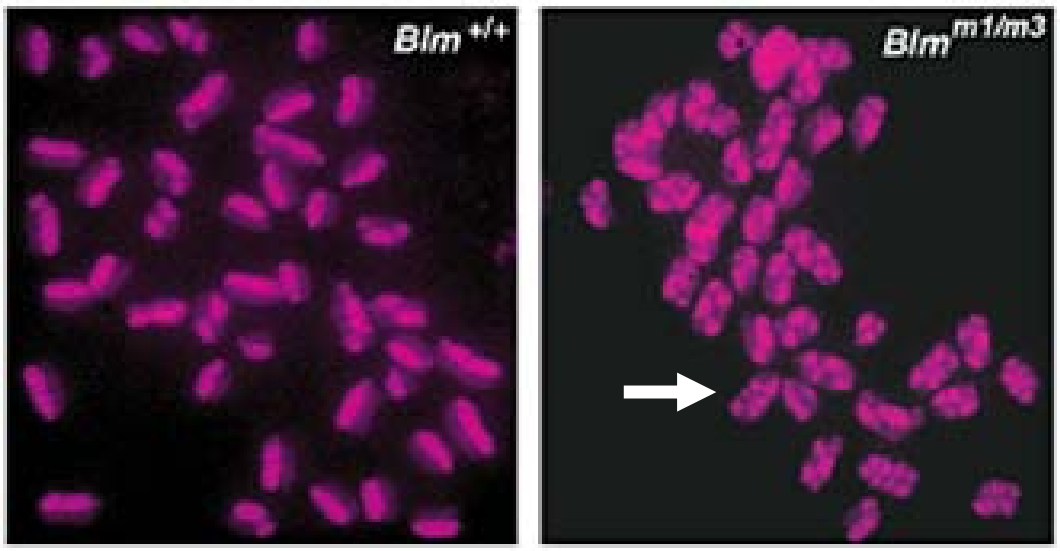


Figure 1-11 ***Blm*** **deficiency affects rates of sister chromatid exchange in ES cells**

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.

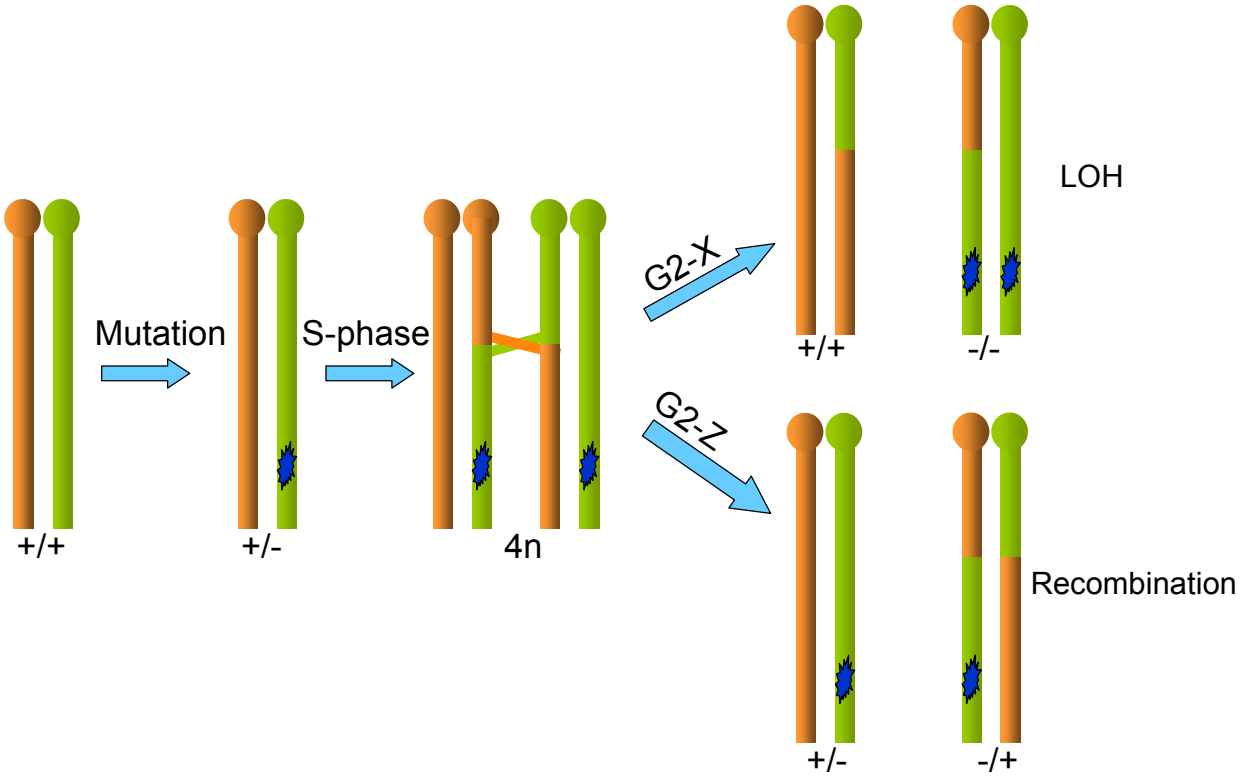


Figure 1-12 Mitotic recombination

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.

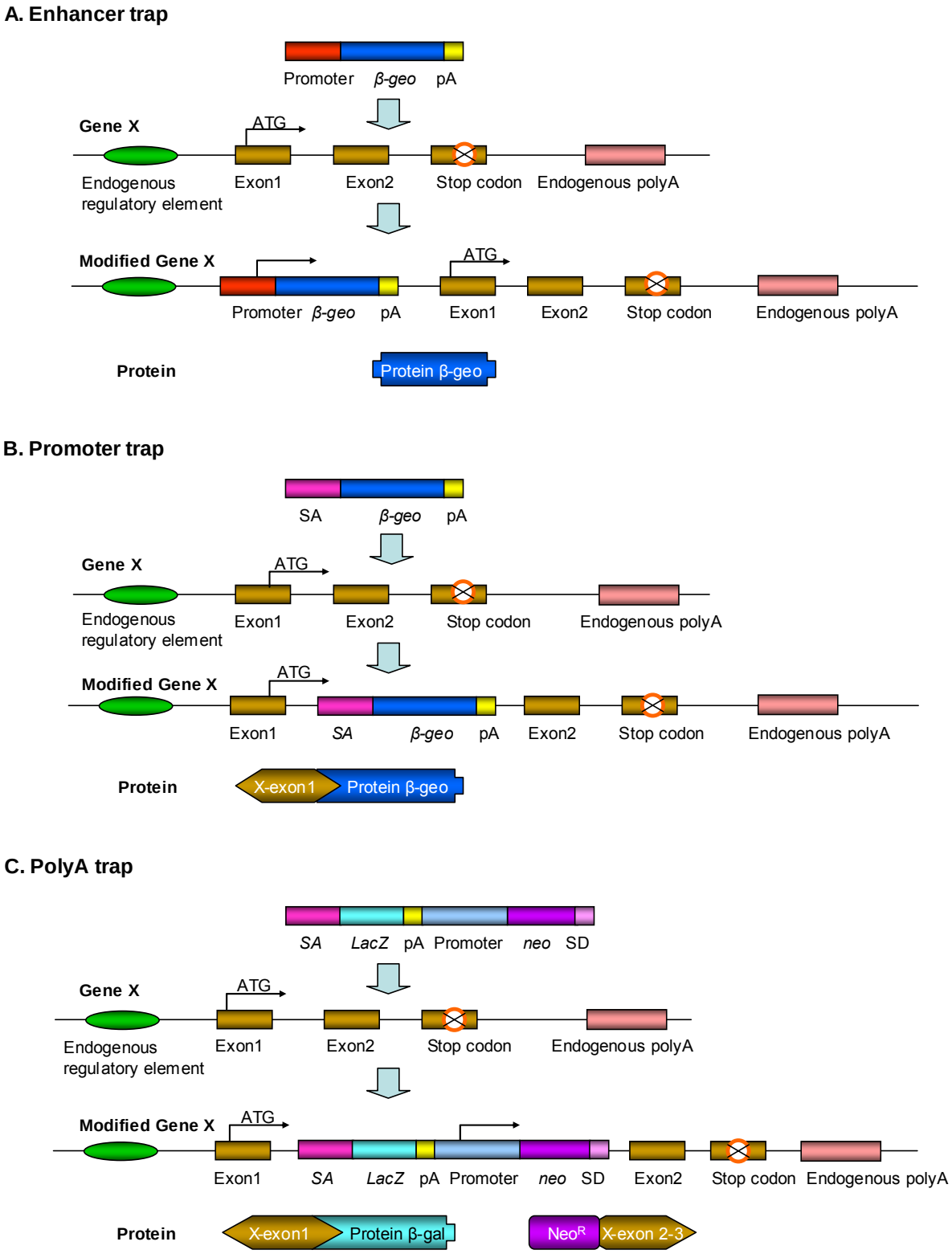


Figure 1-13 Basic gene-trap vectors

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).

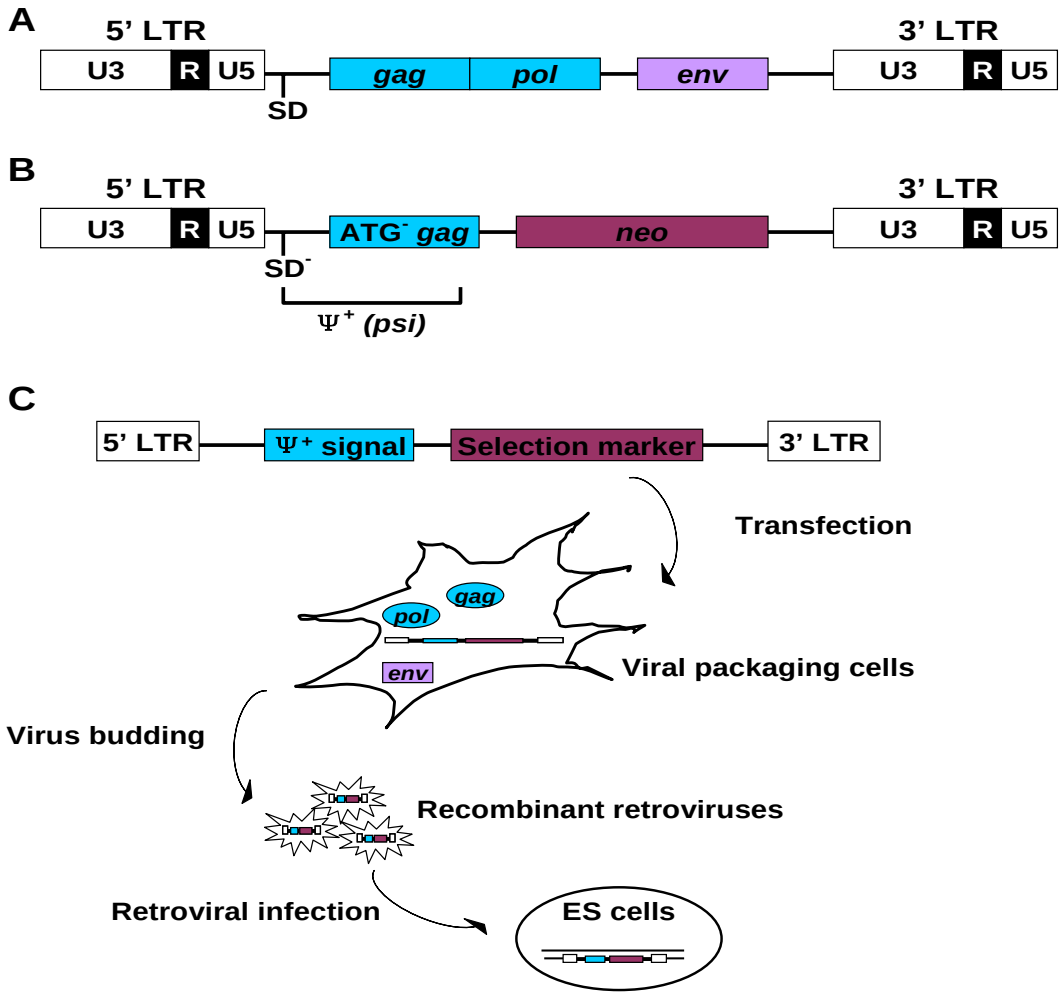


Figure 1-14 Recombinant retroviral vectors and viral production

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.

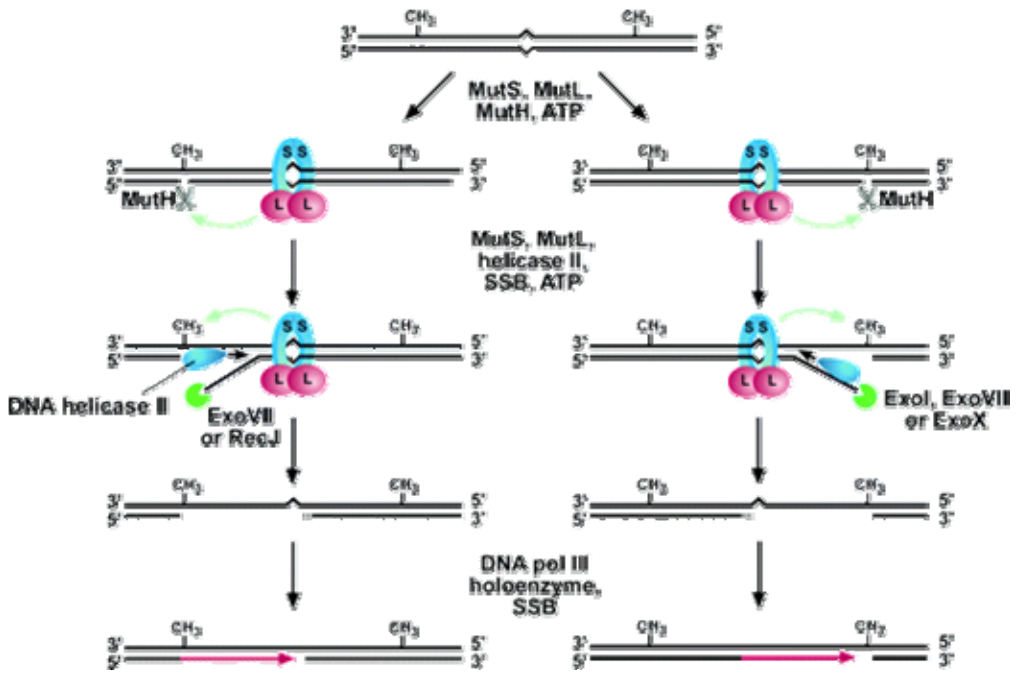


Figure 1-15 ***E. coli*** **mismatch repair system**

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.

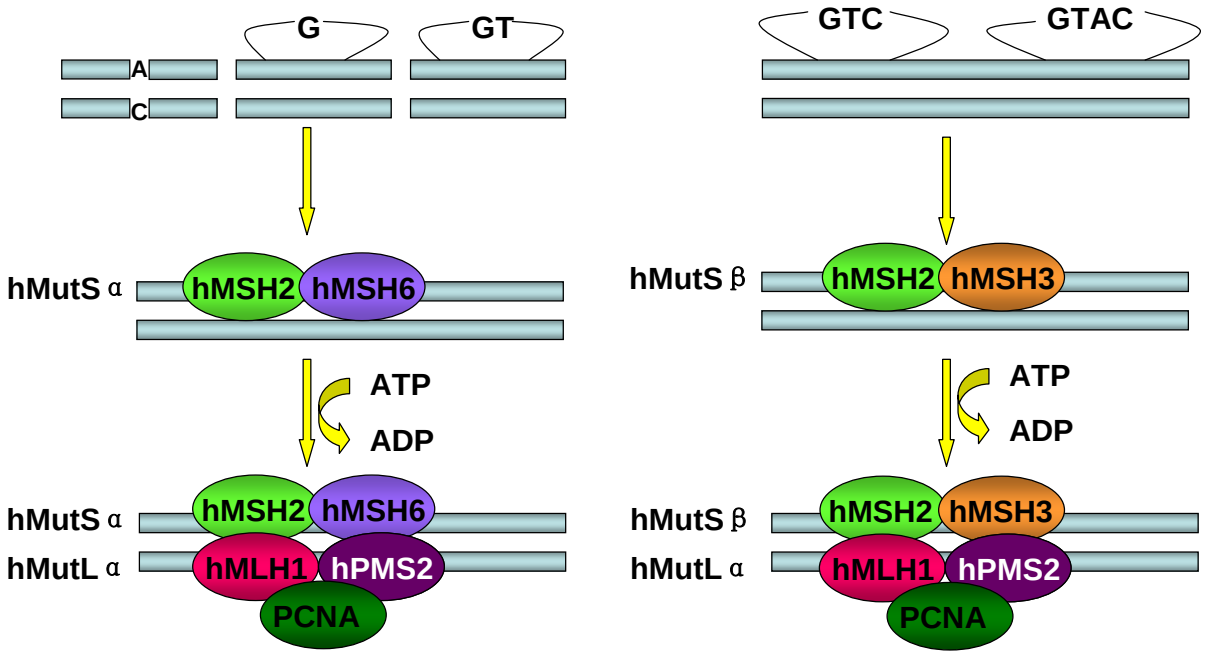


Figure 1-16 Human DNA mismatch recognition preferences

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).

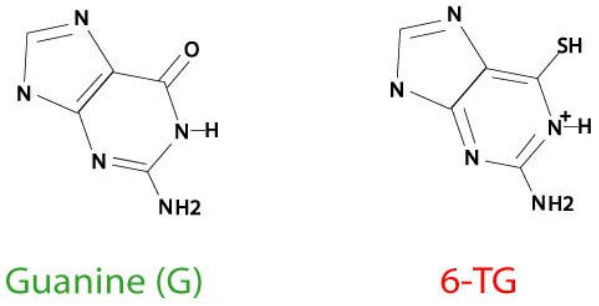


Figure 1-17 The molecules of guanine and its analogue 6-thioguanine (6TG)

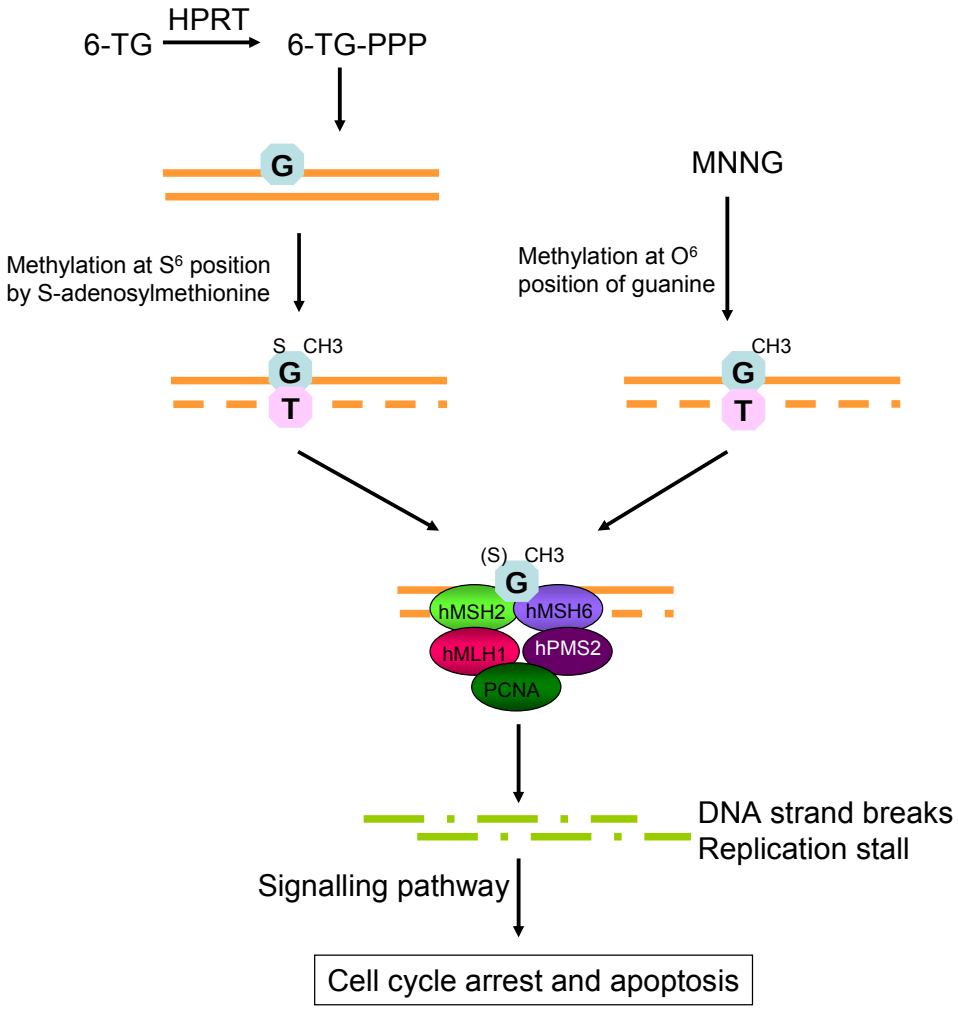


Figure 1-18 Cytotoxicity of MNNG and 6TG

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.

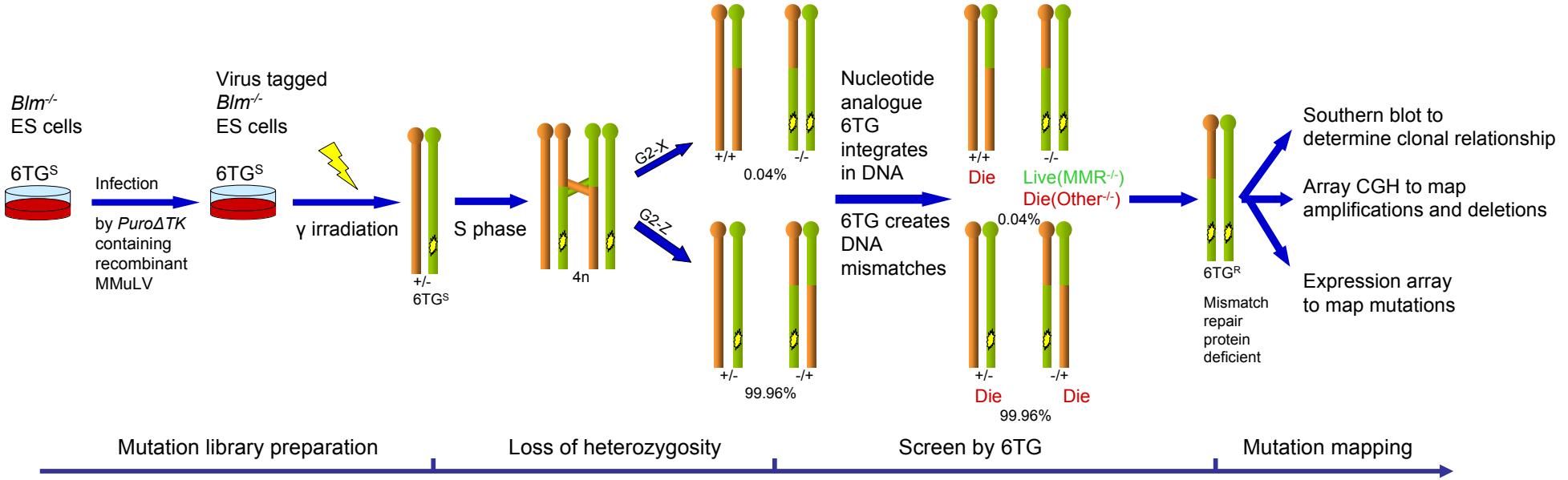


Figure 1-19 A genome-wide recessive screen for DNA mismatch repair genes

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.

CHAPTER 2. MATERIALS AND METHODS

2.1 Cell culture

2.1.1 Mouse ES cell culture

The XY AB1 and AB2.2 cell line were derived from a black agouti 129 Sv embryo described previously (Evans *et al.* 1981). The AB2.2 (129 S7/SvEv$^{Brd\text{-}Hprt^{b\text{-}m2}}$) cell line has an inactivated *Hprt* gene (on the X chromosome) mutated by a recombinant retrovirus (Kuehn *et al.* 1987). These cells and their derivatives were cultured on irradiated (60 Gray) mitotically inactivated monolayer SNL 76/7 STO cells (McMahon *et al.* 1990). The SNL 76/7 cell line is G418^R and expresses leukaemia inhibitory factor (LIF) to maintain ES cells in an undifferentiated state (Williams *et al.* 1988).

ES cells were cultured at 37°C with 5% CO_2 in M15 media (Knockout™ D-MEM (Invitrogen 10829-018) supplemented with 15% foetal bovine serum (Hyclone; CatNo. CH30160.03; LotNo. CPCO285), 2 mM L-glutamine, 50 units/ml penicillin, 40 mg/ml streptomycin and 0.1 mM b-mercaptoethanol (b-ME)). In cell culture, M15 media with supplements was renewed daily if not specified.

2.1.2 *Blm*-deficient ES cell line

The cell line I used in this project is called NGG5-3 (Figure 2-1), which has two targeted *Blm* alleles *Blm*(m1/m3) (Luo *et al.* 2000). The NGG5-3 cell line is derived from AB2.2, an *Hprt*-deficient ES cell line (Kuehn *et al.* 1987). NGG5-3 is cultured on irradiated (60 Gray) SNL 76/7 STO feeder cells. The *Blm*^{tm1Brd} allele contained a *loxP*-flanked *PGK-neo* cassette, which was removed by Dr Ge Guo to generate the *Blm*^{tm4Brd} allele using Cre-*loxP* recombination (Guo 2004).

To enable a screen for genes involved in mismatch repair in *Blm*-deficient cells, the *Hprt* gene product has to be present in the cell line. Thus, an *Hprt* mini-gene was introduced into both autosomal copies of the growth differentiation factor 9 (*Gdf9*) gene loci by sequential gene targeting. This resulted in a null allele of *Gdf9* missing exon 2. As *Blm*-deficient cells have a high rate of LOH, heterozygous alleles may be lost during cell replication. Therefore, two copies of *Hprt* were targeted to *Gdf9*, ensuring that the remaining cells are *Hprt* positive. The *Gdf9* gene has only two exons and is required for sex development in female mice. Importantly, the expression of *Gdf9* should be destroyed and the absence of *Gdf9* does not have any effect on ES cells (Dong *et al.* 1996). The first allele was targeted using Dr Jinwen

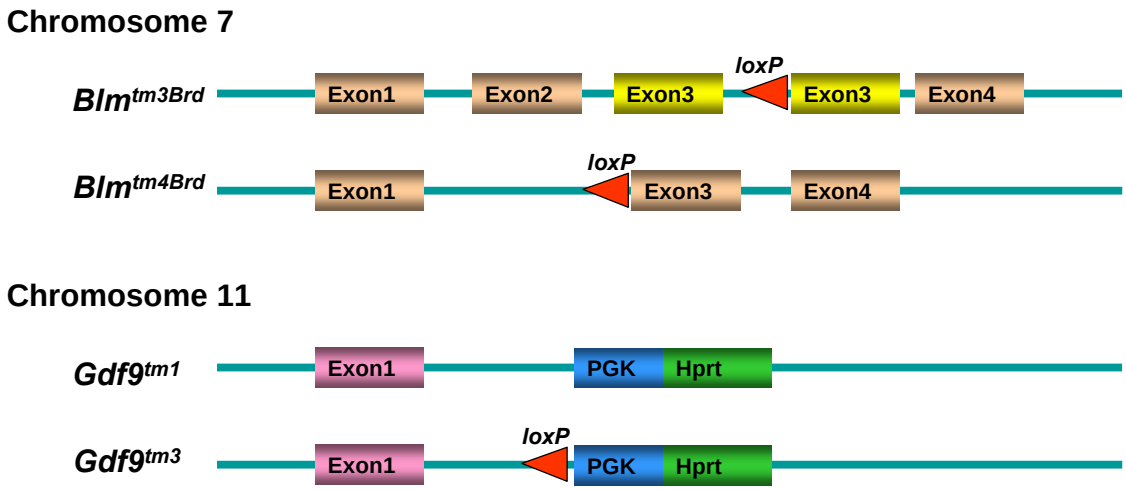


Figure 2-1 Genotype of the *Hprt*$^{+/+}$ *Blm*-deficient ES cell line NGG5-3

In the cell line NGG5-3, both copies of the *Blm* and *Gdf9* genes are null. The *Blm*^{tm3Brd} has all of the endogenous *Blm* exons plus an extra copy of exon 3; the *Blm*^{tm4Brd} allele lacks exon 2. Both mutations cause a frame shift, thus they are believed to be null alleles for *Blm*. Both *Gdf9*^{tm1} and *Gdf9*^{tm3} lack exon 2 of *Gdf9*. The two copies of *PGK-Hprt* mini-genes cause the NGG5-3 cell line to be sensitive to 6TG and resistant to HAT.

Dong's vector. The second targeting vector was modified by Dr Ge Guo so that it had a *loxP*-flanked *PGK-neo* cassette inserted in front of *PGK-Hprt*. The *Neo* selection marker was removed by Cre after the second targeting resulting in the *Gdf9*^{tm3} allele (Guo 2004).

2.1.3 *Dnmt1*-deficient ES cell line

A *Dnmt1*-deficient ES cell line was derived from the previously established *Dnmt1* gene knockout cell line (Guo *et al.* 2004). From the reported *Dnmt1*$^{tm1/tm2}$ cell (Guo *et al.* 2004), a cell line was generated by Dr Wei Wang by removing the neomycin cassette through Cre-*loxP* recombination, which left two alleles with the same structure (*Dnmt1*$^{tm2/tm2}$, Figure 2-2). These have a *loxP* site in both alleles and exons 2–4 deleted. The deletion of exons 2–4 removes a 359-bp coding region, resulting in a frame shift mutation, which can be expected to be a null mutation (Guo 2004). I have targeted one allele of the *Gdf9* locus in this cell line with an *Hprt* mini-gene using the same targeting vector, which produced *Hprt*-proficiency in the *Blm*-deficient cell line NGG5-3 (Dong *et al.* 1996).

2.1.4 Chemicals and media used for ES cell culture

M15 medium: 500 mL Knockout™ D-MEM (Invitrogen 10829-018) is supplemented with 15% foetal bovine serum (Hyclone; CatNo. CH30160.03; LotNo. CPCO285), 2 mM L-Glutamine, 50 units/ml Penicillin, 40 mg/ml Streptomycin and 0.1 mM b-mercaptoethanol (b-ME). Storage: at 4°C.

Foetal bovine serum: Hyclone; CatNo. CH30160.03; LotNo. CPCO285. Storage: at -20°C.

Blasticidin: Blasticidin S HCl (Invitrogen, Cat. No. R210-01), 1000× stock (5 mg/ml) was made in Phosphate Buffered Saline (PBS). After mixing, the 1000× stock solution was sterilized by filtering through a 0.2 µm syringe filter. Storage: at –20°C.

FIAU: 1-(2'-deoxy-2'-fluoro-β-D-arabinofuranosyl)-5-iodouracil, 1000× stock
(200 µM) was made in PBS and 5 M NaOH was added dropwise until it was dissolved. After mixing, the 1000× stock solution was filter sterilized through a
0.2 µm syringe filter. Storage: at –20°C.

G418: Geneticin (Invitrogen, Cat. No. 10131) was bought as a sterile stock solution containing 50 mg/ml active ingredient. Storage: at –20°C.

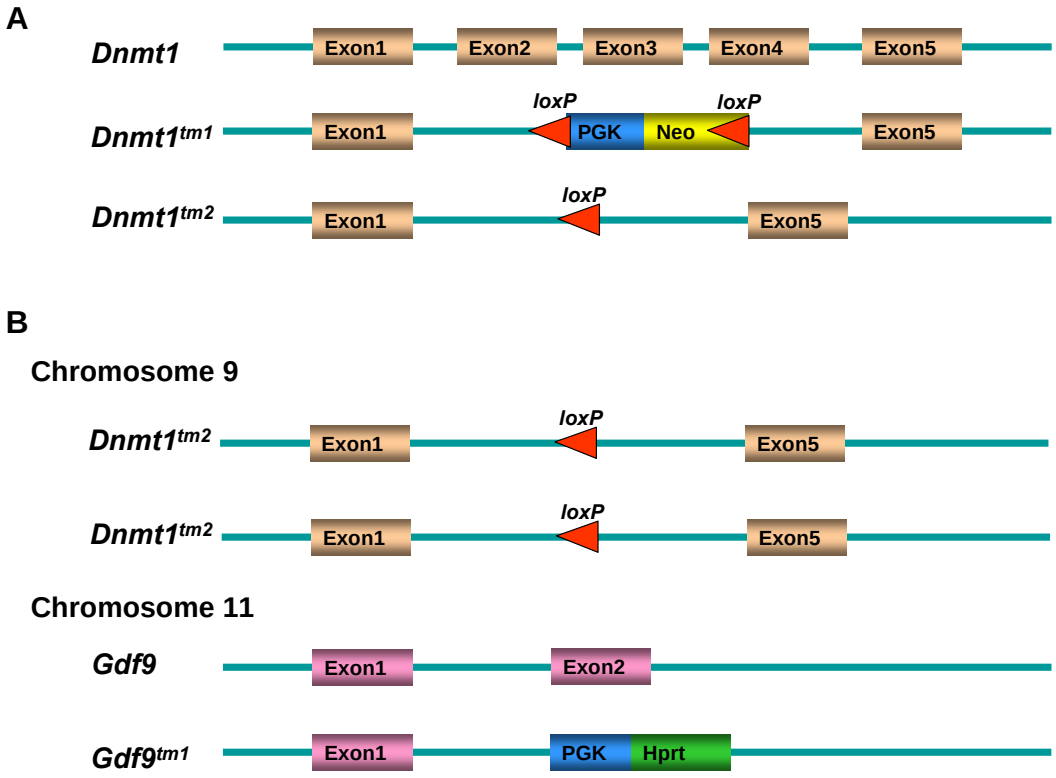


Figure 2-2 Structure of the *Dnmt1* locus in the targeted cell lines

A. Targeting strategy of the *Dnmt1* gene A *loxP*-flanked *PGK-neo* cassette replaces exons 2–4 of the *Dnmt1* gene to generate allele tm1 through homologous recombination. The allele tm2 is generated from tm1 through Cre-*loxP* recombination. **B**. Schematic structure of the *Dnmt1* locus and the *Gdf9* locus in the *Dnmt1*-deficient *Hprt*-proficient cell line.

Puromycin: ($C_{22}H_{29}N_7O_5$.2HCL, Sigma) 1000× stock (3 mg/ml) was made in MiliQ water. After mixing, the 1000× stock solution was filter sterilized through a 0.2 µm syringe filter. Storage: at -20°C.

HAT: a liquid sterile mixture of sodium Hypoxanthine, Aminopterin and Thymidine (50×, Invitrogen, Cat. No. 21060-017). HAT supplement contains 5 mM Hypoxanthine, 20 µM Aminopterin and 0.8 mM Thymidine. Storage: -5°C to -20°C.

HT: a sterile mixture of sodium Hypoxanthine and Thymidine (100×, Invitrogen, Cat. No.11067030). HT supplement contains 10 mM Hypoxanthine and 1.6 mM Thymidine.

6TG: (2-amino-6-mercaptopurine (Sigma)) 5 mM stock was made in PBS and 5 M NaOH was added dropwise until it dissolved. Storage: -20 °C.

PBS (Phosphate Buffered Saline): PBS is prepared in 10 litre quantities, aliquoted into sterile disposable bottles, and stored at room temperature. Mix NaCl (80.0 g), KCl (2.0 g), $Na_2HPO_4.7H_2O$ (10.72 g) and KH_2PO_4 (2.0 g) in 8 Litre of Milli-Q water, then bring the total volume to 10 L. Adjust the PH to 7.2 with a saturated solution of $Na_2HPO_4.7H_2O$. Add phenol red until a peach colour is achieved.

GPS (Glutamine-Penicillin-Streptomycin) 100×Stock Solution: Thaw the 100 ml bottle(s) of 200 mM Glutamine (filtered to be sterile, stored at -20°C). From each bottle, aseptically remove 10 ml of the 200 mM Glutamine to a 15 ml tube. Add the following to the tube and dissolve by mixing: Penicillin (300.0 mg), Streptomycin (500.0 mg). After mixing, filter sterilizes the GPS solution through a 0.2 µm syringe filter back into the original 100 ml bottle of Glutamine. Mix thoroughly. The resulting 100× solution is: 200 mM L-Glutamine; 4,950 U/ml Penicillin; and 500 mg/ml Streptomycin. Aliquot from the original bottle into 25 ml Nunc Universals, label with "100× GPS" and the date, and store at 4°C.

Trypsin Stock solution: This is prepared in 5 L quantities, filter-sterilized, aliquoted into 50 ml tubes, and stored -20°C. Mix the following in Milli-Q water, then bring the total volume to 5 L: NaCl (35.00 g), D-Glucose (5.00 g), $Na_2HPO_4.7H_2O$ (0.90 g), KCl (1.85 g), KH_2PO_4 (1.20 g), EDTA (2.00 g), Trypsin (12.50 g, Invitrogen, Cat. No. 840-7250IL) and Tris Base (15.00 g); these ingredients generally take 30-60 minutes to dissolve thoroughly. Adjust the pH from its initial 8.71 to pH 7.6 with HCl. Add phenol red until a pink colour is achieved.

BME (β-Mercaptoethanol) Stock Solution (100×, 10^{-2} M): Add 72 µL of 14 M β -Mercaptoethanol to 100 ml of PBS and mix well. Filter-sterilize it using a 0.22 µm filter and aseptically decant into 50 ml tubes. Storage: 4 °C.

Gelatin (0.1%): The gelatin is used to treat (coat) tissue culture plasticware so the cells will adhere better to the plate surface. It is prepared in 4 litre quantities, sterilized by autoclaving, aliquoted into sterile disposable bottles, and stored at room temperature. To 4 L of Milli-Q water, add 4.0 g of Gelatin (Sigma; from porcine skin, 300-bloom; cell culture tested). Add phenol red until a bright yellow colour is achieved; swirl to evenly distribute. Autoclave on liquid cycle.

Cell staining buffer: methylene blue 2%, 70% ethanol.

2.1.5 Passaging ES cells

When ES cells were 70-90% confluent on feeder plates, media was replaced 2-4 hours before passaging them. ES cell were washed twice with PBS and trypsin solution added to the plate (50 uL for a well of 96-well plate; 150 uL for a well of 24-well plate; 0.3 mL for a well of 6-well plate; 2 mL for a 10 cm plate). After incubation at 37°C for 10-15 minutes, an appropriate volume (1–5 times volume of trypsin solution used) of M15 was added. Cells were pipetted up and down vigorously to breakup the clumps and re-plate at ratio of 1:3 or 1:4. Cells usually took 2–4 days to reach confluence.

2.1.6 Freezing ES cells

Cells are best frozen at 70–80% confluence. Cells were reefed by M15 media 2–4 hours before they were frozen. Briefly, ES cells were suspended by trypsin according to passaging protocol. Suspended cells were harvested by centrifuge (1000rpm 3 min) and resuspended in one volume of M15 and an equal volume of freezing medium. Usually cells were frozen at density about 10^7 per mL. Freezing media has 60% Knockout™ D-MEM (Invitrogen 10829-018), 20% foetal bovine serum (FBS) and 20% DMSO (Dimethyl Sulfoxide). Cells can be frozen in 96-well / 24-well plates (wells were covered by 0.2 µm filtered mineral oil) or Cryovials at -80°C. If in Cryovials, the temperature decrease is controlled to approximate 1°C per minute. Then Cryovials can be transferred into liquid nitrogen freezer the second day.

2.1.7 Electroporation of DNA into ES cells

DNA used for electroporation was usually prepared by the standard alkaline lysis mini-prep method or by QIAGEN Plasmid Maxi Kit (Cat. No.12163). DNA used for gene-targeting was linearized by enzyme digestion with an appropriate enzyme (NEB) under the conditions recommended by the manufacturers. Before electroporation, DNA was precipitated by ethanol and air-dried in a tissue culture (TC) hood. 5–20 ug of DNA (0.5–1.0 μg/μL) was used in appropriate experiments.

Mouse ES cell electroporation was conducted as described before (Ramirez-Solis *et al.* 1993). Briefly, 70–90% confluent ES cells were fed 2–4 hours before media was removed. The cells were washed with PBS twice and 2 mL of trypsin was applied to a 90 mm plate, which was incubated at 37°C for 15 minutes. Then 8 mL M15 media was added and the cells were pipetted repeatedly to make a suspension of single cells. This mixture was centrifuged at 1000 rpm for 3 minutes and the supernatant was removed. Cells were suspended in PBS at the concentration of 1×10^7 cells/mL. Cells then were transferred into 0.4 cm gap cuvette (Bia-Rad) after mixing with 5–20 ug DNA. The gene pulser (Bio-Rad) was configured at 230 V, 500 μF. After electroporation, cells were plated onto 10 cm feeder plate. Selective drugs were usually applied with M15 media one day after electroporation.

2.1.8 Picking ES cell colonies

Colonies need 8–10 days to grow up from a single cell. They were washed twice by PBS immediately before picking. With a P20 Pipetman®, individual colonies were picked into round bottom wells in 96-well format, which were filled with 50 uL trypsin. Colonies were incubated at 37°C for 10–15 minutes then cell clumps were broken up into single cells by pipetting in 50 uL M15 medium and transferred into 96-well flat bottom feeder plates and cultured in normal ES cell culture conditions.

2.1.9 Production of recombinant retrovirus

The *PuroΔtk* (Chen, Y. T. *et al.* 2000) containing recombinant MMuLV retrovirus was produced by the B4-5 virus producing cell line, a gift from Wei Wang (Wang and Bradley 2007). *PuroΔtk* is a bifunctional fusion protein between puromycin N-acetyltransferase (Puro) and a truncated Herpes Simplex Virus type 1 thymidine kinase (*Δtk*). Murine embryonic stem (ES) cells transfected with *PuroΔtk* become resistant to puromycin and sensitive to 1-(-2-deoxy-2-fluoro-1-beta-D-arabino-furanosyl)-5-iodouracil (FIAU) (Chen, Y. T. *et al.* 2000). B4-5 cells were cultured in M15 ES cell media until they are >80% confluent.

Supernatant was collected after 24 hours (containing MMuLV recombinant virus) was filtered (0.2 μm) before it was used to infect ES cells. Filtered MMuLV containing supernatant can be stored at -80°C. Virus was added in the media then infected cells were selected in puromycin media. Virus titer is calculated as puromycin resistant colonies generated per mL of virus containing supernatant used. The virus titer will vary when different numbers of cells are used for the infection. About 600 infected clones were obtained when 60 μL virus containing supernatant was used to infect 3×10^6 cells. Thus, the virus titer is 10^4 CFU/mL.

2.1.10 Recombination Mediated Cassette Exchange

Recombination mediated cassette exchange (RMCE) is a technology, in which a DNA fragment (cassette) can be introduced into a host genome (exchange) at a specific site through Cre-*loxP* recombination (Figure 2-3). Bacterial Artificial Chromosomes (BACs) can be used for this purpose (Liu *et al.* 2006). Considering the average size of many BACs (~200 kb), it is possible to carry most genes. Transforming a BAC into a mutant genome by RMCE provides a practical way to complement a mutant gene's function.

To introduce single copy BAC DNA into genome, an efficient acceptor site and a selection scheme are needed. The discovery of heterospecific *lox* sites makes it possible to transform a DNA fragment into a defined location in the mouse genome with a specific orientation (Baer *et al.* 2001). A single positive selection can be used to select for BAC insertions, but this does not rule out random integrations of the BAC. RMCE will normally give single copy insertions. Thus, multicopy of the transformed exogenous gene is avoided.

Heterospecific *lox* sites targeted to the X-linked hypoxanthine phosphoribosyltransferase (*Hprt*) locus provide a positive-negative selection scheme. The *Hprt* gene was the first locus mutated in ES cells (Kuehn *et al.* 1987) and it was found to exert a minimal influence on transgene expression (Heaney *et al.* 2004). *Hprt* has also been used as a defined locus to accept a single copy transgene by homologous recombination (Bronson *et al.* 1996; Heaney *et al.* 2004). When an exogenous gene is transformed into *Hprt* locus and destroys *Hprt* gene transcription, cells are resistant to 6TG. The combination of this negative selection for loss of *Hprt* function and positive seletion for the exogenous fragment provide a straightforward selection system to isolate single copy BAC insertions at a specific site. Using RMCE, functional copies of human α Globin and human cardiac sodium channels were used to replace the mouse genes (Liu *et al.* 2006; Wallace *et al.* 2007).

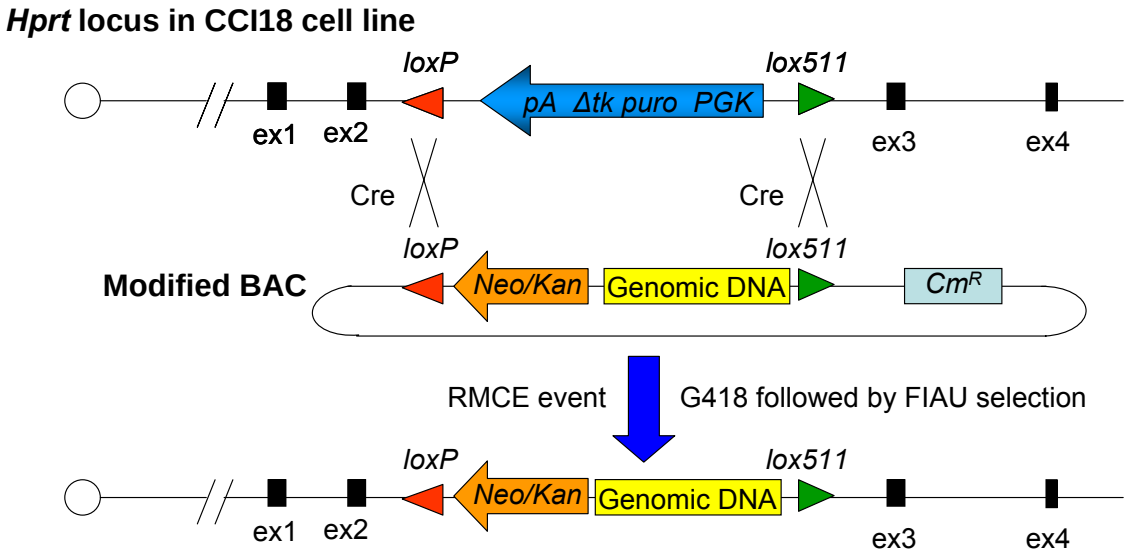


Figure 2-3 The principle of Recombination Mediated Cassette Exchange

Genomic DNA insert from mouse BAC libraries, such as RPCI-23, 24 can be inserted into *Hprt* locus of CCI 18 mouse ES cell line through site directed recombination. BAC: Bacteria Artificial Chromosome. Cm^R: Chloramphenicol antibiotic resistance gene. *Neo/Kan*: *Neo*mycin/Kanamycin resistance gene. *PGK*: phophoglycerate kinase promoter. *PuroΔtk*: a positive-negative selection marker, which is resistant to puromycin and sensitive to FIAU (Chen, Y. T. *et al.* 2000). Ex: Exon.

Dr Prosser in Prof. Bradley's laboratory introduced *PGK-PuroΔtk-pA* cassette flanked by *loxP* and *lox511* (Hoess *et al.* 1986) sites into *Hprt* intron 2 of AB1 mouse ES cell line (McMahon *et al.* 1990). The resulting unpublished cell line CCI18 has the capacity to accept BAC DNA into *Hprt* locus through Cre mediated site directed recombination – recombination mediated cassette exchange (Prosser 2007 unpublished).

Most BAC libraries can be adapted in a single overnight step for RMCE. Mouse (C57BL/6J) BAC libraries RPCI-23, 24 have *loxP* and *lox511* sites, located at either side of genomic DNA insert. Before BACs are electroporated into ES cells, a neomycin/kanamycin resistance gene (*Neo/Kan*) is inserted into the BAC by recombineering. The locus used for this purpose, the *SacBII* gene, is located between the two recombination sites: *loxP* and *Lox511*, thus the *Neo/Kan* gene will be transferred into the *Hprt* locus of CCI18 cell line. After RMCE, positive selection (G418) for the presence of *neomycin* gene and a negative selection (FIAU) for the loss of *tk* transcript can isolate single copy BAC integration events at 100% efficiency. If RMCE is used to introduce a BAC with a functional gene into a mutant cell with a homozygous mutation at the same gene, the incoming gene will complement the mutant gene function, leads to a normal phenotype.

2.2 Gamma Irradiation

Cells were irradiated in Gammacell® 1000 Elite (MDS Nordion). This irradiation generates gamma rays from Caesium^{137} (Cs^{137}). Due to the half life of the isotope, the actual dose decreases with time. However, the half life of Cs^{137} is 30.07 years so the decrease in dose rate is minimal. Doses of ionizing radiation are usually measured by the unit Gray (Gy). 1 Gy corresponds to 1 J/kg of absorbed energy. The conversion factor with the unit "rad" used in the older scientific literature is 1 Gy = 100 rads = 100 cG. The actual dose received by cells using Gammacell® 1000 Elite is at a rate of approximately 12 cG per second.

2.3 DNA methods

2.3.1 Solutions used in molecular experiments

TE: 10 mM Tris-Cl, pH 7.5, 1 mM EDTA in dH_2O.

T0.1E: 10 mM Tris-Cl, pH 7.5, 0.1 mM EDTA in dH_2O.

Cell lysis buffer: 100 mM NaCl, 50 mM Tris pH 7.5, 10 mM EDTA (pH 8.0), 0.5% SDS.

Before use, add Proteinase K to 0.5–1.0 mg/mL.

10× TAE solution: Tris 400 mM, EDTA-Na_2-salt 10 mM and Acetic acid 200 mM, are resolved in dH_2O, adjusted to pH 8.3, autoclave sterilized.

Depurination buffer: 0.25 M HCl in dH_2O.

Denaturation buffer: 0.4 M NaOH, 1 M NaCl in dH_2O.

Neutralization buffer: 0.5 M Tris-HCl (PH 7.4), 1 M NaCl.

Southern blot Hybridization Buffer: PerfectHyb™Plus Hybridization Buffer, Sigma, CatNo.H7033.

Southern blot wash buffer: 0.2×SSC, 0.1% SDS.

2.3.2 PCR

PCR reactions were carried out using Platinum® Taq DNA polymerase (Invitrogen Cat.No.10966-018) if not specified. The reaction mixture were prepared by 5μL PCR buffer (10×, 200 mM Tris-HCl (pH 8.4), 500 mM KCl), 1.5 μL $MgCl_2$ (original 50mM stored at -20°C, final 1.5 mM), 0.4 μL dNTP mixture (original 25 mM stored at -20°C, PH8.0 adjusted by 2N NaOH, final 0.2 mM), 0.5-1 μL Platinum® Taq DNA polymerase, 5–100 ng DNA template, 0.5 μL each of upper and lower primers (original 100 μM stored in water at -20°C, final 1 μM) and appropriate volume of Mili-Q water to bring a final volume to 50 uL.

Reactions were usually designed as: 94°C 2 minutes to denature templates, 25–40 cycles of [94°C 30sec; 60–70°C annealing 45sec; 72°C extension time], followed by 72°C 2 minutes, 4°C 5 minutes. Annealing temperature depends on the primer design. Extension time ensures 1minute for every 1 kb target fragment.

2.3.3 Primers

Lower case usually indicates designed restriction enzyme sites. But in primers zk198.3, 198.4, 199.1 and 199.2, lower case indicates *lox* sites.
zk102.1: 5’- accgCTCGAGTTTTATGGACAG – 3’
zk107: 5’ - CGgaattcAATTCAGAAGAACT - 3’

zk111.1: 5’ – TCCcccgggCCTTTTGTAACCTTTCATAT – 3’

zk111.2: 5’ – GactagtAAGGTGGCATACTTCTCACC – 3’

zk112.1: 5’ – GGggtaccTTTATATCTCCCCGAACCCT – 3’

zk112.2: 5’ – CCatcgatGCAAGTCATACGACTTAGCTAC – 3’

zk113.1: 5’ – CGgaattcAGGGTTGAGTTTCCCTCATC – 3’

zk113.2: 5’ – CCCaagcttGTGTAAGTGTAAGTGCTGTG – 3’

zk114.1: 5’ – GactagtCTCACATTCAAGCCTTGAGCC – 3’

zk114.2: 5’ – GCtctagaCCCTTCCATAGACTCATTTGTG -3’

zk136.1: 5’ – AGCCATcccgggCTTGTTACAACAAAAAGAATGAAA – 3’

zk136.2: 5’ – CCATactagtTGTACACCATATTTTAATAAGTAT – 3’

zk137.1: 5’ – CCTAactagtAGAGGTTGTGCACCGTCATGTGGG – 3’

zk137.2: 5’ – CCAAtctagaCCCCCACTGACGTTAGACGGAATC – 3’

zk142.1: 5’ – CTTTATTCTTGCCATCTGGCTATC – 3’

zk142.2: 5’ – CTATGTAGAACAAAGAAAAAATCA – 3’

zk175.1: 5’ – tttggtaccAAAGTGGTTGAGTAGAAGGGCTTC – 3’

zk175.2: 5’ – tttctcgagCAGGACGTGAATTCTCTGTGCAAA – 3’

zk176.1: 5’ – tttggatccACATGAAGAAGGTCGGTCCGCAGA – 3’

zk176.2: 5’ – tttccgcggCAGAGAAGCTTCTAGGTGTGCCAG – 3’

zk184.1-3’*Blm*TV probe: 3’ – TGAGACAGGATCTTTGACCAGTTTGGCTAG – 3’

zk184.2-3’*Blm*TV probe: 3’ – AACTTGAAGTAACTGAAGTATGTTTTCAT – 3’

zk189: 5’ – CCCCGACTGCATCTGCGTGTTCCAATTC – 3’

zk198.3: 5’ – ATCataacttcgtatagcatacattatacgaagttatC – 3’

zk198.4: 5’ – TCGAGataacttcgtataatgtatgctatacgaagttatGAT – 3’

zk199.1: 5’ – AATTCataacttcgtataatgtatactatacgaagttatG – 3’

zk199.2: 5’ – GATCCataacttcgtatagtatacattatacgaagttatG – 3’

zk212.2-*Hprt*: 5’ – AGAAACATGCCAAGCATGATGGTGTGTCTCA – 3’

zk212.3-*Hprt*: 5’ – TGAACTGTCAAATTGTCCTCCAACAAGT – 3’

zk215-RC-pCMV: 5’ – ATGGGCGGGGGTCGTTGGGCGGTCAGC – 3’

zk216-RC-pCMV: 5’ – GAAATCCCCGTGAGTCAAACCGCTATCCA – 3’

zk217-BSD: 5’ – GGACGGTGCCGACAGGTGCTTCTC – 3’

zk218-BSD: 5’ – TTCGTGAATTGCTGCCCTCTGGTTATGTGT – 3’

zk219-RC-puro: 5’ – CGAGGCGCACCGTGGGCTTGTA – 3’

zk237 *Blm*: 5’ – TAAAGCTAGGCGGGTGTAGCCATGGCGTCT – 3’

2.3.4 Transformation of DNA into *E. coli* strains

A 5 mL *E. coli* (DH10B/EL350/EL250/DY380) (Liu *et al.* 2003) culture in LB media was

produced from a single colony, by shaking at 180 rpm in a 15 mL tube at 30°C (37°C for DH10B strain) overnight. The next day, this was diluted 1:10 or 1:20 to an OD_{600}=0.6–0.8 in 20 mL. The bacteria were allowed to grow for another 1–2 hours. Cells were harvested from LB media by centrifugation at 4000 rpm for 5 min at 0°C then were suspended in 800 µL ice-cold water. Cells were transferred to a 1.5 mL tube on ice followed by another centrifugation at 4000 rpm for 4 min at 0°C. The supernatant was discarded then cell pellets were washed again with ice-cold water followed by another centrifugation. Final cell pellets were suspended in 50 µL ice-cold water. 100 ng BAC (1 µL)/1.0 ng plasmid DNA/20 ng (all quantified by photospectrometer) linearized DNA fragment was mixed well with *E. coli* cells. Linearized DNA fragment was prepared by enzyme digestion followed by agarose gel purification. The mixture was transferred into an electroporation cuvette (1 mm gap). The electroporation machine was set as 1750 V, 25 uF, 200 Ω. The pulse time span is usually 4.0 ms. After electroporation, 1.0 mL LB was added then cells were incubated at 30°C (37°C for DH10B strain) for 1 hour. Cells (20 µL–1mL depends on efficiency of the experiment) were spread on appropriate plate for clonal growth.

2.3.5 Recombineering

2.3.5.1 Principle

The recombineering technique uses homologous recombination as a molecular manipulation tool in *E. coli* (Copeland *et al.* 2001; Yu *et al.* 2000; Zhang *et al.* 1998) and facilitates construction of DNA vectors. Using this technique, targeting vectors can be constructed from BACs within a week. The recombineering system is based on the recombination system from bacteriophage λ to recombine DNA fragments in *E.coli*. Bacteriophage λ contains a recombination system, the Red operon, which utilizes two proteins (Figure 2-4): redα (exo) and redβ (bet). Exo is an exonuclease, which digest linear dsDNA in a 5’ to 3’ direction (Carter *et al.* 1971; Matsuura *et al.* 2001). When the nucleotides are removed, long 3’ ssDNA tails are exposed, which are possibly as long as half the length of the original duplex DNA (Carter *et al.* 1971; Hill *et al.* 1997). Then λ Beta promotes the annealing of the complementary DNA strand, which stably binds to 3’ single-strand DNA (Figure 2-5) (Radding *et al.* 1971), which is greater than 35 nucleotides in length (Mythili *et al.* 1996). It is known that beta protein specifically protects single-stranded DNA from digestion by DNase, thus exogenous DNA cannot be destroyed in *E. coli* (Muniyappa *et al.* 1986). If 3’ ssDNA is annealed to the replication fork, it initiates the recombination process.

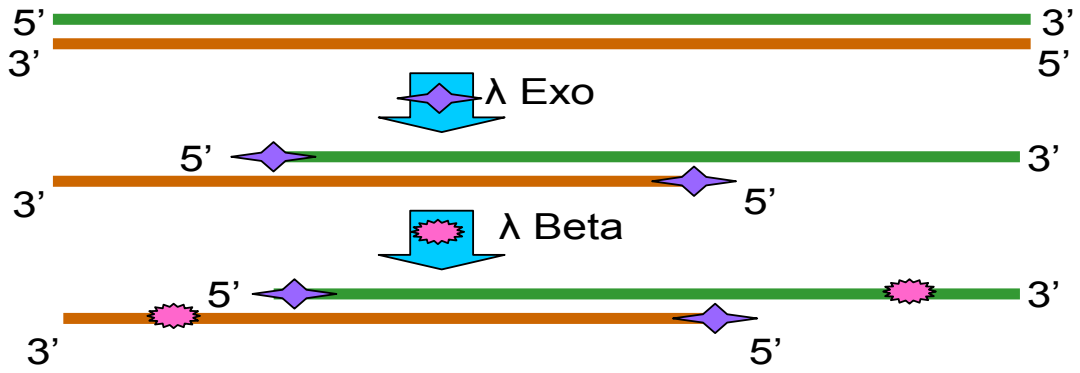


Figure 2-4 The basic functions of Bacteriophage lambda proteins: Exo and Beta

An Exo protein (purple star) digests the 5’ nucleotides of linear dsDNA. The protein Beta (red) binds to the 3’ ssDNA.

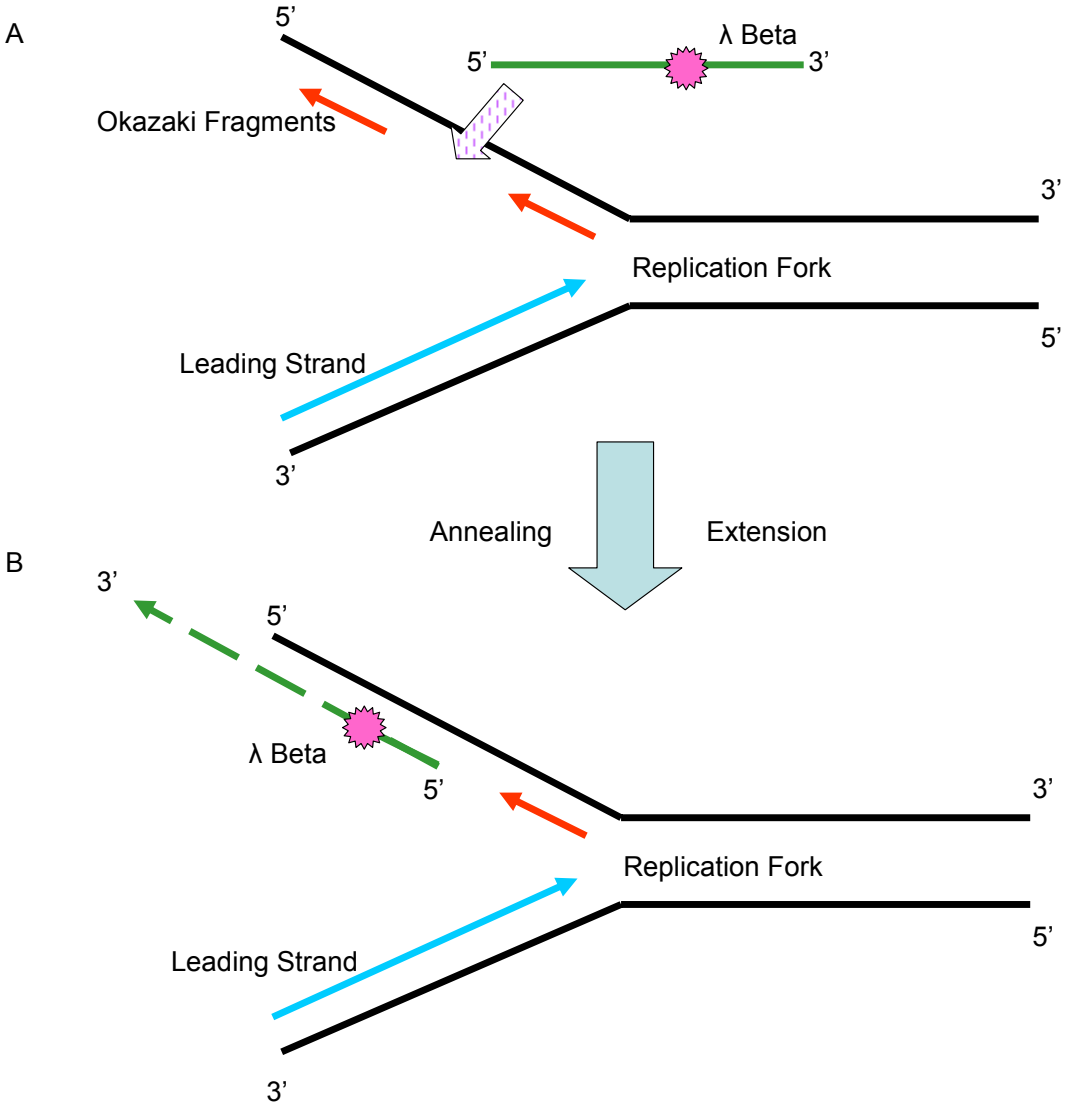


Figure 2-5 The annealing of the ssDNA to the replication fork by lambda Beta protein

A. A protein Beta anneals a 3’ ssDNA tail of a linearized DNA fragment (green line) to the complement lagging strand gap in the replication fork. **B**. This annealed strand starts recombination between linearized DNA fragment and genome DNA (black line).

To make use of recombineering in *E. coli*, Daiguan Yu integrated λ prophage with the λ Red system genes exo, beta and gam into *E. coli* genome. These genes were under the control of powerful λ P_L promoter and the temperature sensitive repressor cl857 (Yu *et al.* 2000). The λ prophage protein expression is tightly regulated at the temperature range from 32°C to 42°C by the repressor cl857 (Figure 2-6). The λ gam protein inhibits the helicase activity and some recombination activities of *Escherichia coli* RecBCD enzyme (Murphy 1991).

Applying the above principle, linearized plasmids can be used to retrieve genomic DNA from Bacterial Artificial Chromosome (BAC) through gap repair in *E. coli* strains containing the integrated λ prophage (Figure 2-7) (Lee *et al.* 2001). A BAC of interest is first transformed into the engineered *E. coli* strain. Transformed single colonies are isolated and confirmed by enzyme digestion and agarose gel electrophoresis. Retrieval plasmids are constructed with two short (50 bp) homology arms (left arm and right arm), which are homologous with the flanking sequence of the relevant fragment of genomic DNA on the BAC. A transformed *E. coli* strain is cultured at 30°C overnight followed by a heating shock at 42°C for 15 minutes to promote λ protein Exo, Beta and Gam expression. These proteins enable recombination between the homologous arms. The result of this gap repair process is to subclone a large genomic DNA fragment from the BAC into a high copy plasmid.

2.3.5.2 Recombineering protocol

Strains EL350/EL250/DY380 were cultured as described in 2.3.4. Before harvesting, cells were transferred into a 42°C water bath for 15min with shaking (180 rpm). Then cells were placed on ice for 5 min to cool. Cells were washed and electroporated as described in 2.3.4.

To retrieve genomic DNA from a BAC, 1–2 μL of linearized and purified retrieval vector (200-500ng) was used. For mini-targeting the retrieval product, 1–2 μL agarose gel purified DNA fragment (20–100 ng) was used in the electroporation.

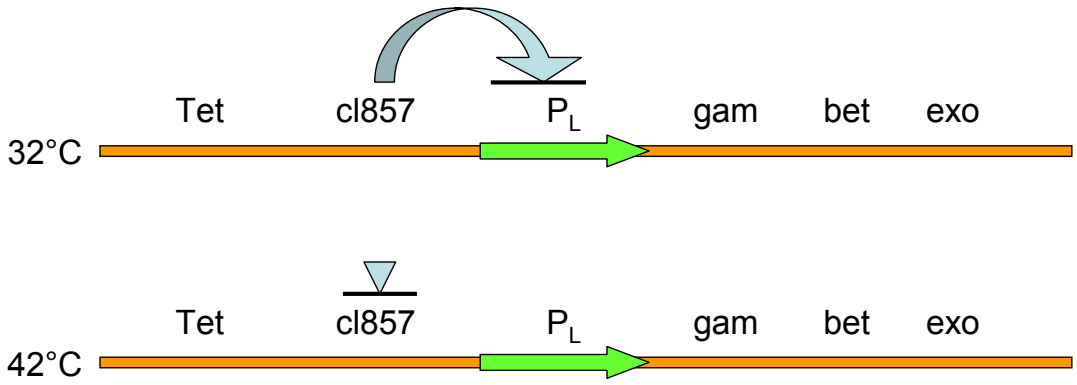


Figure 2-6 The structure and function of the defective lambda prophage

The defective λ prophage contains a Tet selection marker, a cl857 repressor, a PL promoter and λ protein gam, bet and exo. When cells are cultured at 32°C, the cl857 repressor blocks the PL promoter thus repressing the transcription of gam, bet and exo. At 42°C, this repressor is inactivated and the PL promoter initiates expression of gam, beta and exo genes. Tet: tetracycline resistance gene serves as a selection marker.

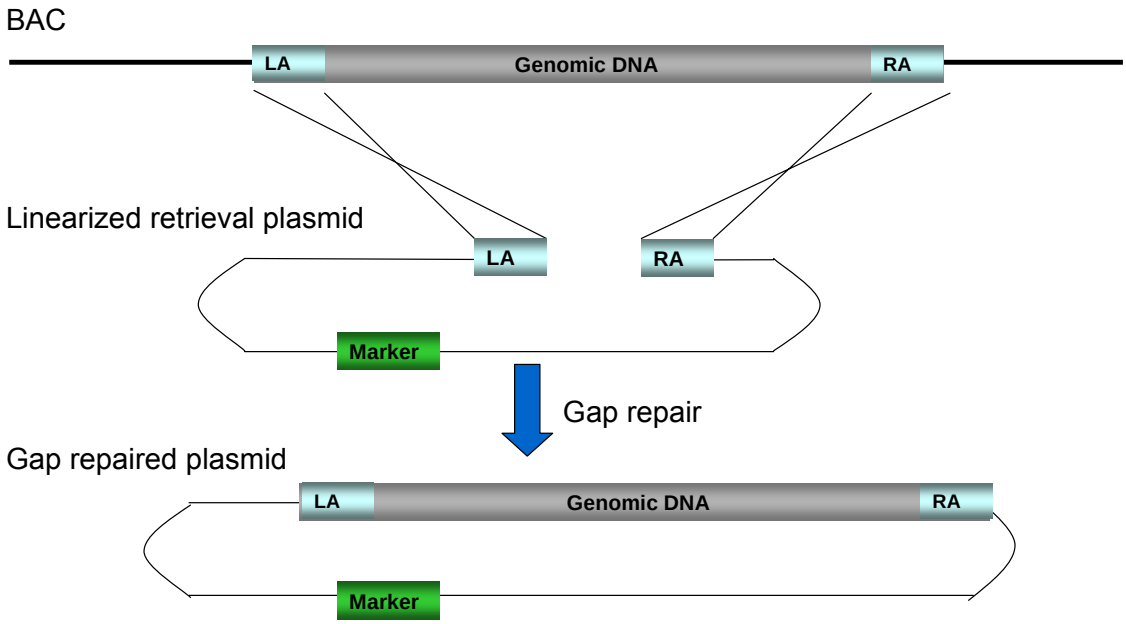


Figure 2-7 Gap repair process

Through the function of λ proteins Exo, Beta and Gam, the gap repair process can recover large fragments of genomic DNA from BACs into high copy number plasmids. The process requires the presence of homology arms (LA and RA, indicates Left Arm and Right Arm, respectively).

2.3.6 Vectors

2.3.6.1 pZK5 – *PuroΔtk/Kanamycin* cassette

As most plasmids carry Ampicillin resistance marker (*Amp*), the *PuroΔtk* cassette was modified to contain a Kanamycin resistance gene (*Kan*) to make the cassette selectable in *E.coli*. The *PuroΔtk* fragment was from the vector YTC-37 (Chen, Y. T. *et al.* 2000).

A Kanamycin resistance cassette was generated by PCR using the pCR 2.1-TOPO vector (Kanamycin resistant, Invitrogen) as template and primers zk107 and zk102.1. The sequences of these oligo nucleotides are shown in 2.3.3. These two primers contain a homologue of pCR2.1-TOPO at 3' and a restriction enzyme digestion site at 5'. Amplified DNA was digested with *Eco* RI and *Xho* I following gel purification. This product was ligated into pBluescript II (pBSII) at the *Xho* I/*Eco* RI sites to produce pZK4 (data not shown).

A *Cla* I-*PuroΔtk-Xho* I and the *Xho* I-*Kan-Eco* RI fragment (from pZK4) were ligated into pBSII to generate vector pZK5 (Figure 2-8). The cassette between *Cla* I and *Eco* RI sites contained the fused positive / negative selection marker *PuroΔtk* for eukaryotic cells and a Kanamycin selection marker (*Kan*) for *E. coli*.

2.3.6.2 pZK9 – Vector to retrieve gene desert DNA from BAC

The retrieval vector pZK9 for the chromosome 6 gene desert targeting was constructed using a pBS *DT-A* backbone (Figure 2-9), which contained the *MC1 DT-A* (Diphtheria toxin A-fragment) cassette (Yagi *et al.* 1990). My colleague Dr Wei Wang inserted a *Kpn* I-*MC1-DT-A-SV40pA-Sph* I fragment into pBS SK(+) *Cla* I-*Eco* RI backbone fragment after both were blunted by T4 polymerase.

The retrieval left arm (PCR product 111) and right arm (PCR 114) were generated by PCR with primers 111.1 and 111.2 for product 111, primers 114.1 and 114.2 for product 114 (Figure 2-10). The sequences of these oligo nucleotides are shown in 2.3.3. The PCR product 111 contained 462 bp chromosome 6 sequence from 10435434 to 10435895 bp and product 114 contained 391 bp chromosome 6 sequence from 10443570 to 10443960 bp. They were digested with *Sma* I/*Spe* I and *Spe* I/*Xba* I respectively and cloned into pBS *DT-A*.

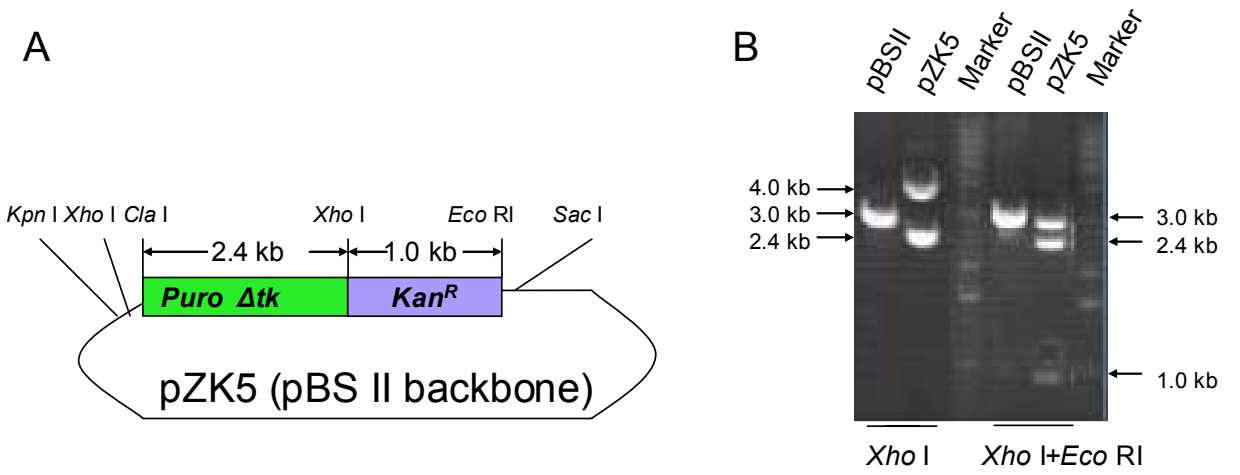


Figure 2-8 **Vector pZK5 – *PuroΔtk/Kanamycin* cassette**

A. Structure of pZK5; **B**. Enzyme digestions confirmed pZK5 was correct as designed. *Xho* I digestion showed pZK5 had 4kb and 2.4kb fragments. *Xho* I+*Eco* RI double digestion showed pZK5 had 1.0kb *Kan*, 2.4kb *PuroΔtk* and 3.0kb pBSII vector.

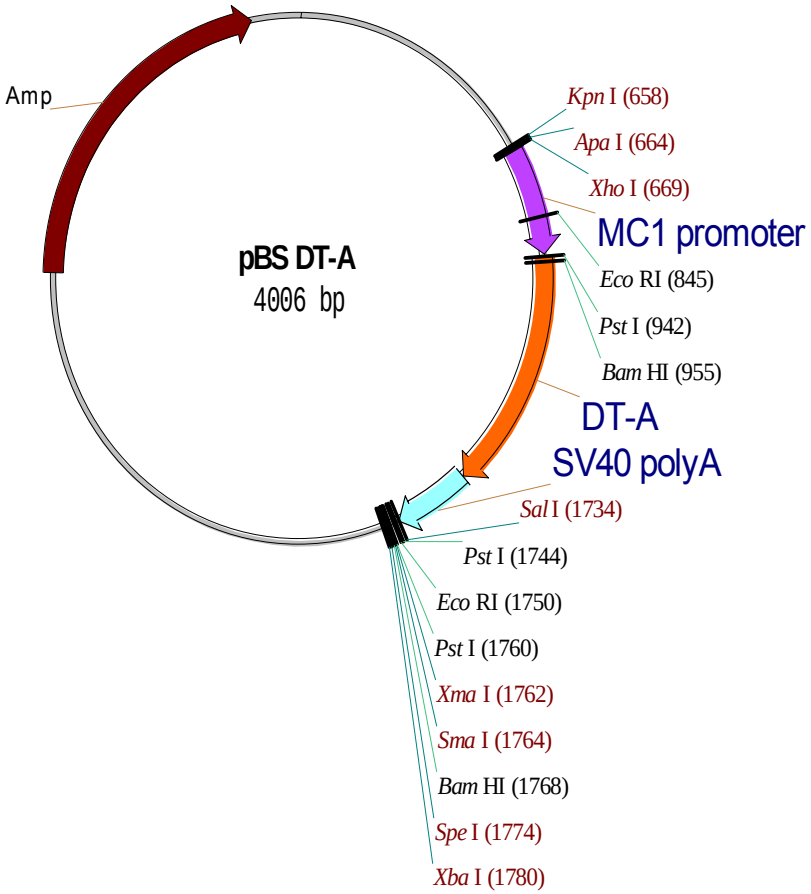


Figure 2-9 **Structure of pBS *DT-A* plasmid**

This vector was used to construct retrieval vectors to retrieve genomic DNA from BACs. The *DT-A* gene was driven by the *MC1* promoter. The SV40 (Simian virus 40) polyA signal was used.

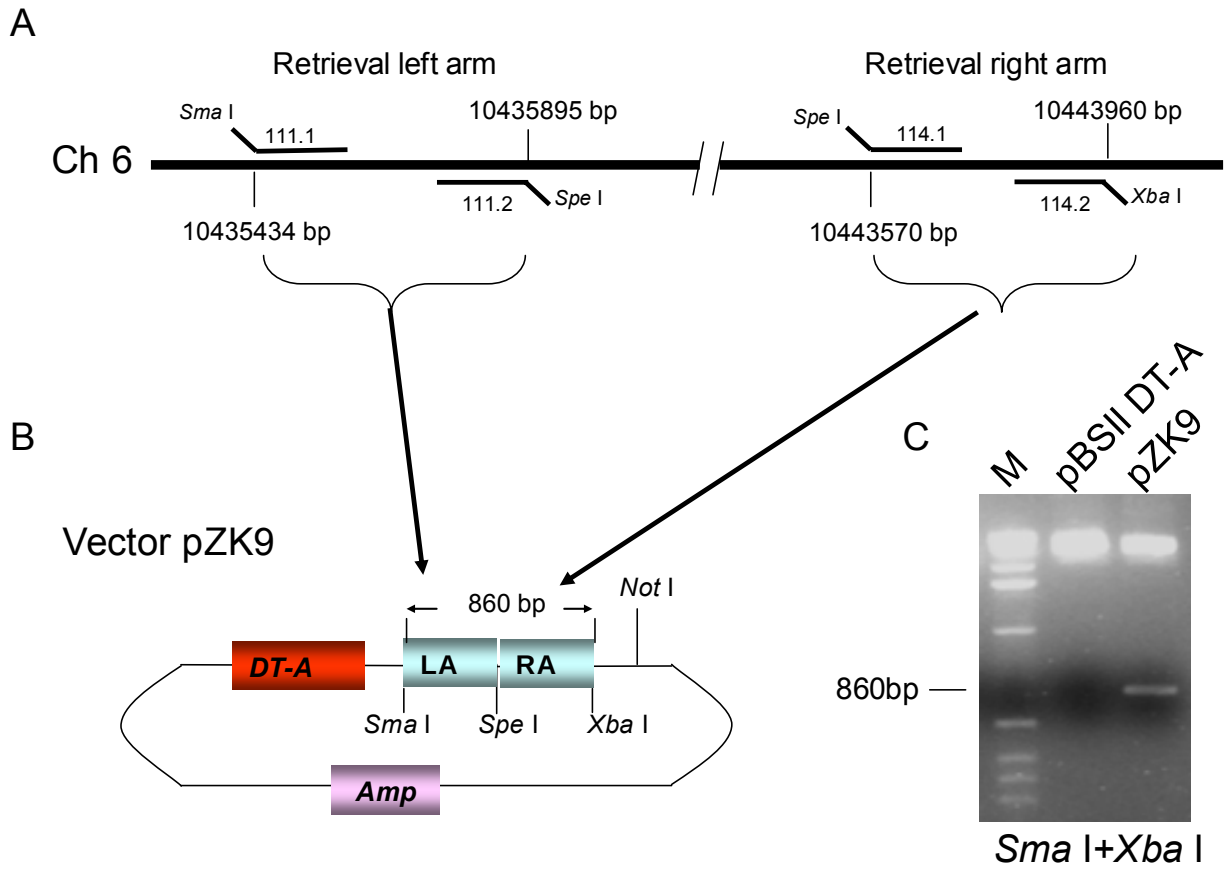


Figure 2-10 Vector pZK9 – to retrieve gene desert genomic DNA from a BAC

A. Position of primers to amplify retrieval homology arms. Part of mouse chromosome 6 is shown here. Primer 111.1 and 111.2 were used to amplify the left retrieval arm. Primer 114.1 and 114.2 were used to amplify the right retrieval arm. Coordinates of start and end point of homology arms are shown in base pairs (bp). *Sma* I, *Spe* I and *Xba* I enzyme sites were synthesized in the 5' of the primers. **B**. Structure of pZK9. pZK9 contained a retrieval left homology arm (LA) and a retrieval right homology arm (RA). The vector contains a unique *Not* I site that will be used to linearize pZK9 prior to targeting experiment. **C**. Enzyme digestion confirmed the pZK9 structure. An ethidium bromide stained agarose gel showing the 860 bp LA/RA fragment cut with *Sma* I/*Xba* I.

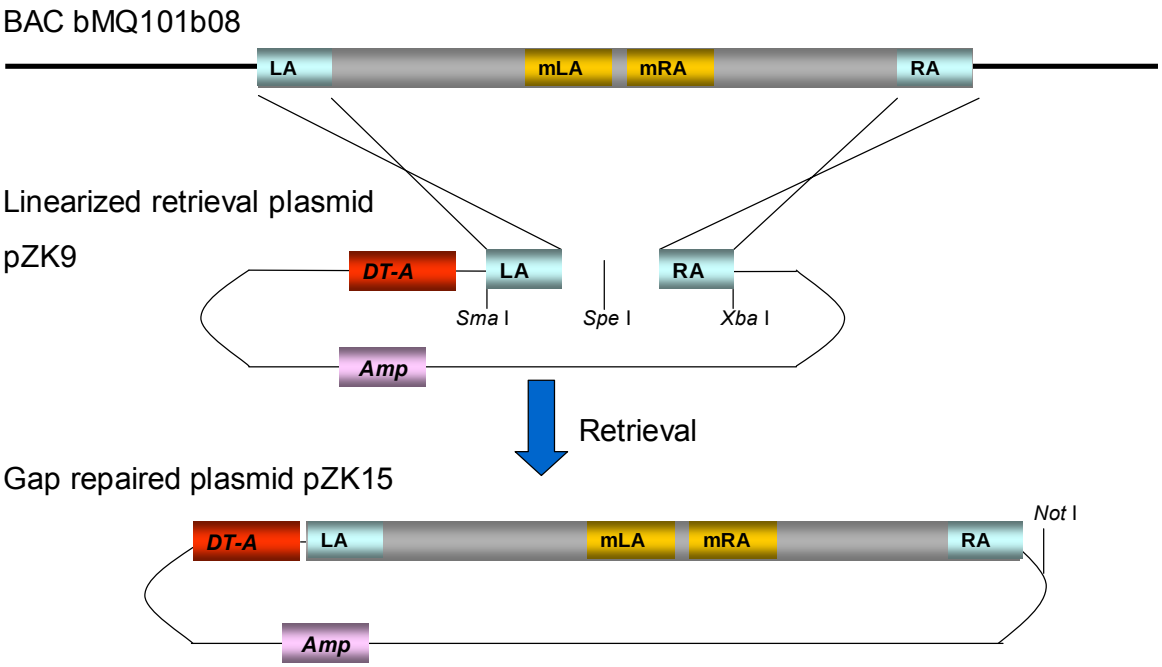
2.3.6.3 pZK15 – Retrieving the homology required for targeting from a BAC

To generate a targeting vector, a large piece of genomic DNA is required for the homology arms. The retrieval plasmid pZK9 has small regions of homology and is designed to recover an 8.5 kb piece of genomic DNA from the BAC bMQ101b08. The BAC was introduced into the EL350 bacteria strain in advance. Transformation was selected in LB agar plates with chlorophenical as the BAC bMQ101b08 contained the chlorophenical marker gene. Correct clones were confirmed by enzyme digestion. To accomplish the retrieval product, pZK9 was linearized with *Spe* I and electroporated into the EL350 bacteria strain. After linearized pZK9 was electroporated into the bacteria, cells were recovered in liquid LB at 30 °C for 45 minutes then placed on Ampicillin agar plates. Ampicillin resistant colonies were picked and the plasmids in those colonies were extracted. Correct pZK9 candidates were identified by a series of enzyme (*Spe* I, *Cla* I, *Eco* RV, *Not* I, *Hind* III and *Sma* I) digestion (Figure 2-11).

2.3.6.4 pZK8 - Chromosome 6 gene desert mini-targeting vector

The mini-targeting vector pZK8 was designed to recombine with the retrieved genomic fragment to generate the chromosome 6 gene desert target vector (Figure 2-12). Two mini targeting homology arms mLA (249 bp) and mRA (286 bp) were amplified using the PCR primer pairs zk112.1 with 112.2 and zk113.1 with 113.3, respectively. The sequences of these oligo nucleotides are shown in 2.3.3. The amplified products were digested with *Kpn* I*/Cla* I and *Eco* RI*/Hind* III, respectively before they were cloned into either side of the *PuroΔtk/Kanamycin* cassette of pZK5.

A.



B.

Expected restriction pattern of pZK15

Restriction Enzyme	Cut	Fragments (kb)
Spe I	3	0.47, 4.2, 7.8
Cla I	1	12.5
Eco RV	3	0.3, 2.3, 9.9
Not I	1	12.5
Hind III	1	12.5
Sma I	2	4, 8.5

C.

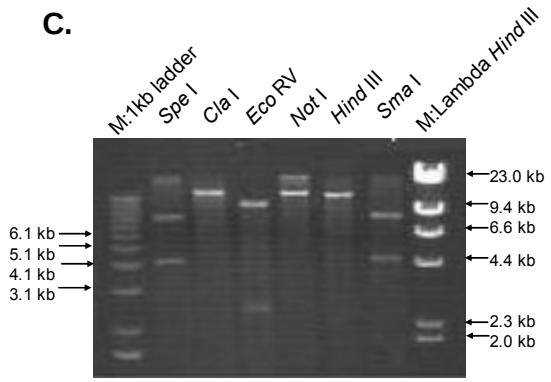


Figure 2-11 Generation of retrieval product pZK15

A. Schematic process of gap repair – retrieving DNA from the BAC bMQ101b08 to form pZK15; **B**. In silico prediction of restricted fragments of pZK15; **C**. Confirmation of one clone of pZK15.

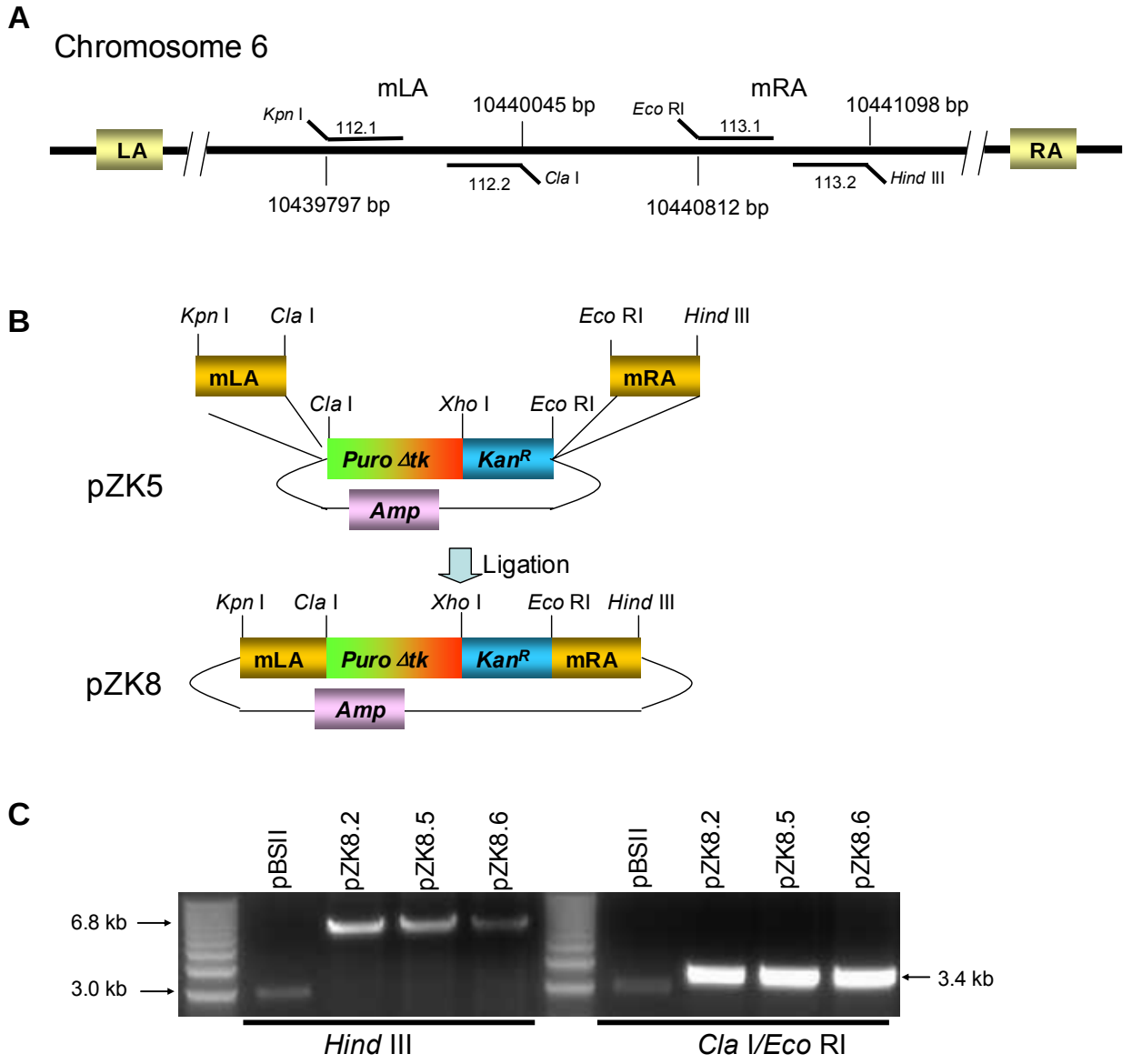


Figure 2-12 Structure of pZK8 – gene desert mini targeting vector

A. Genomic DNA between retrieval left arm (LA) and right arm (RA) of chromosome 6 is shown here. Mini targeting homology arms mLA and mRA were amplified by primer pairs 112.1 with 112.2 and 113.1 with 113.2. Enzyme sites were included in the 5' sequence of primers. Genomic coordinates are shown as base pair number on chromosome 6. **B**. *Kpn* I-mLA-*Cla* I and *Eco* RI-mRA-*Hind* III fragments were cloned one after another into either side of *PuroΔtk/Kanamycin* cassette on pZK5, resulting in pZK8, the gene desert mini targeting vector. **C**. *Hind* III and *Cla* I/*Eco* RI digestion confirmed correct plasmid structure of clones 2, 5 and 6. *Cla* I/*Eco* RI digestion generated two 3.4 kb fragments**.**

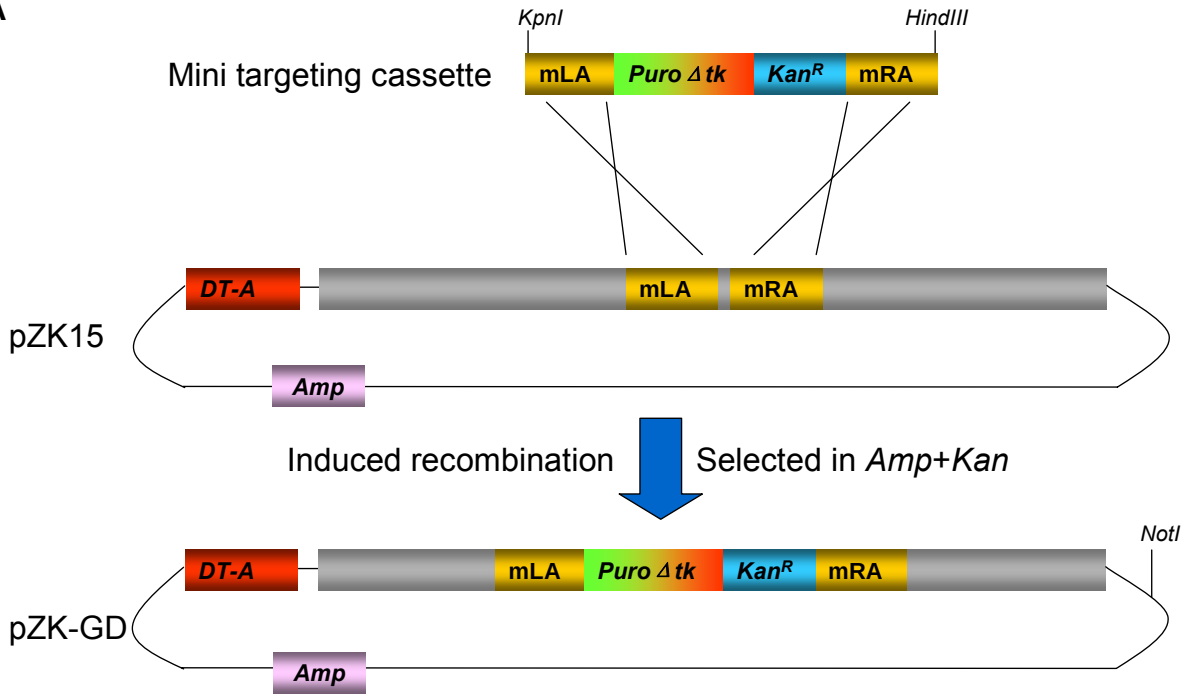
2.3.6.5 Construction of gene desert targeting vector pZK-GD

To generate pZK-GD (Figure 2-13), the mini targeting cassette needed recombine with the retrieved homologous DNA fragment contained in pZK15. The EL350 strain containing pZK15 was cultured in 42°C for 15 minutes and cooled in ice-cold water prior to the transformation of the mini targeting cassette. This temperature shift induces λ protein Exo, Beta and Gam expression and facilitates recombination. Vector pZK8 was digested with *Kpn* I and *Hind* III. The 3.9 kb mini targeting fragment was separated from the vector on an agarose gel. DNA was purified and the fragment was transformed into the EL350 *E. coli* strain containing pZK15 with the retrieved homology arms.

After transformation the *E. coli* cells were plated on Ampicillin/Kanamycin plates. EL350 *E. coli* colonies contained a mixture of the original plasmids and the recombined product. To recover the product, a plasmid mini-prep was performed. The product was re-transformed into DH10B *E. coli* strain and re-plated in Ampicillin/Kanamycin plates.

The correct clones were identified by restricted enzyme digests with *Bam* HI, *Eco* RI, *Eco* RV, *Hind* III, *Pst* I, *Sma* I and *Spe* I (Figure 2-14). The agarose gel pattern was compared to that predicted in silico. This confirmed that pZK-GD clone 31, 34, 35, 103, 105 had the correct structure.

A



B

Restriction Enzyme	Fragments (kb)	
	pZK15	**pZK-GD**
BamHI	2.2, 2.7, 7.5	1.6, 2.2, 2.7, 8.6
EcoRI	0.9, 2.5, 2.8, 6.2	0.035, 0.90, 2.5, 5.5, 6.2
EcoRV	0.3, 2.3, 9.9	0.104, 0.30, 1.24, 1.37, 2.3, 9.8
HindIII	12.5	6.3, 8.9
NotI	12.5	15.2
PstI	0.016, 0.60, 0.69, 0.80, 1.75, 8.7	0.016, 0.60, 0.68, 0.80, 1.0, 1.2. 1.7, 2.1, 6.9
SmaI	0.007, 4.3, 8.2	0.007, 0.9. 1.5, 4.3, 8.4
SpeI	0.47, 4.2, 7.8	0.5, 1.2, 5.6, 7.9

Figure 2-13 Generation of the chromosome 6 gene desert targeting vector - pZK-GD

A. Recombination at the mini-homology arms (mLA and mRA) between the targeting cassette and retrieval product vector pZK15 resulted in the gene desert targeting vector pZK-GD. pZK-GD is resistant to Ampicillin and Kanamycin. **B**. In silico prediction of restriction fragments of pZK-GD compared to pZK15.

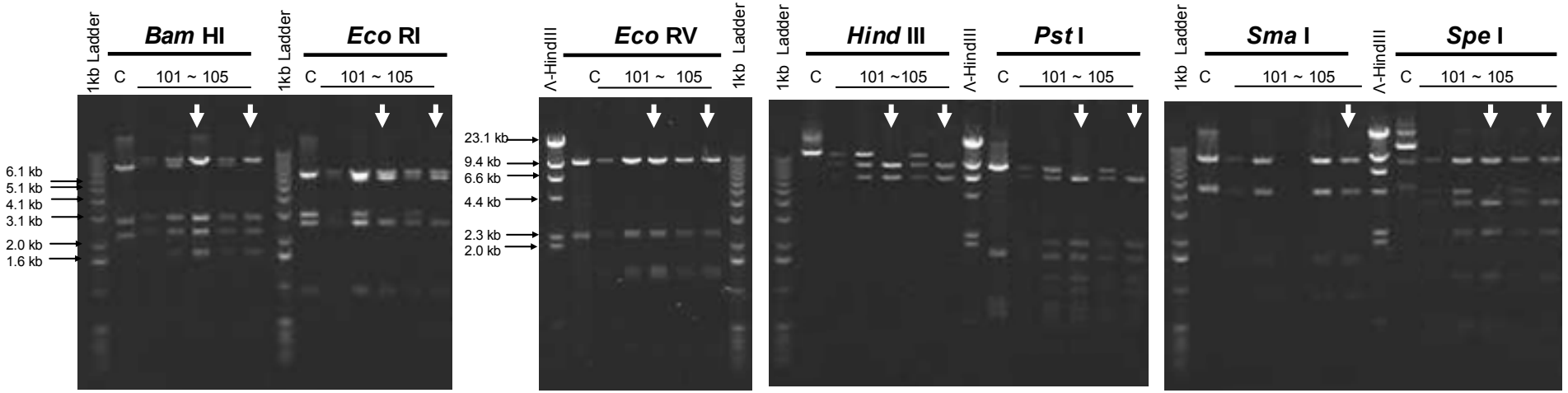


Figure 2-14 Restriction pattern confirmed correct pZK-GD plasmids

Clones 101–105 were digested with *Bam* HI, *Eco* RI, *Eco* RV, *Hind* III, *Pst* I, *Sma* I and *Spe* I. Clones 103 and 105 (white arrow) have the correct restriction fragments compared to the control plasmid pZK15.1 (C).

2.3.6.6 RMCE plasmid with the *Bsd* cassette

A vector was designed (Figure 2-15) to replace the positive/negative selection marker *PuroΔtk* in the BAC acceptor cell line CCI18 with a Blasticidin resistance gene (*Bsd*). The *pCMV-EM7-Bsd-SV40pA* fragment, from pCMV/Bsd (Invitrogen) was isolated by *Xho* I/*Eco* RI double digestion and ligated into pBSII SK+. *LoxP* and *Lox511* adapters were ligated into the vector. The *loxP* adapter was produced by annealing oligo nucleotides zk198.3 and zk198.4. The *Lox511* adapter was produced by annealing oligo nucleotides zk199.1 and zk199.2. The sequences of these oligo nucleotides are shown in 2.3.3. Plasmid pZK33.3, pZK33.4 were confirmed to be correct clones by sequence analysis.

2.3.6.7 Retrieve part of *Blm* gene for targeting from a BAC

To construct a *Blm* gene-targeting vector, homology arms need to be retrieved from a BAC containing the gene. The retrieval vector pZK21 (Figure 2-16) was designed to retrieve the *Blm* gene genomic DNA from the 129S5 BAC bMQ-436e20 by recombineering. The left arm (LA, 500 bp) was amplified from this BAC by the primers ZK136.1 and ZK136.2. And the right arm (RA, 660 bp) was amplified from the same BAC by the primers ZK137.1 and ZK137.2. The sequences of these oligo nucleotides are shown in 2.3.3. The fragments were cut with *Sma* I/*Spe* I (LA) and *Spe* I/*Xba* I (RA), gel purified and ligated in a 3-way reaction with pBS *DT-A* cut with *Sma* I/*Xba* I. This resulting plasmid (pZK21) can be linearized with *Spe* I to facilitate retrieval of the *Blm* gene from the BAC.

2.3.6.8 Retrieval product with *Blm* gene exons 2 and 3 (pZK19)

The recovery of the 10 kb homology arm required for targeting is achieved by gap repair recombinerring. The retrieval vector pZK21 was linearized with *Spe* I and purified by gel purification. The mouse BAC bMQ-436e20 was transferred into EL350 cells first. The cells were isolated on LB agar plates with chlorophenical. Correct clones were resistant to chlorophenical as the BAC bMQ-436e20 contained the chlorophenical marker gene. These clones were confirmed by enzyme digestion. One correct clone was cultured in LB at 30°C and the recombination protiens - λ phage proteins Exo, Beta and Gam were induced to express for 15 minutes at 42°C. These bacteria cells were prepared to be electroporation competent cells then electroporated with 20ng linearized pZK21 plasmid. Cells were recovered in LB at 30°C for 45 minutes then selected in Ampicillin agar plates. The recovered colonies may contain 10 kb recovered *Blm* gene DNA. Plasmid minipreps were performed to candidate clones. Then these clones were screened by the digestion of *Bam* HI*, Eco* RI*, Spe* I and *Xba* I. Correct clones, for example clone 1, 2, 4 and 6 were named

pZK19 and were used to generate homology arms of the *Blm* gene-targeting vector (Figure 2-17).

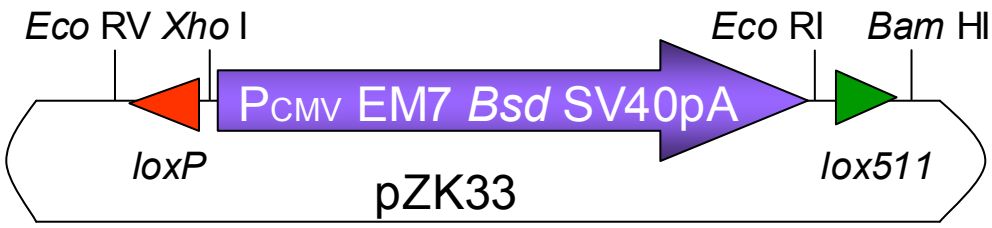


Figure 2-15 **Construction of the RMCE vector pZK33**

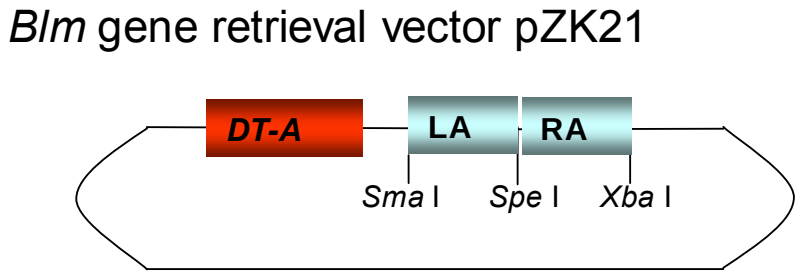


Figure 2-16 ***Blm* gene retrieval vector pZK21**

Retrieval left arm (LA) and right arm (RA) were ligated into pBS *DT-A*. The *Spe* I site was designed to linearize this plasmid for retrieval. *DT-A*: Diphtheria toxin A-fragment.

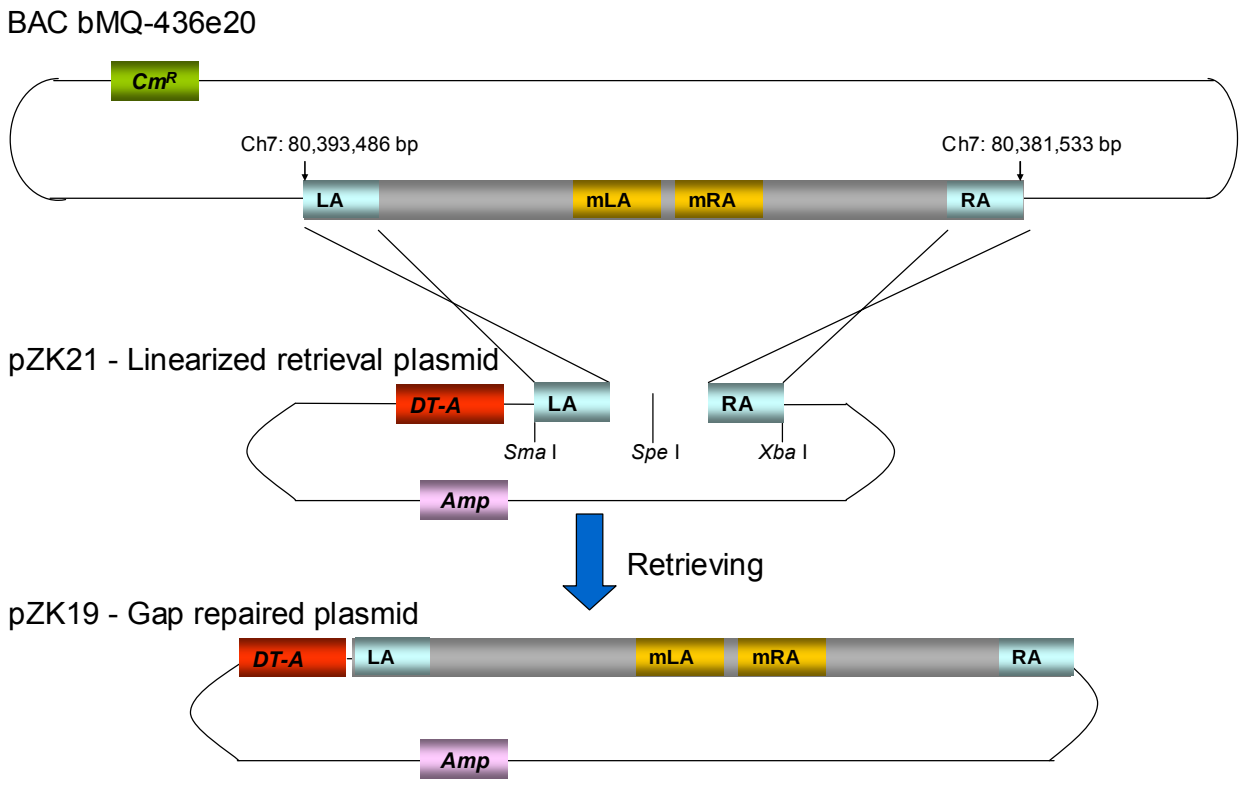


Figure 2-17 Retrieving *Blm* gene-targeting arms from a BAC

Spe I linearized retrieval vector pZK21 was designed to recombine with the BAC bMQ-436e20 and retrieve a fragment of homologous sequence from the BAC.

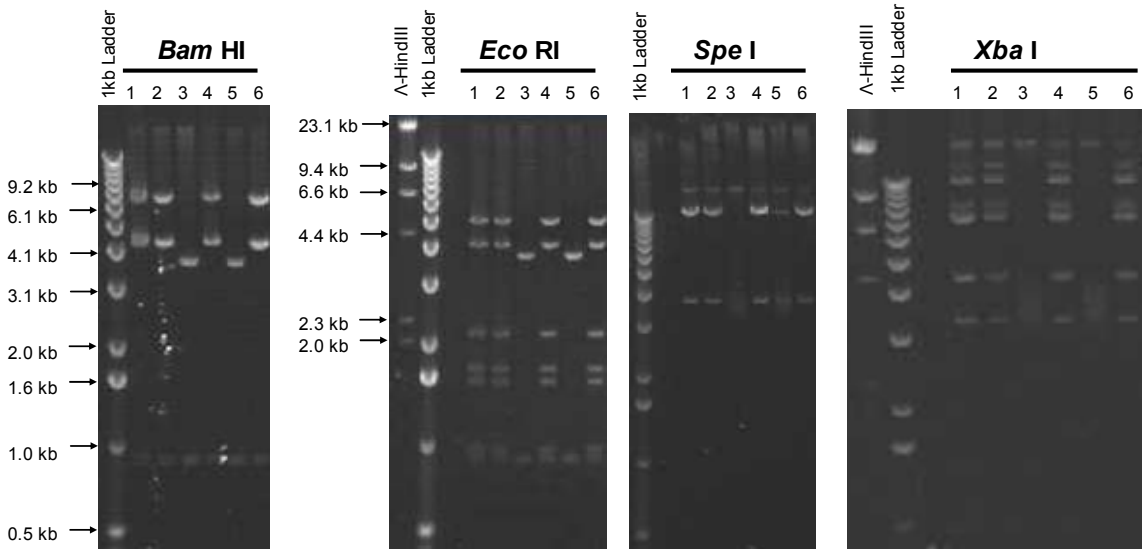


Figure 2-18 Enzyme digestion to identify the correct recombinants

Candidate plasmids of pZK19 were prepared (n=16, 6 are shown here) and digested with several enzymes. Electronic analysis of the vector and genomic sequence (Table 2-1) revealed that clones 1, 2, 4, 6 (shown) and 7–10, 12–16 were correct.

2.3.6.9 *Blm* gene mini targeting vector (pZK31)

To generate the final gene-targeting vector, a *Blm* gene mini targeting cassette is needed. This cassette was designed to contain *PGK/EM7/Neo-MC1-tk* cassette flanked by *FRT* sites. This vector was designed to replace the exons 2 and 3 of the gene with the *PGK/EM7/Neo-MC1-tk* cassette. *PGK* is a eukaryotic promoter and *EM7* is a prokaryotic promoter. Both *PGK* and *EM7* can initiate the expression of the *Neo* cassette but provide different resistance. *PGK-neo* generates resistance to G418 in eukaryotic cells and *EM7-neo* generates resistance to Kanamycin in bacteria. After targeting the *Blm* gene, the *PGK/EM7/Neo-MC1-tk* cassette can be recycled by Flpe then the same gene-targeting vector can be used to target the second allele.

The mini-targeting vector was designed (Figure 2-19) to target the fragment retrieved in pZK19 to generate the final *Blm* gene-targeting vector. The mini left homology arm (mLA) was amplified with primers zk175.1 and zk175.2. The mini right homology arm (mRA) was amplified with primers zk176.1 and zk176.2. The sequences of these oligo nucleotides are shown in 2.3.3. mLA and mRA were cut with *Kpn* I/*Xho* I (LA) and *Bam* HI/*Sac* II (RA), gel purified and ligated into pCOI5 cut with *Kpn* I/ *Sac* II, which contains the *PGK/EM7/Neo-MC1-tk* cassette flanked by *FRT* sites (Haydn Prosser's unpublished construct), through a four-way ligation and selected by Kanamycin.

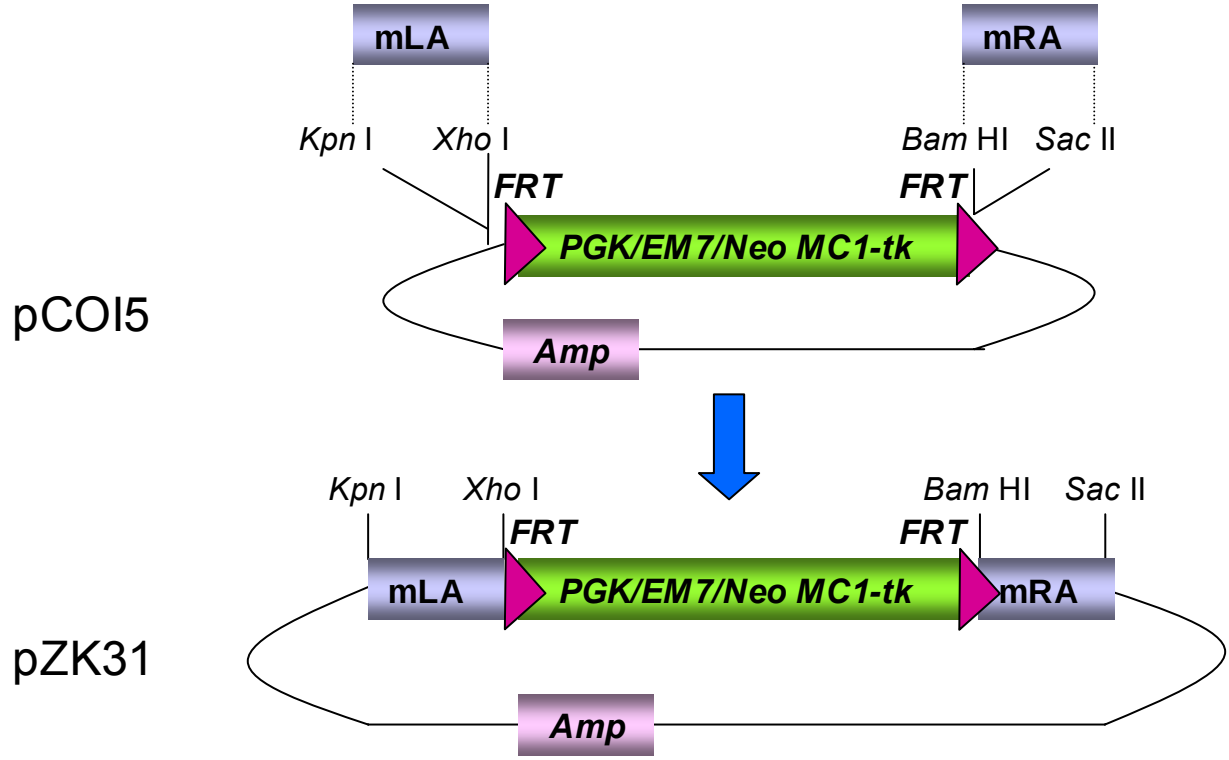


Figure 2-19 Generation of the *Blm* gene mini targeting vector - pZK31

The *PGK/EM7/Neo* and *MC1-tk* cassettes are flanked by two *FRT* sites. The mini targeting left arm (mLA) and mini targeting right arm (mRA) were inserted into pCOI5 backbone through a 4-way ligation into the *Kpn* I and *Xho* I site and *Bam* HI and *Sac* II site, respectively.

2.3.6.10 *Blm* gene-targeting vector (pZK30)

To assemble the *Blm* targeting vector (Figure 2-20), the mini targeting cassette (from pZK31) was isolated by digestion with *Kpn* I/*Sac* II followed by gel purification and electroporated into the EL350 recombinogenic *E. coli* containing the *Blm* homology fragment of pZK19. The *PGK/EM7/neo-MC1-tk* cassette replaced exon 2 and 3 DNA of *Blm* gene and resulted in Kanamycin resistant colonies. The produced plasmid contains a left homology arm (LA, 5 kb) and right homology arm (RA, 4.6 kb) to facilitate homologous recombination of *Blm* gene in ES cells.

The *Blm* targeting vector pZK30 can be selected in Ampicillin and Kanamycin. However, EL350 cells could also contain both of the non-recombined original plasmids, which would also give Ampicillin and Kanamycin resistance. To isolate the recombinant pZK30 plasmid, these plasmids were diluted and retransfected into *E. coli* strain DH10B followed by Ampicillin and Kanamycin selection. Candidate Ampicillin and Kanamycin resistant clones were screened by restriction enzyme digestion to identify those with the correct pattern, which was predicted by electronic sequence analysis (Table 2-1). Several clones were correct. Figure 2-21 shows the correct subclones: pZK30.3.1–30.3.12 and pZK30.4.1–30.4.9.

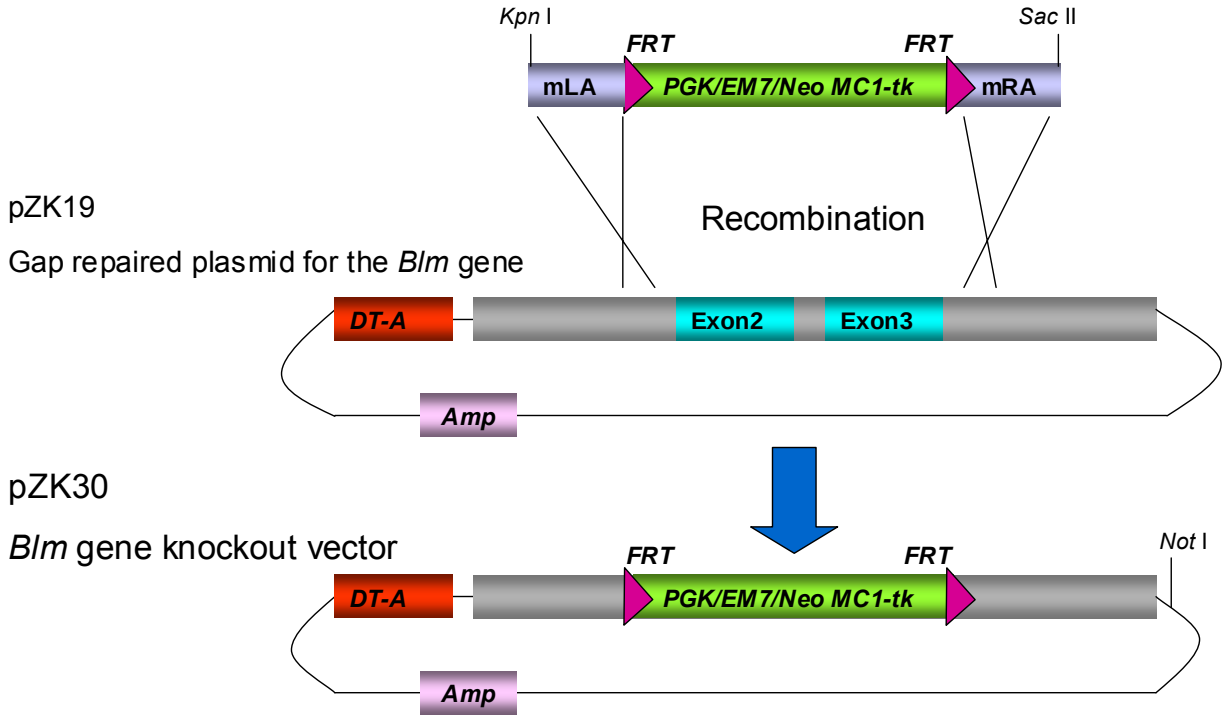


Figure 2-20 Generation of *Blm* targeting vector pZK30

The mini targeting cassette recombines with the *Blm* gene homology fragment in pZK19 to form the *Blm* gene knockout vector pZK30. pZK30 can be selected on Ampicillin and Kanamycin agar plates. The *Not* I site will be used to linearize this gene-targeting vector.

Table 2-1 Electronic sequence analysis of pZK19 and pZK30

Restriction Enzyme	Cuts		Fragments (kb)	
	pZK19	pZK30	pZK19	pZK30
Bam HI	4	5	0.914, 4.248×2, 6.584	0.914, 4.248×2, 3.589, 4.541
Eco RI	7	7	0.905, 0.954, 1.5, 1.7, 2.1, 4.0, 4.9	0.047, 0.905, 0.954, 1.7, 4.0, 4.9, 5.1
Hind III	4	5	0.722, 1.8, 4.0, 9.5	0.722, 1.2, 1.8, 4.4, 9.5
Not I	1	1	15.99	17.54
Xba I	4	7	0.963, 3.3, 4.4, 7.3	0.388, 0.963, 1.5, 1.7, 2.4, 3.3, 7.3

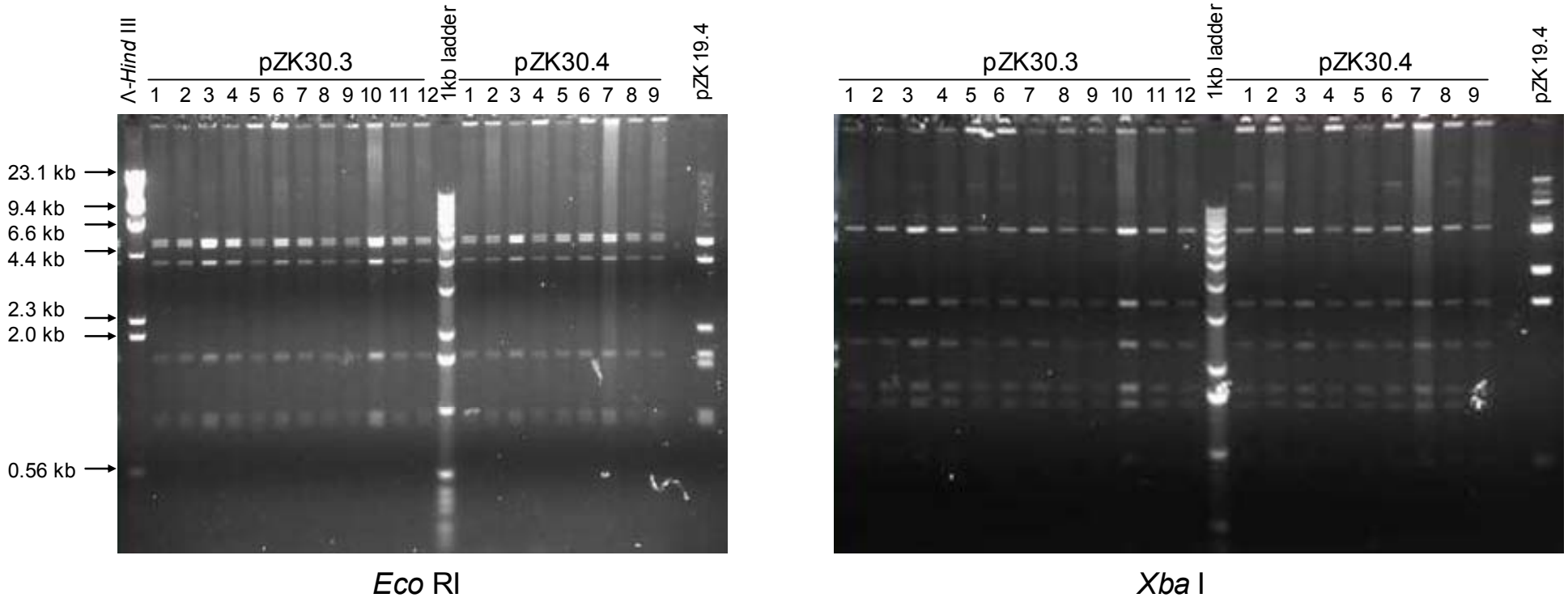


Figure 2-21 Screens for the pZK30 by enzyme digestion

pZK30.3 and pZK30.4 subclones were digested with *Eco* RI and *Xba* I. Subclones 1–12 of pZK30.3 and subclones 1–9 of pZK30.4 were correct. The *Blm* gene retrieval product pZK19.4 served as the control.

2.3.7 Genomic DNA isolation from tissue culture (6-well plate)

Cells in 6-well plates produce a large quantity of DNA, which is sufficient for use in Southern blot or array CGH. When preparing samples of genomic DNA, materials and solutions reserved for genomic use were used to diminish plasmid contamination. These include: Eppendorf tubes, proteinase K, dH_2O, and T0.1E. Genomic pipette tips from which the tips have been snipped off were used to prevent shearing the DNA and keep the average size in excess of approximate 30 kb.

6-well plates were washed twice with PBS before they were lysated by 1 ml of Cell Lysis Buffer. Then the plate was shaken at room temperature at 15-60 rpm for 10 minutes until all the cells detached. Then cell lysate was transferred into 15 ml falcon tubes to be shaken gently at 55°C for 3 hours to overnight. Then two times volume of 100% ethanol was added into the tubes to precipitate DNA. Tubes were inverted a few times to mix well. DNA was spooled onto flame-sealed micropipettes and rinsed once in 70% ethanol, once in 100% ethanol followed by an air dry. Finally, genomic DNA was dissolved in 0.6 mL T0.1E in a sterile Eppendorf tube overnight at room temp or a couple of hours at 55°C if it is hard to dissolve. Genomic DNA can be stored at 4°C for further analysis. Typically, the yield from a confluent 3 cm plate was 100 ug DNA.

2.3.8 Genomic DNA isolation from 96-well plates

Cells cultured in 96-well plates can be used for extracting DNA for Southern blot. When cells were confluent in wells, they were washed twice with PBS before adding 50 µL cell lysis buffer containing freshly prepared 0.5–1 mg/ml Proteinase K. Cell lysates were kept in a humidity chamber at 55°C overnight then 100 µL 100% ethanol was added to each well to precipitate the genomic DNA, which were allowed to stand at room temperature for a minimum of 30 minutes followed by 3 washes with 70% ethanol and 3 washes with 100% ethanol. DNA was left to dry in a 55°C oven for 5 minutes then it was ready to be dissolved for endonucleases digestion. Each well can generate 1–3 µg DNA.

2.3.9 Endonuclease digestion for Southern blotting

10 units of the appropriate enzyme were used in a total reaction volume of 40 µL with the proper reaction buffer according to the instruction from New England Biolabs (NEB). 2–5 µg genomic DNA was digested for each sample and digestion usually lasted overnight at 37°C to fully cut genomic DNA.

2.3.10 Southern blot hybridization

2.3.10.1 Probes

Probes were either PCR amplified from templates or excised from plasmids. They were all separated from other unwanted products by agarose electrophoresis and purified by QIAquick PCR Purification Kit™ (Cat.No. 28104) following manufacture's instructions.

PuroΔtk probe: 1.2 kb DNA fragment from *Pst* I digested plasmid YTC37 (Chen, Y. T. *et al.* 2000).
Δtk probe: 1.0 kb DNA fragment from *Bam* HI/*Xba* I double digested plasmid YTC37 (Chen, Y. T. *et al.* 2000).

2.3.10.2 Southern blotting

1–10 µg genomic DNA was digested with the appropriate restriction enzyme overnight and the DNA fragments were separated by electrophoresis on a 0.8% agarose gel in 1×TAE solution. Following electrophoresis, the gel was sequential soaked in Depurination Buffer for 5–10 minutes and denaturation buffer for 30–60 minutes with gentle agitation. A capillary blot was set up according to the standard protocols with Hybond XL membrane (GE Healthcare RPN203S) and denaturation buffer as the transfer buffer. After an overnight transfer, the membrane was neutralized by neutralization buffer for 5-10 minutes and baked for 1 hour at 80°C.

2.3.10.3 Probe preparation

10–30 ng DNA probe was labelled by Prime-It® II Random Primer Labelling Kit (Stratagene, Cat. No.300385) according to the manufacturer's instructions. Specific activities of ≥1×10^9 dpm/µg were achieved with DNA probes ranging in size from 400 bp to 10 kb. Briefly, 25 ng DNA probe and 10 µL random oligonucleotide primers were mixed in a final volume of 34 µL in dH_2O. Probe and primers were heat at 100°C for 5 minutes and cooled down at room temperature. 10 µL 5× dCTP buffer (provided in kit), 5 µL $dCTP^{32}$ (5 µCi) and 1 µL $Exo^{(-)}$ Klenow (5U/µL, final concentration 0.1U/µL) were added to probe and primers to make 50 µL mixture. The mixture was incubated at 37°C for 10 minutes and purified by ProbeQuant G-50 Micro Columns (GE Healthcare, Cat.No. 27-5335-01). Probes were then denatured at 100°C for 10 minutes and chilled on ice for 5 minutes before use.

2.3.10.4 Hybridization and membrane washes

The blot was pre-hybridized at 65°C for 1 hour with 1 mg single-strand salmon testes DNA (Sigma, Cat. No.D9156) in 10–15 mL hybridization buffer. The labelled probe then was added into the hybridization tube and incubated in a rotating oven at 65°C for 4 hours or overnight. The blot was rinsed by wash buffer at 65°C for 15 minutes and at room temperature for 10 minutes and exposed to X-ray film (Fuji) overnight at -80°C before development.

2.3.11 Array comparative genomic hybridization

2.3.11.1 Principle

Irradiation mutagenesis generates large amplifications and deletions on chromosomes. To identify these, a method to detect genomic DNA copy number variations is needed. Array comparative genomic hybridization (CGH) has become a powerful approach to detect genome-wide chromosomal imbalances. The principle behind this technology is the use of differentially fluorescently labelled samples of DNA and a reference, which are hybridized simultaneously to nucleotide probes on a glass slide (Figure 2-22). The signal of two fluorochromes at each probe location will depend on the copy number of DNAs. Through this method, any gain or loss of DNA sequences, such as deletions, duplications and amplifications can be identified in the whole genome.

The resolution of array CGH ranges from 50bp to 10 Mb. BACs, PACs, PCR amplified DNA/cDNA and oligo-nucleotides are all able to be used as probes in array CGH (Chung *et al.* 2004; Kallioniemi *et al.* 1992; Selzer *et al.* 2005).

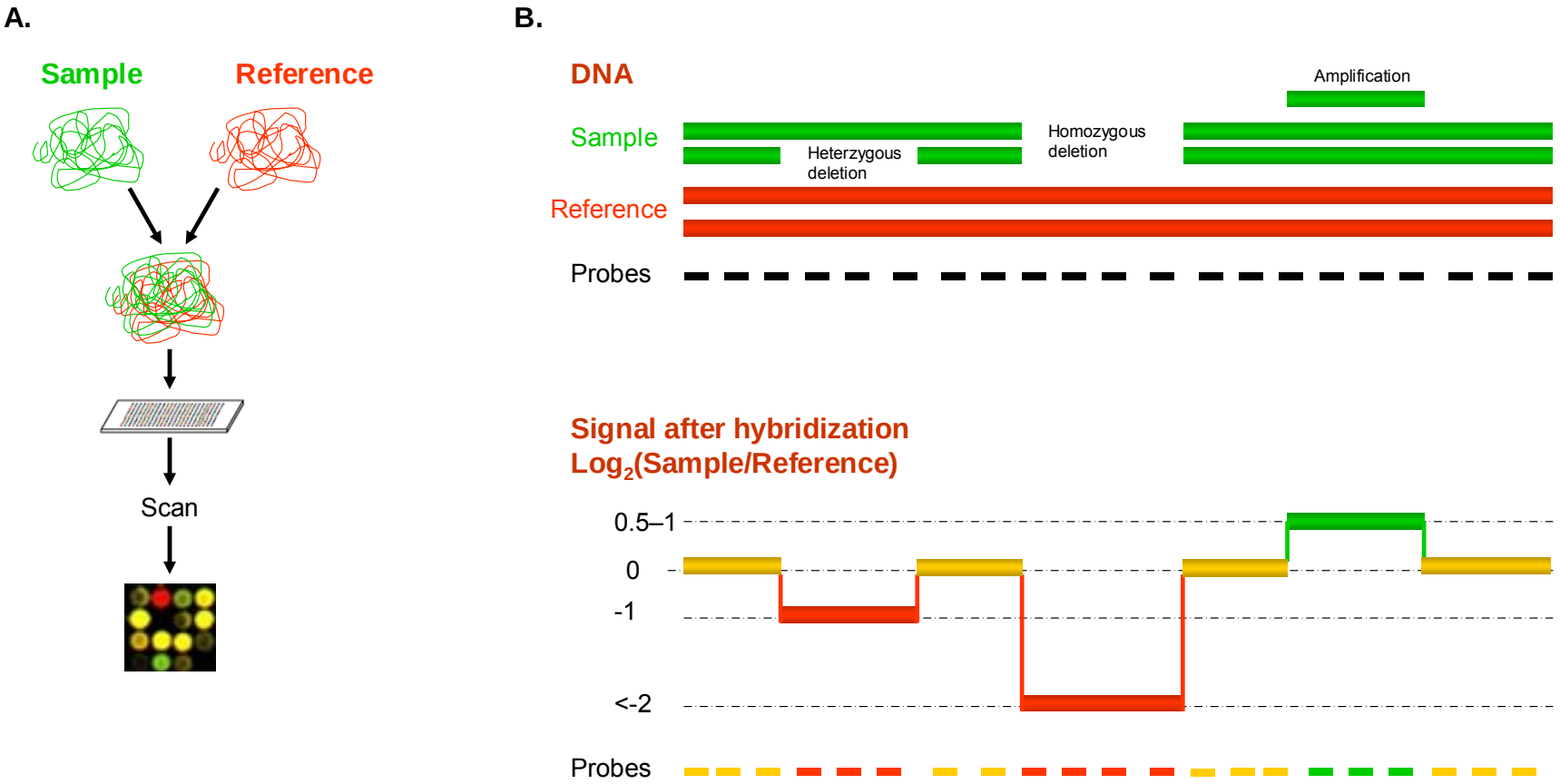


Figure 2-22 The principle of array comparative genome hybridization

A. The schematic view of the pocedures to conduct array CGH. Sample DNA and reference DNA labelled by green and red fluorescence are hybridized with probes arranged on a glass slide. Copy number variation is observed by the ratio of the two probes on the array. **B**. Example of Log_2 signal associated with various copy number alterations. Lack of one or two copies in the sample DNA (green) is reflected by lack of a green signal with the probes, while if a region of the genome is amplified in the sample, the green signal will be stronger than the red control. The $base_2$ logarithm of the signal intensity is used to reflect these alterations.

2.3.11.2 Reagents

For all data the following arrays were used: mWGTP1.1.1 (mouse whole genome BAC tilepath array, contains 2 copies of all probes) and mWGTP1.2.1 (mouse whole genome BAC tilepath array, contains 1 copy of all probes). One copy of all probes includes 18,294 tile path BACs. The original protocol has been reported previously by Chung *et al* (Chung *et al.* 2004)

BioPrime labelling kit (Invitrogen, Cat. 18094-011)

2.5× random primers solution: 125 mM Tris-HCl (pH 6.8), 12.5 mM MgCl2, 25 mM 2-mercaptoethanol, 750 µg/ml oligodeoxyribonucleotide primers (random octamers).

10× dNTP mix: 1 mM dCTP, 2 mM dATP, 2 mM dGTP, 2 mM dTTP in TE buffer, stored at -20

Exo-Klenow Fragment: 40 U/µl, Klenow fragment in 100 mM Potassium Phosphate (pH 7.0), 1 mM DTT, and 50% Glycerol.

1 mM Cy3-dCTP (Amersham, Cat. PA53021)

1 mM Cy5-dCTP (Amersham, Cat. PA55021)

NucAway spin columns (Ambion, Cat. 10070)

Hybridisation buffer : ULTRAHYB (Ambion, Cat. 8670). Aliquots stored at -20°C. Remove and heat to 70ºC prior to use, making sure that all components are in solution.

3 M NaAc pH 5.2 AMBION (Cat. 9740): store at room temperature.

mouse Cot-1 DNA (Invitrogen, Cat. 18440-016): 1 mg/ml in 10 mM Tris-HCl (pH 7.4), 1 mM EDTA. Store at -20°C.

100% Ethanol, 80% Ethanol

Yeast tRNA (Invitrogen, Cat. 15401-029): Store at -20°C.

Cysteamine: Sigma, 0.2 M in Water, filtered to be sterile, pre-warmed at 37°C before use.

2.3.11.3 Random labelling DNA for array CGH

500 ng DNA sample was mixed with 60 µl 2.5× random primers solution to make up final volume of 130.5 µl with water. This DNA/primer mixture was denatured at 100°C for 10 minutes followed by immediately cooling on ice. Labelling reaction (150 µl) was made by adding 15 µl 10× dNTP, 1.5 µl Cy3 or Cy5-labelled dCTP (GE Healthcare, PA53021 and PA55021, 1 mM, final 10 µM) and 3 µl Klenow enzyme on ice (final 0.8 U/µL). Then it was mixed well and kept at 37°C overnight.

To purify labelled DNA, ProbeQuant G-50 Micro Columns (GE, Cat.No. 27-5335-01) were used. The pre-loaded columns were spun at 4300rpm for 1.5 min before loading. ~74 µL labelling reaction per column (2 columns per labelling) were loaded carefully followed by spinning for 2 min at 4300rpm (750 g) to collect the labelled DNA. It then was assessed by a Nanodrop in the “Micro-array” mode to measure Cy3 and Cy5 labelled DNA concentration by detecting the fluorescence signal. Proper parameter will be: Cy3: ~7 pmol/uL, DNA 370 ng; Cy5: ~5 pmol/uL, DNA 350 ng.

2.3.11.4 Hybridization

The following mixes per array were prepared and placed at -70°C for at least 30 min to precipitate fluorescence labelled DNA.

In 2 mL tube:

Cy3-labelled DNA	~180 µl (final ~37 ng/µL)
Cy5-labelled DNA	~180 µl (final ~35 ng/µL)
mouseCot1 DNA	135 µl (final 0.075 µg/µL)
Yeast tRNA	6 µl (final 0.33 µg/µL)
3 M NaAc pH 5.2 (Ambion #9740)	50 µl (final 83 mM)
100% ethanol	1250 µl

Then labelled DNA was centrifuged for 30min at 15000 rpm at 4°C. Precipitated DNA was washed once by 600 µl 80% ethanol followed by a quick centrifuge at 13000 rpm for 5min. Trace volumes of ethanol were pipetted off and a purple pellet was left in the tube. After air drying for approx. 5 min at 70°C, 10ul of Cysteamine and 35µl (or 70µl for the higher resolution arrays) of pre-warmed hybridisation buffer (Ambion) were applied to each pellet

and incubated for 15 min at 70°C with frequent mixing during the first 5 minutes.

The tube was incubated at 37°C for 1–2.5 hours. A gentle pulse spin before 37°C incubation is recommended. Then the contents (~40 µL) in the tube were applied onto a coverslip (24 x 36 mm) and this was placed on the warmed array. The array was hybridized with labelled DNA in slide booster at 37°C for 16–24 hours. 15 uL AdvaSon™ Coupling Liquid for SlideBooster hybridizations was added below the array to distribute DNA on the array evenly. Water was used to maintain humidify in the hybrid booster.

2.3.11.5 Array washing

After the arrays were hybridized they were rinsed in 1×PBS (with 0.8 µM cysteamine) to removed coverslips. Then they were sequentially washed in the following buffers to achieve an increasing stringency. All solutions supplemented with 10 mL 0.2M cysteamine in 2500 mL as an antioxidant.

10 min at room temperature in PBS with 0.1% Tween-20
30 min at 55°C in 0.1X SSC
15 min at room temperature in 1XPBS
2 min at room temperature in PBS with 0.05% Tween-20
30 sec–1 min at room temperature in water

The arrays were centrifuged at 2200 rpm for 3 minutes to dry then they were ready to be scanned.

2.4 RNA methods

2.4.1 RNA isolation

Total RNA was extracted by RNAqueous™ Kit (Ambion) following the manufacturer's instructions. Briefly, cells were cultured in a 24-well plate without feeder cells until confluent. 700 µL lysis/binding solution was added to each well to lyse cells. An equal volume (700 µL) of 64% ethanol was added to the lysates. These mixtures were applied to filter cartridges separately followed by a centrifugation at 10,000 g for 30 seconds. The flow-through was discarded. A first wash was conducted by applying 700 µL wash solution 1 to a filter cartridge followed by a centrifugation at 10,000 g for 30 seconds. The flow-through was discarded. Filter cartridges were then washed twice with wash solutions 2 and 3 with the same wash procedure. Finally, filter cartridges were put into a new collection tube. 50 µL of

elution solution was applied followed by centrifugation at 10,000 g for 30 seconds. The concentration of the eluted RNA was measured by a spectrophotometer and stored at -80°C.

2.4.2 Expression arrays - RNA isolation and sample labelling

RNA was isolated using the RNAqueous™ Kit (Ambion). RNA samples were amplified and labelled with biotinylated uridine triphosphate (UTP) using the Illumina® TotalPrep RNA Amplification Kit (Cat No.IL1791) following the manufacture's instruction (illustrated in Figure 2-23). Briefly, total RNA was reverse transcribed to produce the first strand complementary DNA (cDNA) using oligo d(T) primers with a T7 RNA polymerase binding sequence (T7 promoter) at the primer's 5' end. Then the RNA strand of the RNA-DNA hybrid duplex is degraded by RNaseH. The second strand of cDNA is synthesized by DNA polymerase I (Pol I) and T4 DNA polymerase. A purification step is applied to remove degraded RNA, excess primers, enzymes and salts. Finally, antisense RNA is transcribed and amplified by T7 RNA polymerase. Biotinylated uridine triphosphate (UTP) is used to detect the signal from the complementary RNA. Amplified cRNA is purified to remove enzymes and salts. The biotin labeled cRNA is hybridized with the probes on Illumina® Mouse-6 Expression BeadChip. Each sample was hybridized individually. Wash procedures are applied to remove non-complementary hybridized cRNA. Cyanine3 (Cy3, wavelength 532 nm) conjugated streptavidin is used to visualize the biotin labelled cRNA.

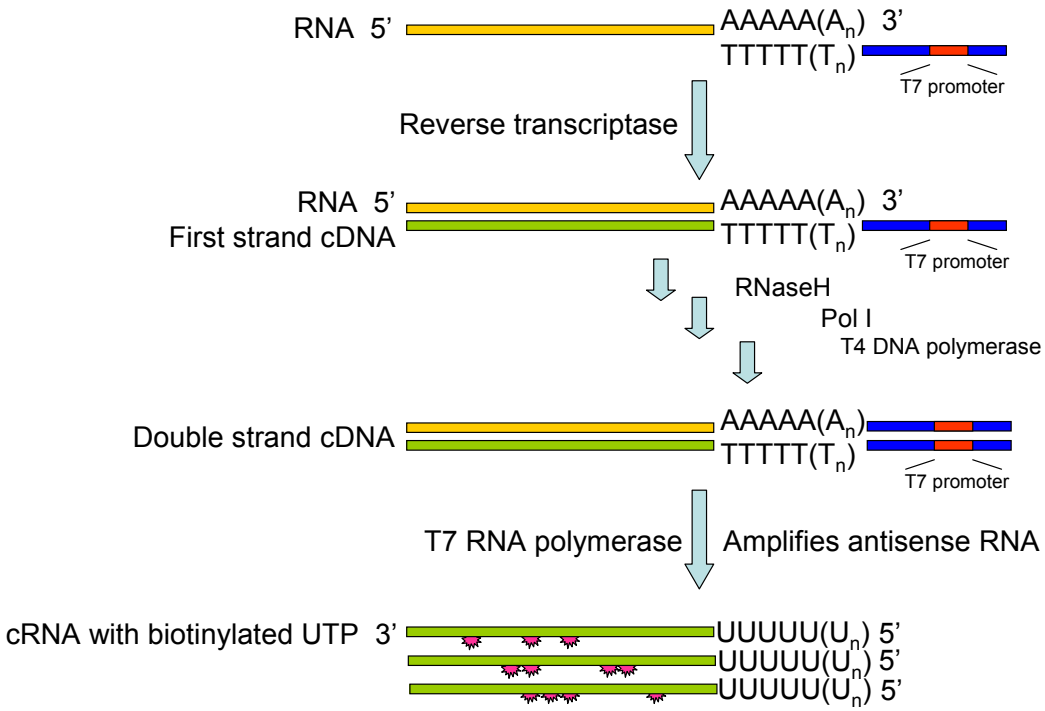


Figure 2-23 RNA amplification and labelling for expression array

Total RNA is reverse transcribed to produce the first strand cDNA from oligo d(T) primers containing a T7 promoter sequence at the 5' end. The RNA strand in the RNA-DNA hybrid duplex is degraded with RNaseH. DNA polymerase I (Pol I) and T4 DNA polymerase then synthesize the second trand cDNA with T7 promoter sequence at the 3' end. Double-strand cDNA is purified to remove RNA, primers, enzymes and salts before antisense RNA is transcribed *in vitro* by T7 RNA polymerase. Uridine triphosphate (UTP) used in the *in vitro* transcription is labeled by biotin (red star). Biotinylated UTP incorporated in the amplified complementary RNA (cDNA) is visualized on the array by using Cy3 conjugated streptavidin. This figure is modified from Van Gelder *et al*. (Van Gelder *et al.* 1990).

2.5 Statistical analysis of arrays

2.5.1 Array CGH

Reciprocal fluorescence array CGH was mainly conducted to compare 6TG-resistant mutant and reference sample (*Blm*-deficient NGG5-3 cell line). DNA of NGG5-3 was obtained from the passage immediately before the starting point of generating the mutation library (retrovirus tagging). Following stardard washing protocols arrays were scanned using an Agilent Technologies DNA microarray scanner. Images were analysed using BlueFuse software. Within an experiment, Log_2(ratio) between -0.29 and +0.29 for the signal from any BAC probe was regarded as no copy number change. When more than one reciprocal fluorescence array CGH were analysed in a batch, raw data was used without excluding any BAC probes. Thus, more copy number changes may be detected in a single mutant genome. However, a large scale normalization analysis has been conducted, generating a better standard to distinguish if a BAC probe was amplified, deleted or its copy number remains the same.

2.5.2 Expression array

2.5.2.1 Detection P-value

A whole genome expression array, Illumina® mouse-6 beadchip was used to attain genome-wide transcriptional information from all types of cells analysed. A detection P-value was calculated for probes on all Illumina® arrays using BeadStudio (version 2.3.41) software. The detection P-value provides a measure of the probability of a measured signal being due to hybridization with Cy3 labelled complementary RNA rather background non-specific binding. There are >700 negative control probes on Illumina® whole genome bead arrays. These negative control probes do not have any specific target transcripts in the mouse transcriptome. Any signals detected with these probes are regarded as non-specific binding. The detection P-value is calculated by ranking the signal of the negative control probes with the signals of other probes. Thus, the lower a detection P-value, the higher probability there is that the linked probe has bound specifically.

2.5.2.2 Combined P-value and transcripts' presence

A combined P-value was obtained for each probe to indicate signal consistency between replicates. AB1 and AB2.2 wild type cells had 2 biological replicates; each mutant as well as *Blm*-deficient cell and *Dnmt1*-deficient cell, had 3 biological replicates. All RNA samples for

replicate samples were isolated independently and array hybridizations were conducted separately. Analysis of each array generated a detection P-value for each probe.

The combined P-value of a probe is a measure of the probability that the signal for a particular probe is a background signal. Thus, if a specific probe's combined P-value is 0.05, it means there is 5% probability that this probe has the same signal level as background. A combined P-value of 0.05 was set to distinguish the transcriptional status of a probe's target sequence. A combined P-value between 0 and 0.05 (exclusive) is interpreted as indicating that the probe sequence is present in the RNA of replicates. A combined P-value between 0.05 (inclusive) and 1 means the relevant probe is absent in the RNA of replicates. This criterion to distinguish presence and absence of transcripts can be raised if more confidence in the expression data is needed.

2.5.2.3 Comparison analysis

Comparison analysis revealed the statistic significance of the relative expression the fold change between two cell lines. Within each pair of comparison, the expression level of each probe is compared between the replicates of one cell line and the replicates of another cell line. The adjusted P-value (adj.P.val) is used to indicate the probability that the two groups of replicates for a given probe are expressed at the same transcriptional level. The lower the adjusted P value, the lower is the probability that the two groups are expressed at the same level. Five thousand probes with the lowest adjusted P-value in each pair of comparison were selected for analysis. Comparison analysis does use the detection P-value and the combined P-value as parameters.

CHAPTER 3. EVALUATING SIZES OF HOMOZYGOUS DELETIONS

3.1 Introduction

Ionizing radiation has been used to generate random deletions in mouse ES cells and mice. Combining gene targeting technique and negative selection strategy, mutant cells with deletions at specific loci can be isolated (Kushi *et al.* 1998; Schimenti *et al.* 2000; Thomas *et al.* 1998; You *et al.* 1997). Kushi has measured the sizes of deletions generated by X-rays around the *Hprt* (Hypoxanthine-guanine phosphoribosyltransferase) locus on the X chromosome of the mouse E14 ES cell line (Kushi *et al.* 1998). He found that these deletions ranged from 200 to 700 kb (kilobase pairs). Schimenti used F1 hybrid mouse ES cells and made the deletions selectable by targeting a negative selection cassette Herpes Simplex Virus thymidine kinase (*HSV-tk*) to 3 loci on chromosome 5. The targeted ES cell clones were irradiated with 4 Gray of gamma rays and clones with deletions were isolated by negative selection for loss of the *HSV-tk* gene in FIAU (1-2'-deoxy-2'-fluoro-beta-D-arabinofuranosyl-5-iodouracil) (Schimenti *et al.* 2000). Using this approach, heterozygous deletions up to approximate 70 Mb of proximal mouse chromosome 5 were generated in cells. Sizes of the deletions were measured by examing the presence of the microsatellite markers. Forty per cent (11/26) of these clones injected were able to produce germ line chimeras.

However, these approaches can only measure the sizes of heterozygous deletions. As I planned to generate homozygous deletions in *Blm*-deficient cells, the sizes of this type of deletions need to be determined for at least three main reasons: First to determine if deletions can be detected using a 200 kb resolution array CGH; second these data are needed to estimate the size of the library; and the response of the *Blm* cells to irradiation was unknown. If homozygous deletions cover any viability associated genes, cells cannot survive then these deletions will not be observed. Therefore, a gene rare region is more appropriate to conduct experiments to measure sizes of homozygous deletions.

3.1.1 Gene deserts

A gene desert is defined as a genomic region larger than 500 kb in length without a gene. About 20–25% (~605 Mb) of the human genome is gene desert regions. This number falls to 208 Mb if a gene desert is defined as a region larger than 1 Mb without a gene (Nobrega

et al. 2003; Venter *et al.* 2001). Although gene deserts are devoid of genes, in many cases conserved elements can be identified across species. These elements have been shown to have enhancer activity in humans by constructing transgenic mice carrying a beta-galactosidase reporter system and these 300–800 bp conserved sequences upstream of a mouse heat shock protein 68 minimal promoter (Nobrega *et al.* 2003). Despite this, another study from the same group demonstrated that large homozygous deletion of gene desert regions does not have any impact on mouse development (Nobrega *et al.* 2004). Gene desert deletions (1.5 Mb, 0.8 Mb on chromosome 3, 19, respectively) were generated by Cre mediated recombination between targeted *loxP* sites. Mice homozygous for either deletion were viable and fertile without any distinguishable physiological features compared to wild type mice. A gene desert spanning 2.7 Mb (9.2–11.9 Mb position) on mouse chromosome 6 was chosen for this study (Figure 3-1). This region is highly conserved with the human genome (Figure 3-2, Figure 3-3). Like the mouse region, these regions in the human genome are gene poor, too.

In this chapter, a strategy was developed which could determine sizes of homozygous deletions generated by irradiation without complications arising from deleting essential genes. In order to do this, a design similar to that employed by Kushi *et al*. 1998 and Schimenti *et al*. 2000 was used. Briefly, a positive-negative selection marker (*PuroΔtk*) was targeted into a gene desert. A targeted single clone was expanded then irradiated. After irradiation, cells were selected in puromycin for the presence of the *PuroΔtk* marker. Then cells were cultured for multiple generations to allow LOH to occur. As a result of LOH, homozygous deletions at the targeted locus led to the loss of the *PuroΔtk* marker gene, therefore these cells can be selected in FIAU. Homozygous deletions can be analysed by comparative genomic hybridization (CGH) array.

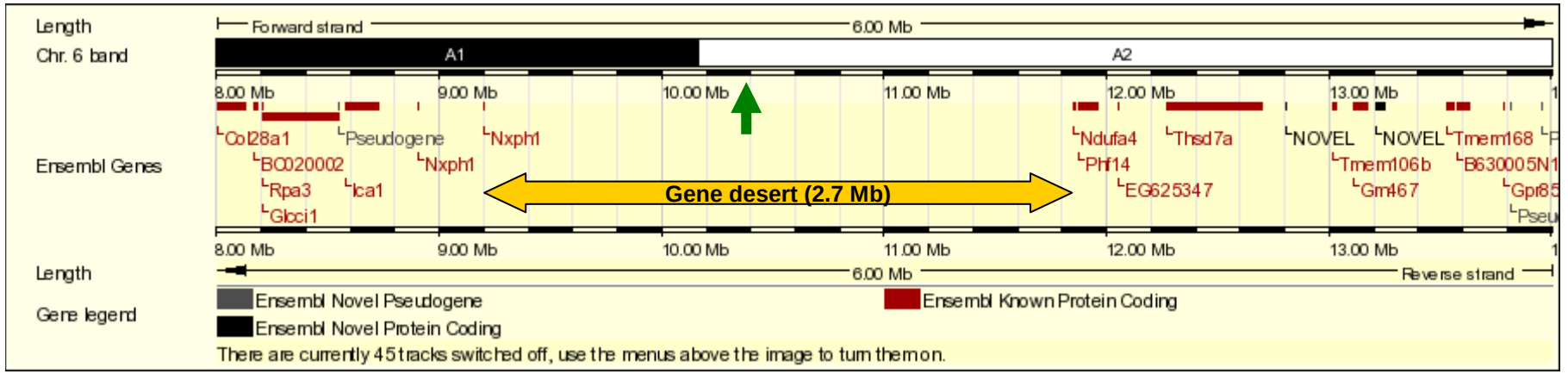


Figure 3-1 A gene desert on chromosome 6

The figure shows the 8–14 Mb region of mouse chromosome 6. From Ensembl gene *Nxph1* to *Ndufa4*, there is a 2.7 Mb gene desert region. The green arrow indicates the locus selected for targeting. From NCBI m36, Ensembl release 46.

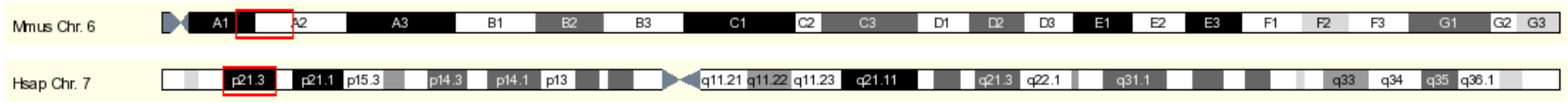


Figure 3-2 Conserved linkage between mouse chromosome 6 gene desert region and human genome

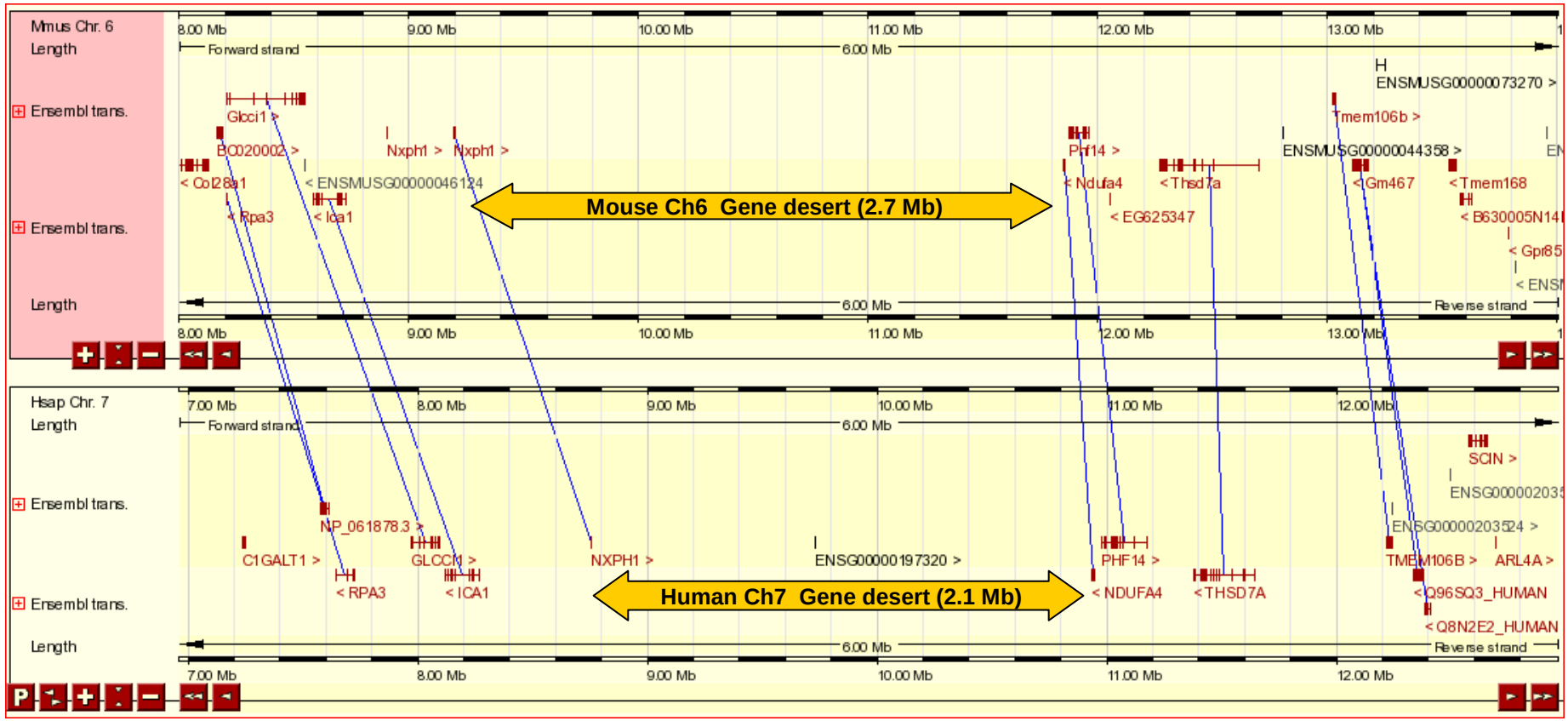


Figure 3-3 Syntenic genes around the chromosome 6 gene desert

The upper panel and the lower panel show 8 to 14 Mb of mouse chromosome 6 and 7 to13 Mb of human chromosome 7, respectively. Blue lines indicate conserved genes in these genomes. Sequence alignments from Ensembl also indicate the gene deserts in the above regions (yellow arrows) are conserved. From NCBI m36 (mouse) and NCBI 36 (human), Ensembl release 46.

3.1.2 A strategy to isolate mutants with homozygous deletions

To determine the relationship between radiation dose and deletion length in a neutral part of the genome, a positive-negative selection (PNS) marker (*PuroΔtk*) (Chen, Y. T. *et al.* 2000) was targeted into 10,440,045bp–10,440,812bp (replacing 766 bp sequence, NCBI m36, Ensembl release 46) in the gene desert region on chromosome 6 in the NGG5-3 *Blm*$^{-/-}$ cells.

The *PuroΔtk* cassette is selectable in both a positive and negative direction. Targeting events can be selected in puromycin media. Heterozygous deletions with loss of the marker can be directly selected in FIAU. Similarly, deletions of the targeted allele can be selected in FIAU after loss of heterozygosity (LOH).

Homozygous deletions of a gene desert region can be selected through the strategy illustrated in Figure 3-4. When the *PuroΔtk* cassette is targeted on one allele of the gene desert, targeted colonies can contain wild type cells. The colonies were expanded in a high concentration of puromycin media so that the contaminating wild type cells are killed because they do not contain the *PuroΔtk* cassette. After expansion the cells are irradiated to generate mutations, including deletions. A second round of puromycin selection is then initiated. Two populations of cells will survive and they are the cells with *PuroΔtk* at one allele and a deletion in the other allele on the homologous chromosome of the gene desert region, and the cells with the *PuroΔtk* cassette and mutations at other loci. By culturing the population for 15 days, sister chromatid exchange and LOH events occur. Homozygous deletions at the targeted locus will result in loss of the *PuroΔtk* cassette. A rough estimation of the frequency of sister chromatid exchange on chromosome 6 resulting in homozygous deletions at the targeted locus is about 7% (10Mb/150Mb, length of the proximal region to the gene desert/the whole length of chromosome 6). These cells are viable in media containing FIAU (Figure 3-4). The lengths of deletions can be identified using array CGH. However, other types of cells without the *PuroΔtk* cassette could be selected by FIAU, too. For example, cells with either heterozygous or homozygous deletions on the same chromosome as the gene desert will be FIAU-resistant; and some cells with deletions in other chromosomes can also be resistant to FIAU.

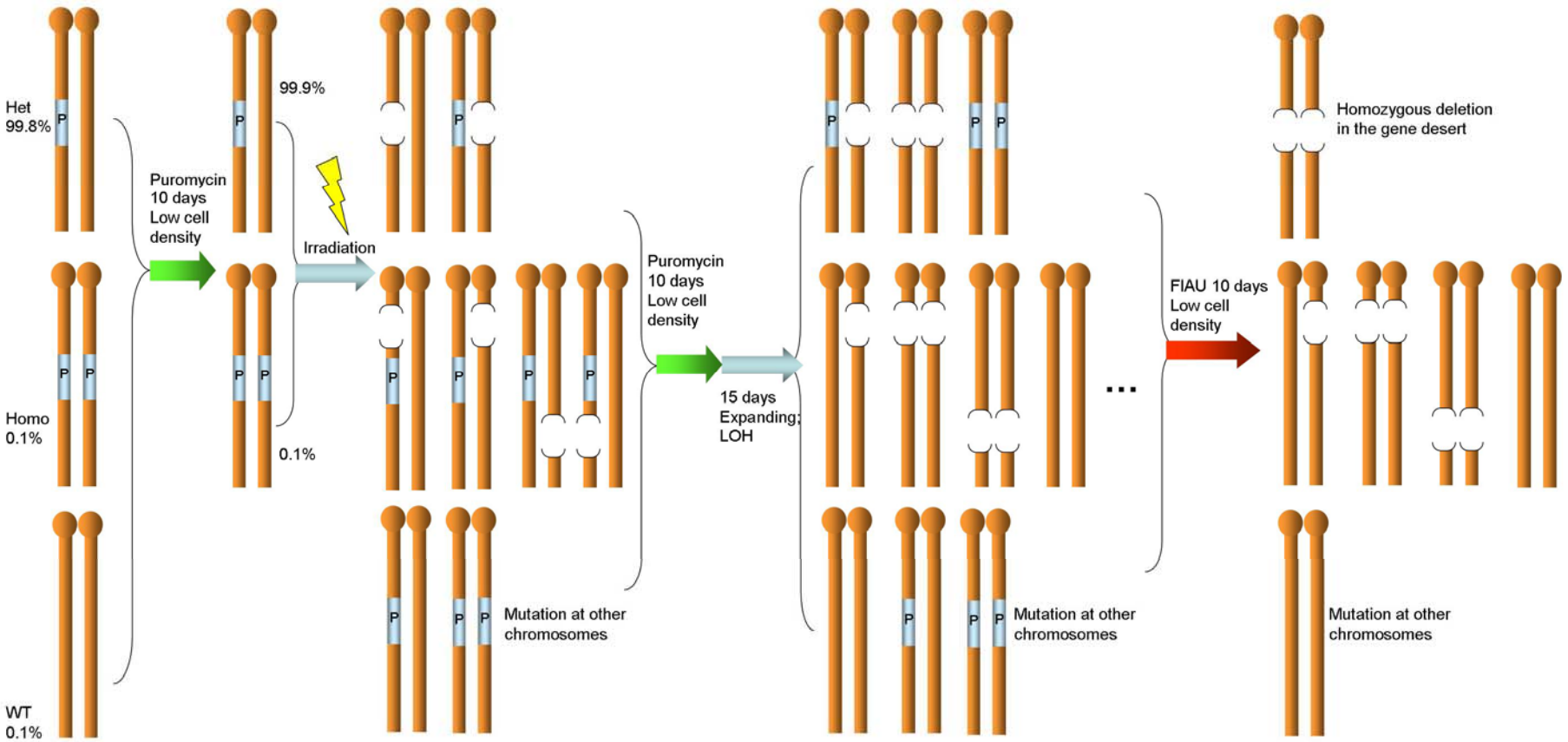


Figure 3-4 Design to generate homozygous deletions in gene desert region

The starting population is a targeted clone, which has the *PuroΔtk* cassette at the gene desert region. After 10 days puromycin selection on low density plated cells, wild type cells are removed. Irradiation generates deletions. Another 10 days of puromycin selects only those cells with *PuroΔtk* at one allele and a deletion in the other allele of the gene desert region or those cells with a deletion elsewhere. Sister chromatid exchange occurs and LOH in 15-day culture then FIAU selects homozygous deletions in the gene desert in spite of the presence of cells with both heterozygous and homozygous deletions at other loci.

3.2 Results

3.2.1 Chromosome 6 gene desert knock-in cell ZK2.1

The chromosome 6 gene desert targeting vector pZK-GD31 was constructed by recombineering and used for targeting in the *Blm*-deficient ES cell line NGG5-3 (Guo *et al.* 2004). pZK-GD31 was linearized with *Not* I, checked on an agarose gel to confirm the digestion, precipitated with ethanol and redissolved in water prior to electroporation.

Approximately ten million *Blm*-deficient ES cells (NGG5-3) and 20 µg linearized pZK1 were used in the electroporation. After electroporation, cells were plated onto 90mm feeder plates and puromycin selection was initiated 24 hours later. After 10 days selection, puromycin resistant colonies were visible. There were about 300 colonies per 10^7 cells plated. A total of 96 colonies were clonally isolated and screened for targeted clones by Southern blot analysis.

Using a 5' external Southern blot probe, targeted and wild type alleles can be distinguished. This 718 bp probe was amplified from genomic DNA (10434089 to 10434806 bp, NCBI m36, Ensembl release 46, primers 142.1 and 142.2, sequences at section 2.3.3). Genomic DNA was digested with *Hind* III, fractionated on a 0.8% agarose gel and then transferred to Hybond XL membrane and hybridized using a standard protocol. From 96 clones, one targeted clone (ZK2.1 G4) was identified with both a 15.3 kb wild type band and an 11.7 kb targeted band (Figure 3-5).

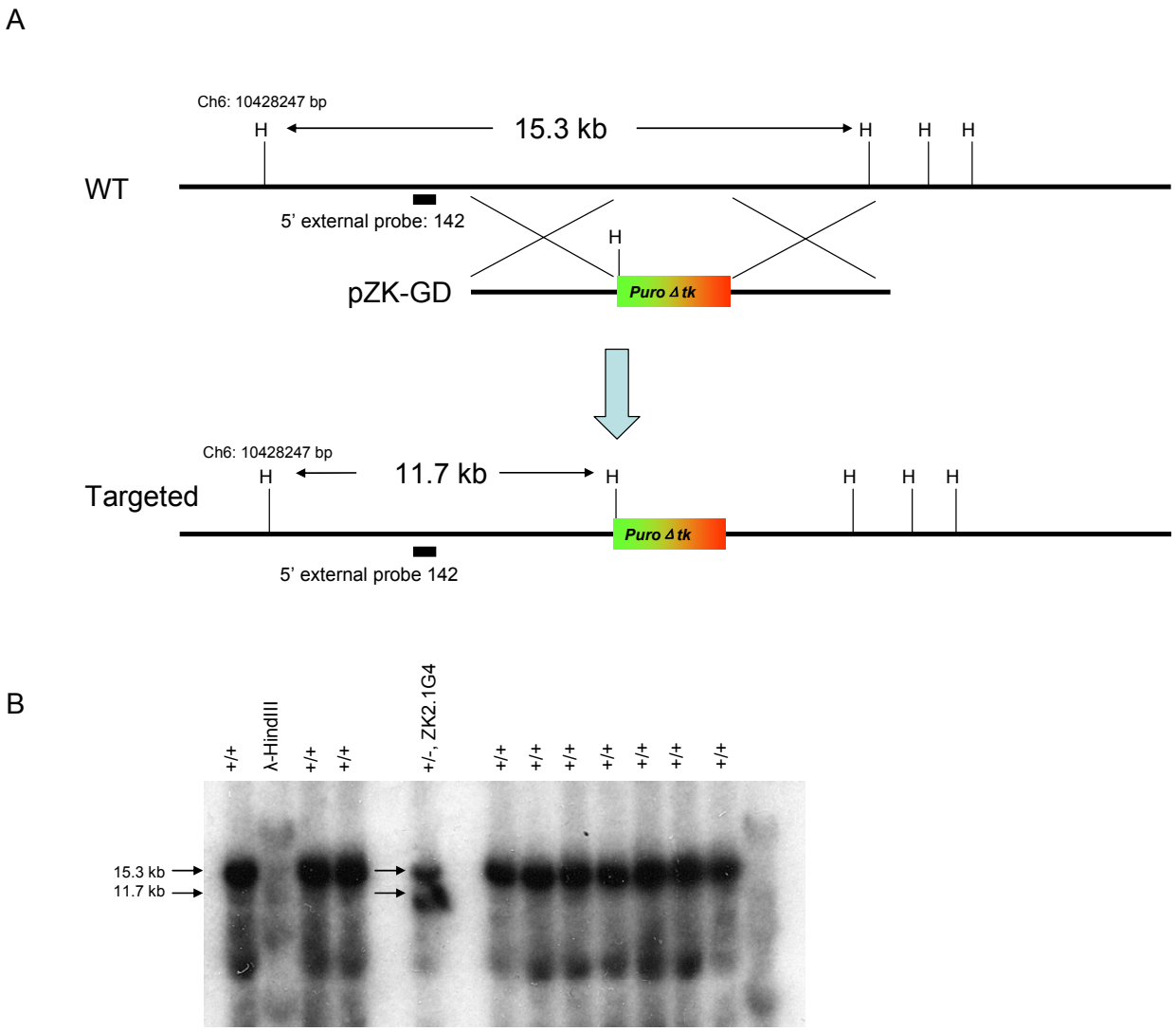


Figure 3-5 Chromosome 6 gene desert targeting

A. Schematic figure of Southern blot strategy for detecting gene targeting. DNA was digested with *Hind* III (H). Using a 5' external probe (probe 142), the wild type fragment is 15.3 kb while the targeted fragment is 11.7 kb. **B**. Southern blot screen targeted clones. Clone ZK2.1 G4 has both a wild type allele and a targeted allele.

3.2.2 Testing background FIAU resistance levels prior to selection for homozygous deletions

The targeted ES cell clone ZK2.1G4 was used to determine sizes of homozygous deletions. Clone G4 was cultured and expanded in puromycin media at low cell density for 10 days to eliminate trace numbers of wild type cells then the cells were pooled together. Before the purified ZK2.1G4 population was irradiated, one million cells were plated and cultured 90mm plates with FIAU media as a control. This control showed that FIAU-resistant background was high. A potential reason for the high background was the purity of the clone plated in FIAU. Therefore the cell line was cloned by dilution and single cell clones were isolated. Twelve clones were expanded and the targeted allele was confirmed by Southern blot. The FIAU resistance background was examined by a double selection in the puromycin and FIAU media. $1–2\times10^6$ cells were plated at the beginning of the double selection. The degree of the double resistance background in different clones varied as shown in Figure 3-6. Among clones 4, 6, 7, 8, 11 and 12, puromycin and FIAU-resistant background is between 10^{-4} and 10^{-2} per plated cell.

Because these 12 clones were subcloned, the chance of impurity was very low. In addition, none of them showed full resistance to either FIAU or puromycin+FIAU. One explanation was that mutations, which occurred during cell culture lead to the FIAU or puromycin+FIAU resistance. To test this assumption, another electroporation of the gene desert targeting vector was performed in both wild type and the *Blm*-deficient cells (NGG5-3) to generate random insertions. Transfected cells were selected in puromycin and two sets of 96-well plates were clonally isolated. ZK2.9, ZK2.10 (*Blm*$^{-/-}$ background) and ZK11, ZK12 (wild type background), were tested for resistance to puromycin and FIAU (Figure 3-7). As expected all (100%) of the transfected puromycin^R clones in the control AB1 cell line (wild type) were sensitive to FIAU. However, half of the puromycin resistant clones isolated in the *Blm*-deficient cells were resistant to FIAU though levels varied from 10^{-5} to 10^{-2}. This indicates that the *PGK-PuroΔtk*-pA cassette was intact and functional in wild type cells. While *PGK-puro* was still functional in the *Blm*$^{-/-}$ cells, the *Δtk* component seemed to have lost function in some of the cells in half of the clones. However, not all the cells in each clone were FIAU-resistant since individual colonies could be seen in individual wells instead of a lawn of resistant cells, suggesting that the *PGK-PuroΔtk*-pA cassette was not lost at the stage when it was integrated into cell genome, rather this occurred during the expansion of the clone. Because of this, FIAU selection could not be used to isolate homozygous mutations in the *Blm*-deficient ES cells.

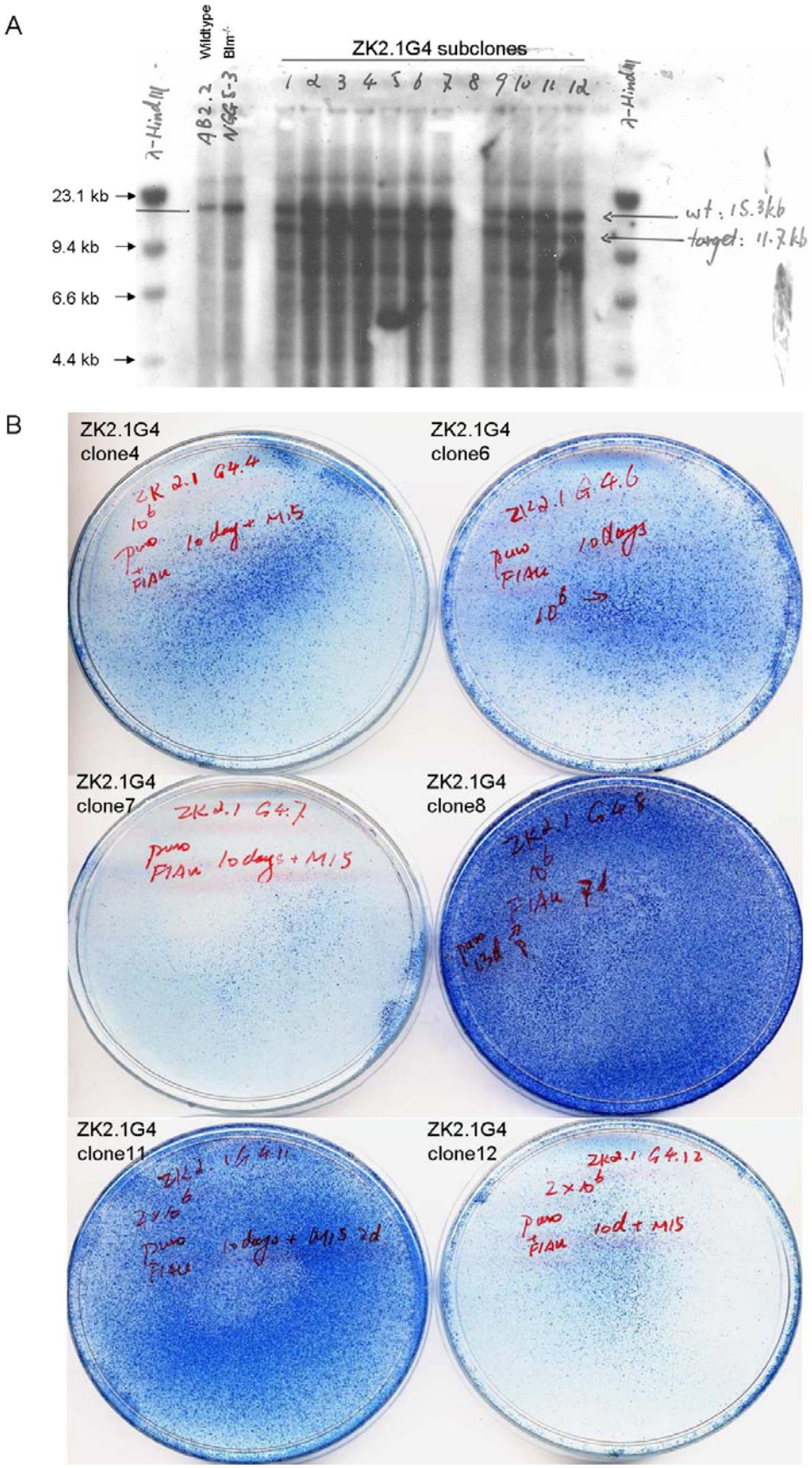


Figure 3-6 **Gene desert targeted clone with resistance to puromycin and FIAU**

Figure 3-6 desert targeted clone with resistance to puromycin and FIAU

A. Southern blot analysis confirmed subclones 1–12 from ZK2.1G4 gene desert targeted cell. Wild type cell (AB2.2) DNA and *Blm*-deficient cell (NGG5-3) DNA were used as controls. Both controls have the wild type allele at 15.3 kb. Subclones 1–12 all have the wild type allele and the targeted allele (11.7 kb). **B**. Puromycin and FIAU double resistance background of two subclones of gene desert targeted cells. 2 million cells of both clone 11 and 12 and one million cells of clone 4, 6, 7 and 8 were plated in 90mm plate. These subclones were cultured in puromycin+FIAU media for 10 days then cells were grown in M15 normal media for 2 days before they were stained by methylene blue. The clones have different levels of background.

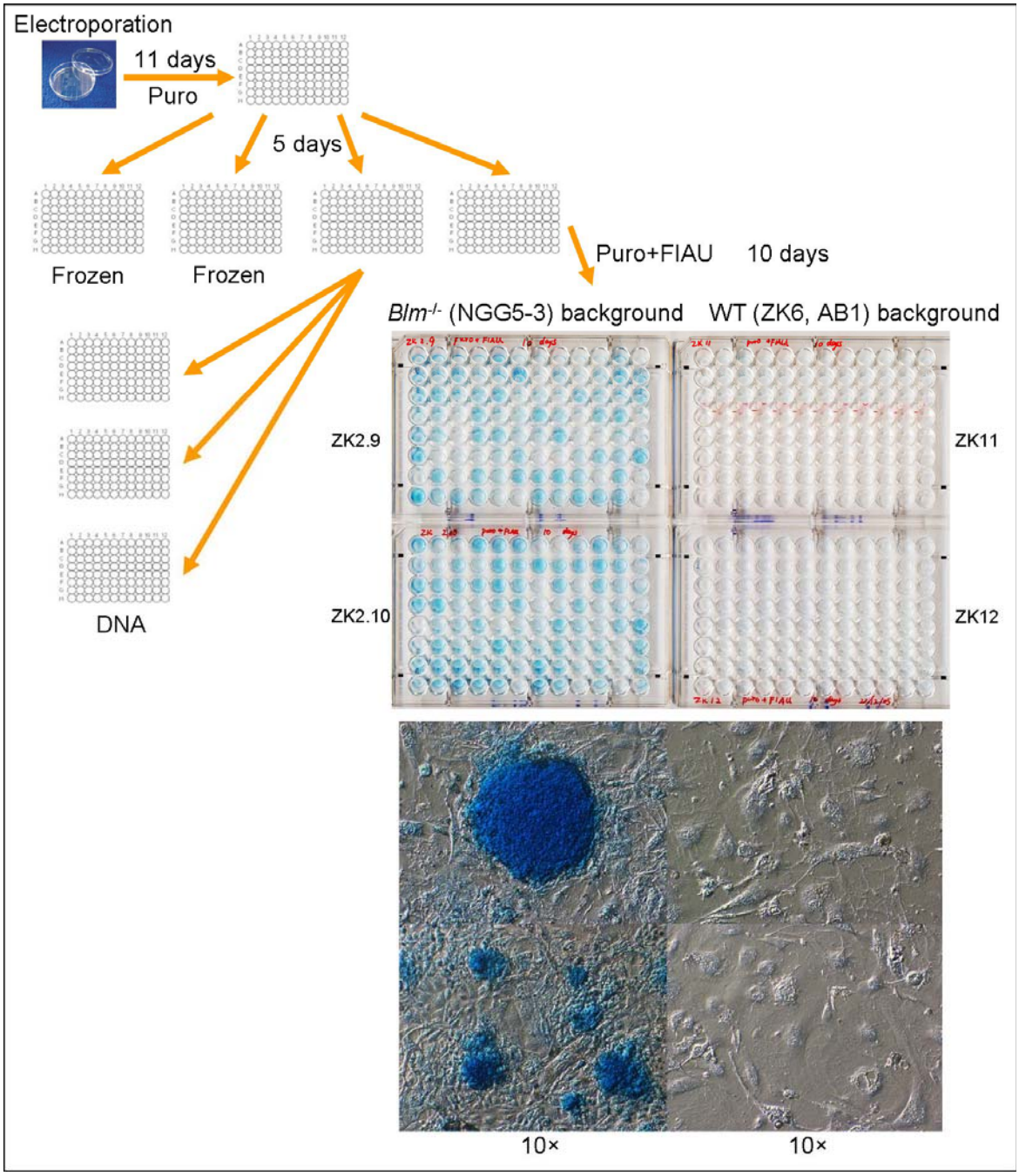


Figure 3-7 Puromycin and FIAU double resistant background in NGG5-3 cells

The gene desert targeting vector pZK-GD.31 was electroporated into wild type (ZK6) and *Blm*-deficient (NGG5-3) cells, which were selected in puromycin media for 11 days. Then colonies were picked up, passaged and duplicates were plated for selection in (puromycin+FIAU) media. Cells were stained with methylene blue. Resistant clones could be seen in half of the wells in the *Blm*-deficient cell plates, but not in wild type cell plates.

3.2.3 Puromycin and FIAU double resistant clones contain the *PuroΔtk* cassette

It is known that the *Blm*-deficient ES cells have an elevated rate of mitotic recombination between non-sister chromatids (Luo *et al.* 2000). Thus, the occurrence of puromycin and FIAU double resistant clones in gene desert targeted *Blm*-deficient cells but not wild type cells can be explained by loss-of-function mutations of the *Δtk* part of the *PuroΔtk* cassette, such as deletion or point mutations, which must leave the *PGK-puro* cassette functionally active and result in lost of the function of *Δtk*. To test this hypothesis, *Eco* RI digested DNA samples from 22 double resistant gene desert targeted clones and the *Blm*-deficient cells were hybridized with the *PuroΔtk* probe and the *Δtk* probe separately (Figure 3-8). All DNA samples of these mutants, except the *Blm*-deficient cell DNA, showed the expected band after Southern blotting, suggesting the *PuroΔtk* fusion gene did not experience any large fragment loss. Another possible explanation of the puro^R, FIAU^R phenotype is the *Δtk* cassette contains small frame-shift mutations, which did not affect the function of puromycin resistance cassette. Therefore these clones remain puromycin and FIAU resistant. However, the frequency of FIAU resistance in puro^R cells is about 10^{-5} to 10^{-2} (Figure 3-7), which is much higher than the spontaneous mutation rate of about 5×10^{-6} per locus per generation in normal cells and the LOH rate of ~4.2×10^{-4} events per locus per cell per generation in the *Blm*-deficient ES cells. Thus, small frame-shift mutations may not be the reason of the fluctuating frequency of puromycin and FIAU double resistant cells.

One other hypothesis is raised after considering that the fluctuating frequency of the double resistant cells indicates the loss-of-function of the *Δtk* part was generated in the process of cell culture and the original mutants regenerated various numbers of offspring. Such a phenomenon is similar to the LOH events of a pre-existing heterozygous mutation. Therefore, it is assumed that a gene required for FIAU metabolism was heterozygously mutated in the *Blm*-deficient cells. Homozygous mutations of this gene led by LOH can result in FIAU-resistance although the *Δtk* part of the *PuroΔtk* fusion gene may remain functional. And LOH can occur anytime during cell culture so that the number of FIAU^R cells varies. The NGG5-3 *Blm*-deficient ES cell line was derived from wild type ES cells after multiple gene targeting experiments and subclonings, thus this cell line is very likely to have acquired mutations prior to my experiments.

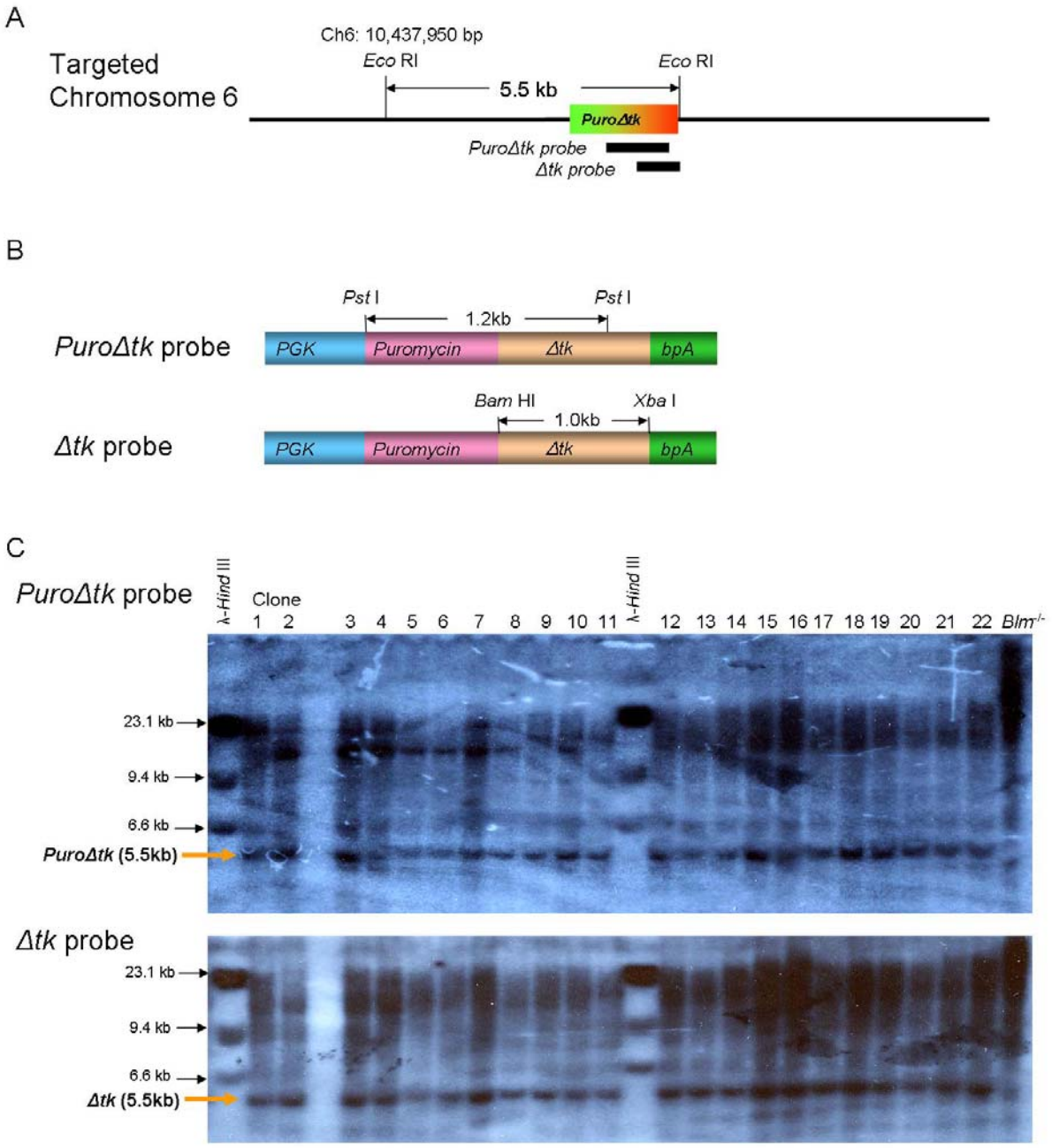


Figure 3-8 **Southern blot analysis confirms the presence of *PuroΔtk* cassette**

Figure 3-8 Southern blot analysis confirms the presence of *PuroΔtk* cassette

A. Southern blot strategy to identify if the *PuroΔtk* cassette is intact. The *PuroΔtk* targeted allele is shown. After *Eco* RI digestion, hybridization with either *PuroΔtk* probe or the *Δtk* probe will generate a 5.5 kb band. **B**. Generation of the *PuroΔtk* probe or the *Δtk* probe. The *PuroΔtk* probe is a 1.2 kb fragment cut by *Pst* I from vector YTC37 (Chen, Y. T. *et al.* 2000). The *Δtk* probe is a 1.0 kb fragment cut by *Bam* HI and *Xba* I from vector YTC37. **C**. Southern blots show that 22 puromycin and FIAU double resistant gene desert targeted clones contain the whole *PuroΔtk* cassette. DNA from these clone and the *Blm*-deficient cells was hybridized with the *PuroΔtk* probe or the *Δtk* probe. The 5.5 kb band can be seen in all double resistant clones but not in *Blm*-deficient cells. There is a weak background band between 9 to 24 kb in all these cells. And its existence was also observed in other experiments when hybridizing the *PuroΔtk* probe with DNA of cells derived from the NGG5-3 *Blm*-deficient cell line (Figure 7-11–14).

3.3 Discussion

The positive-negative marker *PuroΔtk* was successfully targeted into chromosome 6 of *Blm*-deficient mouse ES cells. One targeting event was found in 96 puromycin resistant colonies. This targeting efficiency is much lower than what was typically achieved in the laboratory (10–50%). The low targeting efficiency was probably caused by the accessibility of the chromatin in the gene desert region. This locus is near the Centromere (proximal 10Mb out of 150 Mb) of mouse chromosome 6 and is repetitive. This might also play a role in the low targeting efficiency.

It was expected that targeting of *Δtk* would facilitate the selection of deletions in FIAU. The selection of such deletions relies on a low background of loss of *Δtk*. However, a high background was observed in the gene desert targeted cells. One explanation of this high background was the impurity of the isolated targeted clone. Therefore this targeted cell line was subcloned by dilution. Twelve subclones with the confirmed gene desert targeted allele also showed similar resistance to FIAU. Because these clones were not fully resistant to FIAU, loss of *Δtk* function did not occur at an early time point. The frequency of the puromycin and FIAU double resistant background in these subclones is between 10^{-4} and 10^{-2}. It was assumed loss of *Δtk* function occur during expansion of the clones. To test this assumption, the gene desert targeting vector was electroporated into wild type cells and *Blm*-deficient cells. Transfected cells were selected in puromycin and 192 colonies from each genetic background were clonally isolated. All of the transfected puromycin^R clones in the control AB1 cell line (wild type) were sensitive to FIAU, which was expected. Some cells in half of the puromycin resistant clones isolated from the *Blm*-deficient cells were resistant to FIAU. This indicates that the *PGK-PuroΔtk*-pA cassette was intact and functional in wild type cells. However, the *PGK-puro* component was functional in the *Blm*$^{-/-}$ cells but the *Δtk* component had lost its function in some cells of the half of the clones. The loss-of-function of the *Δtk* component occurred during expansion instead of prior to homologous recombination or random integration because individual FIAU-resistant colonies appeared in FIAU selection instead of a lawn of resistant cells. Therefore, the puromycin and FIAU double resistance must have a genetic basis, which occurs in the process of cell growth. In addition, these events are independent of genomic loci because they were found in both targeted cells and random integrated cells. Moreover, it is associated with the *Blm* deficiency.

To examine if this genetic events occurred at the *PuroΔtk* locus, a Southern blot analysis

was conducted. DNA of twenty-two double resistant clones from the gene desert targeted clone ZK2.1G4 was hybridized with both the *PuroΔtk* probe and the *Δtk* probe. The analysis confirmed the presence of the whole fusion *PuroΔtk* gene in those mutants. However, this does not rule out the possibility of small frame-shift of the *Δtk* cassette, though this is unlikely to occur with such a frequency.

Another hypothesis to explain the presence and the fluctuating frequency of the double resistant cells is that these phenomena are due to the LOH events of a pre-existing heterozygous mutation within FIAU metabolism pathway. The NGG5-3 *Blm*-deficient ES cell line had experienced four homologous recombination experiments and a number of subcloning procedures, thus this cell line is very likely to have acquired mutations prior to my experiments. LOH events of this mutation can cause a homozygous mutation, which lead to the resistance to FIAU in cells. When a cell with such a homozygous mutation expand, more and more cells are resistant to FIAU although the number of these cells are determined by the time point of the LOH event and the doublings of the cells after the LOH event. However, this process does not necessarily involve the inactivation of the *Δtk*.

CHAPTER 4. ATTEMPT TO GENERATE A *Blm*-DEFICIENT CELL LINE WITH BAC ACCEPTOR

4.1 Introduction

Homozygous deletions can occur in a *Blm*-deficient mutation library, followed by irradiation and the loss of heterozygosity (LOH). To complement mutant gene function, integrating BAC genomic DNA into a specific locus is a possible approach. BACs are able to carry original genomic sequences and cis-regulatory elements of a gene. Thus, the expression of the introduced gene is more likely expressed in its original status. In addition, single copy transgenes through BACs avoid unpredicted effects resulted from various copy numbers. In this chapter, the strategy to generate a *Blm*-deficient ES cell line with a BAC acceptor at the *Hprt* locus is described.

The gene-targeting technique was used to knockout the *Blm* gene. Through homologous recombination, exons 2 and 3 of the *Blm* gene were replaced by a positive and negative selection cassette *PGK/EM7/neo-MC1-tk* which was flanked by *FRT* sites. This cassette can be selected in *E. coli* by Kanamycin through expression of the *Neo* cassette driven by the *EM7* promoter; it can also be selected by G418 through the expression of the *Neo* cassette driven by the *PGK* promoter. The whole selection marker can be removed by Flpe through the flanking *FRT* sites. Loss of *MC1-tk* can be selected with FIAU. The deletion of exon 2 and 3 of the *Blm* gene disrupts the downstream reading frame and results in loss-of-function of the BLM protein.

The strategy pursued in this project is to generate homozygous mutations with large deletions which can be identified by array comparative genomic hybridization (CGH). However, there will not be enough evidence to conclude which single gene is associated with the phenotype given that deletions range from several hundred kilo base pairs to million base pairs. Recovering the normal phenotype by expression of candidate gene facilitates the identification of the causal mutation. When a normal gene, introduced from outside of a cell, converts a mutant phenotype to a normal phenotype, it can be concluded that the gene is responsible for the mutant phenotype. There are several ways to introduce a gene into cells including: I. Integrating cDNA into a genome to achieve stable expression. II. Introducing the genomic locus, for example by integrating bacterial artificial chromosomes (BACs) into a genome to achieve stable expression.

A cDNA can complement a mutant's function by providing the correct transcripts. Using this principle, a technology called expression cloning been used in many systems, for instance to clone the co-receptor of human immunodeficiency virus (Deng *et al.* 1997) and the co-receptor of hepatitis C virus (Evans *et al.* 2007). In these projects, cDNA libraries were transformed into a specific cell line, which should not be infectable by the type of virus. Cells which gain the ability to be infected after cDNA transformation can be used to identify the specific cDNA, which codes for the protein which restores viral infection. The drawback of expression cloning is that some genes have very long transcripts therefore cDNAs of these genes are under represented in cDNA libraries, which limits the applicability of this method. In addition, some other reasons limit the application of the expression cloning, for instance, not all cDNAs express appropriately in cell lines; and alternative splice structures can be lost through cDNA expression; and physiological expression levels are hard to achieve.

BACs have insertions with lengths ranging from 100-200 kb of genomic DNA thus are long enough to host most genes with their regulatory elements. Thus, introduction of BACs into a cell can restore gene function when they are integrated into the host genome. As a variety of BAC libraries are ready to use, introducing a BAC into a cells is a convenient approach to restore the normal function of mutant genes.

Cells with a BAC acceptor locus can efficiently integrate large DNA fragments into the genome to provide stable gene expressions. This kind of integration can be achieved either randomly or locus specifically. Random integration is easier to do and saves time. But it can result in multiple copies of transformed genes (Antoch *et al.* 1997) and cannot ensure intact DNA integration (Chandler *et al.* 2007). Also insertion sites vary thus gene expression levels are much more variable (Korn *et al.* 1992; Nobrega *et al.* 2003). Locus specific integration can be achieved by recombination mediated cassette exchange (RMCE) (Baer *et al.* 2001), which generates a single copy integration of the exogenous DNA. RMCE requires a recombinase such as Cre. A genomic DNA fragment flanked by two hetero-specific *lox* sites can be replaced by a plasmid or BAC DNA fragment flanked by the same two *lox* sites. In this way, the expression of the integrated gene can be well controlled.

The murine X-linked *Hprt* locus is a relatively neutral site for RMCE since mutations in this locus do not impair mouse viability (Kuehn *et al.* 1987; Kushi *et al.* 1998). In addition, it has been shown that BACs can be integrated into the mouse *Hprt* locus and transmitted through germ line transmition. Mice with such a BAC expresses in the expected tissue specific

manner (Heaney *et al.* 2004). I planned to modify an existing BAC acceptor cell line CCI18 (Prosser 2007 unpublished). This cell line is wild type for *Blm* function thus it is necessary to generate a *Blm*-deficient cell line with the ability to integrate genomic DNA fragments from a BAC into the *Hprt* locus. In the CCI18 cell line, a *loxP*- and *Lox511*- flanked *PuroΔtk* cassette was targeted to intron 2 of the *Hprt* gene through homologous recombination. Although this is inserted in the gene the function of *Hprt* is not disrupted.

As I needed to use the *PuroΔtk* cassette in the following retroviral infection experiments, I replaced the *PuroΔtk* cassette in the CCI18 cell line with a blastistidin selection cassette (*Bsd*) through RMCE. Then the *Blm* gene was targeted by a targeting vector with a *FRT* sites flanked *PGK/EM7/neo-MC1-tk* cassette. By recycling the selection marker through Flpe recombinase, both alleles of *Blm* gene can be mutated.

The Mouse BAC libraries RPCI-23/24 have genomic DNA fragments flanked by the *loxP* and *Lox511* sites. After introducing a eukaryotic selection marker, for example the *neomycin* gene between two *lox* sites, mouse genomic DNA in these BAC libraries can be easily introduced into *Hprt* locus by Cre mediated RMCE. A single copy of genomic DNA can be integrated at the *Hprt* locus, resulting in a loss-of-function of *Hprt*. RMCE cells can be isolated by G418 for integration of *neomycin* gene and by 6TG for the inactivation of the *Hprt* gene. Regulated by the integrated gene's own regulatory elements, functional expression will be achieved.

4.2 Results

4.2.1 Generation of the ZK6 wild type BAC acceptor cell line

The ZK6 *Bsd* BAC acceptor cell line was generated by replacing *PuroΔtk* in the CCI18 cell line through recombination mediated cassette exchange between the cassette provided by pZK33 and the *PuroΔtk* cassette at the *Hprt* locus (Figure 4-1). Bcyd selection results in two types of cells: correct RMCE events and the random integration of the *Pcmv-EM7-BSD-SV40pA* cassette. As a control the *Bsd* RMCE plasmids pZK33.4 and pZK33.3 were transformed into 1–3×10^7 CCI18 cells separately without CRE. After electroporation the cells were selected in Bcyd and FIAU. Five and one colony grew in the above plates, respectively, which mean background is low.

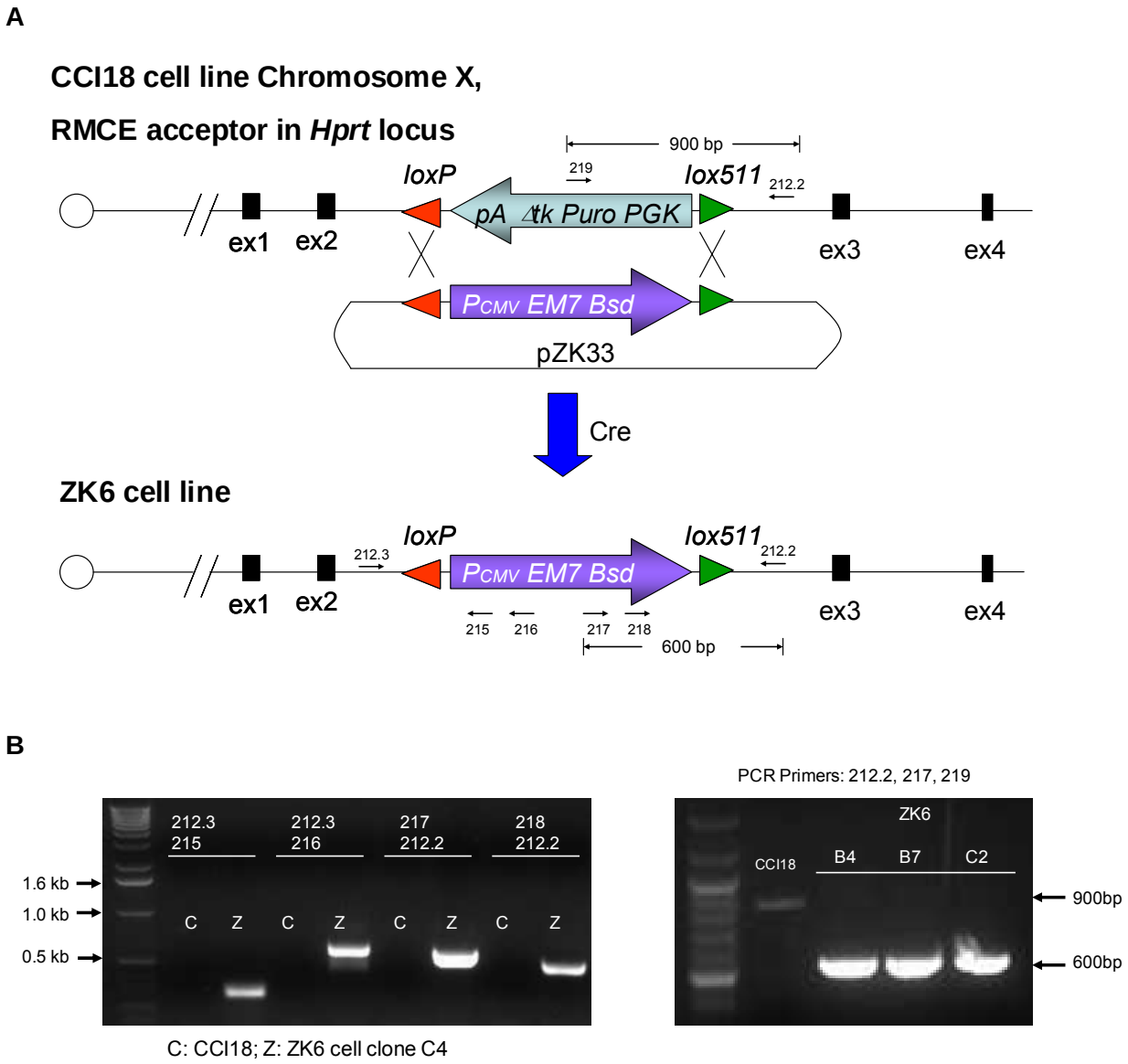


Figure 4-1 Generation of ZK6 BAC acceptor cell line

A. Schematic of RMCE event and position of genotyping primers. CCI18 contains the *PuroΔtk* cassette while ZK6 contains the *CMV-EM7-Bsd* cassette. Ex: exon; **B**. PCR results using the primer pairs 212.3/215, 212.3/216, 212.2/217, 212.2/218 confirmed ZK6 clone C4 has the correct junction fragment compared with CCI18 cells. PCR results using primers 212.2, 217 and 219 confirmed the genotype of CCI18 and ZK6 clones B4, B7 and C2.

To replace the *PuroΔtk* cassette by RMCE, a Cre expression plasmid pCAG-Cre was transformed into CCI18 cells by electroporation. The *loxP-Bsd-lox511* cassette was then introduced with Cre expression plasmids by electroporation in two experiments between one and five days later. These time courses were used to explore optional conditions for a high percentage of RMCE events. Each electroporation was split into two plates for selection in Bcyd and FIAU or Bcyd only. Colonies growing in Bcyd and FIAU media will represent those with the designed RMCE events plus random insertions in cells, which have lost *tk* expression. Colonies growing in Bcyd media only will be those with RMCE events plus those with a random insertion of the *Bsd* cassette.

The results of this assessment are illustrated in Table 4-1. After 5 days of the first introduction of the Cre recombinase, a higher RMCE frequency could be achieved. In the first experiment, cells were electroporated with pCAG-Cre one day before the RMCE construct and the pCAG-Cre (the 2nd time) were introduced. In this case, 10% of the Bcyd resistant clones were RMCE events. In the cells transfected with pCAG-Cre 5 days before the RMCE construct and the pCAG-Cre (the 2nd time) were introduced, 100% of the Bcyd resistant clones were RMCE events.

As designed, this RMCE introduces a *Bsd* resistant gene and removes the *PuroΔtk* gene, therefore Bcyd and FIAU double resistant clones should have the desired genetic change. These clones were confirmed by PCR genotyping. Primers were designed as shown in Figure 4-1. All primers were tested and confirmed that they can specifically amplify DNA fragments using CCI18 cells or pZK33 *Bsd* RMCE plasmid as templates in the preliminary experiments (data not shown). When RMCE does not occur, primer 212.2 in *Hprt* and the primer 219 in the *PuroΔtk* cassette can amplify a 900 bp fragment by PCR (Figure 4-2). After correct RMCE, the PCR product generated with 212.2 and 219 in the CCI18 cells should not be generated. Endogenous *Hprt* primers 212.2 and 212.3 can amplify several junction DNA fragments with the primers 215, 216, 217 and 218, which are in the *Pcmv-EM7-Bsd-SV40pA* cassette. This was confirmed using CCI18 and the clone C4 of ZK6 cell (Figure 4-1).

68 Bcyd and FIAU double resistant clones were genotyped by PCR. Sixty five of them had a 600 bp PCR product from correct RMCE using primers 212.2, 217 and 219 (Figure 4-2). When using CCI18 *PuroΔtk* BAC acceptor cells as the template, this 600 bp junction PCR product was not produced. These 65 double resistant clones can be regarded as different clones of ZK6 cell line, named as ZK6: A1–A4, A6–A12, B1–B12, C1–C5, C7–C12, D1–D2, D4–D12, E1–E12, F1–F8.

Table 4-1 The efficiency of RMCE

Time	Cell number per transformation	$Bcyd^R$ colony	$Bcyd^R$+$FIAU^R$ colony	$(Bcyd+FIAU)^R/Bcyd^R$
1 day	1.2×10^7	180	18	10%
5 day	3.0×10^7	41	68	~100%

The different sets of plasmids were transfected into CCI18 cells. Then cells were grown in normal media for 1 or 5 days before applying the drug for selection. 2-3 days additional Bcyd selection was used to maintain experiment time course comparable. These data indicated cells with transformed Cre plasmid 5 days before the introduction of RMCE cassette had increased RMCE percentage.

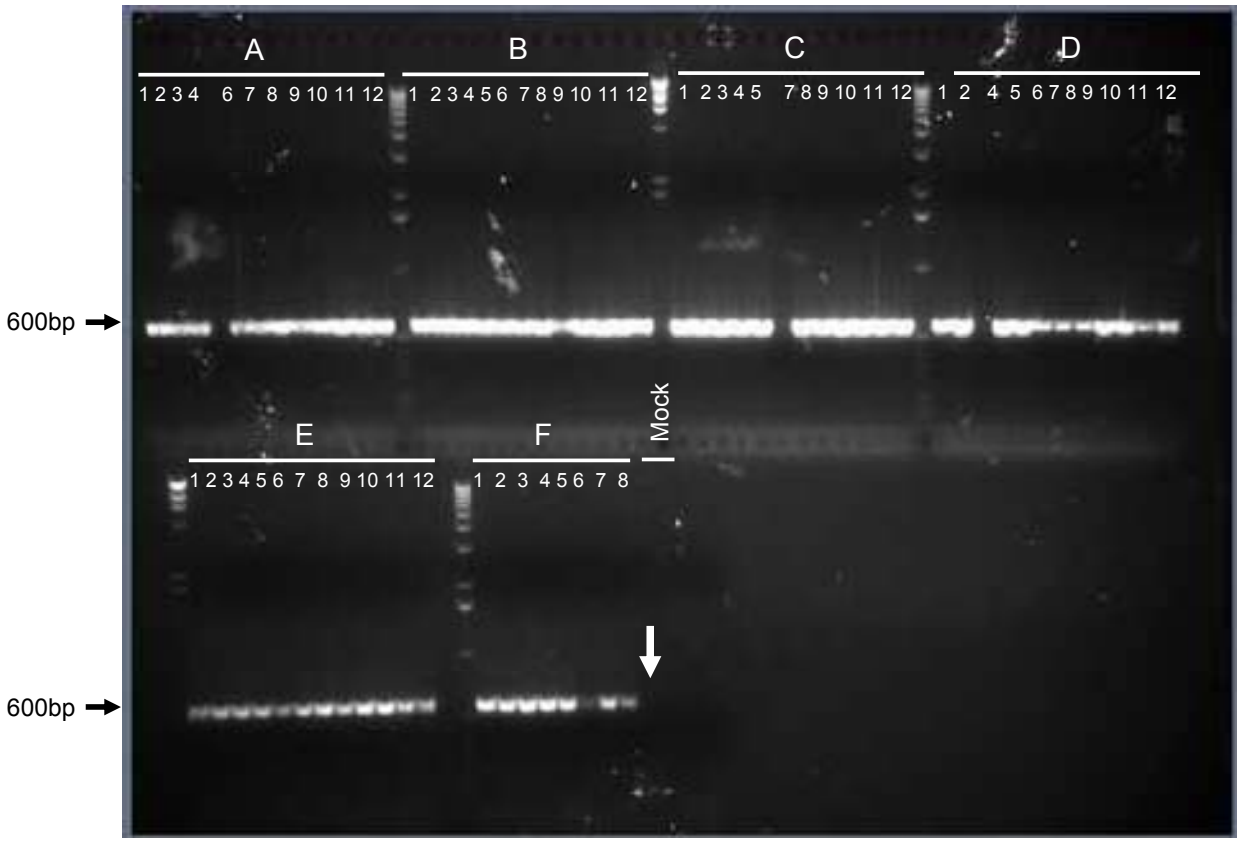


Figure 4-2 Genotype of all ZK6 BAC acceptor clones

Sixty eight Bcyd and FIAU double resistant clones were genotyped by a PCR strategy using primers 212.2, 217 and 219. Sixty five ZK6 clones, except the clones A5, C6 and D3, appeared positive by PCR genotyping with a 600 bp band according to primer design in Figure 4-1. CCI18 (mock) cell should have a 900 bp band (white arrow), which however is very weak here. This band can be seen in Figure 4-1 B.

4.2.2 Confirmation that RMCE can be achieved

BACs from library RPCI-23, 24 (http://bacpac.chori.org/home.htm) are suitable resources to complement normal function of mutant mouse genes. Prior to RMCE, these BACs need to be modified by recombineering to carry a *Neo*mycin/*Kanamycin* marker for both prokaryotic and eukaryotic selection. The *loxP* and *Lox511* sites on the backbone pBACe3.6 (Frengen *et al.* 1999) enable the large genomic DNA fragments to be introduced into the X-linked *Hprt* locus of ZK6 by RMCE (Figure 4-3). The suitability of my cell line ZK6 for RMCE was confirmed by PhD student Qi Liang who introduced fragments of the *p53* gene into the *Hprt* locus of ZK6. This RMCE event was selected in G418 first for 5 days for the integration of the *neomycin* cassette. Then cells were selected in 6TG containing media because 3 exons of *p53* interrupt the transcript of the *Hprt* gene which results in 6TG resistance. $6TG^R$ cells were confirmed to have gained the *p53* gene fragment and removed the P_{CMV}-*EM7-Bsd* cassette.

4.2.3 *Blm*$^{+/-}$ BAC acceptor cell lines

To generate a *Blm*-deficient BAC acceptor cell line with ability to incorporate BAC fragments through RMCE, the *Blm* gene-targeting vector was electroporated into the ZK6 cell line. After G418 selection clones were picked into 96-well plates. Half of the volume from each well was expanded and the other half was used to genotype the targeting event by PCR (Figure 4-4).

PCR primers were tested before being used for genotyping. A total of ~1000 $G418^R$ clones were generated in 3 electroporations. Two hundred eighty-eight clones were picked and screen by PCR. Three were positive. One of them was subcloned and confirmed by Southern blot analysis (Figure 4-4).

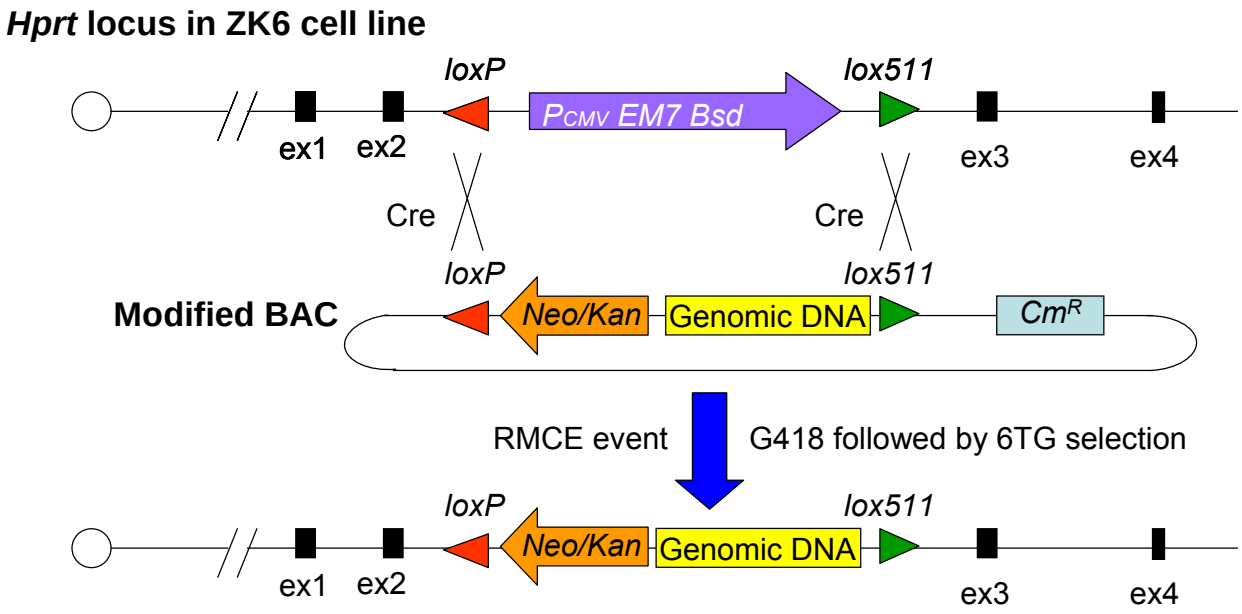


Figure 4-3 The principle of RMCE in ZK6 cell line

Recombination between *Lox* sites is catalysed by Cre. As a result of RMCE, genomic DNA (inserted a *Neo*mycin/*Kanamycin* cassette) is introduced into the *Hprt* locus of ZK6. RMCE events can be selected in G418 and 6TG, for the introduction of the *Neomycin* cassette and loss-of-function of *Hprt* function, respectively. These two selective chemicals can be used sequentially (G418 first) to avoid interfering with cell viability. Ex: Exon. *Neo*/*Kan*: *Neo*mycin/*Kanamycin* cassette.

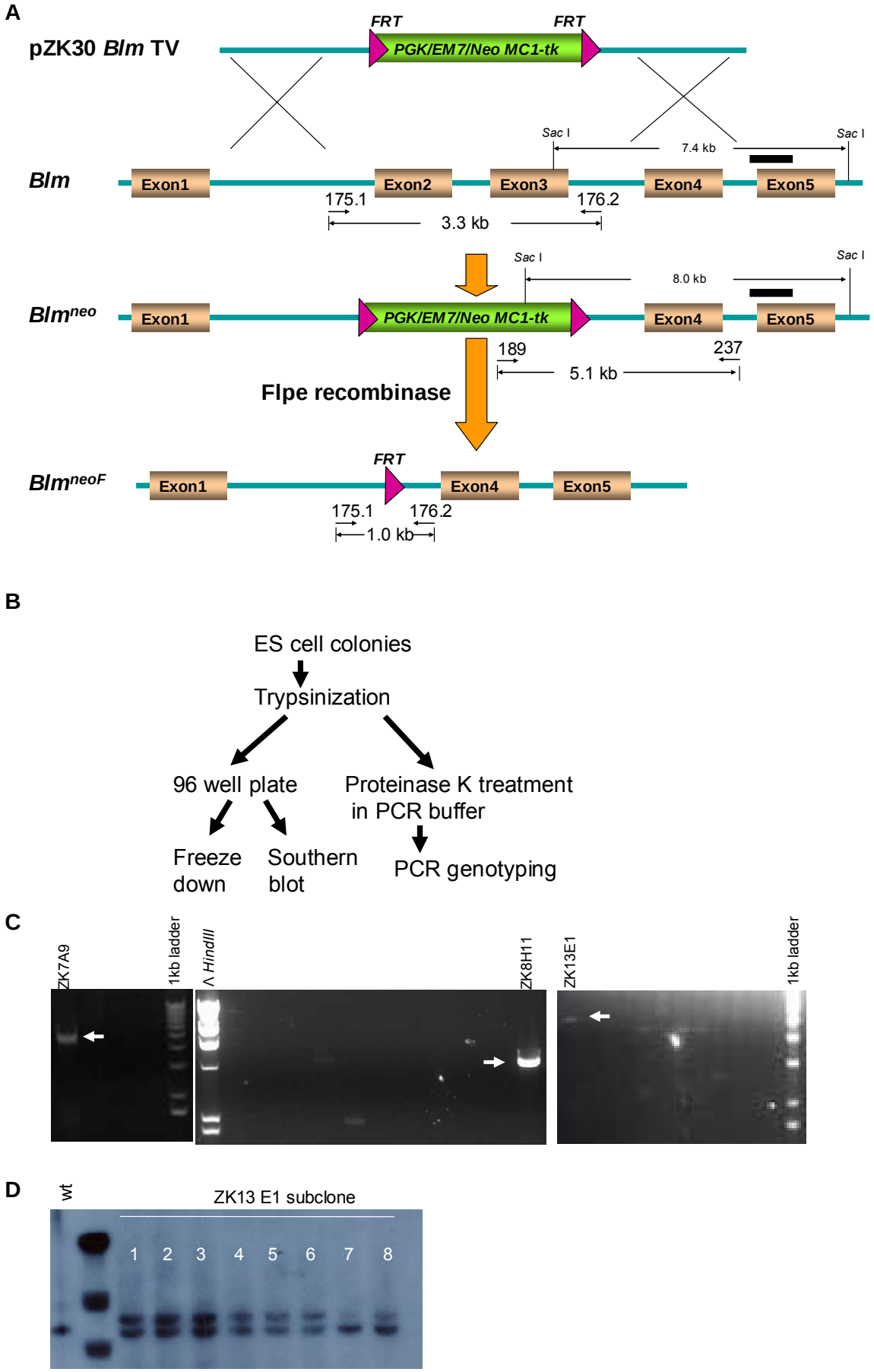


Figure 4-4 ***Blm*** **gene-targeting strategy**

Figure 4-4 *Blm* gene-targeting strategy

A. Schematic view of the *Blm* gene-targeting vector (pZK30 *Blm*TV), wild type allele (*Blm*), targeted allele (Blm^{neo}) and neomycin pop-out allele (Blm^{neoF}). pZK30 *Blm* gene-targeting vector contained the selection cassette *PGK*/*EM7*/*neo-MC1-tk*, which was flanked by *FRT* sties. Exon 2 and 3 in the wild type *Blm* allele were replaced to generate the Blm^{neo} allele. After Flpe recombinase mediated recombination, the selection marker could be popped out, leaving a *FRT* site at the allele Blm^{neoF}. Primers 189, 237, 175.1 and 176.2 were designed to genotype cells. *Sac* I digestion was used to distinguish wild type and targeted alleles by Southern blot. Black bar was probe 185. **B**. Experiment design after colonies were picked. When colonies were trypsinized, a half volume of cells were harvested and lysed. Crude DNA served as a template in PCR genotyping. The other half volume of cells was expanded in 96-well plates and passaged for DNA extraction and subsequent Southern blot analysis. **C**. PCR genotyping showed that clones ZK7A9, ZK8H11 and ZK13E1 had the correct products. PCR primers 189 and 237 amplify a 5.1 kb band from the targeted allele. **D**. Southern blot analysis confirmed that ZK13 E1 subclones 1–6 were $Blm^{+/-}$ BAC acceptor cells. Wild type cells showed one 7.4 kb band, while targeted alleles showed an 8.0 kb band. The probe was zk185 (the black bar in a.).

4.2.4 Using Flpe/*FRT* to remove the selection marker

The selection marker had to be popped out from *Blm*$^{+/-}$ cells by Flpe to reuse the same targeting vector for targeting the second allele of the *Blm* gene. Before doing this, the targeted clones ZK7A9, ZK8H11 and ZK13E1 needed to be subcloned as 1–10% cells in the original colonies may not be targeted and thus will contribute significantly to the background. 5' and 3' PCR genotyping confirmed that the single cell subclones ZK7A9.1and ZK8H11.2 had correct alleles (Figure 4-5). To test the *MC1-tk* function, ZK7A9.1 and ZK8H11.2 were plated in FIAU (Figure 4-6). Both clones appeared to be FIAU-resistant. Therefore, second round of gene-targeting was carried out to isolate a new *Blm*$^{+/-}$ clone which resulted in clone ZK13E1. This was single cell subcloned as before and Southern analysis confirmed that all the subclones were *Blm*$^{+/-}$ (Figure 4-4), but these clones were also FIAU-resistant. The plasmids used in these transformations were subcloned and then transformed into CCI18 cells to test whether their selection markers were functional. The result indicated that they all contained functional *neomycin* and *HSV-tk* gene (Figure 4-7). This proved that the selection cassettes in the targeting vector were functional. Thus, the parental cells of the heterozygously targeted *Blm*$^{+/-}$ cells may contain a pre-existing mutation in the FIAU metabolism pathway, which blocks the killing effect of FIAU. This phenomenon is similar in that in Chapter 3.

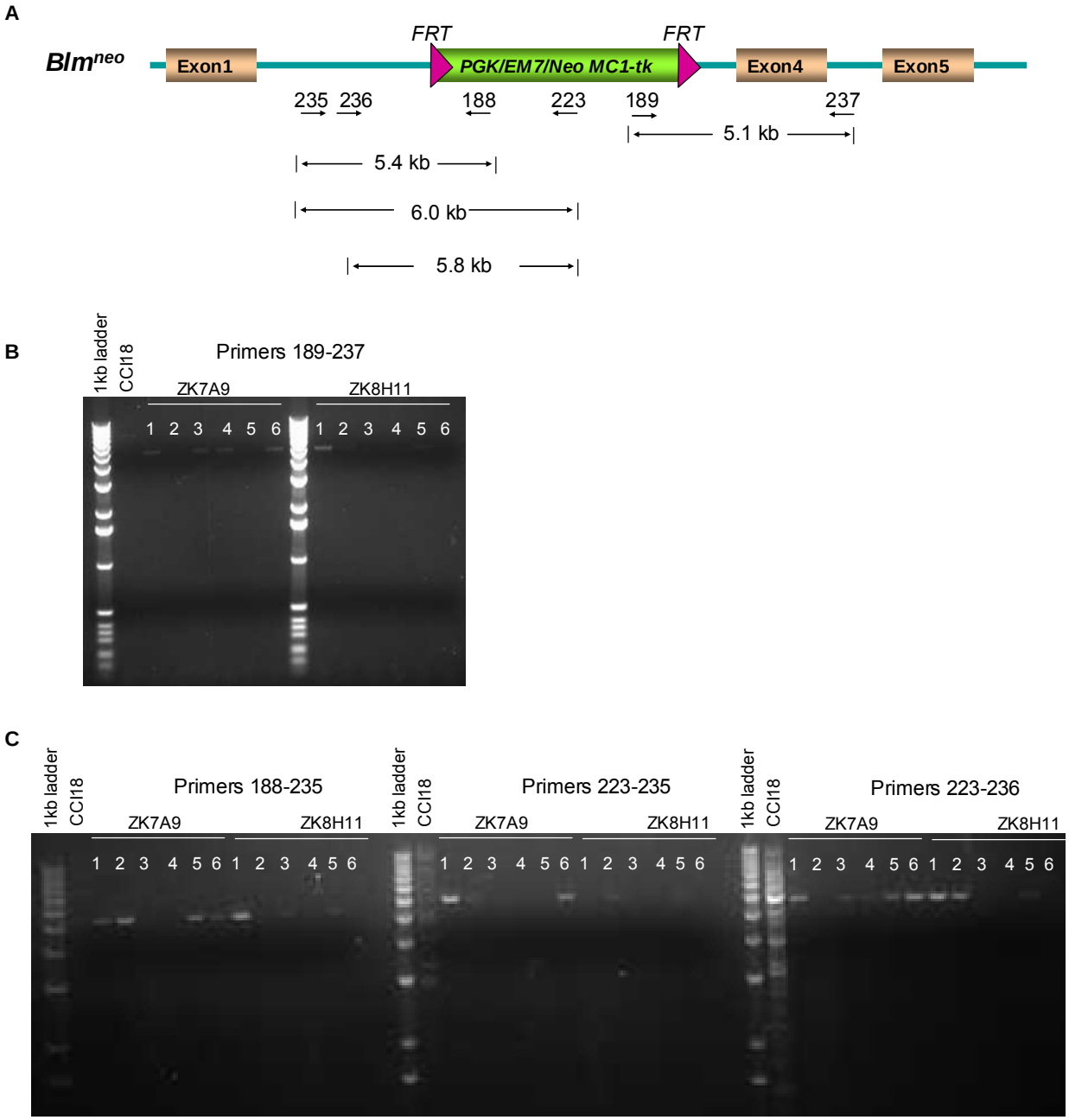


Figure 4-5 PCR genotype of ZK7A9 ZK8H11 subclones

A. Schematic view of primer design to PCR genotype pZK30 targeted *Blm*^{neo} allele. **B**. 3' PCR genotype. Primer pair: zk189-237. Expect a 5.1 kb product. **C**. 5' PCR genotype. Primer pairs: zk188-235, zk223-235 and zk223-236. Expected products were 5.4 kb, 6.0 kb and 5.8 kb in length, respectively.

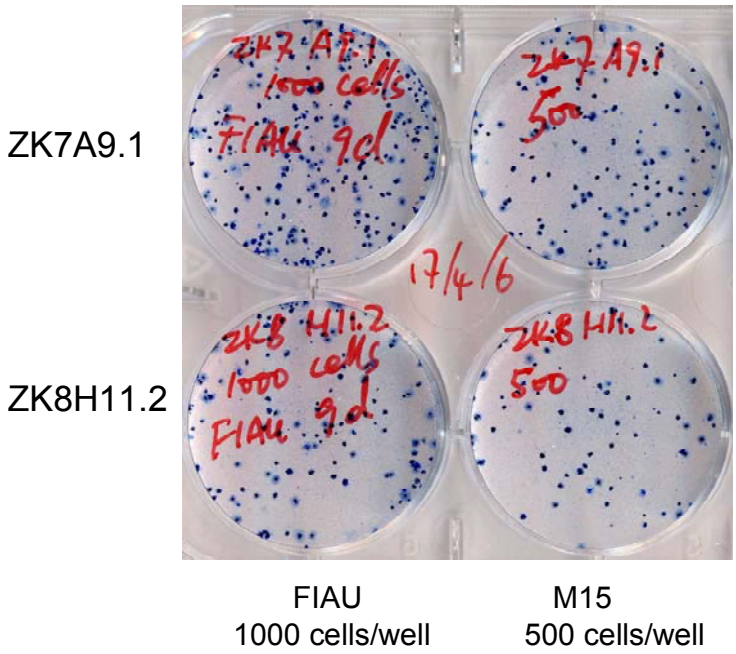


Figure 4-6 FIAU-resistant background of ZK7 and ZK8 cells

Two subclones of *Blm* heterozygous targeted cells ZK7A9.1 and ZK8H11.2 were cultured in M15 normal media (500 cells) and FIAU (1000 cells) in 30 mm wells. Cells were stained by methylene blue after nine days culture. Both clones are resistant to FIAU.

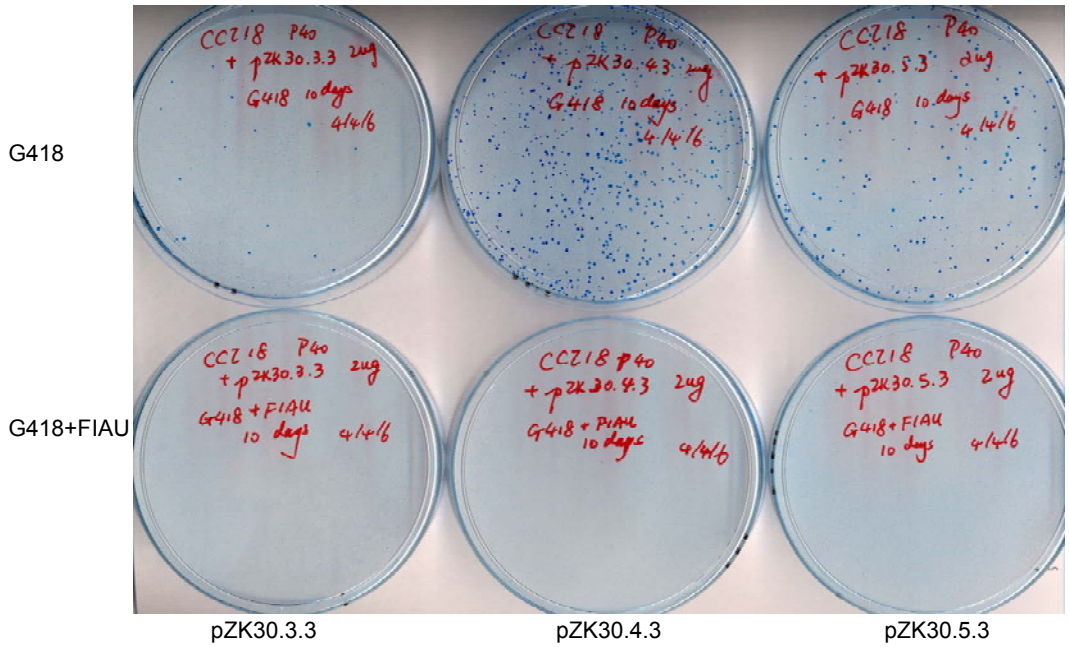


Figure 4-7 Selection confirmed functional markers in *BIm* targeting vectors

Not I linearized plasmid pZK30.3.3, 30.4.3 and 30.5.3 (4ug each) were transformed into CCI18 cells. Each electroporation was split into two plates evenly for either G418 selection or G418+FIAU selection in the following 10 days. Cells were stained by Methylene Blue (2%). Three electroporations produced G418 resistant colonies but not G418+FIAU-resistant colonies, which indicated that 3 types of plasmids had functional *Neomycin* and *HSV-tk* genes.

4.2.5 Generation of the *Blm*^{neoF} allele

Since the isolated *Blm*$^{+/-}$ clones ZK7A9, ZK8H11 and ZK13E1 were resistant to FIAU, it was impossible to select for FIAU^{R} clones resulting from the loss of the marker following Flpe mediated recombination. However, considering the efficiency of Flpe-mediated recombination is around 1%, the clones that have ocurred Flpe-mediated recombiantion can still be achieved and isolated without the FIAU selection.

The clonally purified derivatives of the targeted clones ZK13E1.1, ZK13E1.2 and ZK13E1.3 were electroporated with the Flpe expression plasmid pCAG-Flpe, resulting in ZK15, ZK16 and ZK17, respectively. A PCR genotyping strategy (Figure 4-4, A) using the primers zk175.1 and zk176.2 confirmed that clones ZK15 C11, F10, G1 and ZK16 H2 had the desired popped out *Blm*^{neoF} allele (Figure 4-8). Although PCR using these two primers could also generate a 3.3 kb wild type fragment, the short extension time of PCR did not allow this to occur.

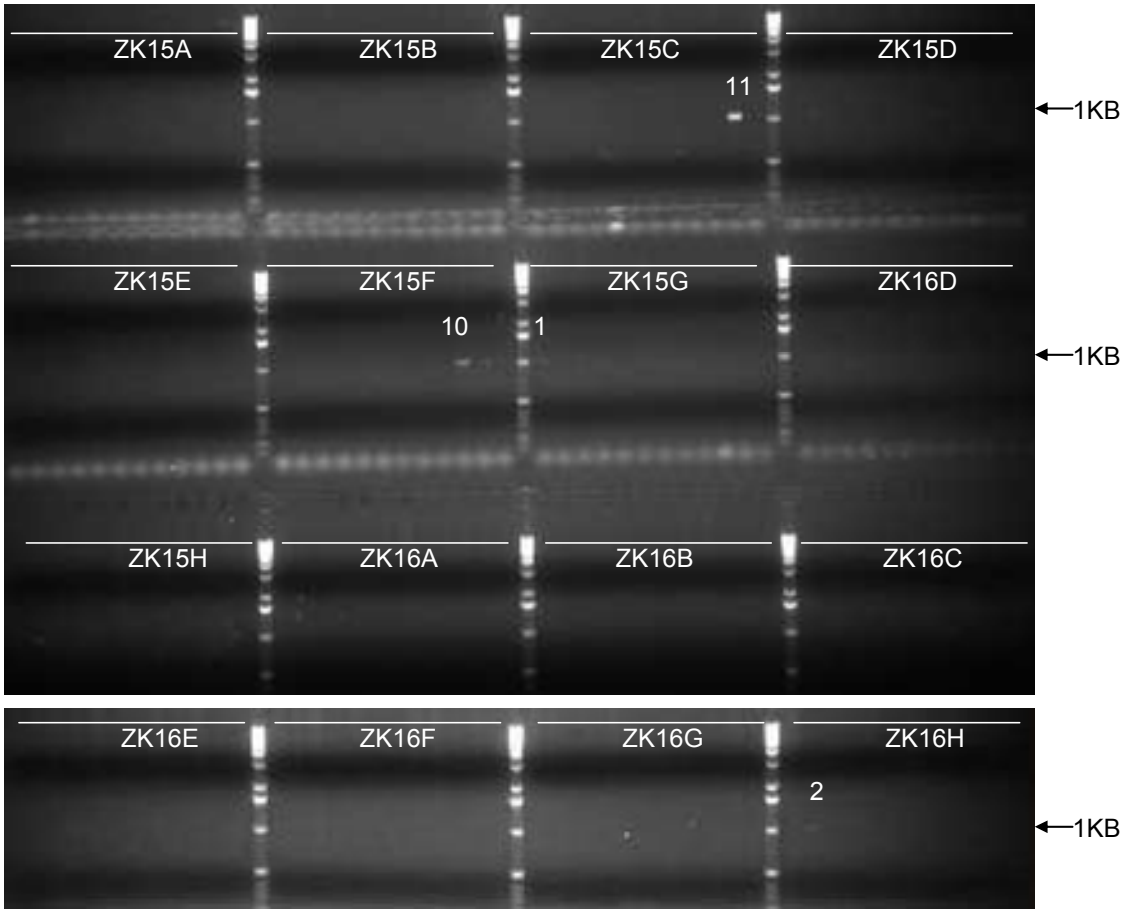


Figure 4-8 PCR screen of ZK15 and ZK16 clones

1kb PCR products were expected by amplifying DNA with primers zk175.1 and zk176.2. Clone ZK15 C11, F10, G1 and ZK16 H2 contained a 1kb band.

4.3 Discussion

A *Blm*-deficient ES cell line with a BAC acceptor locus would be a valuable resource to study many biological pathways, which are active in mouse ES cells. Following the successful screens done with *Blm*-deficient ES cells (Guo *et al.* 2004; Wang and Bradley 2007), more loss-of-function screens can be designed and implemented using this system. To generate a *Blm*-deficient ES cell line with a BAC acceptor, an existing wild type BAC acceptor cell line (CCI18) was used. CCI18 has a *loxP* and *lox5-11* flanked *PuroΔtk* cassette, which was replaced by a *loxP* and *lox5-11* flanked *Bsd* cassette through recombination mediated cassette exchange (RMCE). The two experiments within different time frames showed that the RMCE efficiency can vary from 10% to 100% depending on the time point of introducing RMCE cassette with more Cre (Table 4-1). When the RMCE construct and the Cre plasmid were electroporated 5 days later than the first introduction of the Cre recombinase, 100% $Bcyd^R$ and $FIAU^R$ cells were correct RMCE events. This increased efficiency of RMCE may be achieved by the accumulated Cre protein in the cells.

By switching the markers from *PuroΔtk* to *Bsd*, a MMuLV recombinant virus containing the *PuroΔtk* cassette can be used to infect cells and to generate molecular tags in the future. A *Blm* gene-targeting vector (pZK30) was constructed to delete the exon 2 and 3 of the gene. With a *PGK/EM7/neo-MC1-tk* cassette flanked by *FRT* sites which can be selected in *E. coli* and eukaryotic cells and recycled using Flpe mediated recombination. Although the *PGK/EM7/neo-MC1-tk* cassette was shown fully functional when random integrated into the wild type cells (Figure 4-7), the targeted clones appeared to have lost the function of *HSV-tk*. This inactivation of the *HSV-tk* cassette may be due to the genome instability caused by *Blm* heterozygosity. It has been known that chromosomal breakage in human spermatozoa was increased in two out of three fathers (Mutant *BLM* carriers) of persons with Bloom syndrome (Martin *et al.* 1994). Moreover, the carrier frequency (1.8%, 23/1244) of the mutant BLM^{Ash} allele in colorectal cancer samples is higher than that (0.8%, 85/10,099) in historical controls from healthy people (Gruber *et al.* 2002). And mice heterozygous for the *Blm* gene demonstrated the *Blm* haploinsufficiency developed twice the number of intestinal tumours when crossed with the mice carrying a mutation of the *Apc* tumour suppressor gene (Goss *et al.* 2002). The dis-function of the *HSV-tk* can be resulted from small mutations causing reading frame shifts. The mechanism of the spontaneous inactivation of the *HSV-tk* cassette may be similar or associated to that of the *Δtk* cassette in the *Blm* deficient background, described in Chapter 3.

When Flpe was used to remove and recycle the *PGK/EM7/neo-MC1-tk* cassette, FIAU selection could not be used to isolate cells which had excised the *PGK/EM7/neo-MC1-tk* cassette. A PCR strategy was used to screen a large number of clones to identify those with the popped out allele. These two cells are heterozygous for the *Blm* gene and ready to be targeted again to generate a homozygous *Blm* mutant cell line with a BAC acceptor.

CHAPTER 5. CELL VIABILITY AND CHROMOSOME REARRANGEMENTS AFTER GAMMA IRRADIATION

5.1 Introduction

When a genetic screen is conducted it is critical to analyse enough independent mutations to recover genes in the pathway of interest. The chance of recovering a mutation will depend on the complexity of the library. Ideally independent mutations in the same gene will be recovered since this will help to validate the role of that gene in the specific pathway. An average complexity of 5 mutations per gene is adequate for most purposes.

The complexity of the library will depend on the number of mutations per cell and the number of cells in the library. In this chapter, experiments were conducted to measure ES cell survival rates and mutation frequencies after irradiation. Cell viability data at a series irradiation doses enables the complexity of the mutation pools to be established. A whole mutation library is ideally composed of separated mutation pools. There are two advantages to do this.

First, independent mutations at the same locus can be distinguished quickly. It is known that irradiation can generate mutations at a rate of ~10^{-4} per locus per generation in mice (Asakawa *et al.* 2007; Malling *et al.* 1977). Assuming the same mutation rate, it can be assumed that approximately ten thousand clones can provide enough mutations to cover the whole mouse genome once. If a genetic screen is designed and implemented properly in mutation libraries, in which mutations cover the whole genome several times, individual mutant cells with different mutations at the same locus can be identified. A fast way to distinguish that these are different is to conduct a screen in separated mutation pools so that mutants are independent between pools. Mutant cells with the same gene mutated can be quickly recognized as independent if they are from separated mutant pools.

Second, screening in separated mutation pools prevents daughter cells with one mutant allele to be dominant among the isolated mutants. As I use *Blm*-deficient ES cells to generate homozygous mutations through the mechanism of loss of heterozygosity (LOH), LOH events may occur at any time during the cell culture. When a homozygous mutant cell occurs several generations earlier than the others, it will become dominating in the homozygous mutant population after several cycles. As a result, it will be more difficult to

identify most of the other mutants. If mutants are segregated and replicate within smaller mutant pools, a single mutant cell can only be dominant in one specific pool, without making it difficult to identify mutations in other mutant pools.

Experiments in this chapter also characterized the size and frequency of deletions produced by irradiation in 129 inbred mouse ES cells. It is known that irradiation generates chromosome rearrangements, for instance amplifications and deletions. Deletions are required to generate loss-of-function mutations. Deletions generated by irradiation had been examined in a number of species including *Drosophila* and mice by cytogenetic analysis (Brewen *et al.* 1973; Searle *et al.* 1974; Traut *et al.* 1971). More recently the negative selection cassette Herpes Simplex Virus type 1 thymidine kinase (*HSV-tk*) was targeted to specific sites in mouse ES cells and deletion sizes were measured after irradiation and negative selection (Chick *et al.* 2005; Thomas *et al.* 1998; You *et al.* 1997). However, these studies limited the analysis to specific loci. If any local genes are essential for cell viability, cells with these genes deleted cannot survive thus they will not be recovered for analysis, thus the deletion size may be biased. Additionally, it is not known the size and number of deletions that can be generated in a single ES cell genome. To answer this question, both wild type cells and *Blm*-deficient cells were irradiated at a series of dosages. The resulting chromosome deletions, as well as amplifications were analysed using a 200 kb resolution comparative genome hybridization (CGH) array. This technology is used in human genetics (Oostlander *et al.* 2004; van Beers *et al.* 2006). For instance array CGH was used to identify copy number variations among 270 healthy human beings, from four populations with ancestry in Europe, Africa or Asia (Redon *et al.* 2006). Considering the availability coverage and resolution, a 200 kb resolution array CGH was used in this study.

5.2 Results

5.2.1 Cell viability after irradiation

To measure cell viability after gamma irradiation, experiments were performed on cells with different genotypes and with a series of irradiation dosages. The AB1, AB2.2 and the *Blm*$^{-/-}$ cells were derived from the 129 S7/SvEv mouse. Here I consider the AB1 and AB2.2 cell lines to be as wild type cells, while the *Blm*-deficient ES cells are the NGG5-3 cells from Guo, G., derived from AB2.2 (129 S7/SvEv$^{\text{Brd-}Hprt\text{b-m2}}$) with the genotype *Blm*(m3/m4), *Gdf9*(m1/m3).

Cell viability was assessed by a colony survival assay. When cells grew to confluence on 90

mm feeder plates, they were collected by trypsinization and irradiated for different lengths of time and then re-plated onto 30 mm feeder plates. Since the percentage of cells surviving varied over several orders of magnitude, the cells were plated at different densities. Non-irradiated cells were plated at 500 cells per 30 mm well to determine their plating efficiency. After staining surviving colonies, plating efficiency was calculated as: Plating efficiency = Number of survival colonies/500. Plating efficiencies differed between independent experiments. Usually the plating efficiency was between 20-44% (Table 5-1).

The survival frequency of irradiated cells was calculated as: Survival frequency = Number of survival colonies/(Number of plated cells × plating efficiency). Survival frequency was normalized to 100% if the cells were unirradiated. Survival rates of the wild type cells and the *Blm*-deficient cells of different experiments are shown in Table 5-2 and Table 5-3, respectively and are summarized in Table 5-4. Survival rates of both types of cells were around 20% at 2 Gray, 10% at 3 Gray, 5% at 4 Gray, 3% at 5 Gray, 1% at 7 Gray and 0.1% at 10 Gray. In general, *Blm*-deficient cell are more resistant to irradiation at all doses, which can be seen in Figure 5-1.

Table 5-1 ES cell plating efficiency

Experiment	Number of survival colonies	Plating efficiency	Average plating efficiency
1	104	0.21	0.21
2	103	0.21	0.21
3	127	0.20	0.20
4	193	0.44	0.44
5	134	0.27	0.27
6	161	0.32	0.32
7	120	0.24	
	136	0.27	0.26
8	131	0.26	
	107	0.21	0.24
9	124	0.25	
	133	0.27	0.26
10	147	0.29	
	162	0.32	0.31
		Total average	**0.27**

ES cell plating efficiency was calculated in ten experiments. Experiments 7–10 have two replicated wells of cells. In each 30mm well, 500 cells were plated in normal M15 media and cultured for 10 days. Survival colonies were stained by methylene blue. The plating efficiency = Number of survival colonies/500. Survival rates of the wild type cells and the *Blm*-deficient cells were calculated in the experiments 4, 5, 6 and 10.

Table 5-2 Wild type ES cell survival rates following 0-10 Gray gamma irradiation

Dose (Gy)	Cells plated	Cells surviving	Plating efficiency	Survival rate	Average survival rate	SD
0	–	–	–	100%	100%	–
2	2000	107	31%	17%		
		111		18%		
		118		19%	18%	0.01
3	2000	49	31%	8%		
		56		9%		
		79		13%	9.9%	0.025
4	5000	65	31%	4.2%		
		68		4.4%		
		81		5.2%	4.6%	0.005
5	10^4	76	31%	2.5%		
		82		2.6%		
		96		3.1%	2.7%	0.003
7	5×10^4	80	32%	0.50%	0.50%	–
10	10^5	3	31%	0.01%		
		12		0.04%		
		18		0.06%		
	10^5	18	32%	0.06%		
		19		0.06%		
		25		0.08%	0.05%	0.0002

Survival rates of the wild type cells (AB1 and AB2.2) after irradiation at 2–10 Gray are shown. Survival frequency was normalized to 100% if the cells were unirradiated. Three–six independent experiments were performed at 2, 3, 4, 5 and 10 Gray. Standard deviations (SD) of the survival rates at these doses are also shown. One experiment was performed at 7 Gray. Plating efficiencies were calculated from experiment 6 and 10 (Table 5-1).

Table 5-3 ***Blm*-deficient ES cell survival rates following 0-10 Gray gamma irradiation**

Dose (Gy)	Cells plated	Cells surviving	Plating efficiency	Surviving rate	Average survival rate	SD
0	–	–	–	100%	100%	–
2	2000	220	44%	25%		
		220		25%		
		238		27%	26%	0.01
3	2000	157	44%	18%		
		159		18%		
		188		21%	19%	0.02
4	5000	175	44%	8.0%		
		186		8.5%		
		191		8.7%	8.4%	0.004
5	10^4	173	44%	3.9%		
		174		4.0%		
		184		4.2%	4.0%	0.001
7	5×10^4	221	27%	1.6%	1.6%	–
10	10^5	14	44%	0.03%		
		29		0.07%		
		60		0.14%		
	10^5	37	27%	0.14%		
		38		0.14%		
		40		0.15%	0.11%	0.0005

Survival rates of the *Blm*-deficient cells (NGG5-3) after irradiation at 2–10 Gray are shown. Survival frequency was normalized to 100% if the cells were unirradiated. Three–six independent experiments were performed at 2, 3, 4, 5 and 10 Gray. Standard deviations (SD) of the survival rates at these doses are also shown. One experiment was performed at 7 Gray. Plating efficiencies were calculated from experiment 4 and 5 (Table 5-1).

Table 5-4 Summary of ES cell survival rates following gamma irradiation

Dose (Gray)	0	2	3	4	5	7	10
WT (AB1/AB2.2)	100%	18±1%	9.9±2.5%	4.6±0.5%	2.7±0.3%	0.50	0.05±0.02%
***Blm*$^{-/-}$ (NGG5-3)**	100%	26±1%	19±2%	8.4±0.4%	4.0±0.1%	1.6	0.11±0.05%

The survival rates ± standard deviation (SD) of the wild type cells (AB1/AB2.2) and the *Blm*-deficient cells (NGG5-3) are compared at different doses of gamma irradiation. Survival frequency was normalized to 100% if the cells were unirradiated. Data from Table 5-2 and Table 5-3. n≥3 (except at 7 Gray).

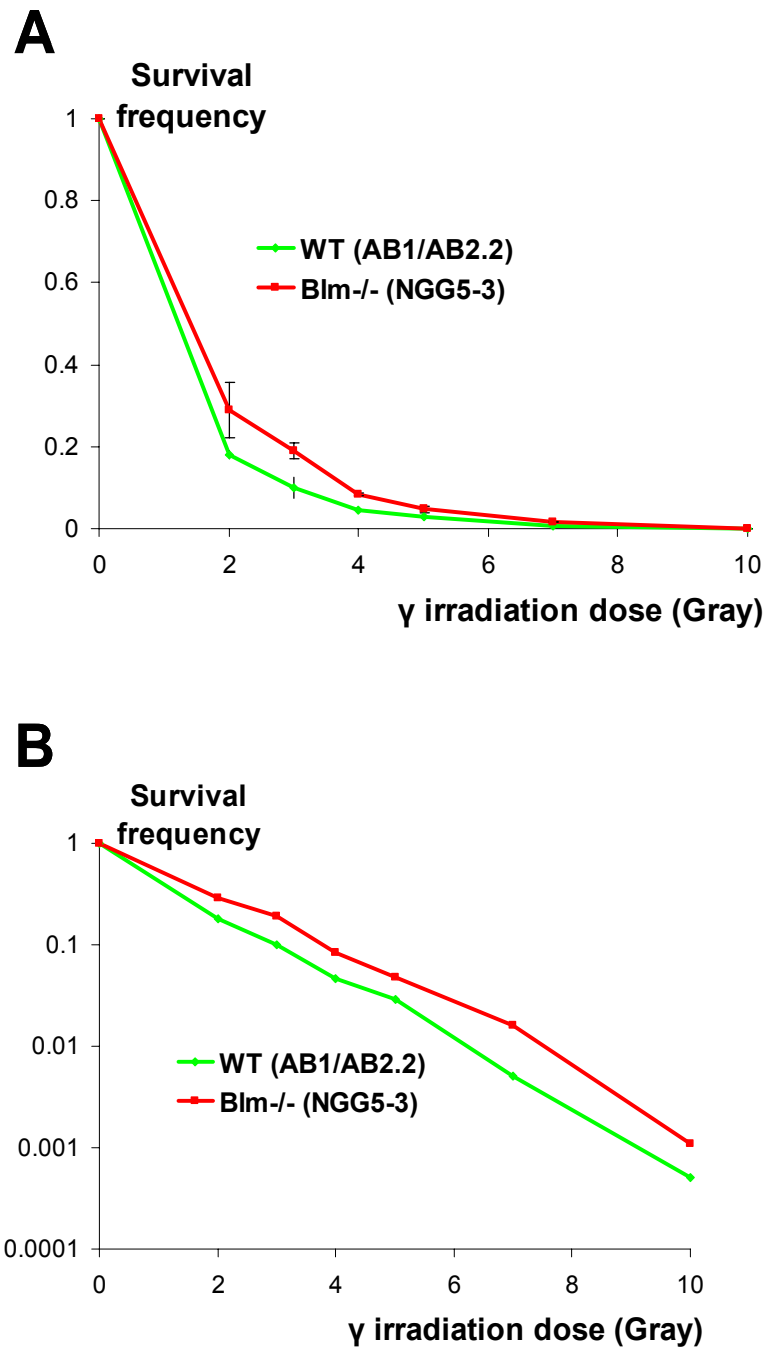


Figure 5-1 Gamma irradiation caused viability change in ES cells

Linear (**A**) and logarithmic (**B**) plots of survival rates. Error bars in (**A**) indicate standard deviation. WT, wild type ES cells (AB1 or AB2.2); *Blm*$^{-/-}$, *Blm*-deficient AB2.2 ES cells (NGG5-3).

5.2.2 Irradiation generates megabase sized chromosomal alterations

To measure large chromosome amplifications and deletions generated by gamma irradiation, both wild type (ZK6) and *Blm*-deficient cells (NGG5-3) were irradiated at 2, 5 and 7 Gray. The *Blm*-deficient NGG5-3 cell was also irradiated at 10 Gray. After irradiation, single colonies were isolated and expanded to extract DNA, which was analysed on the 200 kb resolution CGH array. The cost of array CGH is high, thus only total 11 samples of wild type cells and 15 samples of the *Blm*-deficient cells were analysed. The following samples were analysed: clones WT2G(4), WT5G(4) and WT7G(3) were wild type cells irradiated at 2, 5 and 7 Gray, respectively; clones Blm2G (3), Blm5G(3), Blm7G(3) and Blm10G(6) were the *Blm*-deficient cells irradiated at 2, 5, 7 and 10 Gray, respectively. The reference DNA used for these experiments was from ZK6 or NGG5-3 cells passaged immediately before irradiation.

To obtain convincing data, a dye-swap hybridization was necessary to be performed for each sample. By using the BlueFuse software, the Log_2(sample DNA signal)/(reference DNA signal) or the Log_2(reference DNA signal)/(sample DNA signal) of each BAC probe were measured. The CGH array contains 18,294 unique BAC probes, however, signals of BAC probes with high variations in the dye-swap arrays were excluded from the analysis. The Log_2(ratio)s of high quality signals from BAC probes were plotted against their position in the mouse genome. Considering the Log_2(sample DNA signal)/(reference DNA signal) is -1 ($Log_2(1/2)$) for a heterozygous deletion in a diploidy cell and -0.42 ($Log_2(3/4)$) for a deletion in a tetraploidy cell, it was necessary to set up threshold to distinguish deletions, duplications (amplifications) and unchanged DNA. Generally, we used the thresholds of 0.2999 and -0.2999 of Log_2(ratio) for nomal array analysis in our laboratory. These thresholds can leave buffer space between Log_2(ratio) of unchanged DNA and duplications/deletions. The Log_2(ratio) above 0.2999 indicates a duplication or an amplification. The Log_2(ratio) between -0.2999 and +0.2999 indicates unchanged DNA copy number. The Log_2(ratio) below -0.2999 indicates a deletion.

The array profile of the clones WT2G2, WT5G4, WT7G2, Blm2G2, Blm5G1, Blm7G1 and Blm10G1 are shown in Figure 5-2 to Figure 5-8. The copy number changes of these clones and the other clones are summarized in Table 5-5 and Table 5-6. In the wild type cells irradiated at 2–7 Gray (Table 5-5), 9% (1/11) clones carry large deletions. Deletions are usually heterozygous, ranging from 0.1 (detection limit of the array CGH) to 5 Mb. However, deletions as small as 0.1 Mb are detected through the copy number change of a single BAC

probe, thus this type of deletions are not very convincing. About one third of wild type cells surviving 7 Gray irradiation carry deletions, and none of the eight wild type cells surviving 2–5 Gray irradiation contains detectable deletions. About 27% (3/11) of the wild type cells irradiated at 2–7 Gray have duplications, ranging from 0.2–14 Mb. None of the four wild type clones surviving 2 Gray irradiation carries duplications.

Besides of the normal heterozygous deletions and duplications, array CGH can also detect other types of chromosomal rearrangements including whole chromosome gain/loss or segmental chromosomal gain/loss (Table 5-5 and Table 5-6). Thus, in the array analysis of the irradiated cells, these types of chromosomal rearrangement, which occurred spontaneously or induced by irradiation, can also be anticipated to be observed. The mutant cell WT2G2 may be a tetraploid because the Log_2(ratio) of the whole chromosomes 2 and 5 is near -0.2 and that of the chromosome 11 is near 0.2, suggesting the chromosomes 2 and 5 are probably 3N and the chromosome 11 is 5N. However, the Log_2(ratio) of 0.2 is also lower than 0.32, the value of Log_2(5/4), thus the WT2G2 cells might be mosaic. Similar chromosomal abnormalities were also observed in Blm5G1, Blm5G2, Blm5G3, Blm7G3, Blm10G4, Blm10G5 and Blm10G6. The chromosome Y in the mutant WT7G2, Blm7G1, Blm10G5 and Blm10G6 were lost and some one- to two-BAC-sized deletions in other chromosomes are associated with the loss of chromosome Y. The genomic loci of these BACs may not be correct thus need to be relocated to chromosome Y. On the contrary, the mutant of Blm2G3 has one more copy of chromosome Y, with the average Log_2(ratio) of about 1, than the *Blm*-deficient cells.

About 62% (7/11) of the wild type cells irradiated at 2–7 Gray does not carry any chromosome rearrangements. In the irradiated *Blm*-deficient ES cells (Table 5-6), 53% (8/15) of 2–10 Gray irradiated cells have heterozygous deletions, ranging from 0.3–46 Mb. The frequency of deletions is higher in 7 and 10 Gray irradiated cells (67%, 6/9) than in 2 and 5 Gray irradiated cells (33%, 2/6). Sixty-seven per cent (10/15) of 2–10 Gray irradiated *Blm*-deficient cells obtain duplications, ranging from 0.1–24 Mb. Considering deletions and duplications together, thirty-three per cent (5/15) cells contain the both types of chromosome rearrangements after 2–10 Gray irradiation, and 93% (14/15) of the 2, 5, 7 and 10 Gray irradiated *Blm*-deficient cells have either deletions or duplications or both.

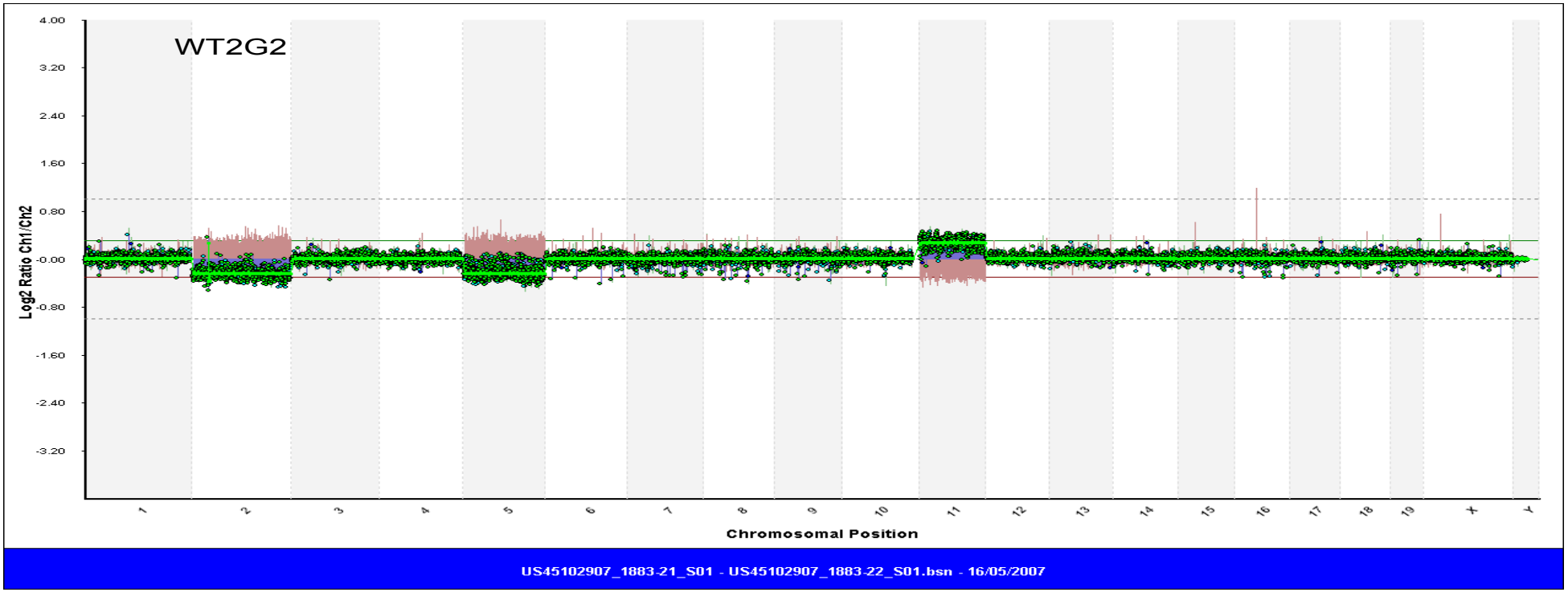


Figure 5-2 Array CGH analysis on WT2G2 – a wild type clone surviving 2Gy irradiation

BACs were aligned on the X axis in ascending order according to their genomic position in every chromosome. Chromosomes were also placed in ascending order followed by X and Y. A dot indicates the middle point position of a given BAC. The Y axis indicates the Log2(mutant DNA signal/reference DNA signal) of a given BAC probe. This Log_2(ratio) indicates a duplication (amplification) when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. The grey dotted lines are for Log2(ratio) references of either +1 or -1. This clone might be a tetraploid because the Log2(ratio) of chromosomes 2, 5 and 11 are near either 0.3 or -0.3, indicating the signal ratio of sample to reference is close to 5/4 or 3/4. In the 4N genome, one copy of the chromosomes of 2 and 5 is probably lost and one copy of chromosome 11 is probably duplicated.

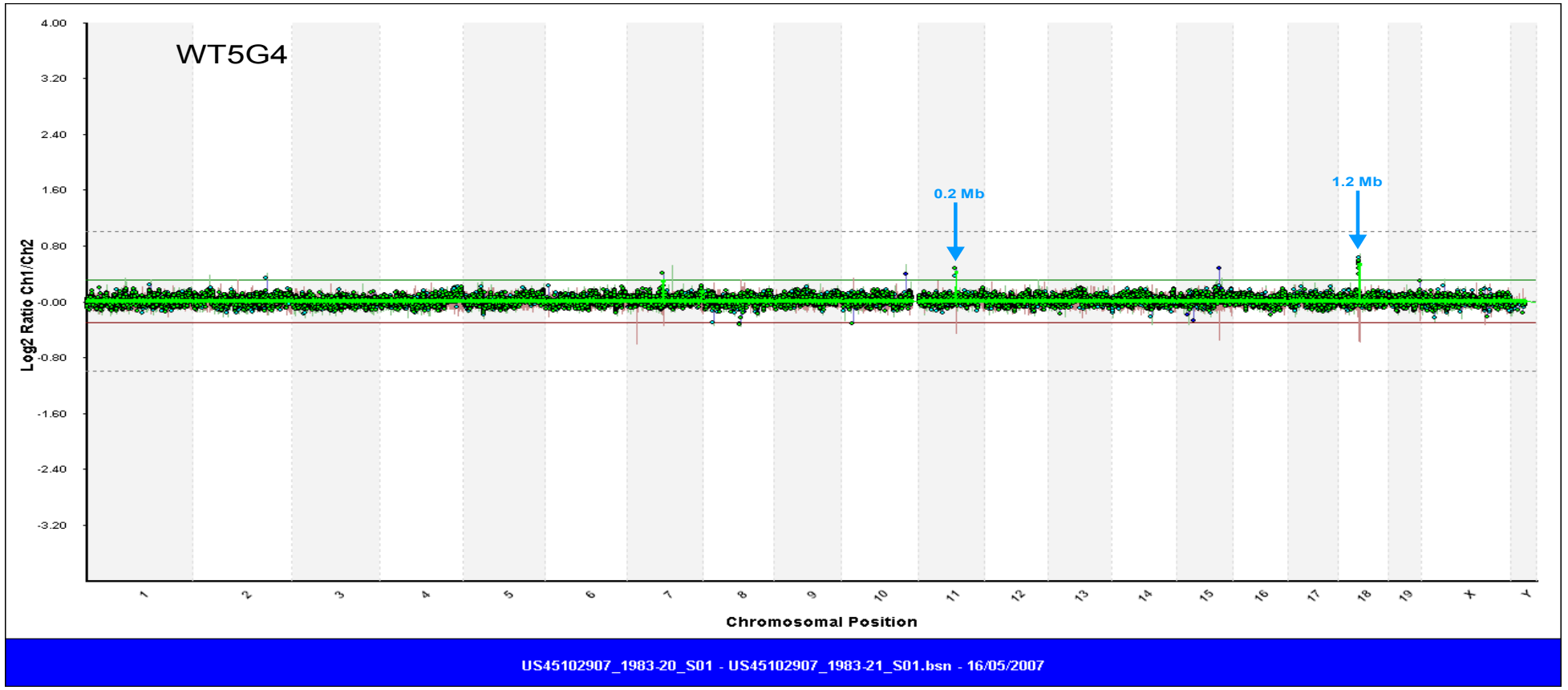


Figure 5-3 Array CGH analysis on WT5G4 – a wild type clone surviving 5Gy irradiation

The locations of two duplications of the clone WT5G4 are indicated by blue arrows. One is on chromosome 11 (0.2 Mb) and the other is on chromosome 18 (1.2Mb).

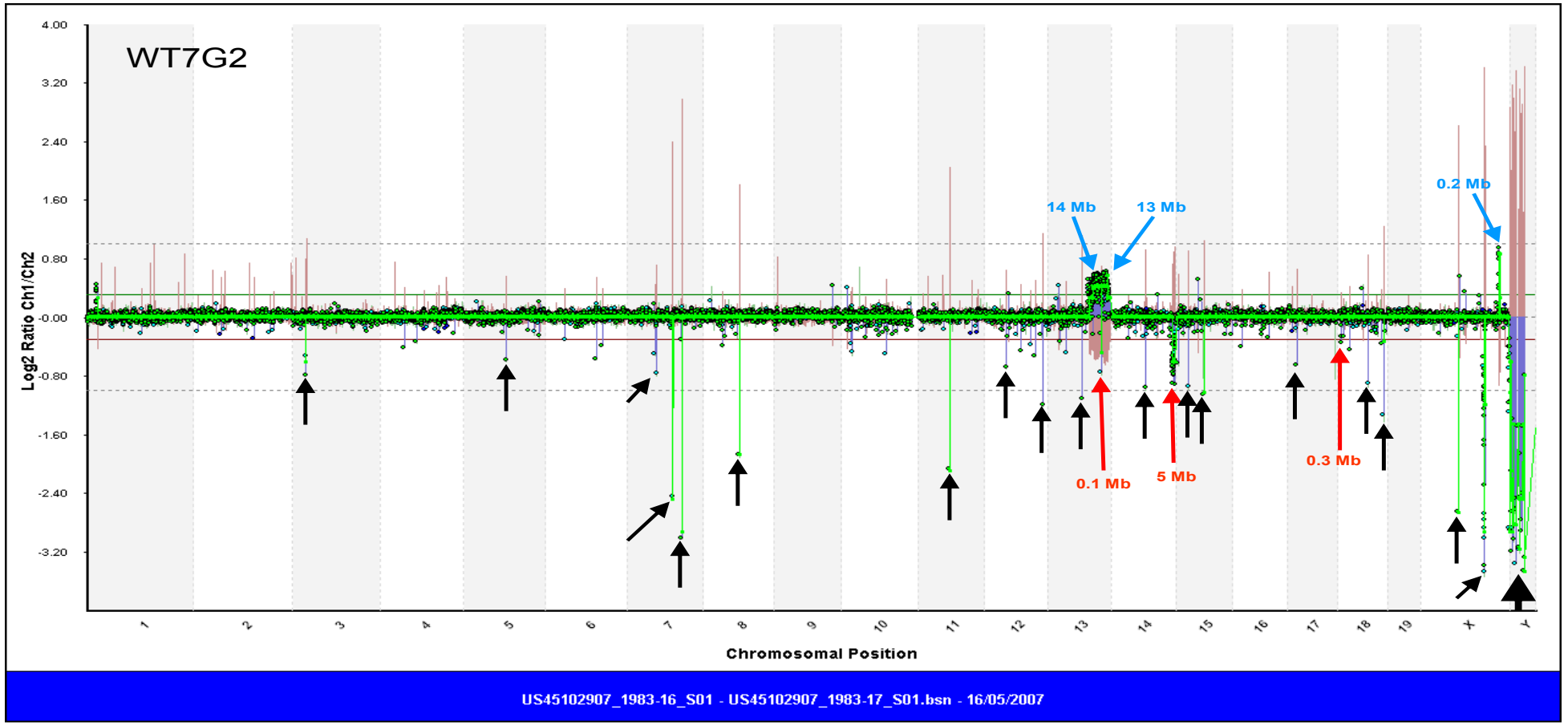


Figure 5-4 **Array CGH analysis on WT7G2 – a wild type clone surviving 7Gy irradiation**

The locations of the deletions and duplications of the clone WT7G2 are indicated by red arrows and blue arrows, respectively. Three heterozygous deletions are on chromosome 13 (0.1 Mb), 14 (5 Mb) and 18 (0.3 Mb). Two duplications are on chromosome 13 (13 and 14 Mb) and one is on chromosome X (0.2 Mb). A two-BAC-sized deletion on chromosome 3 (black arrow) and many one-BAC-sized deletions (black arrows) on chromosomes 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are probably associated with the loss of chromosome Y because the same pattern of BAC deletions were also observed in other mutant clones, which lost chromosome Y.

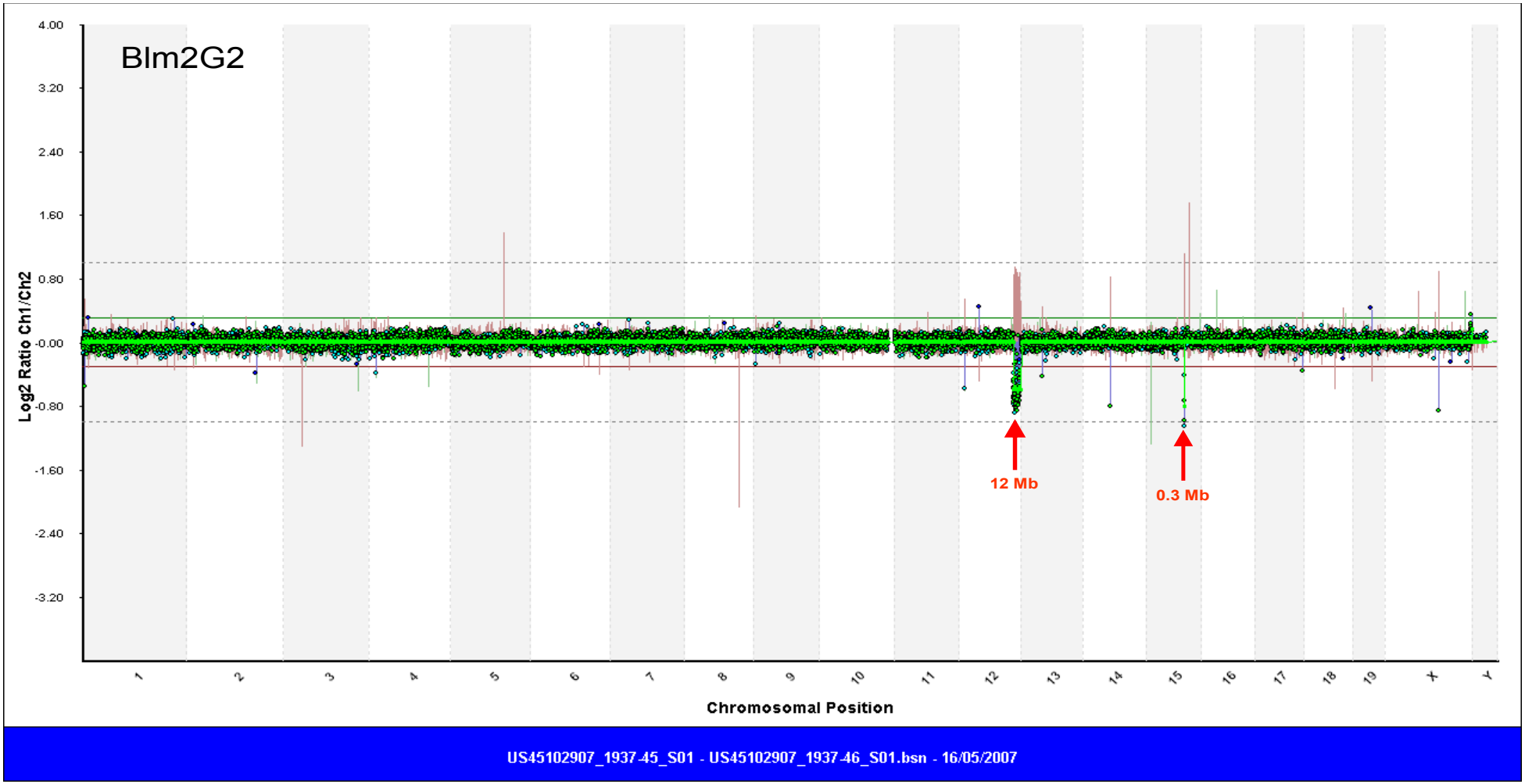


Figure 5-5 **Array CGH analysis on Blm2G2 – a *Blm*$^{-/-}$ clone surviving 2 Gy irradiation**

The locations of two heterozygous deletions of the clone Blm2G2 are indicated by red arrows. One is on chromosome 12 (12 Mb) and the other is on chromosome 15 (0.3 Mb).

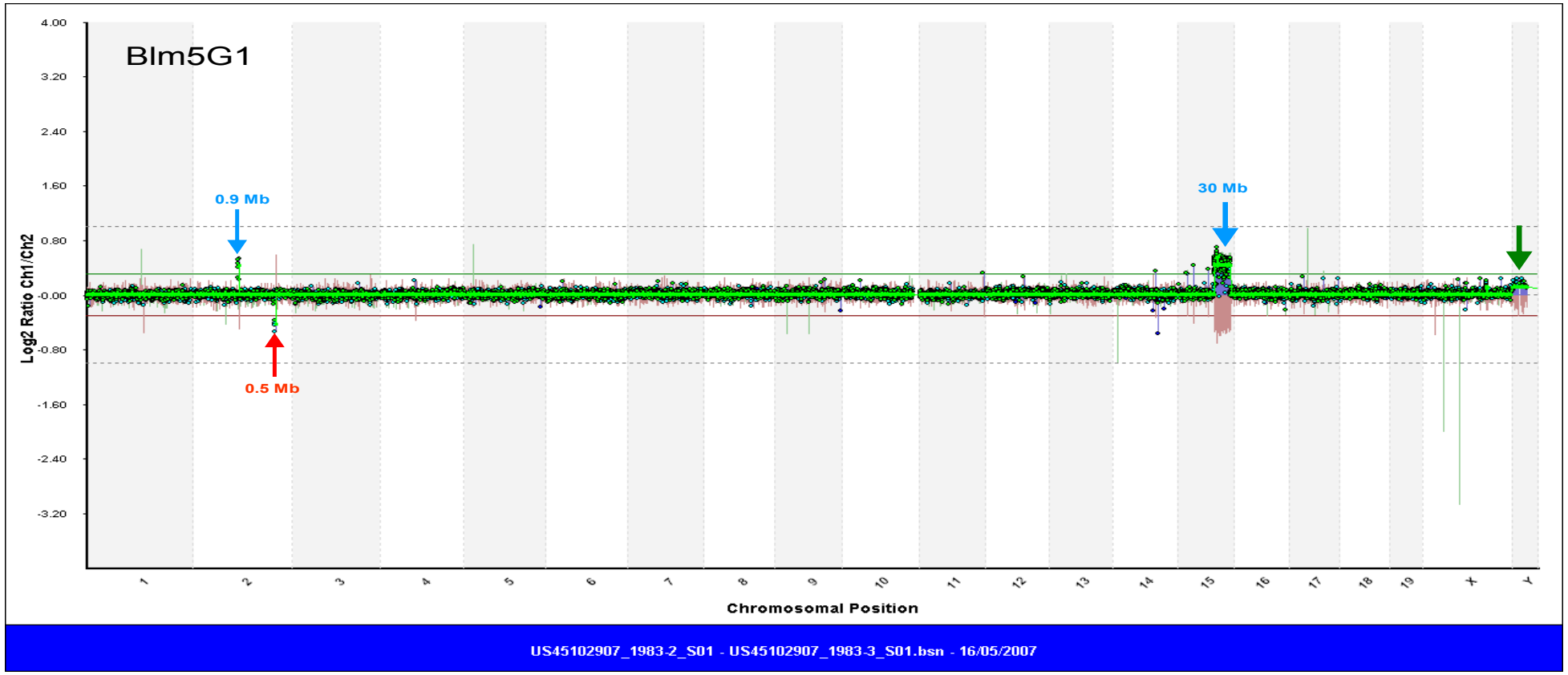


Figure 5-6 **Array CGH analysis on Blm5G1 – a *Blm*$^{-/-}$ clone surviving 5 Gy irradiation**

The locations of deletions and duplications of the clone Blm2G2 are indicated by red arrows and blue arrows, respectively. There are one heterozygous deletion on chromosome 2 (0.5 Mb) and two duplications on chromosome 2 (0.9 Mb) and chromosome 15 (30 Mb). A green arrow shows that the copy number of the clone is slightly higher on chromosome Y than that of the reference cells, indicating this clone is a mosaic population with the cells containing a normal chromosome Y and the cells containing a duplicated chromosome Y.

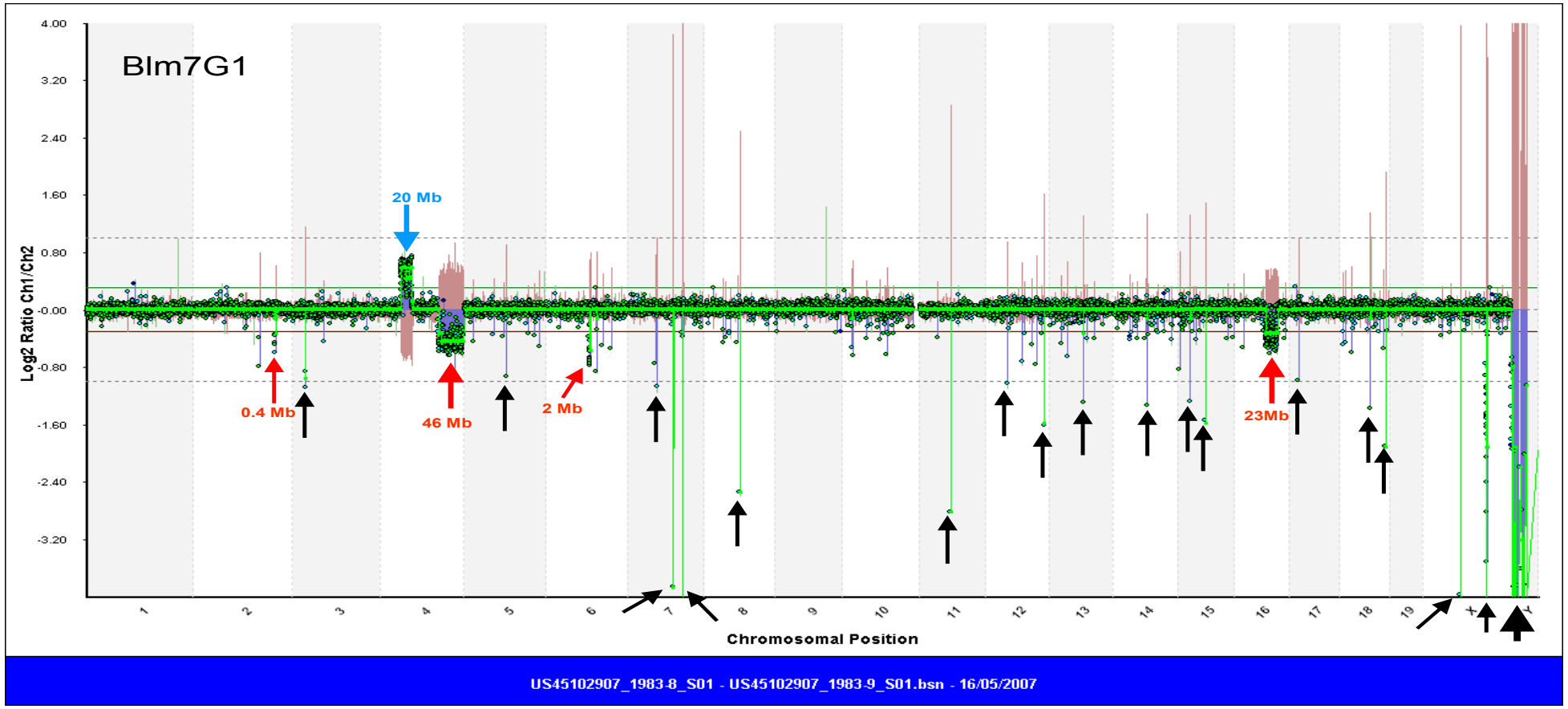


Figure 5-7 **Array CGH analysis on Blm7G1 – a *Blm*$^{-/-}$ clone surviving 7 Gy irradiation**

The locations of the deletions and duplications of the clone Blm7G1 are indicated by red arrows and blue arrows, respectively. Four heterozygous deletions are on chromosome 2 (0.4 Mb), 4 (46 Mb), 6 (2 Mb) and 16 (23 Mb). A duplication is on chromosome 4 (20 Mb). A two-BAC-sized deletion on chromosome 3 (black arrow) and many one-BAC-sized deletions (black arrows) on chromosomes 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are probably associated with the loss of chromosome Y because the same pattern of BAC deletions were also observed in other mutant clones, which lost chromosome Y.

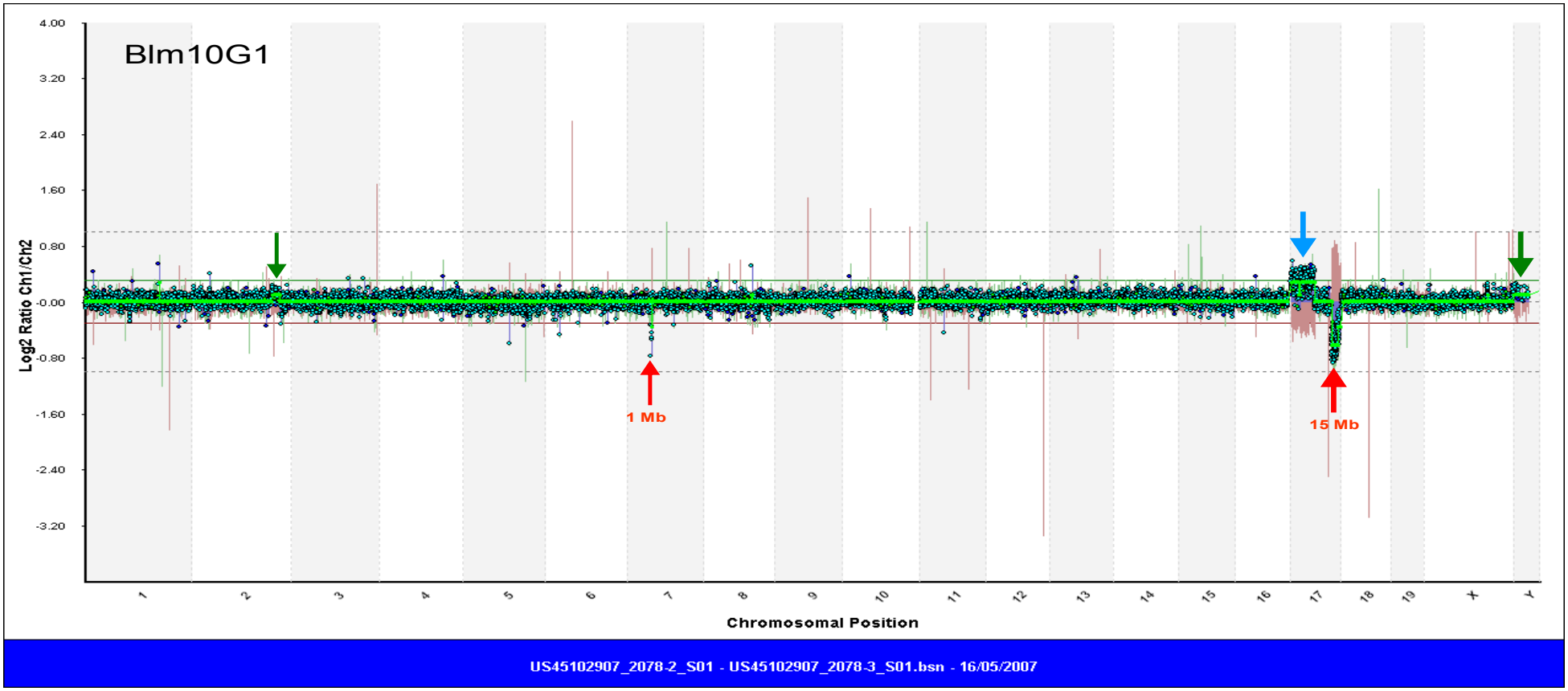


Figure 5-8 **Array CGH analysis on Blm10G1 – a *Blm*$^{-/-}$ clone surviving 10 Gy irradiation**

The locations of deletions of the clone Blm10G1 are indicated by red arrows. Two heterozygous deletions are on chromosome 7 (1 Mb) and chromosome 17 (15 Mb). A blue arrow indicates a segmental change of chromosome 17. The Log_2(ratio) of this region is near 0.3, suggesting the clone might be a tetraploid with a segmental duplication. Two green arrows show that the copy number of the clone is slightly higher in part of chromosome 2 and the whole chromosome Y, indicating this clone is a mosaic population similar to the mutant clone Blm5G1.

Table 5-5 Summary of chromosome changes in wild type cells caused by irradiation

Dose (Gray)	Clone	Number of deletions	Deletion range (Mb)	Abnormal chromosomes Log_2(ratio)<0	Number of duplications	Duplication range (Mb)	Abnormal chromosomes Log_2(ratio)>0
2	WT2G1	0	-	-	0	-	-
	WT2G2	0	-	Ch 2 and 5 (Log_2≈-0.2)	0	-	Ch 11 (Log_2≈0.2)
	WT2G3	0	-	-	0	-	-
	WT2G4	0	-	-	0	-	-
5	WT5G1	0	-	-	0	-	-
	WT5G2	0	-	-	0	-	-
	WT5G3	0	-	-	6	0.03(2), 0.05, 0.08, 0.2(2)	-
	WT5G4	0	-	-	2	0.2, 1.2	-
7	WT7G1	0	-	-	0	-	-
	WT7G2	3	0.1, 0.3, 5	Ch Y (loss)	3	0.2, 13, 14	-
	WT7G3	0	-	-	0	-	-

Data based on selective analysis by BlueFuse software. All BACs with either Log_2(ratio) between -0.29 and +0.29 or standard deviation above 0.2 in fluorescence swap experiments were excluded.

Table 5-6 Summary of chromosome changes in *Blm*$^{-/-}$ cells caused by irradiation

Dose (Gray)	Clone	Number of deletions	Deletion Range (Mb)	Abnormal chromosomes Log_2(ratio)<0	Number of duplications	Duplication Range (Mb)	Abnormal chromosomes Log_2(ratio)>0
2	Blm2G1	0	-	-	0	-	-
	Blm2G2	2	0.3, 12	-	0	-	-
	Blm2G3	0	-	-	7	0.1(2), 0.2(3), 1.5, 1.9	Ch Y (Log_2≈1)
5	Blm5G1	1	0.5	-	2	0.9, 30	Ch Y (Log_2≈0.2)
	Blm5G2	0	-	-	2	1.2, 1.7	Ch Y (Log_2≈0.2)
	Blm5G3	0	-	-	2	0.1, 0.9	Ch Y (Log_2≈0.2)
7	Blm7G1	4	0.4, 2, 23, 46	Ch Y (loss)	1	20	-
	Blm7G2	0	-	-	4	0.1, 0.5, 1.3, 2.1	-
	Blm7G3	2	2, 5.4	Ch 3, 4, 16, 17 and Y (Log_2≈-0.2)	1	0.1	Ch 5, 13 (Log_2≈0.2)
10	Blm10G1	2	1, 15	-	0	-	-
	Blm10G2	0	-	-	1	0.5	-
	Blm10G3	2	2, 13	-	1	0.5	-
	Blm10G4	3	2, 6, 7	Ch 2, 3, 4(part), 7(part), 13(part) and 17(part); (Log_2≈-0.2)	0	-	Ch Y (Log_2≈0.2)
	Blm10G5	2	6, 7	Ch Y (loss)	0	-	-
	Blm10G6	0	-	Ch Y (loss)	1	0.5	-

Data based on selective analysis by BlueFuse software. All BACs with either Log_2(ratio) between -0.29 and +0.29 or standard deviation above 0.2 in fluorescence swap experiments were excluded.

The deletion and duplication data of Table 5-5 and Table 5-6 are summarized in Table 5-7. The chromosomal abnormality is usually caused by other cellular factors instead of DNA strand breaks induced by irradiation. And chromosomal duplication and loss are so large that counting these abnormalities will distract the analysis of the sizes of the duplications and deletions induced by irradiation. Therefore, in these summarized data, abnormal chromosomes are not included. Analysis from both the wild type and the *Blm*-deficient (Table 5-7) cells exhibit that the average 0–2 deletions and 0–2 duplications were generated in each 2–10 Gray irradiated cell. Despite the similarity of the number of deletions between the two cell types, the *Blm*-deficient cells contain much larger deleted regions than that in the wild type cells. About 10–26 Mb DNA sequence was deleted after irradiating the *Blm*-deficient cells at 7 and 10 Gray but only ~3 Mb DNA sequence was deleted after irradiating the wild type cells at 7 Gray. Four out of four wild type cells surviving 2 Gray irradiation do not carry DNA duplications while the average length of duplicated regions per cell is 1.4 Mb in the *Blm*-deficient cells. There are 0.5 Mb (5 Gray) and 9.1 Mb (7 Gray) duplicated chromosome regions per wild type cell. And there are 12 Mb (5 Gray), 8.0 Mb (7 Gray) and 0.3 Mb (10 Gray) duplicated DNA regions in each *Blm*-deficient cell.

Table 5-7 **Summary of deletions and duplications in wild type cells and *Blm*-deficient cells generated by 2-10 Gray gamma irradiation**

A

Dose (Gray)	Number of samples	Average number of deleted regions ± SD	Average length per deletion (Mb)	Average length of deleted regions per cell (Mb)
2	4	0 ± 0	–	0
5	4	0 ± 0	–	0
7	3	1 ± 2	1.8	1.8

B

Dose (Gray)	Number of samples	Average number of duplicated regions ± SD	Average length per duplication (Mb)	Average length of duplicated regions per cell (Mb)
2	4	0 ± 0	–	0
5	4	2 ± 3	0.3	0.5
7	3	1 ± 2	9.1	9.1

C

Dose (Gray)	Number of samples	Average number of deleted regions ± SD	Average length per deletion (Mb)	Average length of deleted regions per cell (Mb)
2	3	1 ± 1	6.2	4.1
5	3	0 ± 1	0.5	0.2
7	3	2 ± 2	13	26
10	6	2 ± 1	6.6	9.8

D

Dose (Gray)	Number of samples	Average number of duplicated regions ± SD	Average length per duplication (Mb)	Average length of duplicated regions per cell (Mb)
2	3	2 ± 4	0.6	1.4
5	3	2 ± 0	5.8	12
7	3	2 ± 2	4.0	8.0
10	6	1 ± 1	0.5	0.3

A. Deletions in the wild type cells; **B**. Duplications in the wild type cells; **C**. Deletions in the *Blm*-deficient cells; **D**. Duplications in the *Blm*-deficient cells. SD: standard deviation.

5.3 Discussion

5.3.1 Survival frequency of ES cells after irradiation

As expected the survival rate fell as the dose of gamma irradiation increased. The rates obtained for wild type cells are consistent with published reports (You *et al.* 1997). These data enable me to extrapolate the number of independent clones in each pool of ES cells. Combined with sizes and frequency of deletions generated by a certain dose of irradiation, the genome coverage of the mutations can be calculated in this project.

Although cells with deficiency of DNA repair proteins, such as the human autosomal recessive disorder *xeroderma pigmentosum* (XP) cells can have increased sensitivity to killing exposure to a variety of DNA-damaging agents, including UV irradiation. The deficiency of the murine Bloom syndrome protein in ES cells exhibit slightly higher resistance to irradiation compared to the wild type cells. The repair of double-strand breaks (DSB), which is induced by irradiation, is usually executed through two major pathways: non-homologous end joining (NHEJ) and homologous recombination (HR). Several other sub-pathways may also exist, since many proteins have been reported to be involved in DSB repair and the coordinated functions of these proteins are necessary for successful DSB repair (Lieber *et al.* 2003; West 2003). It is known that BLM responds quickly to DNA DSB by colocalizing with phosphorylated histone H2AX (γH2AX) and ATM kinases at sites of laser light-induced DNA double-strand breaks. However, the recruitment of BLM to the damaged sites is not dependent on the presence of ATM or any of RAD17, DNA-PKcs, NBS1, XRCC3, RAD52, RAD54 and WRN (Karmakar *et al.* 2006). In *BLM*-deficient cells, phosphorylated histone H2AX (γH2AX), Chk2 and ATM checkpoint proteins can colocalized in nuclear foci following DSB (Rao *et al.* 2007). In addition, microhomology elements are not used to DSB repair in *Blm*-deficient cells (Langland *et al.* 2002). After DSB repair, more mutations were found in Bloom syndrome deficient cell extracts. This result is consistent with my results shown by array CGH that *Blm*-deficient cells carry more duplications and deletions after irradiation.

5.3.2 Megabase duplications and deletions generated by irradiation

After irradiation at certain dosage, a higher proportion of the *Blm*-deficient cells carry chromosome duplications and deletions compared with wild type cells (Table 5-5, Table 5-6) although the average number of duplications and deletions is similar between two genetic backgrounds. The average 0–2 deletions and 0–2 duplications were generated in each of the wild type and the *Blm*-deficient (Table 5-7) cells after irradiation at 2–10 Gray. The *Blm*-

deficient cells contain larger deleted regions than that in the wild type cells. About 10–26 Mb DNA sequence was deleted after irradiating the *Blm*-deficient cells at 7 and 10 Gray but only ~3 Mb DNA sequence was deleted after irradiating the wild type cells at 7 Gray. When the irradiation dose is between 2 to 7 Gray, deletion and duplication can be very large, for example there is a 46 Mb deletion and a 28 Mb duplication in Blm7G1. These large mutations make the process to map the mutated genes harder than that mutated by smaller mutations. Although the number of samples is relatively low, these deletion/duplication data roughly reflected the frequency and the size of deletions induced by irradiation. I decided to use 10 Gray to make mutation libraries as it generates reasonable size of mutations (~10 Mb per cell) at a high percentage (67%, 4/6). Thus, there is no need to select for mutations before screening by 6TG. Ten Gray irradiated cells retain around 10 Mb deleted regaions, so 450 independent survival clones will be enough to make mutations cover the whole mouse genome.

5.3.3 Chromosomal abnormalities

Besides of the heterozygous deletions and duplications induced by irradiation, whole chromosome gain/loss or segmental chromosomal gain/loss were also observed in the mutants. These can be either spontaneous or irradiation-induced. The copy number change of the whole chromosomes might be spontaneous. Chromosomes can be lost or gained in tetraploid cells, resulting in three copies or five copies of individual chromosomes. The array Log_2(ratio) of these chromosome gain should be able to be distinguished with trisomies. However, it will be hard to do so if a trisomy occur in the process of cell culture and lead to a mosaic cell population. Trisomy 8 and trisomy 11 are commonly observed in a large proportion of cultured ES cell lines (Sugawara *et al.* 2006) and thus were expected to be seen in the mutant cells. The clone WT2G2 might have a trisomy 11. Chromosome Y can be lost in diploid cells, which was also observed by my colleagues in their projects.

On the other hand, irradiation can generate large segmental chromosome gain and loss. During S/G2 phase, a deletion or a duplication of a segmental chromosome can result in 3N or 5N of the region and be kept in a tetraploid cell. If such a cell divides normally, a mosaic colony can be generated so that the Log_2(ratio) of the changed region is nearer to zero compared to those in pure colonies.

CHAPTER 6. PILOT MISMATCH REPAIR SCREEN

6.1 Introduction

Before the mutation load of the 10 Gray irradiated cells was known (Chapter 5), a small-scale experiment was designed to test the possibility of screening DNA mismatch repair mutants generated by 10 Gray gamma irradiation in the NGG5-3 *Blm*-deficient cells (Figure 6-1). In this experiment, a recombinant Moloney Murine Leukemia Virus (MMuLV) containing the positive/negative selection marker *PuroΔtk* was used. The negative selection cassette carried by the provirus facilitates selection for deletions after irradiation because a deletion that removes the *PuroΔtk* cassette can be directly selected in FIAU. Following infection, puromycin selection was implemented to isolate infected cells, which retained the *PuroΔtk* gene thus were resistant to puromycin. Then the cells were irradiated and selected in FIAU media to ensure all surviving cells lost the functional *PuroΔtk* cassettes. In principle, FIAU selection avoided the presence of non-mutated cells in the survival pool. However, the *PuroΔtk* cassette can also be lost by LOH, which is elevated in the *Blm*-deficient cells and is likely to be elevated further by irradiation. The clonal survival in FIAU was assessed and $FIAU^R$ clones were examined for deletions using a 200 kb resolution CGH array. While this assessment was in progress, the population of FIAU resistant clones were cultured for long enough to accumulate LOH events and were then selected in 6-thiogunine for MMR mutants. $6TG^R$ clones were isolated and also analysed by array CGH technology.

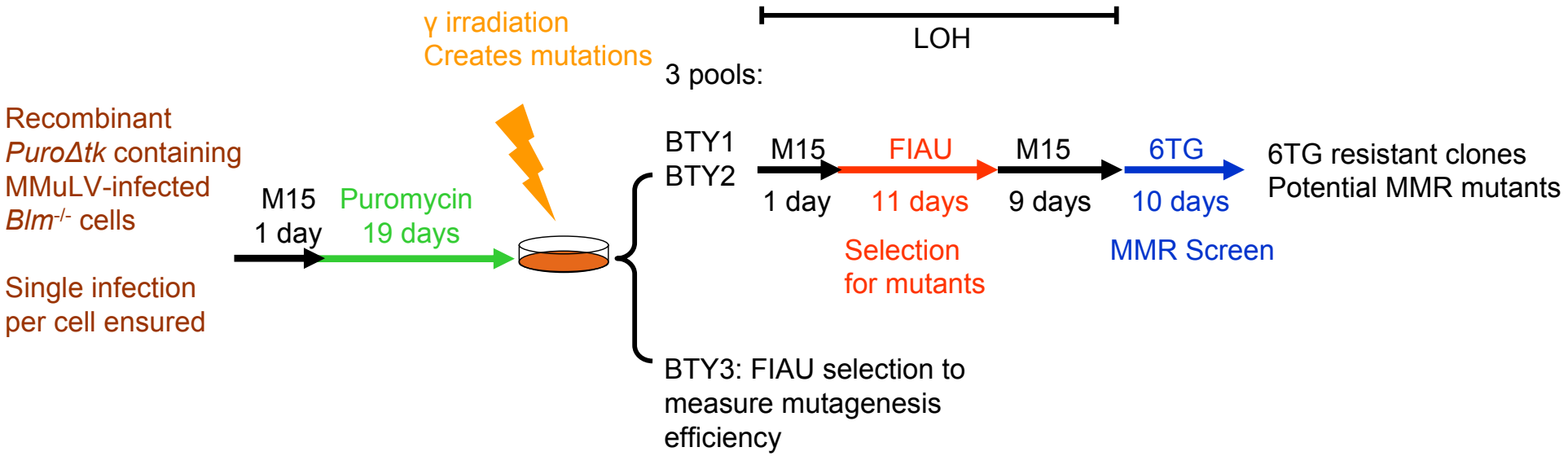


Figure 6-1 Pilot screen on *Blm*$^{-/-}$ cells carrying selected mutations generated by gamma irradiation

The *Blm*-deficient cells were infected with a *PuroΔtk* containing recombinant MMuLV. Experiments were controlled to ensure each cell had a single infection, which means one copy of the *PuroΔtk* gene was present. Single infected clones were generated after the puromycin selection. These clones were expanded followed by a 10 Gray irradiation. Cells were then split into three pools BTY1, 2 and 3. Pools BTY 1 and BTY2 were cultured in FIAU to obtain cells, which mainly had lost the functional *PuroΔtk* cassette. Then these cells were cultured for another 9 days in M15 media to allow LOH to occur. Cells were screened in 6TG for MMR-deficient mutants. Pool BTY3 was selected in FIAU to measure the complexity of the mutation pool.

6.2 Results

6.2.1 The procedure of the pilot screen

The *Blm*-deficient cells were infected with a *PuroΔtk* recombinant retrovirus (MMuLV). The *PuroΔtk* cassette provides infected cells with the resistance to puromycin and the sensitivity to FIAU. By infecting three million cells with 50 μL viral supernatant (virus titer is 10^4 cfu/mL), each infected cell was ensured to contain one copy of the *PuroΔtk* gene and five hundred infected clones were recovered. These infected clones were cultured in puromycin to avoid loss of the *PuroΔtk* gene, which can be resulted from spontaneous mutation and LOH. These cells were then expanded to 2×10^7 followed by a 10 Gray irradiation. After the irradiation, cells were then equally split into three pools BTY1, 2 and 3 and maintained as pools. All the three pools were cultured in FIAU media to isolate mutants which lost *PuroΔtk* function.

The cells in the pool BTY3 were used to measure the complexity of the mutation pool (6.2.2) and the efficiency of irradiation as a mutagen (6.2.4). After FIAU selection, cells in the pool BTY3 were stained by methylene blue and surviving cells were counted. Before the staining, five $FIAU^R$ clones were picked and expanded for a copy number variation analysis using array CGH.

The cells in the pools BTY1 and BTY2 were cultured for another 9 days in M15 media to allow LOH to occur. Cells were then screened in 2 μM 6TG at various cell densities (Table 6-1) for 10 days and eight clones were recovered (named as ITC1–8). General analysis of these mutant clones will be described in the section 6.2.5.

A parallel experiment under the same experimental condition was conducted on ES cells with a non-mutated *Blm* locus, the BAC acceptor cell line ZK6. The mutant pools generated in the ZK6 genetic background were called ATY1, 2 and 3. Five clones of pool ATY3 were also picked for copy number variation analysis.

Table 6-1 $6TG^R$ clones isolated from pools BTY1 and BTY2

Pools	Cell density (cells/90 mm plate)	$6TG^R$ clone ID
BTY1	2×10^6, 5 plates	ITC1–6
	5×10^6, 1 plate	-
	10^7, 1 plate	ITC7
BTY2	2×10^6, 5 plates	-
	5×10^6, 1 plate	-
	10^7, 1 plate	ITC8

The mutation pools BTY1 and BTY2 were selected in 6TG at a series of cell densities. Total 25×10^6 cells of each pool were selected. Eight $6TG^R$ clones were isolated, seven from the pool BTY1 and one from the pool BTY2.

6.2.2 The complexity of mutant pools

In this pilot experiment, the multiplicity of infection was controlled so that each cell would only be infected once because the number of plated cells (10^6 of both the wild type cells and the *Blm*-deficient cells) overwhelmed the active virus particles (500). The wild type (ZK6) and *Blm*$^{-/-}$ cells (NGG5-3) infected with recombinant virus were named AT and BT, each containing 500 single infection events. AT and BT cells were kept in puromycin for 19 days before 2×10^7 of each of them were irradiated at 10 Gray. Immediately after irradiation the cells were divided into three pools.

The complexity of each mutant pool was assessed. The 10 Gray survival frequency and FIAU-resistant colony number were calculated for the ATY3 and BTY3 pools (Table 6-2). The FIAU-resistant background was also assessed. Although LOH events could be reduced by maintaining puromycin selection until irradiation, it was not possible to select for FIAU function. One major cause of background is the low fidelity of reverse transcriptase (Monk *et al.* 1992). Thus, the *Δtk* cassette might carry point mutations causing cells to be resistant to both puromycin and FIAU. The puromycin and FIAU double resistant background was assessed in ATY3 and BTY3 pools. About 0.7% of infected wild type cells and 0.4% of infected *Blm*$^{-/-}$ cells should be puromycin^R and FIAU^R due to spontaneous mutation, accounting for 10 to 20 of the surviving FIAU-resistant clones. Excluding the puromycin^R and FIAU^R backgrounds, 14% of ATY3 clones and 25% of BTY3 clones are FIAU-resistant. About 96% of FIAU^R clones had lost FIAU sensitivity due to mutation or loss of the *Δtk* cassette.

6.2.3 Genomic changes after irradiation and FIAU selection

Five clones from each pool of ATY3 and BTY3 were picked for copy number variation analysis before methylene blue staining. The reference DNA was from either the ZK6 cell line (for analysis of ATY3 clones) or the *Blm*-deficient NGG5-3 cell line (for analysis of BTY3 clones). DNA from ZK6 and NGG5-3 was extracted from the passage immediately before the irradiation. The results (Table 6-3) showed that eight out of ten clones (3 from ATY3 and 5 from BTY3) carried at least one deletion after irradiation at 10 Gray. Every FIAU^R *Blm*-deficient clone contained 2 to 11 deletions ranging from 0.1 Mb to 17 Mb, although the one-BAC-sized deletions (0.1 Mb) are not convincing data.

Table 6-2 FIAU^R cells in ATY3 and BTY3

Mutation pool/ Genotype	ATY3/ WT (AB1)		BTY3/ *Blm*$^{-/-}$ (NGG5-3)	
Irradiated cell plated	6.6×10^6	5×10^5	6.6×10^6	5×10^5
10 Gray survival frequency	-	0.03%	-	0.06%
Surviving cells after Irradiation	2000	155	4000	303
Number of FIAU^R clone	287	-	984	-
FIAU^R background	14%	-	25%	-
(Puro and FIAU)^R background *	0.7%	-	0.4%	-
Number of (Puro and FIAU)^R cells	10	-	20	-

FIAU-resistant clones in the ATY3 and BTY3 pools are analysed. Pools ATY3 and BTY3 had 6.6×10^6 cells prior to gamma irradiation. In this experiment, survival frequencies of 10 Gray irradiation was measured by counting surviving colonies of 5×10^5 cells. These survival frequencies were then used to estimate the number of surviving cells of each pool. Two thousand wild type cells and four thousand *Blm*-deficient cells survived in each pool. After FIAU^R colonies were stained, 287 and 984 FIAU^R colonies were in plates of ATY3 and BTY3, respectively. * The background of puromycin and FIAU double resistant cells in pools was measured from the AT (Passage 4) and BT (Passage 4) cells, which were used to generate ATY1–3 and BTY1–3 mutation pools on the same day.

Table 6-3 Summary of chromosome changes of random picked clones from pool ATY3 and BTY3

Genotype	Clone	Number of deletion	Deletion range (Mb)
Wild type (ZK6) Pool ATY3	A1	0	-
	A2	7	0.1(6), 2.2
	B1	4	1(2), 1.6, 2
	C1	0	-
	D1	1	6
Blm$^{-/-}$ (NGG5-3) Pool BTY3	B2	2	0.1, 14
	C2	7	0.1(2), 0.5, 4.1, 1.8, 1.9, 6.9
	D2	3	0.1, 0.2, 5.5
	D11	6	0.1(4), 0.7, 14
	D12	11	0.1(6), 0.7, 1.0, 17, 15, 12

Five FIAU^R clones from each ATY3 and BTY3 pools were analysed using CGH arrays. Numbers and lengths of the deletions are described here. The number in the brackets indicates the frequency of the deletion.

6.2.4 Efficiency of gamma irradiation

The ratio of both FIAU^R and (puro and FIAU)^R colonies in the populations before and after 10 Gray irradiation were measured in multiple experiments using the parental cells AT and BT (Table 6-4). Before irradiation, about 3-4% of both AT and BT cells were FIAU-resistant. This number was roughly the same as the ratio of puromycin and FIAU double resistant cells. At this stage, all puromycin and FIAU double resistant events arose from replication errors of reverse transcriptase on the *Δtk* cassette while FIAU^R clones likely arose from LOH events. After the irradiation at 10 Gray, FIAU-resistant cells accounted for 13–14% of the surviving cells in both AT and BT cells. Assuming that the proportion of spontaneous FIAU^R and puro^R+FIAU^R cells remains the same before and after irradiation, the effect of irradiation can be assumed to represent the increase in events over this background. The data in Table 6-4 illustrates that the propoition of FIAU^R and puro^R+FIAU^R clones increased after irradiation to 10–15% of the population. This is somewhat lower than that suggested by the data in Table 6-2. This difference could be a variation between observation of a single experiment and the mean value of several experiments. The proportion of the puro and FIAU double resistant clones is about 10%, up from 3–4% before irradiation. This increase may be caused by deletions at the *Δtk* part of the puro*Δtk* gene or LOH events of a pre-existing mutation of a gene on the FIAU metabolism pathway, like what was discussed in Chapter 3. The proportion of FIAU^R cells rose from 3–4% to 13–14%. The same cause of the appearance of the puro and FIAU double resistant clones contributes to this increase. In addition, it seems 3–4% of the population are puromycin-sensitive and FIAU-resistant. These cells account for the real loss of the puro*Δtk* gene, which can be resulted from the irradiation-induced large deletions or LOH events that replaced part of the chromosome containing the puro*Δtk* gene with a part of the homologous chromosome.

6.2.5 Clone ITC8 was 6TG-resistant

After the pilot MMR screen, eight 6TG^R clones (ITC1–8) were isolated and expanded. Their drug resistance profiles were re-tested (Figure 6-2). Clones ITC1–7 were sensitive to 6TG, although a small number (1–20) of 6TG^R colonies can be seen in ITC4, 6 and 7. These single colonies may indicate weak 6TG resistance of clone ITC4, 6 and 7. Clone ITC8 appeared resistant to 6TG. This clone was analysed by array CGH technique (section 6.2.6).

Table 6-4 The ratio of FIAU-resistant and (Puro and FIAU) double resistant colonies before and after irradiation

		FIAU^R	(Puro and FIAU)^R	N	Passages
AT	**Before IR**	3.3 ± 1.3%	3.1 ± 1.8%	n=2	5–6
	After IR	13 ± 3%	10 ± 1%	n≥5	4–6
BT	**Before IR**	3.9 ± 0.5%	3.8 ± 0.3%	n=2	5–6
	After IR	14 ± 3%	11 ± 2%	n≥5	4–6

The frequencies of FIAU-resistant cells and puromycin- and FIAU-double resistant cells in the AT and BT cells infected by the *PuroΔtk* containing MMuLV were measured before and after 10 Gray irradiation (IR). The frequencies are indicated as the mean value of frequencies ± standard deviation.

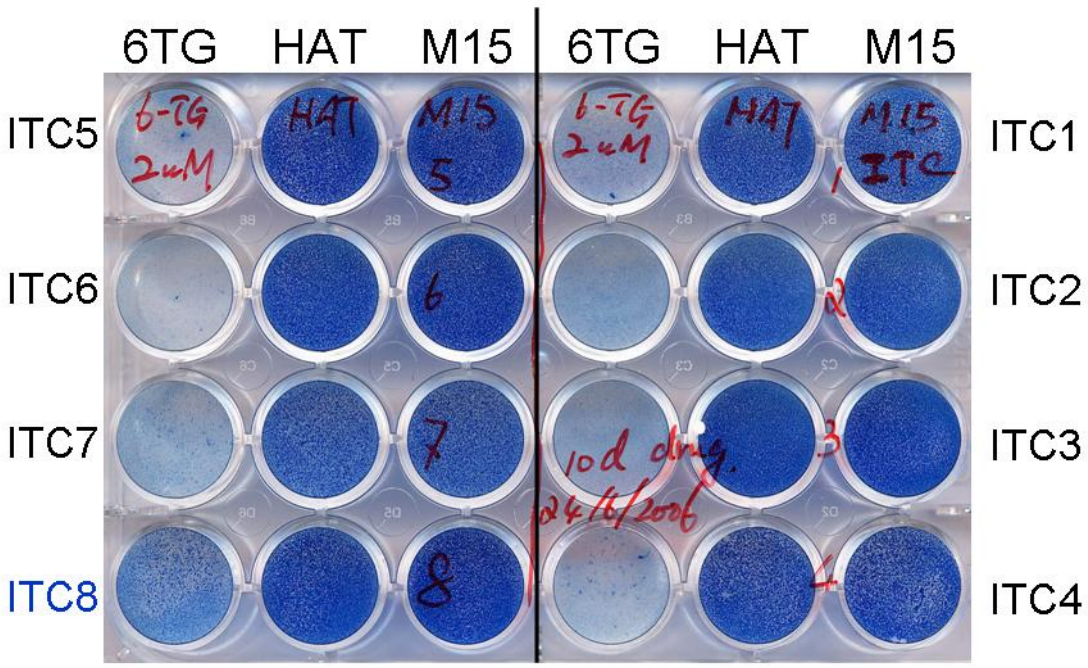


Figure 6-2 Phenotype confirmation of ITC clones

Cells from each of the clones ITC1–8 were distributed evenly into 3 wells, cultured in either normal media (M15), HAT media or 6TG media (2 μM) for 10 days. Then all the wells were stained by methylene blue (2%). Clones ITC1–7 were sensitive to 6TG, although a small number (1–20) of 6TG^R colonies can be seen in ITC4, 6 and 7. Clone ITC8 was proved resistant to 6TG. All of the eight ITC clones are HAT resistant (*Hprt*$^{+/-}$ or *Hprt*$^{+/+}$).

6.2.6 ITC8 carries a homozygous deletion covering the *Msh2* and *Msh6* loci

Array CGH analysis was conducted on ITC8 DNA with reference DNA from the NGG5-3 *Blm*$^{-/-}$ cells (Figure 6-3). A 1.8 Mb deletion on chromosome 17 with a Log_2(ratio) of -1.7 was detected. This ratio suggests that this is a homozygous deletion. This deleted region contains the DNA mismatch repair genes *Msh2* and *Msh6*. The deletion is complex and does not appear to be homozygous across the whole region. It appears that in one chromosome, 11.4 Mb distal region (including genes *Msh2* and *Msh6*) of chromosome 17 was deleted; and in the other chromosome, 2.8 Mb DNA including genes *Msh2* and *Msh6* was deleted. This 2.8 Mb region in the middle of the 11.4 Mb region is homozygously deleted. The isolation of this mutant clone and its configuration by array CGH confirmed various aspects of the experimental design: 1. It confirmed irradiation is efficient to generate mutations in ES cells; 2. It confirmed the screening conditions to isolate MMR gene deficient mutants from mutation pools; 3. It confirmed array CGH technique is adequate to identify deletions at MMR loci. Therefore a genome-wide MMR screen was designed and implemented (Chapter 7). However, it was noticed that the deletion in the mutant clone ITC8 is not homozygous but a complex, suggesting not only homozygous deletions caused by LOH, but also mutation complexes can be identified in the large-scale MMR screen.

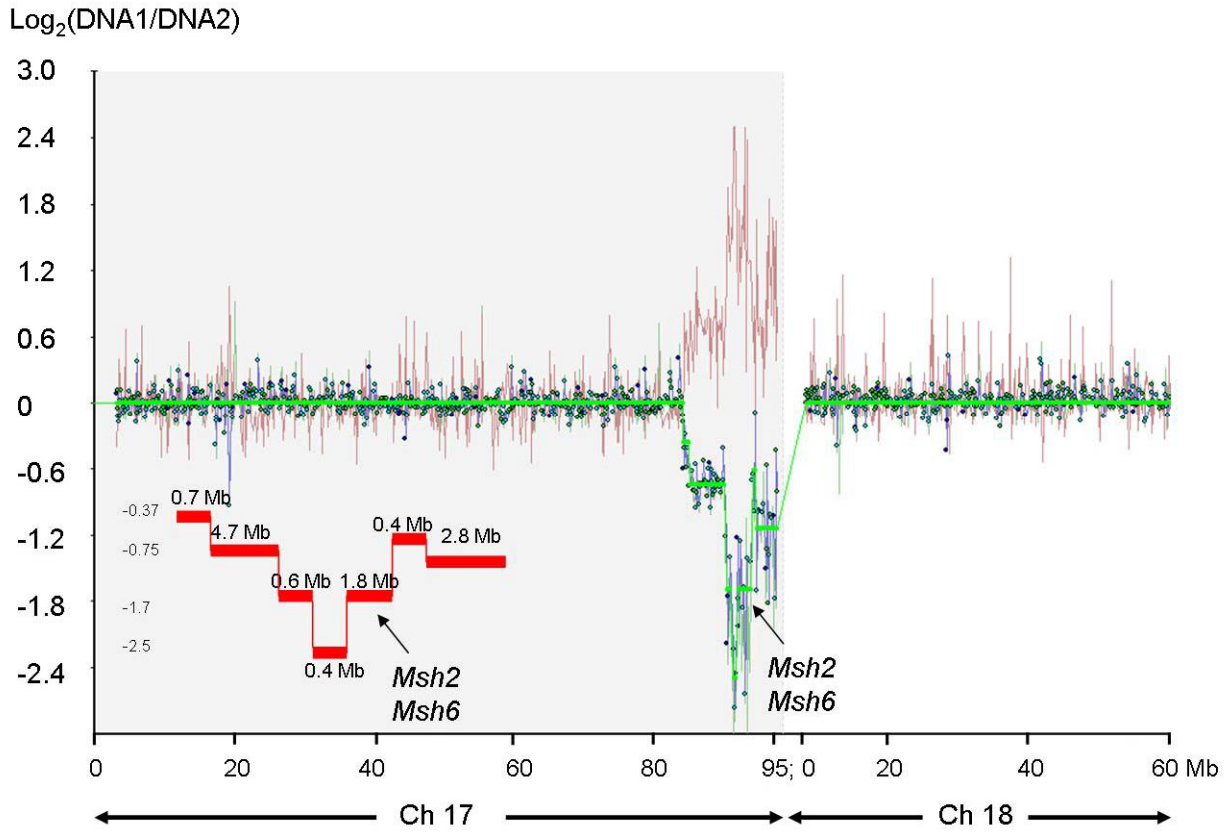


Figure 6-3 Array CGH analysis of 6TG^R clone ITC8

The array CGH profile is shown for part of chromosome (Ch) 17 and 18. The Log_2(ratio) of the mutant DNA to the *Blm*-deficient cell DNA was calculated for each BAC probe and plotted in this region. A blue line connects the nearest BACs to each other. A light green line indicates the average Log_2 (mutant DNA signal/reference DNA signal). A dark red line indicates the average Log_2 (reference DNA signal/mutant DNA signal). The red bars are a schematic illustration of the deletion. At -1.7 of the Log_2 ratio, a 1.8 Mb deletion contains the *Msh2* and *Msh6* genes.

6.3 Discussion

At the time of designing this pilot screen, it was not known that nearly 100% of surviving cells treated by 10 Gray irradiation contain mutations. Therefore it was needed to test the possibility and the procedures of isolating MMR-deficient mutants in an irradiation-induced mutation library. This pilot study used the *PuroΔtk* containing recombinant MMuLV to facilitate a preselection for mutants. Five hundred single infected cells were isolated and expanded. These cells were irradiated at 10 Gray and selected in FIAU for the mutations destroying the *Δtk* cassette at the proviral loci. In the parallel experiments on wild type cells, three out of five FIAU-resistant cells contain deletions, ranging from one deletion to seven deletions in each mutated cell. All five randomly chosen $FIAU^R$ cells contain deletions, ranging from 2 to 11 deletions per cell. These data confirmed that FIAU selected *Blm*-deficient cells all carry deletions thus were ready for MMR screen. However, it could not be assumed that all mutations only occurred around the proviral loci because we know now that the average size of all deleted regions in one survival cell may be as long as 10 Mb and the average number of deletions per cell is about two (Chapter 5). The proportion of $FIAU^R$ and $(puro+FIAU)^R$ cells were measured before and after irradiation at 10 Gray. About 13–14% of the cells surviving irradiation are FIAU-resistant, up from 3–4% before irradiation. Three to four per cent surviving cells are $puromycin^S$ and $FIAU^R$, which account for the real loss of the *PuroΔtk* cassette, probably due to deletion or LOH.

After FIAU selection, these *Blm*-deficient mutant cells were cultured for another nine doublings to accumulate LOH events and generate homozygous mutations. Eight $6TG^R$ clones were isolated. One of the eight mutants, ITC8, is strongly resistant to 6TG. Using array CGH, the muant clone ITC8 was analysed and both copies of the *Msh2* and *Msh6* genes were found to have been lost. This indicated a high likelihood of isolating homozygous deletions within the DNA mismatch repair pathway. In addition, the array CGH experimental data demonstrated that it was qualified to detect large chromosome copy number variations generated by irradiation. As this was a pilot screen, array CGH analysis on the other mutants was not performed.

About 14% ((287-10)/2000) of the wild type cells and 24% ((984-20)/4000) of the *Blm*-deficient cells, both containing the *PuroΔtk* cassette, lost the function of the *Δtk* cassette after 10 Gray irradiation (Table 6-2). That means around 10–20% of the mouse genome (30–60Mb) was affected by 10 Gray gamma irradiation. This number is too high to believe. To further validate the possibility, several experiments can be designed. One of these is to

use single targeted cells with the *PuroΔtk* positive/negative selection cassette to conduct similar experiments. Different targeted loci can be used but experiments have to be conducted separately. These targeted cells can be irradiated then selected for loss-of-function of the *Δtk* cassette. This will avoid the complexity of measuring mutagenesis efficiency in a pool. Secondly, array CGH with a higher resolution can be used. The array CGH I used was a 200 kb resolution array CGH. This array cannot provide more accurate data for deletions and duplications. Currently I assume that irradiation generates a huge number of small deletions and other mutations, which are beyond the current detection limit of array CGH. By using arrays with a resolution of 50 bp or 5 kb, more small deletions and duplications should be identified. Thirdly, the average length of deleted regions per cell and the average length of amplified regions per cell were obtained from 3–6 samples at each dose (Chapter 5). The sizes of these sample groups were not big enough to make these data convincing. Thus, more survival clones need to be analysed to validate these data.

This pilot screen demonstrates several facts of successfully using irradiation mutagenesis to conduct a genome-wide recessive MMR screen. 1. 10 Gray irradiation can efficiently generate mutations in *Blm*-deficient ES cells; 2. the *Blm*-deficient cells are efficient in generating homozygous mutations. 3. MMR-deficient mutants can be isolated by 6TG selection. 4. 200 kb resolution array CGH can be used to identify homozygous deletions generated by irradiation. However, there are some notable drawbacks of irradiation. Combining these results with data from Chapter 5, I found that irradiation can generate multiple duplications and deletions in a single cell. This will increase the difficulty of identifying which gene is responsible for the 6TG resistance. Especially, large heterozygous deletions in one allele may be coupled with small mutations in the other allele. When the small mutations are not identified by array CGH, the phenotype causing the mutation cannot be detected. Thus, in the genome-wide screen, expression array technique may be used to complement the analysis using array CGH. Expression array can generate data of transcriptional variations. It can identify which genes lose their transcripts even though these genes are not regarded as deleted according to array CGH. Nonetheless, an expression array analysis also has some drawbacks in identifying 6TG resistance related mutations. As the transcription of downstream regulated genes can be shut down by the effect of other mutations, downstream regulated genes may be regarded as mutated. A possible way to solve this problem is to combine data from both array CGH and expression. It is highly likely that genes which are deleted and transcriptionally silenced in a mutant retain 6TG resistance, causing mutations at MMR loci.

CHAPTER 7. GENOME-WIDE MISMATCH REPAIR SCREEN

7.1 Introduction

The data of the pilot screen described in Chapter 6 indicated that cells with mutations in mismatch repair (MMR) genes could be identified in a *Blm*-deficient ES cell mutation library generated by irradiation. It also showed that array CGH with 200 kb resolution was able to scan the mouse genome for deletions induced by irradiation. Thus, in this chaper a new mutation library with higher genome coverage was generated and screened (Figure 7-1). In this screen, we expected to identify mutants of known MMR genes and discover new ones. Prior to generating the mutation libraries, the cells were tagged with a recombinant MMuLV retrovirus to avoid parallel analysis of daughter clones. This is particularly important because array analysis is very expensive and it is much better to analyse more clones than the same clone multiple times. The diverse genomic locations of these molecular tags facilitate the identification of relationships between clones in the same pool. Seven mutation pools were generated by irradiating the MMuLV tagged *Blm*-deficient cells (NGG5-3) at 10 Gray. Screening these individual pools has at least two advantages: Firstly, it is easy to distinguish independent mutants among pools. If a gene is very important to the MMR system and its mutants are able to be isolated, mutants of this gene can be expected in more than one pool due to the high genome coverage of irradiation-induced mutations. Mutation pools were generated and screened separately, thus clones in one pool are independent from clones in another pool. Secondly, individual mutation pools prevent one mutant from dominating the whole mutation library. The screen is targeted for homozygous deletions generated through loss of heterozygosity, which may occur at anytime in cell culture. Therefore, mutants for which LOH occurred earlier can be many times more abundant than mutants for which LOH occurred later. A single mutant with an early LOH event generating a homozygous mutation can replicate to be approximately 1000 (2^{10})–10,000 (2^{14}) times more than any other mutants. By using several mutation pools, this phenomenon, though it occurs, will not affect the process of isolating mutants in other pools.

Based on the previous data of irradiation survival frequency, the number of cells to be irradiated can be calculated to achieve a required number of independent mutants. This must be less than the total number of the tags in each pool. This ensures that each surviving clone has a unique tag and can be distinguished. It is known from Chapter 5 that two-thirds of surviving cells contain deletions and each surviving cell has ~10 Mb deleted regions following a 10 Gray irradiation. The haploid mouse genome is 3,400 Mb in length

thus one thousand surviving clones in a pool ensure that deletions can cover the mouse genome twice. After irradiation, the mutant pools were expanded for 15 cell doublings which leaves enough time for non-sister homologous chromatid exchange to generate LOH events and also to allow the decay of messenger RNA and proteins of mutated MMR genes. The mutation library was then selected in 6TG media to isolate candidate MMR-deficient clones.

Array CGH analysis is able to detect genomic heterozygous and homozygous deletions generated by irradiation. Homozygous deletions at known MMR genes and novel genes are anticipated in the mutants. Observation of homozygous deletions at known MMR loci, for instance *Msh2* and *Msh6*, would confirm the reliability of the screen system. At the same time, new mutant genes should be identified. However, heterozygous deletions, recognized by array CGH, are also worth further consideration as they may lead to loss-of-function mutations. Small mutations, occurring in the other non-deleted homologous DNA sequence of the so-called heterozygous deletion, can also cause loss-of-function mutations, such as frame shift mutants.

In the mutation library induced by irradiation, MMR mutants may have mutations, such as small mutations or translocations, which cannot be detected by the 200 kb array CGH, but these can result in abnormality at the transcriptional level. Thus, mutant clones were also analysed using the Illumina® Mouse-6 Expression BeadChip. Small mutations can cause loss-of-function mutations through the mechanism of nonsense-mediated decay (NMD). NMD is an RNA quality surveillance pathway, by which mRNAs harbouring premature termination (nonsense) codons (PTC) can be selectively degraded (Chang *et al.* 2007). The mRNA molecules with PTCs can generate deleterious functions if they are translated to truncated proteins. It is necessary to remove these mRNAs to avoid deleterious functions. Irradiation can introduce premature termination codons in mRNA through reading frame shifts. The absence of its mRNA led by NMD can be detected by expression array. Moreover, reading frame shifts caused by translocations can also result in PTCs and the degradation of its mRNA through the NMD pathway. Another application of expression array analysis is the identification of genes which are regulated directly or indirectly by defects in MMR.

To summarise, large homozygous and heterozygous deletions can be identified by array CGH; the silenced genes caused by smaller mutations and translocations can also be detected by the expression array. By combining two array analysis approaches, novel MMR genes and components in the MMR-associated pathway can be identified.

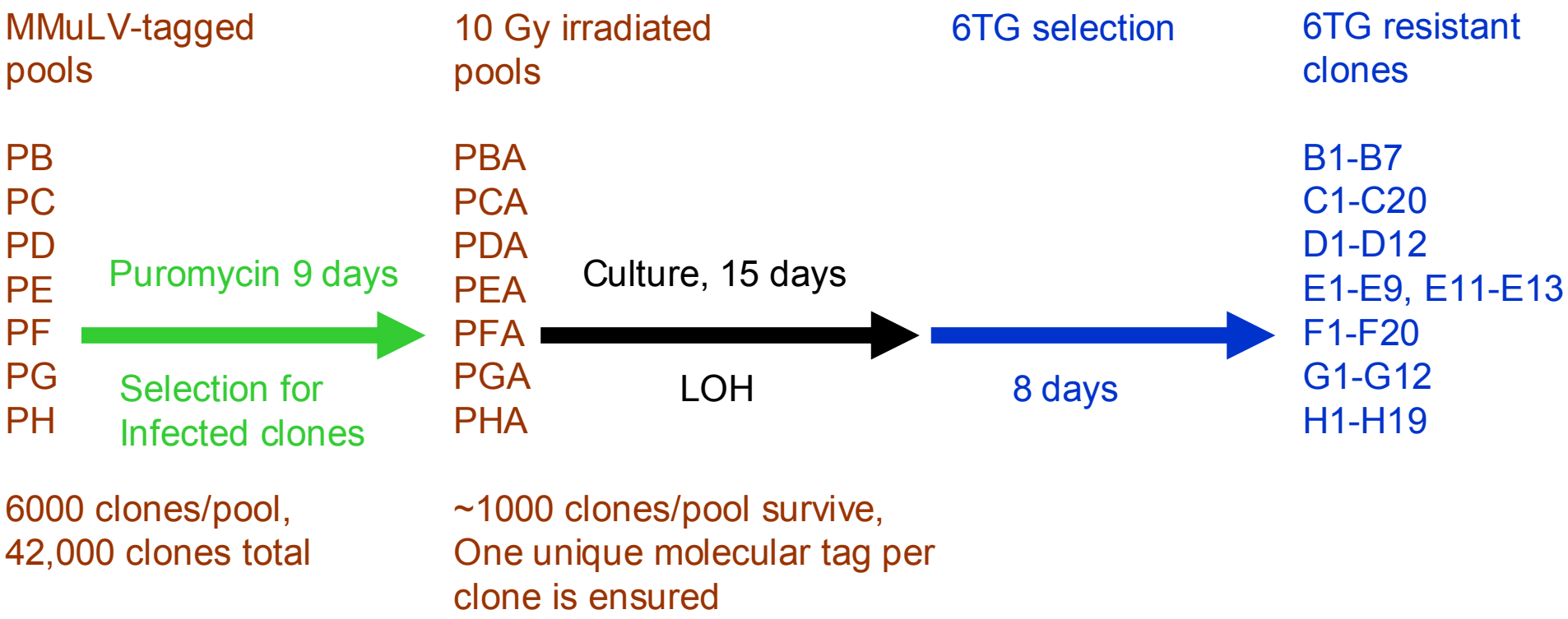


Figure 7-1 The procedure of a genome-wide MMR screen on the MMuLV-infected, irradiated ES cell library

A recombinant retroviral vector (MMuLV) containing the *PuroΔtk* selection marker was used to infect *Blm*-deficient ES cells in seven pools (PB–PH). Cells were selected in puromycin for nine days. The ratio of virus to cells was controlled to ensure that each cell was only infected once. Each pool consisted of 6,000 independently infected cells. When these cells grew to form colonies, they were trypsinized and irradiated separately, which resulted in ~1000 clones in each pool. These were expanded for 15 days in non-selective media and then selected in 6TG media. The resulting MMR-deficient candidates are termed B–H followed by clone numbers.

7.2 Results

7.2.1 Use of MMuLV to provide molecular tags in mutation pools

The *Blm*-deficient ES cells (NGG5-3) were infected in seven pools (PB–PH) by a recombinant MMuLV retrovirus containing the *PuroΔtk* cassette. The ratio of recombinant virus to ES cells was controlled so that each cell was infected only once. The recombinant MMuLV was generated by a virus producing cell line B4-5 (Wang and Bradley 2007). This cell line was cultured in non-selective ES cell media for one day then the media was fitered to remove any cells. When the cells were 70–80% confluent, they are trypsinized and plated with an appropriate volume of media containing the virus. Infected cells were selected in puromycin and the resisting colonies composed of mutation pools. The viral titer was determined by plating three million cells with 60μL media containing the virus, followed by puromycin selection. The resulting colonies were stained by methylene blue and counted. Under this condition, 607 infected colonies were identified, which gives a titre of 10,000 viral particles per mL. The 600μL media containing the virus was used to infect three million *Blm*-deficient cells per pool and approximately 6000 colonies per pool were achieved. Each pool were irradiated at 10 Gray to generate mutation pools according to the data in Chapter 5, survival frequency of the *Blm*-deficient cells irradiated at 10 Gray was ~0.1%. Therefore, I expect that mutation pool has a complexity of 800 clones.

Given that there were 6000 independent retroviral insertions per pool before cells were expanded for irradiation. Thus, in the 800 colonies of each pool after the irradiation, each should carry a unique molecular tag provided by the recombinant MMuLV. Thus, clonal relationships within a pool can be determined by Southern blot analysis using a *PuroΔtk* probe and an enzyme that generates a junction fragment between the provirus and the genomic DNA. If the fragments identified by Southern blot analysis are the same between two $6TG^R$ clones, it is highly likely that they are derived from the same clone. Relationships can be verified later by the pattern of genomic alterations.

7.2.2 Gamma-irradiation mutation library

A dose of 10 Gray was used to irradiate 4×10^6 cells per pool. Cells were plated in one 90mm plate per pool. This resulted in around one thousand independent mutant clones per pool, which is the complexity of the pool (Table 7-1).

Each mutant clone formed a colony after 10 days' culture in non-selective media. Colonies of each pool were trypsinized, passaged and cultured for another five days in non-selective

media. Then mutation pools were ready for 6TG selection as 15 cell doublings should generate at least one LOH event of each mutation. Ten million cells of each mutation pool were selected in 6TG at the density of two million cells per 90 mm plate.

Table 7-1 Complexity of each mutation pool

	Pools						
	PB	**PC**	**PD**	**PE**	**PF**	**PG**	**PH**
Colony number (10^5 cells/well)	21	19	20	19	25	33	26
Survival frequency	0.021%	0.019%	0.020%	0.019%	0.025%	0.033%	0.026%
Pool complexity	840	760	800	760	1000	1320	1040

Cell pools PB–PH were irradiated at 10 Gray. After irradiation, 100,000 cells were plated in a well to calculate irradiation survival frequency for each mutation pool. According to these data, about 1000 (760–1320) independent clones were obtained in each mutation pool.

7.2.3 Assessment of screen background

In order to control for the presence of existing MMR mutants in the *Blm*-deficient parental cells (NGG5-3) used to construct the irradiation library, the cells were plated in 6TG under the conditions used to screen the library pools (Figure 7-2). This cell density was selected for two reasons: 1. This number of cells will not grow to confluency during the screen process; 2. This cell density will select against *Hprt*-deficient mutants. *Hprt*-deficient ES cells are resistant to 6TG, however, at this high cell density, surrounding *Hprt*-proficient cells can pass 2'-deoxy-6-thioguanosine-triphosphate, a product with a phosphoribosyl group transferred by hypoxanthine guanine phosphoribosyl transferase, into the *Hprt*-deficient cell, thus killing the *Hprt*-deficient cells. Therefore, *Hprt*-deficient mutant clones can be removed from the $6TG^R$ mutant clones. In *Hprt*-proficient cells, 6TG can integrate into DNA strand as a nucleotide analogue and be methylated by an intracellular methyl group donor, S-adenosylmethionine (SAM) and forms S^6-methylthioguanine [^{SMe}G] (Karran *et al.* 2003; Swann *et al.* 1996). The resulting mismatch is then recognized by MutSα (MSH2 and MSH6) and the cell is led to cell cycle arrest and apoptotic cell death. When an MMR component, such as MSH2 protein, is deficient, cells survive in 6TG.

To measure the background of this MMR screen, four million parental *Blm*-deficient cells were selected in 2 μM 6TG at the density of 2×10^6 cells per plate for 8 days. These cells were cultured for another five days in non-selective media (M15) to allow potentially surviving clones to form bigger colonies. These two plates were stained by methylene blue to visualize colonies. A single colony-like blue spot was observed in one plate and none was observed in the other plate. According to this result, a maximum of 2–3 pre-existing MMR mutant might be isolated in each mutation pool. If a similar number of clones were recovered from one pool, these clones would possibly be false-positives. However, since the mutations in each pool were designed to cover the mouse genome three times and each heterozygous mutation can lose its heterozygosity in 12 cell doublings ($-Log_2(4\times10^{-4})$), in theory, each original mutant can have less than 8 (2^3) homozygous mutant clones after 15 cell doublings. Therefore the real positive clones in each pool should be numerous, thus this level of background should not interfere with the results.

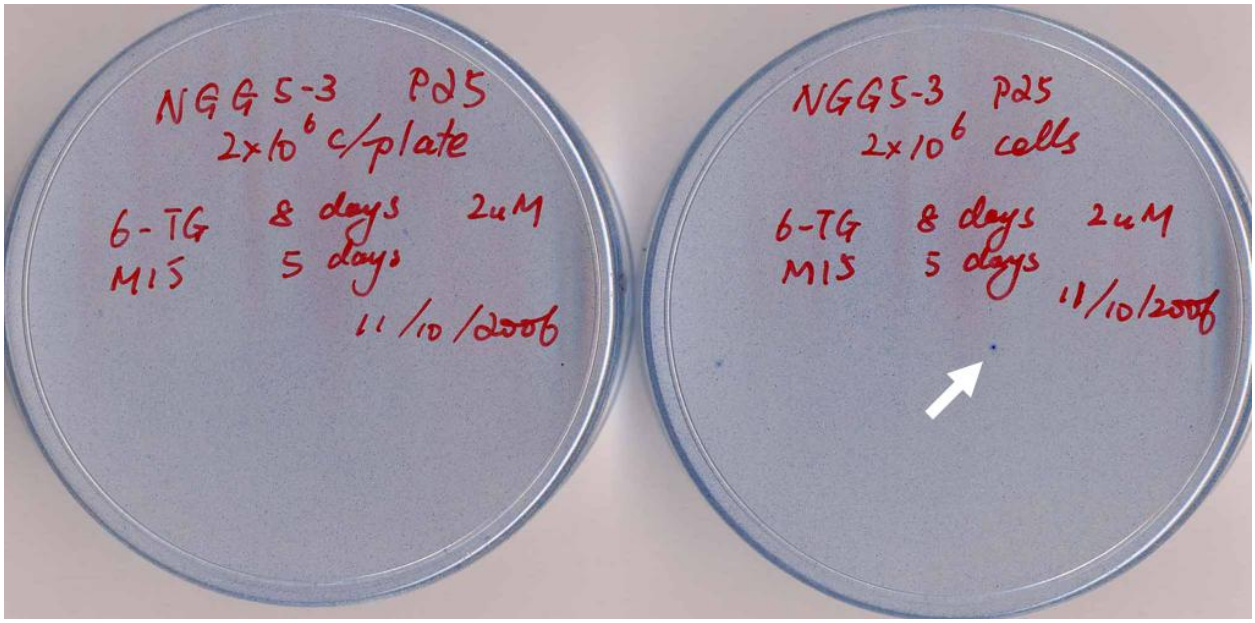


Figure 7-2 Clean background of the 6TG screen on *Blm*$^{-/-}$ cells

Two million NGG5-3 *Blm*-deficient cells were plated in each 90 mm plate. After 24 hours, cells were selected in 2 μM 6TG for eight days and recovered in M15 normal media for another five days. Then plates were stained by methylene blue. Only one colony-like blue spot is in both plates (indicated by a white arrow).

7.2.4 6TG-resistant clones

In total, 103 $6TG^R$ clones were identified from the seven pools. These were isolated in 96-well plates and named B–H followed by a clone number. These clones were distributed between 7 mutant pools with 7–20 clones per pool (Table 7-2). The 6TG resistance of each clone was re-tested in a 24-well plate by replica plating in 2 concentrations of 6TG (either 1 or 2 µM) as well as HAT (Figure 7-3–9). After selection, the cells were stained by methylene blue. MMR mutants should be resistant to 6TG and HAT, whereas *Hprt*-deficient mutants will be resistant to 6TG and sensitive to HAT. As expected, the results showed no *Hprt*-deficient mutants. Of the clones, 49 were strongly resistant to 6TG, 29 were weakly resistant and 25 were false positives since they were 6TG sensitive (Table 7-2). The weak resistance to 6TG may be due to a dose-dependent effect of heterozygous deletions of MMR genes.

Several selection controls were plated in parallel. The NGG5-3 cell line is the *Blm*-deficient parental cell line of the mutants with the genotype of $Blm^{-/-}Hprt^{+/+}$. With the presence of the *Hprt* gene, it was resistant to HAT but sensitive to 6TG (Figure 7-3, Figure 7-6). The NGG5-3 cell line was derived from the NN5 cell line, which was generated from the double targeted cell *Blm*(m1/m3) (Luo *et al.* 2000). NN5 was generated from the *Hprt*-deficient AB2.2 cell line and is thus $Blm^{-/-}Hprt^{-}$. Without the functional *Hprt* gene product, NN5 was sensitive to HAT but resistant to 6TG (Figure 7-8). The cell line ww56+*Hprt* ($Dnmt1^{-/-}Hprt^{+}$) has a double targeted *Dnmt1* gene and a heterozygously targeted *Gdf9* gene with an *Hprt* selection marker. As described by Guo, a *Dnmt1*-deficient cell was resistant to 6TG (Guo *et al.* 2004). With an *Hprt* gene, it was resistant to HAT, too (Figure 7-3, Figure 7-6). As expected, all control cell lines demonstrated the correct phenotype in hypoxanthine/aminopterin/thymine (HAT) or 6TG.

Table 7-2 6TG resistant mutant clones in the mutation library

A

Pool	PBA	PCA	PDA	PEA	PFA	PGA	PHA
Number of 6TG^R clones observed	7	20	12	13	20	12	19

B

Pool	**Phenotypes**			**Total number of clones**
	Strong	**Weak**	**Sensitive**	
PBA	2	3	2	7
PCA	7	6	7	20
PDA	9	2	1	12
PEA	6	4	3	13
PFA	15	5	0	20
PGA	4	2	6	12
PHA	6	7	6	19
Total number of clones	**49**	**29**	**25**	**103**

C

Clone ID	**Strong**	**Weak**	**Sensitive**	**Total number of clones**
B	6;7	1;2;3	4;5	7
C	1;2;4;5;10;15;17	3;11;13;14;18;19	6;7;8;9;12;16;20	20
D	1;2;3;4;5;7;8;9;12	6;11	10	12
E	1;3;4;5;9;14	2;7;11;13	6;8;12	13
F	1;2;3;4;6;8;11;12;13;15;16;17;18;19;20	5;7;9;10;14	0	20
G	6;8;9;10	1;4	2;3;5;7;11;12	12
H	1;3;5;6;13;14;	2;4;7;8;9;11;12;	10;15;16;17;18;19	19
Total number of clones	**49**	**29**	**25**	**103**

A. Number of clones isolated from each mutation pool; **B**. 6TG resistance phenotypes of primary clones; **C**. Scoring of 6TG resistance phenotypes of individual clones.

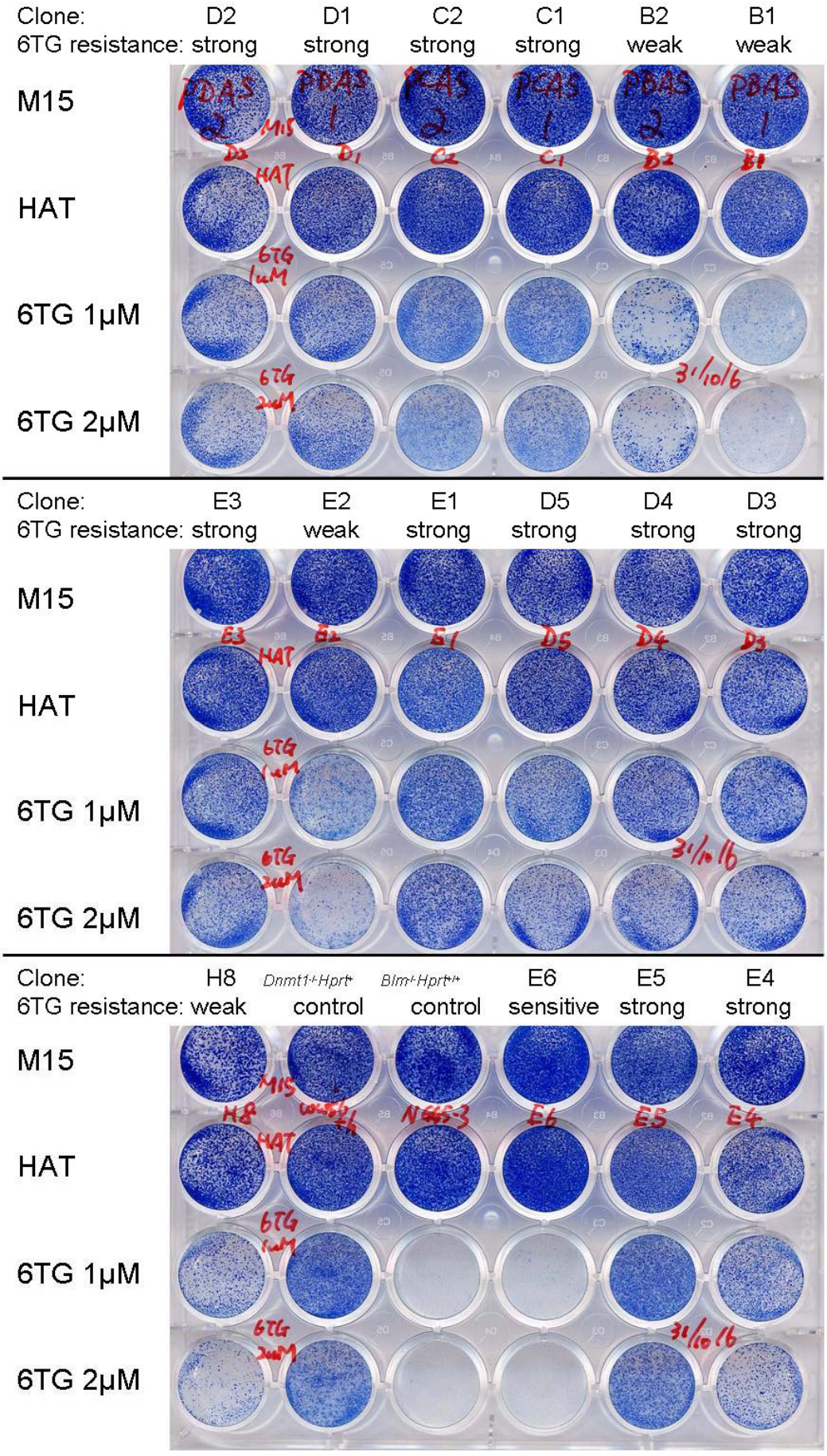


Figure 7-3 **Phenotype of mutant clones**

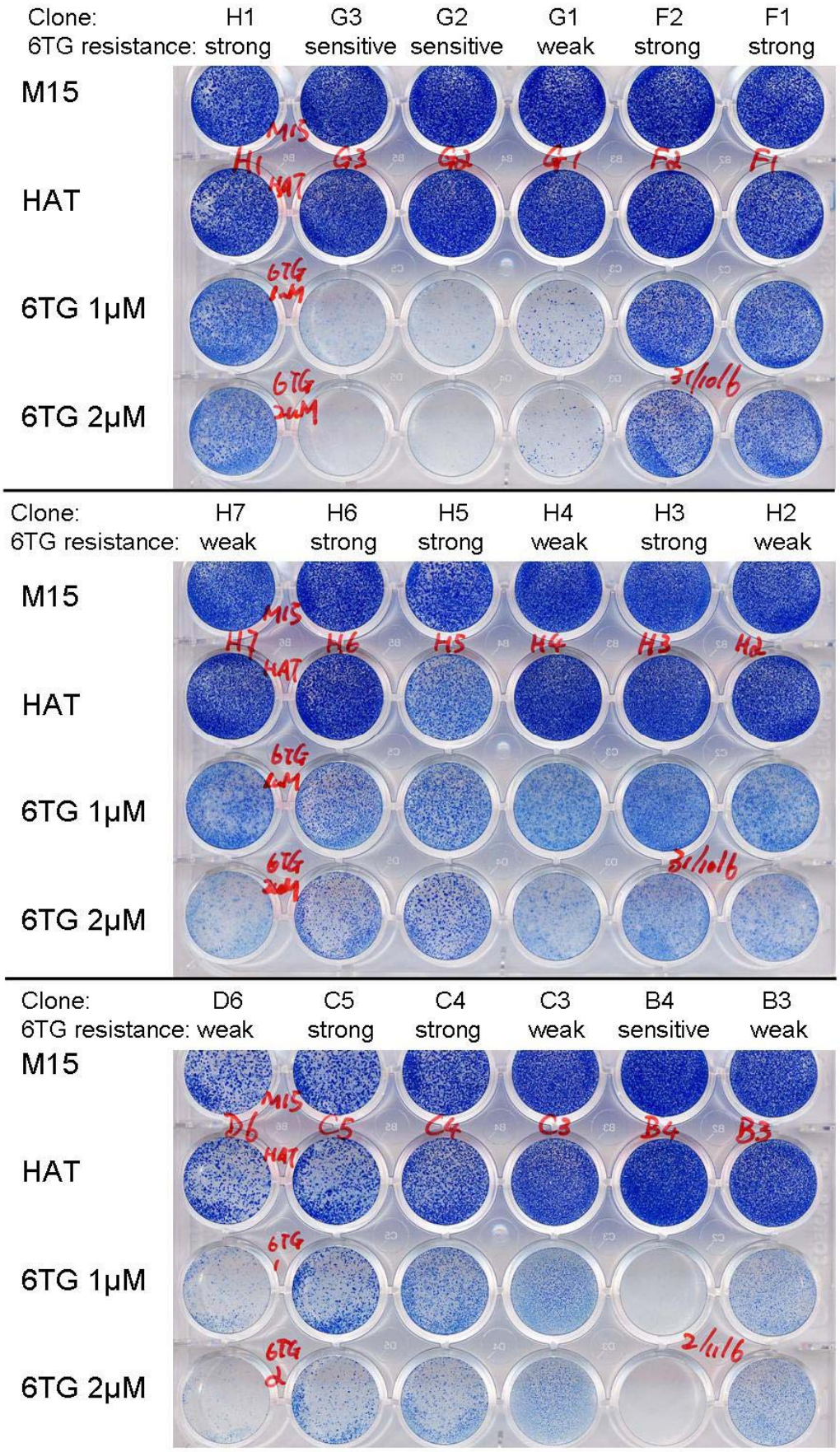


Figure 7-4 **Phenotype of mutant clones**

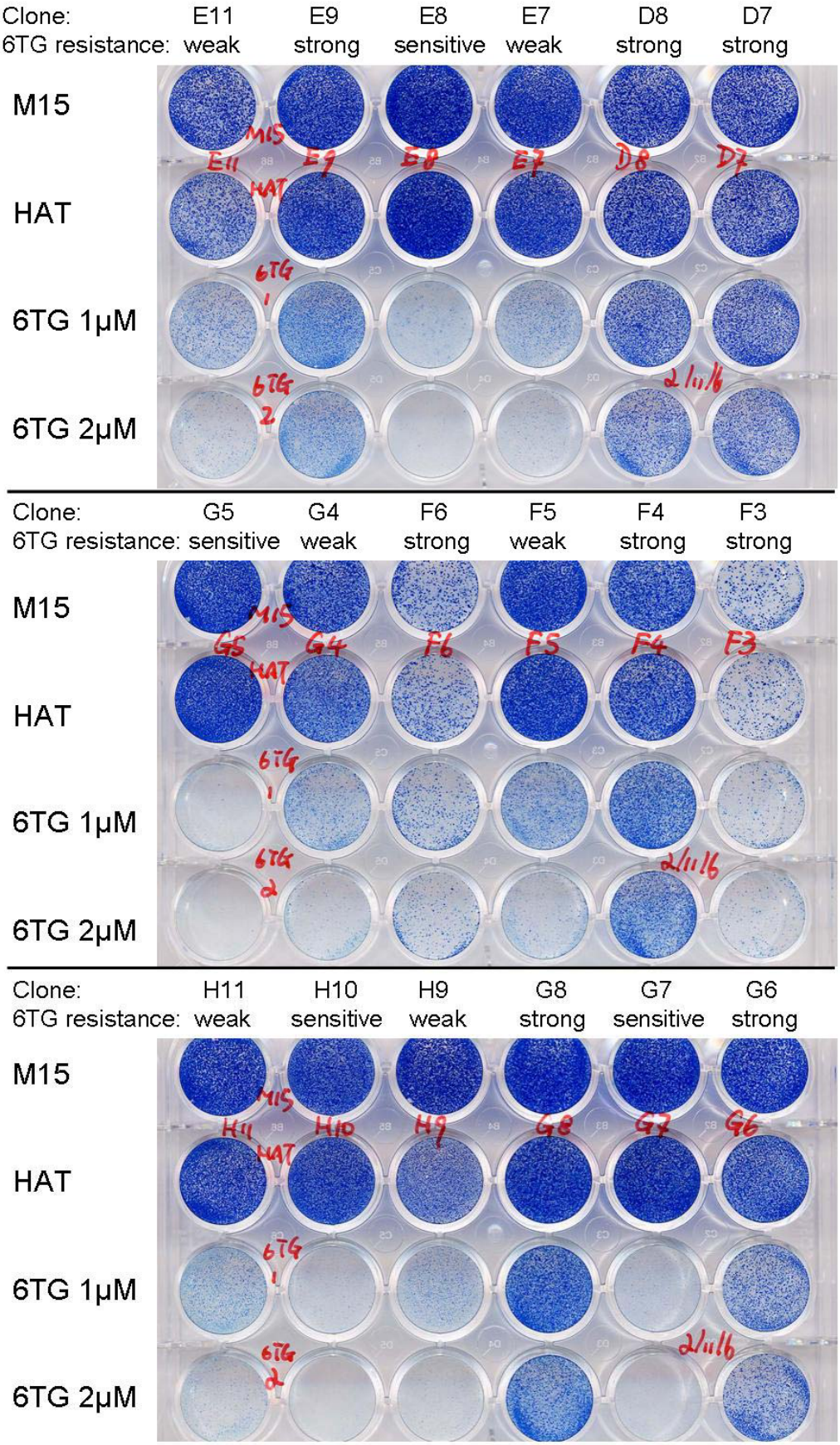


Figure 7-5 **Phenotype of mutant clones**

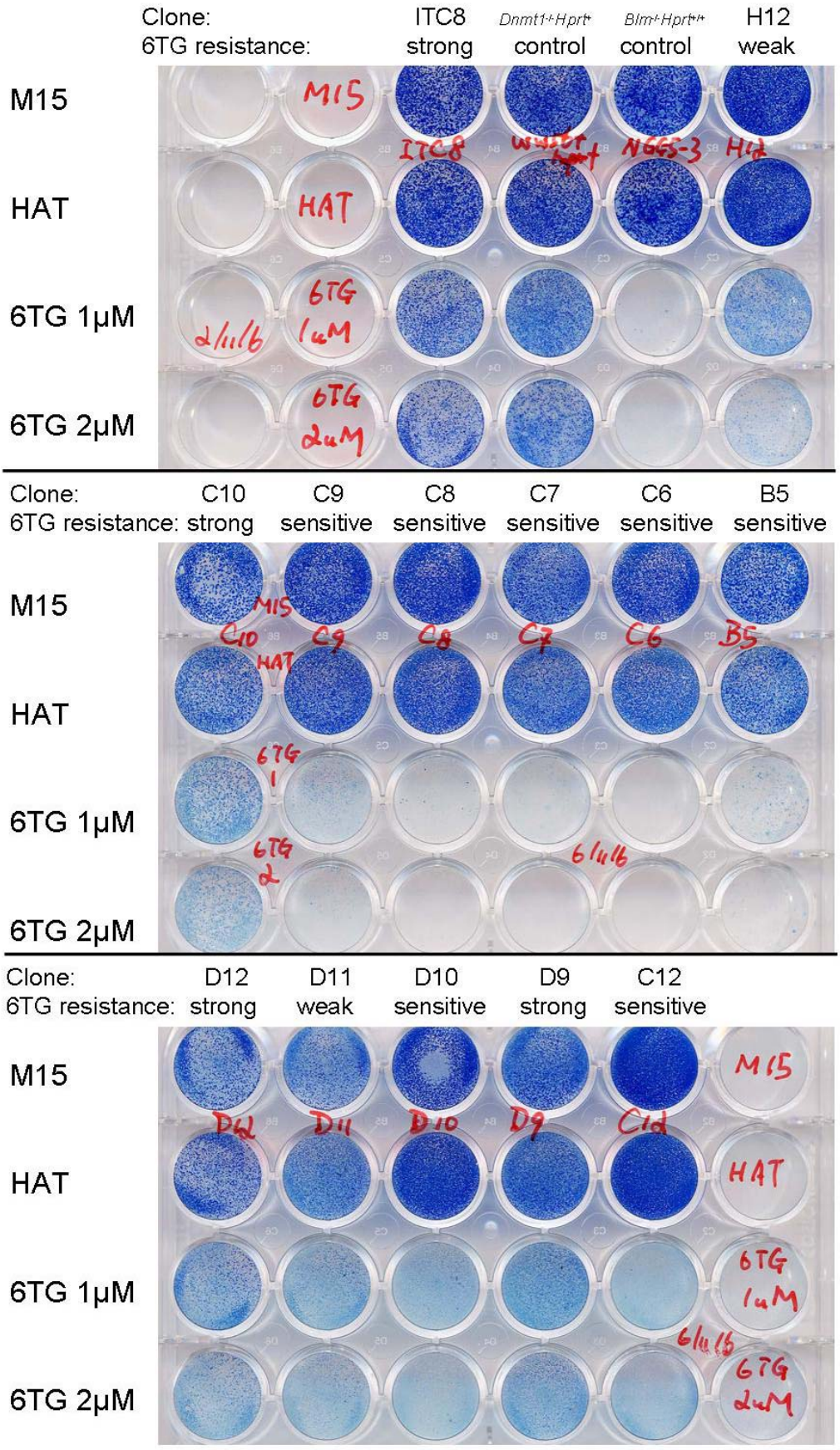


Figure 7-6 **Phenotype of mutant clones**

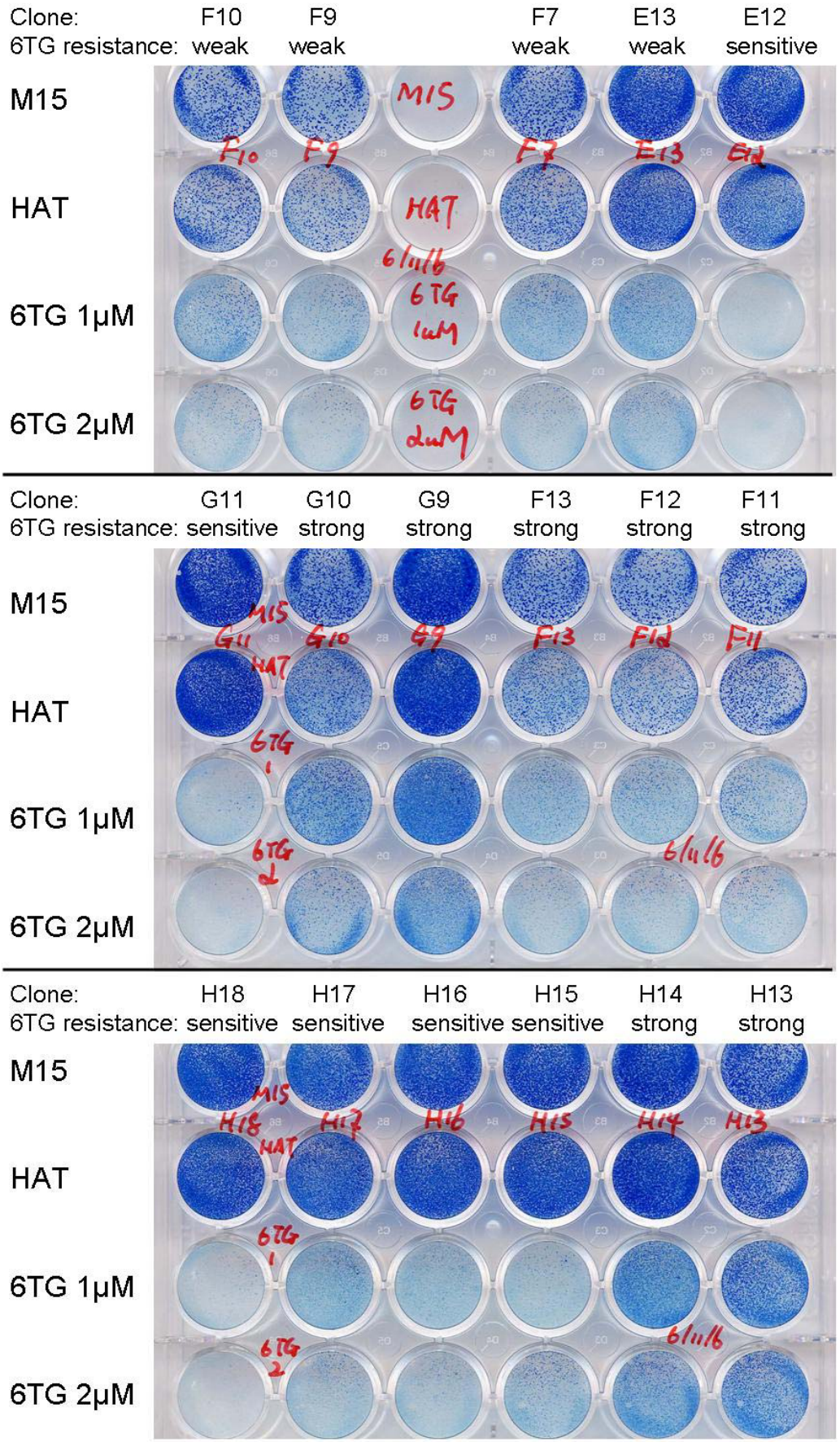


Figure 7-7 **Phenotype of mutant clones**

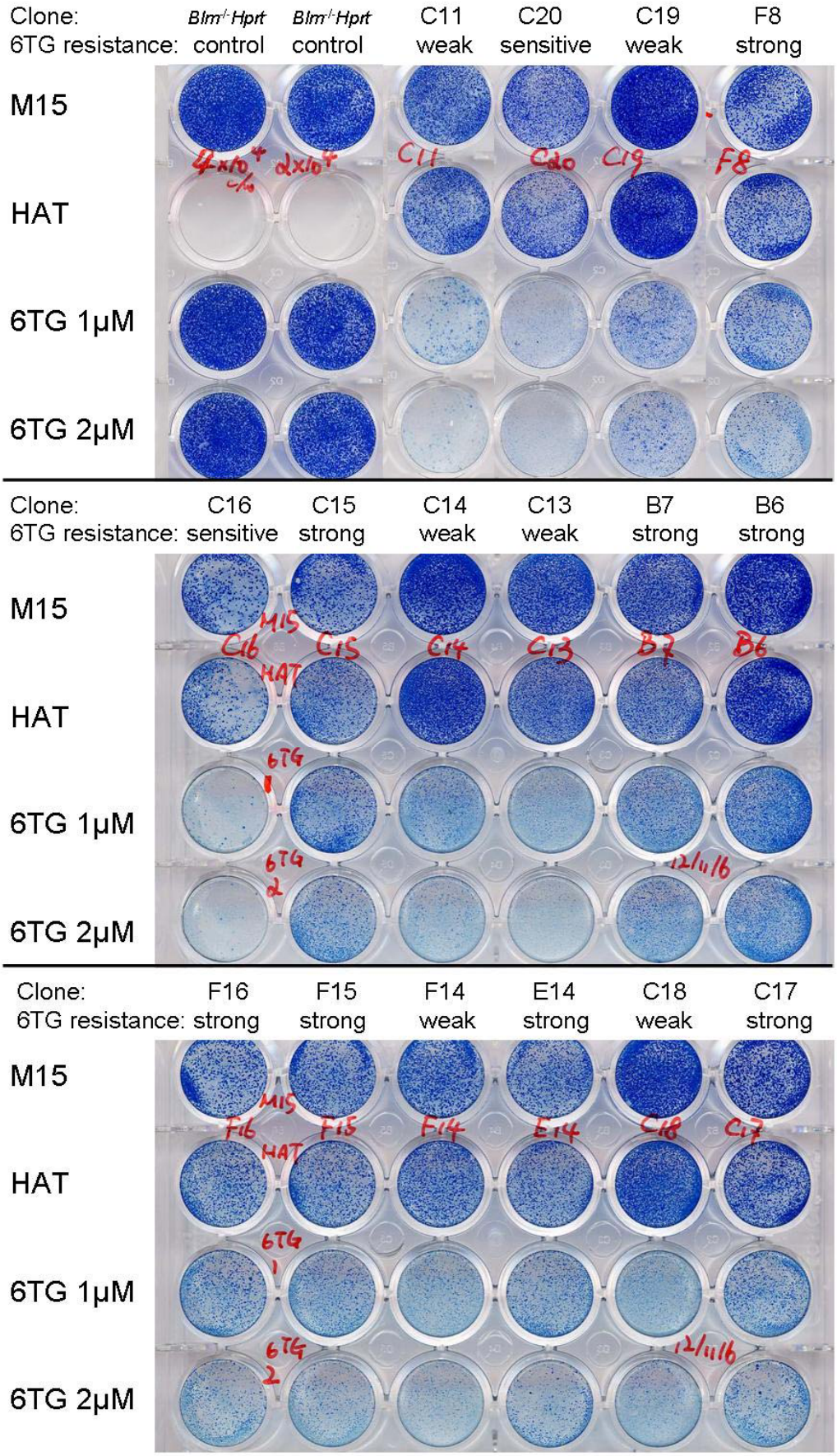


Figure 7-8 **Phenotype of mutant clones**

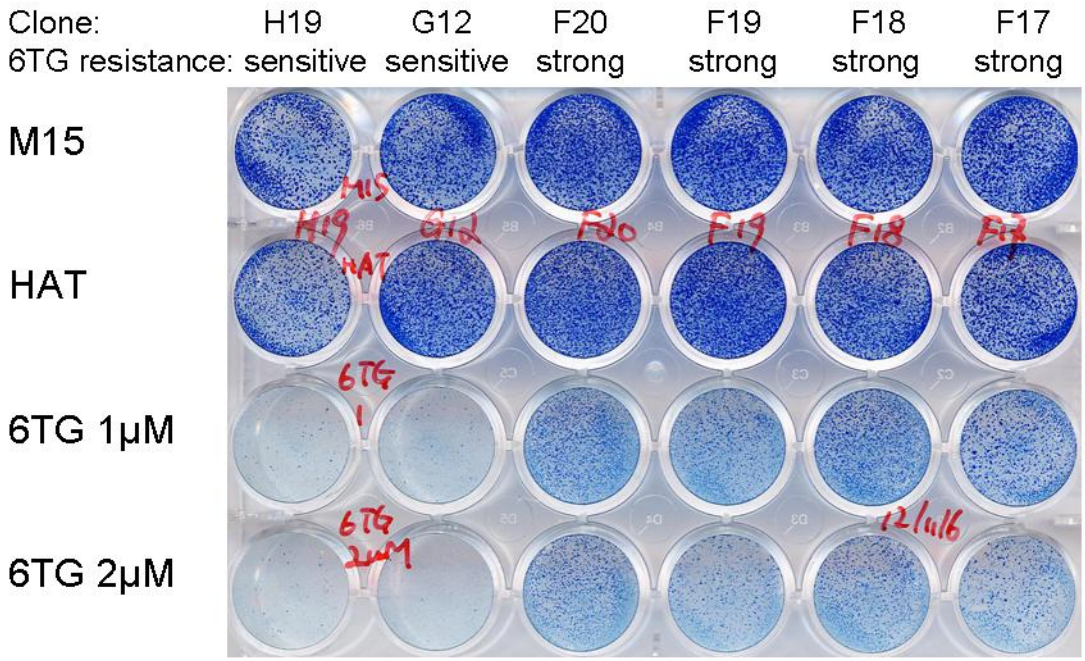
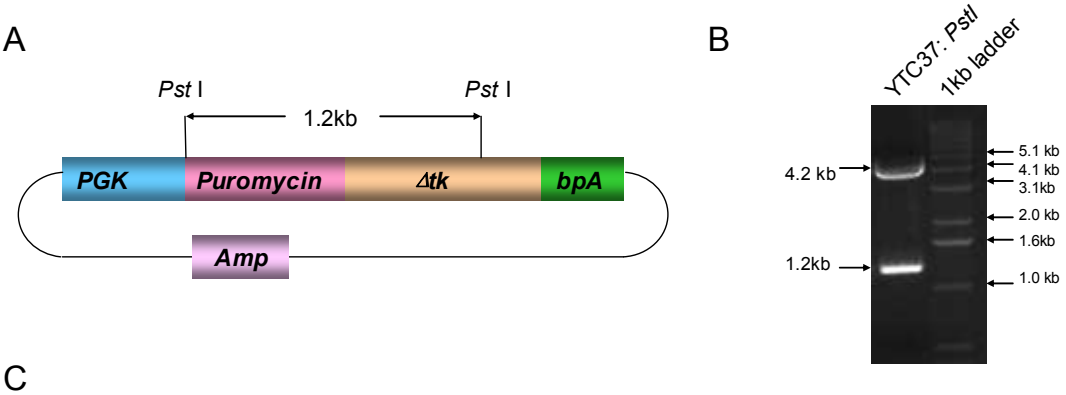


Figure 7-9 **Phenotype of mutant clones**

7.2.5 Clonal relationship of 6TG^R clones

The clonal relationships within a mutation pool were established by Southern blot analysis using a *PuroΔtk* probe (Figure 7-10). The *PuroΔtk* probe is a 1233 bp *Pst* I fragment from plasmid YTC37 (Chen, Y. T. *et al.* 2000) used to construct the retroviral vector. Genomic DNA from the clones was digested with *Hind* III, which cuts once in the provirus, thus each proviral insertion will generate a different-sized restrict fragment depending on the position of the *Hind* III site in the flanking genomic DNA. Independent clones will have different fragment sizes on the Southern blot. According to Southern blot analysis (Figure 7-11–13), 32 redundant mutants clones were identified. Twenty-seven unique MMR-deficient candidates were identified with strong resistance to 6TG (Table 7-3).



C

AGCCAACGCCACCATGGGGACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCCGGGCCGT
ACGCACCCTCGCCGCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGACCCGGACCGCCACATCGAGCG
GGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGG
CGCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCA
TGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAG
GAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGT
GCTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCC
CCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGTGCCCGAAGGACCGCGCGCAGCCTGGTGCATGA
CCCGCAAGCCCGGTGCCGGATCCATGCCCACGCTACTGCGGGAATATATAGACGGTCCTCACGGGATGGGGAAAAC
CACCACCACGCAACTGCTGGTGGCCCTGGGTTCGCGCGACGATATCGTCTACGTACCCGAGCCGATGACTTACTGG
CAGGTGCTGGGGGCTTCCGAGACAATCGCGAACATCTACACCACACAACACCGCCTCGACCAGGGTGAGATATCGG
CCGGGGACGCGGCGGTGGTAATGACAAGCGCCCAGATAACAATGGGCATGCCTTATGCCGTGACCGACGCCGTTCT
GGCTCCTCATATCGGGGGGGAGGCTGGGAGCTCACATGCCCCGCCCCCGGCCCTCACCCTCATCTTCGACCGCCA
TCCCATCGCCGCCCTCCTGTGCTACCCGGCCGCGCGATACCTTATGGGCAGCATGACCCCCCAGGCCGTGCTGGC
GTTCGTGGCCCTCATCCCGCCGACCTTGCCCGGCACAAACATCGTGTTGGGGGCCCTTCCGGAGGACAGACACATC
GACCGCCTGGCCAAACGCCAGCGCCCCGGCGAGCGGCTTGACCTGGCTATGCTGGCCGCGATTCGCCGCGTTTAC
GGGCTGCTTGCCAATACGGTGCGGTATCTGCA

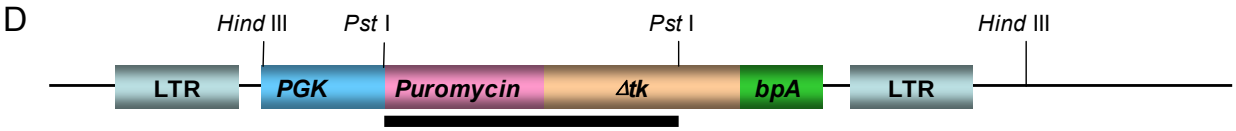


Figure 7-10 Design of the *PuroΔtk* probe and the Southern blot strategy to establish clonal relationships

A. Schematic view of *Pst* I sites in the *PuroΔtk* cassette. **B**. Isolation of the *PuroΔtk* probe. Plasmid YTC-37 was digested with *Pst* I, resulting in two fragments. The 1.2 kb fragment was purified for use as a Southern blot probe later. **C**. The sequence of the *PuroΔtk* probe from the DNA fragment between two *Pst* I sites of YTC-37 plasmid. **D**. Southern blot analysis strategy to establish clonal relationships between clones. A locus with the provirus and flanking genomic DNA is shown. By cutting genomic DNA of a mutant clone with *Hind* III, the DNA fragment containing the *PuroΔtk* cassette and the nearest *Hind* III site in the flanking DNA will hybridize with the *PuroΔtk* probe (black bar).

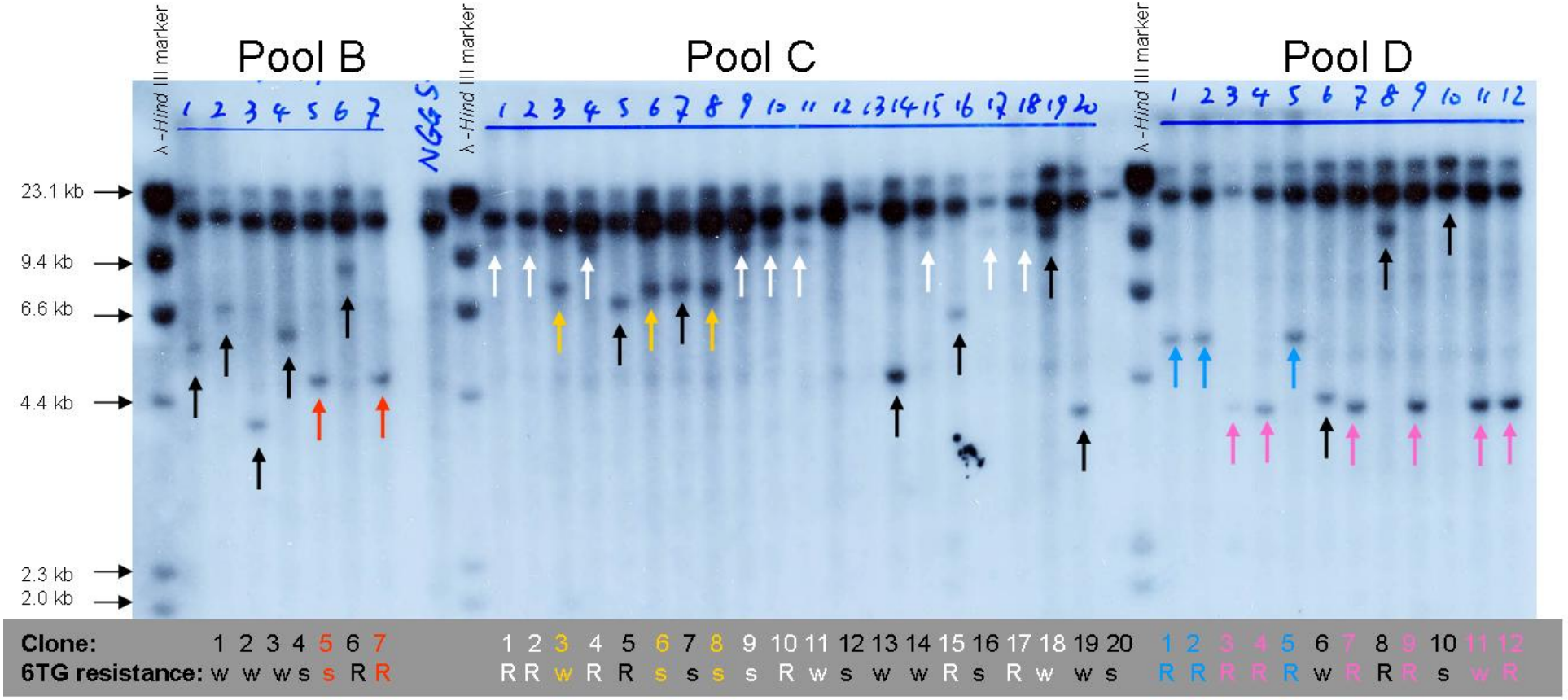


Figure 7-11 **Southern blot to establish clonal relationships of 6TG^R clones in mutation pools B, C and D**

Figure 7-11 Southern blot to establish clonal relationships of 6TG^R clones in mutation pools B, C and D

DNA of 7 mutants from pool B, 20 mutants from pool C and 12 mutants from pool D were digested with *Hind* III. There is a strong background band below the 23.1 kb marker. It exists in the control cell line (NGG5-3) and all mutants. When Blasting the *PuroΔtk* probe sequence in the mouse genome (NCBI m37, Ensembl release 47), there was no homologous sequence longer than 21 bp. Thus, this background band comes from shared vector sequences from the multiple targeting experiments used to construct NGG5-3. The junction fragments are indicated by arrows. The identical size of mutants' junction fragments within a mutation pool suggests clonal relationships and these fragments are indicated by arrows of a non-black colour. Different colours indicate different clonal relationships. Black arrows indicate clones without an obvious clonal relationship with others. Mutants C12 and C13 do not have obvious junction fragments. These two clones' junction fragments may be too close to the background band to be seen. The name of each mutant clone and its 6TG resistance is indicated at the bottom of the figure. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

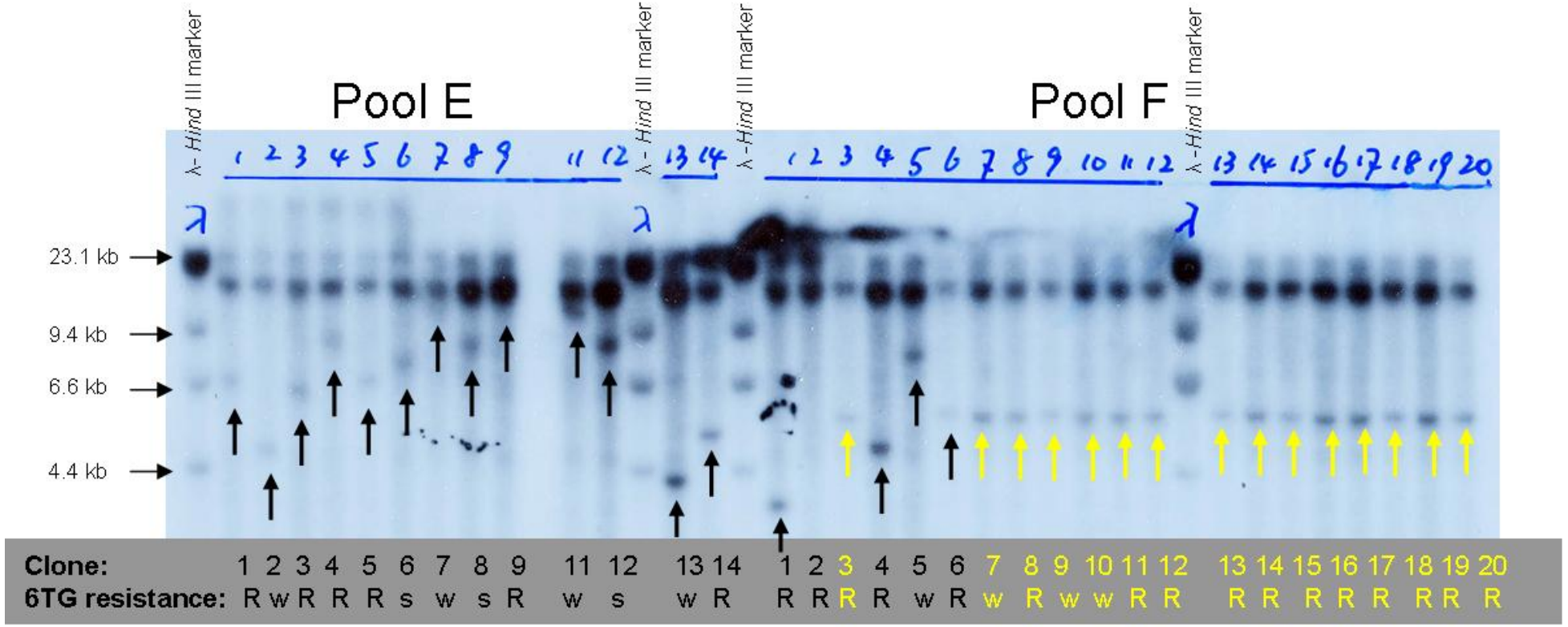


Figure 7-12 **Southern blot to establish clonal relationships of 6TG^R clones in mutation pools E and F**

Figure 7-12 Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools E and F

DNA of 13 mutants from pool E and 20 mutants from pool F digested with *Hind* III and hybridized with the *PuroΔtk* probe. Black arrows indicate clones without an obvious clonal relationship with others. Yellow arrows indicate identical junction fragments in size. Mutant F2 does not have an obvious junction fragment. Its junction fragments may be too close to the background band to be observed. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

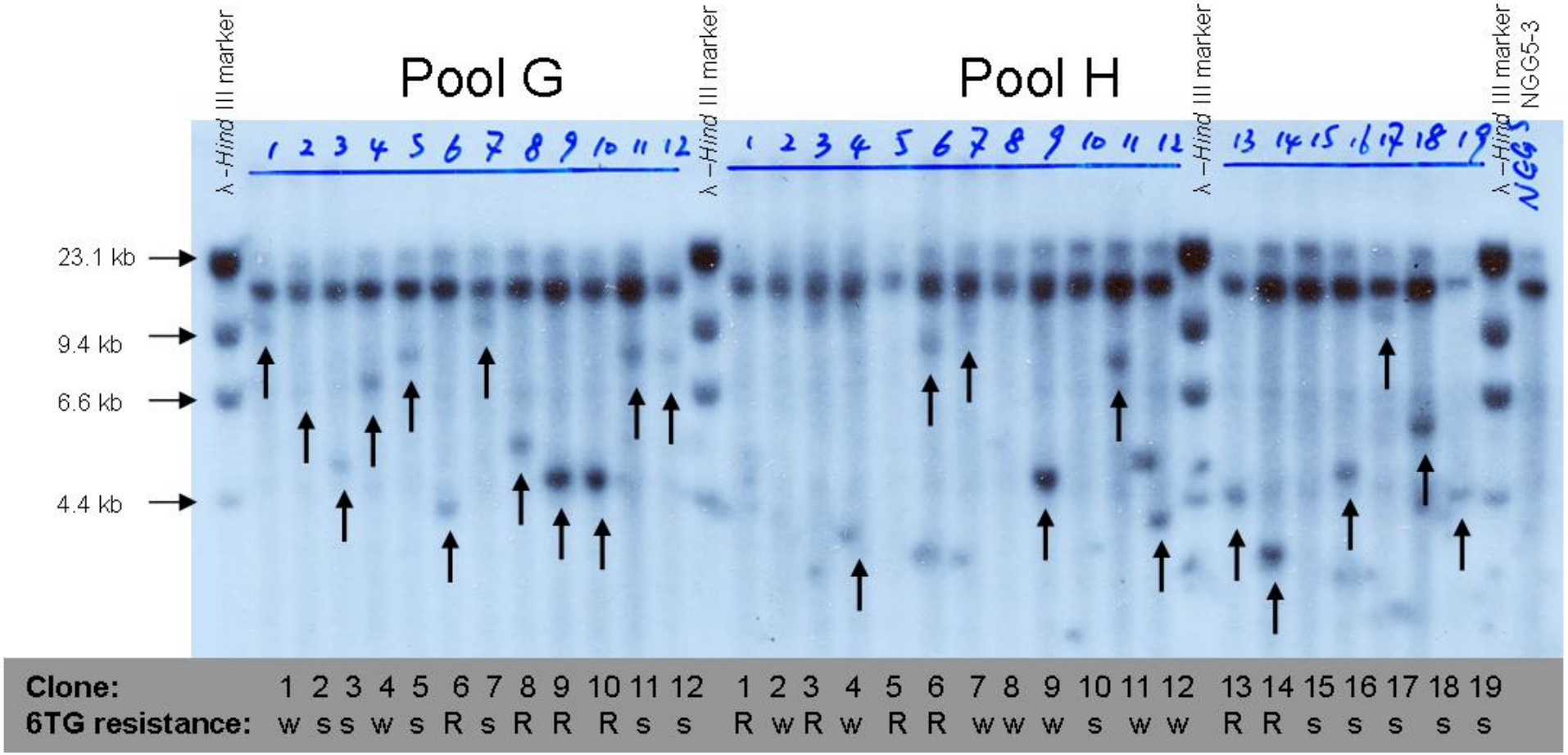


Figure 7-13 Southern blot to establish clonal relationships of 6TG^R clones in mutation pools G and H

Figure 7-13 Southern blot to establish clonal relationships of $6TG^R$ clones in mutation pools G and H

DNA of 12 mutants from pool G and 19 mutants from pool H was digested with *Hind* III and hybridized with the *PuroΔtk* probe. The junction fragments are indicated by arrows. There are no junction fragments that are identical in size. Mutants H1–3, 5, 8, 10 and 15 do not have obvious junction fragments. Their junction fragments may be too close to the background band to be seen. (R: strongly 6TG-resistant; w: weakly resistant to 6TG; s: sensitive to 6TG.)

Table 7-3 **Redundant mutant clones and unique mutant clones**

Pool	Number of unique clones	ID of unique clones	Number of redundant clones	ID of redundant clones	Number of unique clones with strong 6TG resistance	ID of unique clones with strong 6TG resistance
B	6	1–4, 6–7	1	5	2	6, 7
C	8	1, 3, 5, 7, 14, 16, 19–20	10	2, 4, 6, 8–11, 15, 17, 18	2	1, 5
D	5	1, 4, 6, 8, 10	7	2–3, 5, 7, 9, 11–12	3	1, 4, 8
E	13	1–13	0	None	5	1, 3, 4, 5, 9
F	6	1, 2, 4–6, 16	14	3, 7–15, 17–20	5	1, 2, 4, 6, 16
G	12	1–12	0	None	4	6, 8, 9, 10
H	19	1–19	0	None	6	1, 3, 5, 6, 13, 14
Total number	69	69	32	32	27	27

In pools B–H, redundant mutant clones were identified by the clonal relationship established through Southern blot analysis. One clone of a number of redundant clones is chosen to be a representative to be analysed. These representatives and unique clones are chosen for array CGH and expression array analysis.

7.2.6 Array CGH analysis

A total number of thirty-four mutant clones were analysed by array CGH to identify copy number changes of genomic DNA. Twenty-seven are strongly 6TG resistant unique clones (Table 7-3) and seven are weakly 6TG resistant clones (B2, B3, C3, D6, E2, F9 and G4). The reference DNA for these experiments was genomic DNA from NGG5-3, extracted from the same passage of cells used for the virus tagging experiment. This is important because cells accumulate mutations during cell culture. The arrays used for this experiment were mouse tiling BAC clones with a total 18,294 clones. A dye swap experiment or two duplicate array experiments were preformed on each pair of mutant DNA and reference DNA. Each data point is derived from either a dye swap (two data) or two duplicates of a given BAC. Due to the high cost of array CGH, no more replicating experiments were performed. To obtain more confident data, adjacent BACs were used to estimate range of variation and to excluded low quality data points. An average Log_2(ratio) was calculated for a continuous region including BACs with similar Log_2(ratio). Within an experiment, the Log_2(ratio) between -0.29 and +0.29 for the signal from any BAC probe was regarded as no copy number change. The position of a BAC probe on the array is recorded as the position of its middle point (in the base pair of the chromosome). This method is not accurate to base pairs resolution because we cannot identify which part of a BAC sequence is contributing signal on the array. Thus, we have to bear in mind that even if two cell lines have the same BAC(s) deleted, they will not have the same starting and ending base pairs.

7.2.6.1 *Msh2* and *Msh6* homozygous deletions

As previously described, cells deficient for the MMR genes *Msh2, Msh6, Mlh1* or *Pms2* are resistant to 6TG (Abuin *et al.* 2000; Karran *et al.* 1996). In addition, *Dnmt1* deficiency gives 6TG resistance (Guo *et al.* 2004). Therefore, homozygous deletions of these genes were expected in the mutants. Of the thirty-four clones (27 strongly $6TG^R$, 7 weakly $6TG^R$) analysed, six clones (B6, D1, D4, D8, F4 and H14) were identified with homozygous deletions on chromosome 17 between 87.3 Mb and 91.7 Mb (Table 7-4). These deletions covered the genes *Msh2* and *Msh6*. Five clones (B6, D4, D8, F4 and H14) deleted both genes while one (D1) deleted *Msh6* only (Figure 7-14). In addition, nine clones (C1, C3, C5, D6, F9, F16, G8, G9 and H13) contain heterozygous deletions covering both *Msh2* and *Msh6*.

The six mutants with homozygous deletions covering *Msh2* and *Msh6* genes are from four mutation pools and they are not clonal according to the previous data of clonal relationships

built by Southern blot analysis. These deletions range from 2.6 to 4.4 Mb (Table 7-4). The average size of a homozygous deletion is 3.6 Mb. The Log_2(ratio) ranges from -2.21 to -1.16 and the average Log_2(ratio) is -1.67. The array CGH profiles of two of these mutants (D1 and D4) are shown in Figure 7-15 and Figure 7-16.

Mutant clone D1 was identified in which the *Msh6* gene was homozygously deleted without affecting the *Msh2* gene (Figure 7-15). This clone D1, from the mutation pool D, is a fully 6TG resistant mutant. It covers a 3.3 Mb deletion and the average Log_2(ratio) of the BACs in this region is -1.61. In addition to this homozygous deletion, other deletions are observed on chromosomes 12 (14 Mb) and 16 (1.0 Mb), and duplications are observed on chromosomes 10 (2.1 Mb), 13 (51 Mb) and 17 (1.8 Mb, 2.3 Mb). If *Msh6* was not a known MMR gene, it would be easy to identify which deleted gene in this clone results in the 6TG resistance, because the other regions are heterozygous.

In the Figure 7-16, the Log_2 deviation on chromosome 17 indicates a homozygous deletion including *Msh2* and *Msh6* genes. This clone (D4) has also lost chromosome Y and duplicated half of chromosome 1. The single BAC homozygous deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are observed in many samples, these are associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly allocated to other chromosomes.

The tile path BACs at the *Msh2* and *Msh6* loci are shown in Figure 7-17. The zoomed-in figure of *Msh2* and *Msh6* homozygous deletions are shown in Figure 7-18 and the schematic view of these deletions is shown in Figure 7-19. This figure illustrates that the minimal deletion of these clones includes both the *Msh2* and *Msh6* genes, and the minimal common homozygous deletion (2 Mb) of the six clones. Thus, in the absence of any knowledge about these genes I would have identified them by my strategy. These clones confirmed the screen strategy and the capacity to identify deletions generated by irradiation through array CGH technology, it also confirmed the ability of the *Blm*-deficient system to generate homozygous deletions. It is noticed that all homozygous deletions do not cover the region more distal than 91.7 Mb. This is probably because the deficiency of the *Adcyap1* gene (93.6 Mb) may result in cell lethality, implied by high postnatal mortality in its knockout mice (Gray *et al.* 2001).

Table 7-4 Summary of homozygous deletions covering *Msh2* and/or *Msh6*

Clone	Deleted region starting BAC (Mb)	Deleted region ending BAC (Mb)	Deletion length (Mb)	Average Log_2(ratio) over deleted region
B6	RP24-315D17 (88.1)	RP23-340K6 (91.7)	3.6	-1.16
D1	RP23-476A24 (88.4)	RP23-340K6 (91.7)	3.3	-1.61
D4	RP24-343P24 (87.6)	RP23-340K6 (91.7)	4.1	-1.58
D8	RP23-10B20 (87.8)	RP24-82E21 (90.4)	2.6	-2.21
F4	RP23-243O16 (87.9)	RP23-340K6 (91.7)	3.8	-2.09
H14	RP24-107B10 (87.3)	RP24-494L17 (91.7)	4.4	-1.36

Five clones (B6, D4, D8, F4 and H14) deleted the both *Msh2* and *Msh6* genes while one (D1, blue font) deleted *Msh6* only. The sized, starting point, ending point and the average Log_2(ratio) over deleted region are shown here. These homozygous deletions may start or end at the same BAC(s), thus the genomic position of the BACs can be the same. However, the position of a BAC is annotated by the middle point position of this BAC, and a BAC probe does not have to be hybridized in full size to contain a signal. Therefore, the genomic positions of the starting and ending BACs of a deletion can not indicate the actual starting and ending points of a deletion. Genomic position from Ensembl release 48 (NCBI m37).

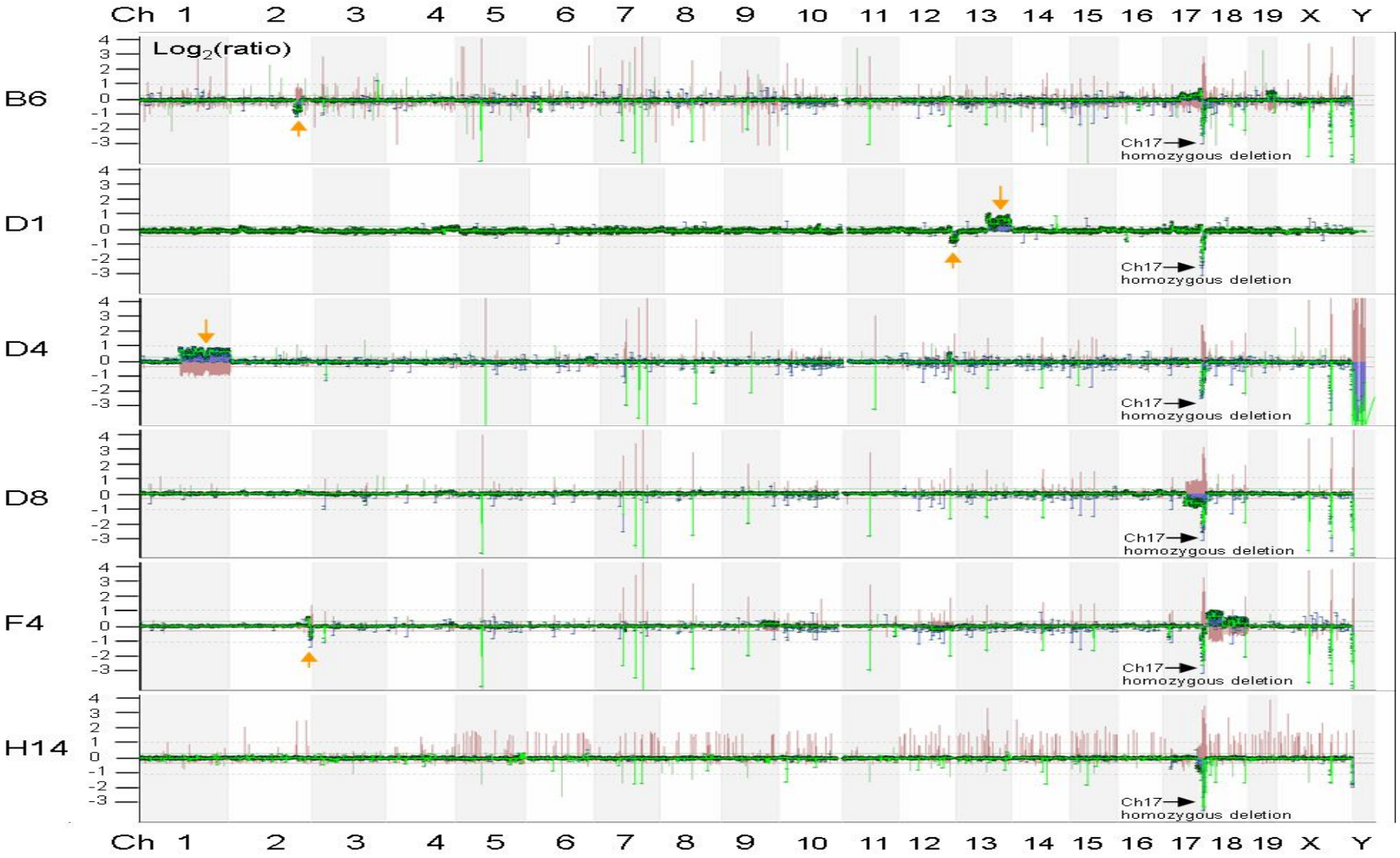


Figure 7-14 ***Msh2*** **and** ***Msh6*** **homozygous mutations**

Figure 7-14 *Msh2* and *Msh6* homozygous mutations

Six mutant clones with strong 6TG resistance were analysed by array CGH. Chromosomes were also placed horizontally in ascending order followed by chromosomes X and Y. A dot indicates the middle point position of a given BAC. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of the BAC probes. This Log_2(ratio) indicates a duplication or amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. Besides the homozygous deletion covering *Msh2* and *Msh6*, there are other genomic DNA copy number changes. For example, clone B6 has a heterozygous deletion on chromosome 2 (yellow arrow), clone D1 has a duplication on chromosome 13 and a heterozygous deletion on chromosome 12 (yellow arrow), clone D4 has distal half of chromosome 1 duplicated (yellow arrow), and clone F4 has a heterozygous deletion on chromosome 2 (yellow arrow). Some small homozygous deletions (single BAC size or two-BAC size) on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are visible in the clones B6, D4, D8 and F4. A common change for these four clones is that they lost all or a big part of chromosome Y. It is assumed that these small deletions should belong to chromosome Y but were incorrectly annotated to other chromosomes.

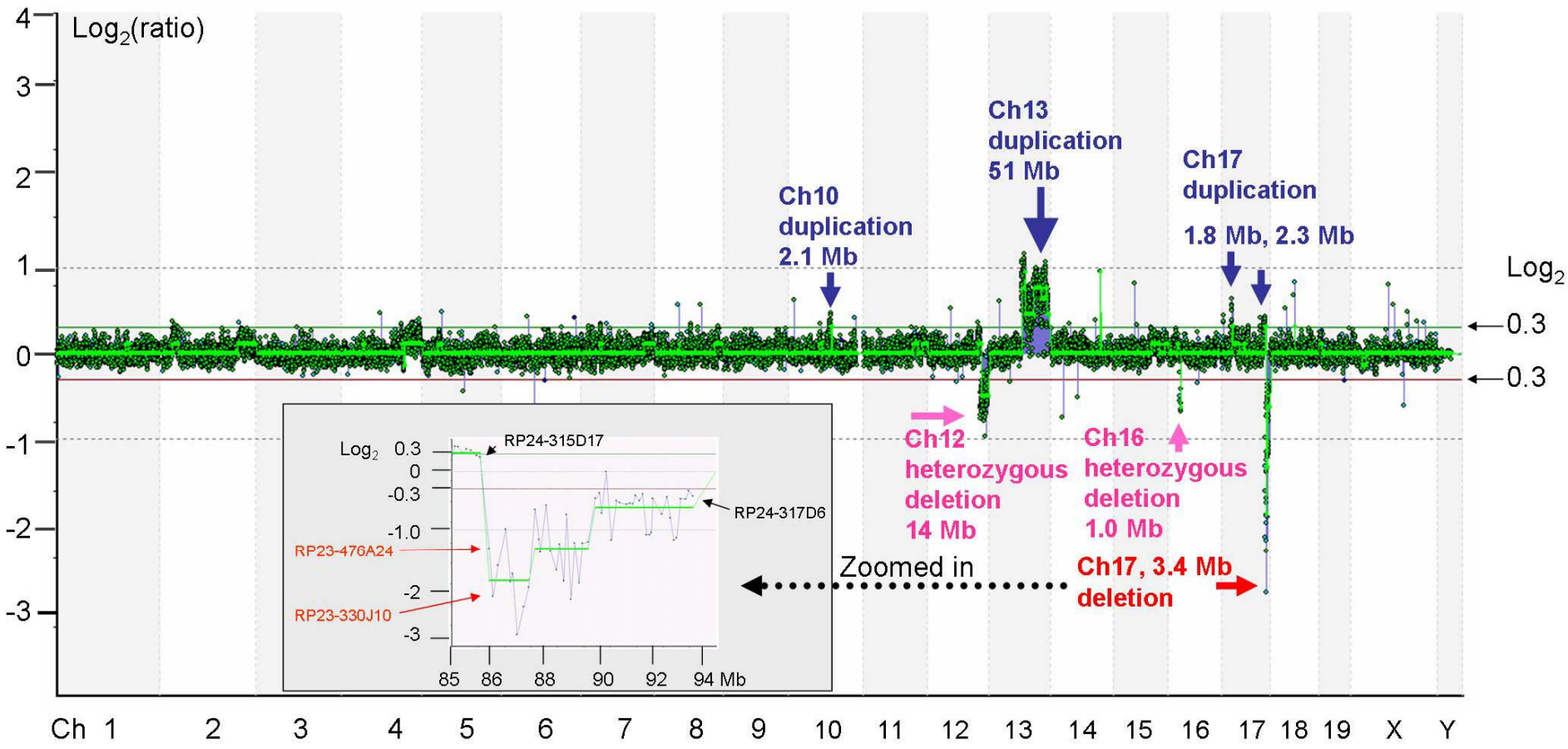


Figure 7-15 **Array CGH profile of mutant D1 with a homozygous deletion at *Msh6***

Figure 7-15 Array CGH profile of mutant D1 with a homozygous deletion at *Msh6*

The Log_2(ratio) of mutant:*Blm*-deficient cell DNA was calculated for each BAC and plotted along each mouse chromosome. Chromosomes were also placed in ascending order followed by X and Y. Each dot indicates a given BAC's middle point position. The Y axis values the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of the BAC clone probes. A red arrow indicates a homozygous deletion including gene *Msh6* on chromosome 17. A zoomed-in picture shows the details of this deletion. BAC RP23-330J10 and RP23-476A24 (red font) are in a homozygously deleted region which includes *Msh6*. BAC RP24-315D17 (containing gene *Msh2*) is duplicated with Log_2(ratio) of ~0.3. BAC RP24-317D6 is the most distal BAC on chromosome 17 in the 200 kb resolution array CGH. In addition, there are deletions (pink arrows) on chromosomes 12 and 16 and duplications (blue arrows) on chromosomes 10, 13, and 17. Some BACs with blue lines are regarded as noise signals because these signals cannot be replicated. The grey horizontal dotted lines are for Log_2(ratio) references of either +1 or -1.

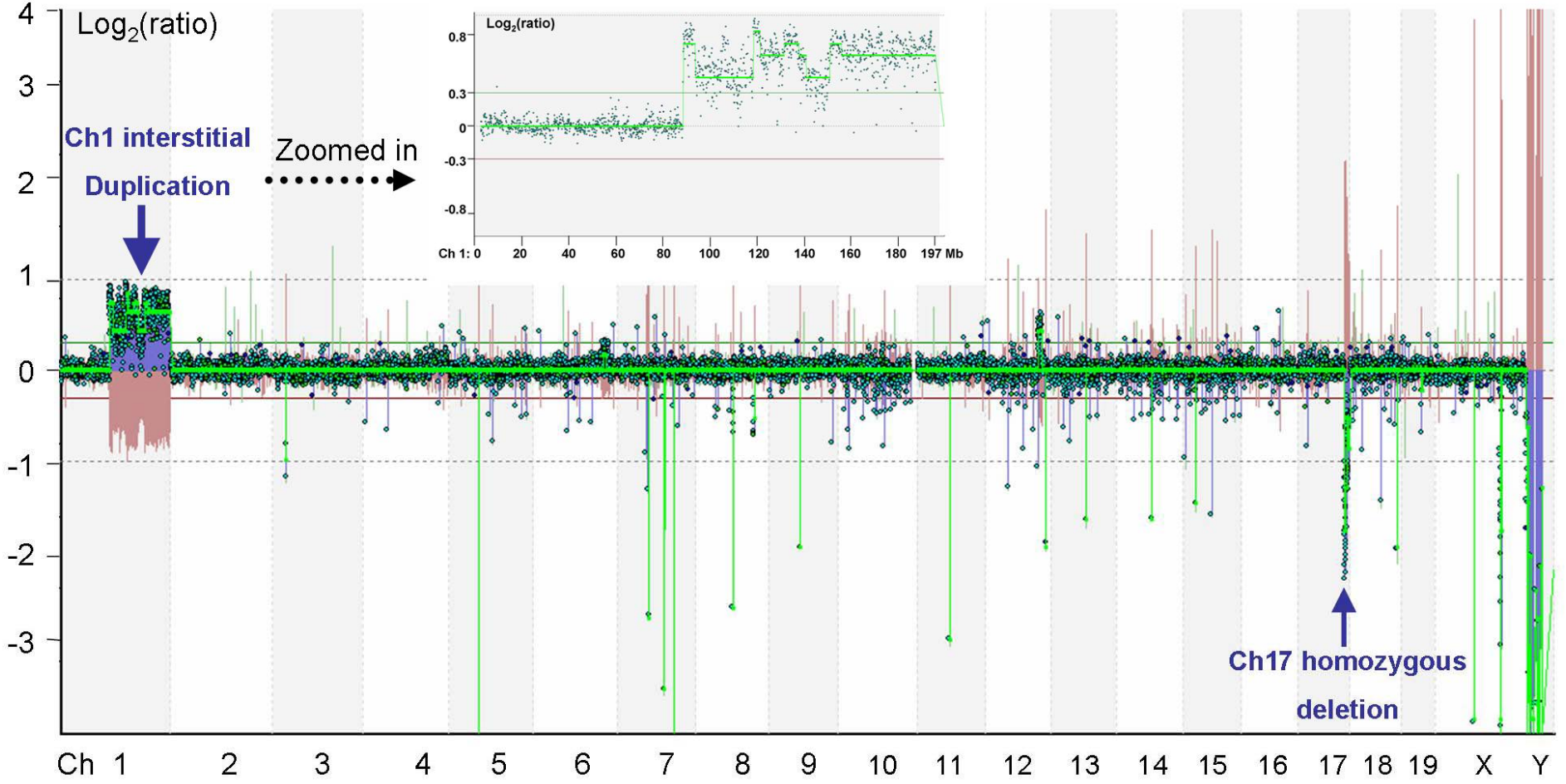


Figure 7-16 **Array CGH analysis of mutant D4**

Figure 7-16 Array CGH analysis of mutant D4

Mutant clone D4 is presented to illustrate how data from array CGH analysis was interpreted. BACs are aligned on the X axis in ascending order according to their genomic position in every chromosome. Chromosomes are also placed in ascending order followed by X and Y. A dot indicates the middle point position of a given BAC. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. This Log_2(ratio) indicates a duplication or amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. The green line below to -2 on chromosome 17 indicates a homozygous deletion including genes *Msh2* and *Msh6*. A zoomed-in picture shows half of chromosome 1 is duplicated. This clone has lost the Y chromosome and the distal half of chromosome 1 is duplicated. There are some single BAC deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X. In many experiments, these deletions were found to be associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly annotated to other chromosomes. The grey dotted lines are for Log_2(ratio) references of either +1 or -1.

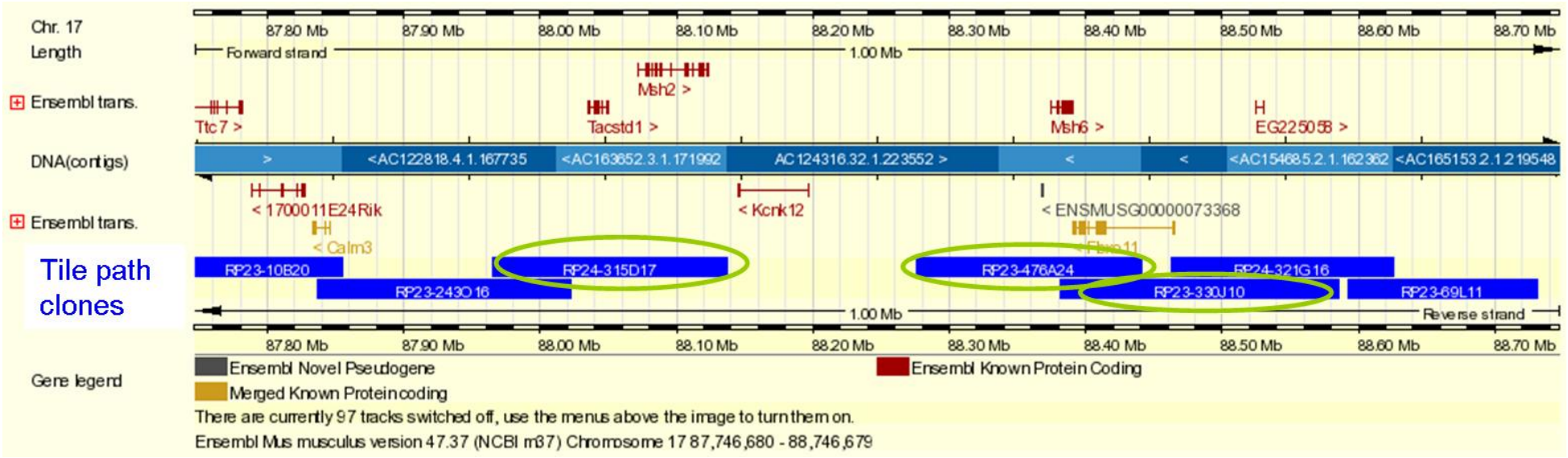


Figure 7-17 Tile path BACs at *Msh2* and *Msh6* loci

A snapshot of the 1.0 Mb chromosome region between 87,746,680–88,746,679 bp on mouse chromosome 17 is shown here (NCBI m37, Ensembl release 47). The Ensembl known protein coding sequences are shown in red. Blue boxes indicate BACs used on the tile path array CGH (200 kb resolution). BAC RP24-315D17 (one of the green circles) covers gene *Msh2*. BAC RP23-476A24 and RP23-330J10 (the other two green circles) cover gene *Msh6*. There is a gap between BAC RP24-315D17 and RP23-476A24.

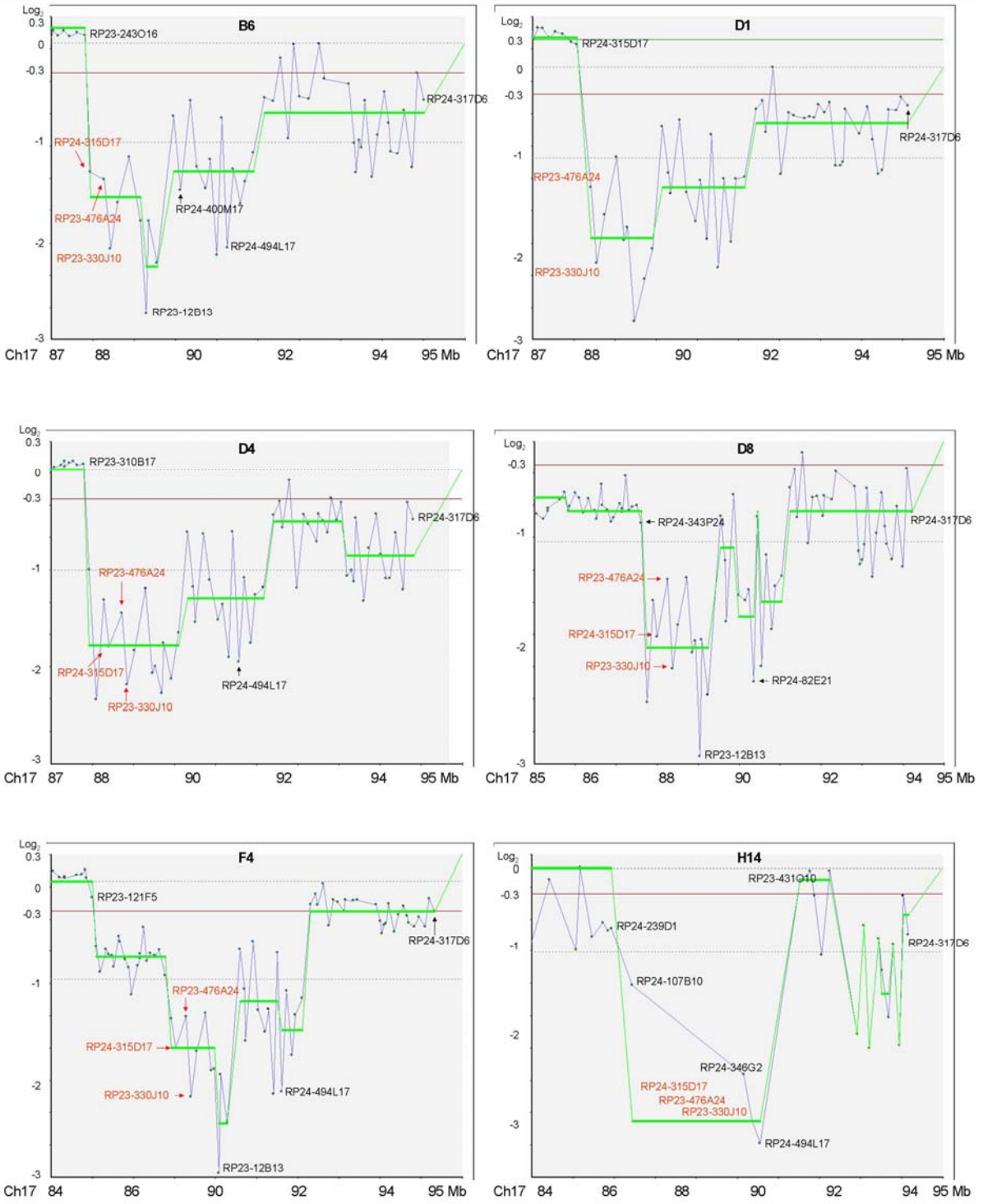


Figure 7-18 Zoomed-in regions of *Msh2* and *Msh6* homozygous deletions

Figure 7-18 Zoomed-in regions of *Msh2* and *Msh6* homozygous deletions

Array CGH profiles of the mutant clones B6, D1, D4, D8, F4 and H14 are shown on part of chromosome 17. The Log_2(Mutant DNA signal/Reference DNA signal) is calculated for each BAC and plotted against their genomic positions. A blue line connects the nearest BACs. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. BAC RP24-315D17 (red font except in D1) covers the gene *Msh2*, while BACs RP23-476A24 and RP23-330J10 (red font) represent a region covering the gene *Msh6*. Log_2(ratio) of these three BACs is below -1 (except RP24-315D17 in D1) therefore the two genes are regarded as homozygously deleted. However, the two homologous chromosomes are not homozygous. Some BACs nearby are annotated in a black font. BAC RP24-317D6 is the most distal BAC of chromosome 17 on the 200 kb resolution array CGH.

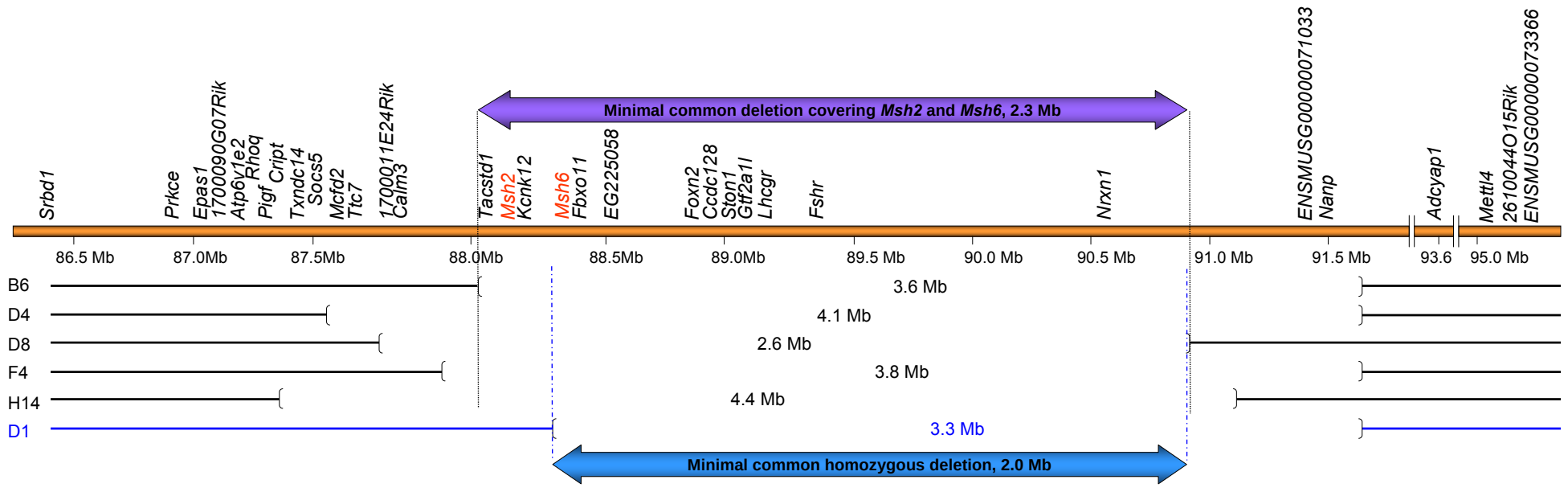


Figure 7-19 Schematic view of *Msh2* and *Msh6* homozygous deletions in six mutants

The region of 86.5–95.1 Mb of mouse chromosome 17 is shown with genes annotated. *Msh2* and *Msh6* loci are shown in a red font. Genomic DNA of the mutant clones B6, D4, D8, F4, and H14 is shown in black lines and genomic DNA of the mutant D1 is shown in blue lines. Homozygous deletions with their sizes in this region are shown in brackets. The first five mutant clones contain homozygous deletions covering *Msh2* and *Msh6* genes. The deletion in clone D1 does not contain *Msh2*. The minimal common deleted region in the first five mutants is 2.3 Mb in size and the minimal common homozygous deletion in these six clones is 2.0 Mb in size.

7.2.6.2 *Msh2* and *Msh6* heterozygous deletions

In addition to the six clones with homozygous deletions, nine clones were identified which had heterozygous deletions of the same region on chromosome 17. These deletions are believed to be heterozygous because the Log_2(Mutant DNA signal/Reference DNA signal) is higher than -1 and below zero. Their Log_2(ratio) is actually between -0.4 and -0.6 (Figure 7-20). The nine mutants are C1, C3, C5, D6, F9, F16, G8, G9 and H13. Their array CGH profiles are shown in Figure 7-20 for the whole genome and in Figure 7-21 for chromosome 17. These mutants are isolated from five mutation pools (C, D, F, G and H). Six of them (C1, C5, F16, G8, G9 and H13) are strongly resistant to 6TG. Three of them (C3, D6 and F9) are weakly resistant to 6TG. Among these mutants, the retroviral tagging revealed that F9 and F16 are derived from the same clone thus their array profiles are identical. This validates the reliability of the array CGH method. Within fifteen deletions and duplications in these two mutants, half of them are the same and the other half deletions and duplications have Log_2(ratio) which is close to either -0.3 or 0.3, the threshold set up for the BlueFuse software. Therefore small fluctuation of Log_2(ratio) resulted in different call of regions. However, this can be calibrated using manual examination.

These nine deletions range from 13.3 Mb to 46 Mb in size (Table 7-5), much larger than the sizes of homozygous deletions covering the *Msh2* and *Msh6* genes (Table 7-4). The average length of heterozygous deletion is 28 Mb. Except one clone, G9, deletions in the eight clones extend to the last BAC probe on chromosome 17, RP24-317D6, suggesting that these eight clones lost the whole distal part of chromosome 17. These clones also have other duplications and deletions in their genome. Although these mutant clones still have one copy of genes *Msh2* and *Msh6* in the genome, six of them have shown strong 6TG resistance. It is assumed that the undeleted allele of either *Msh2* or *Msh6* has a small mutation, which results in a loss-of-function of either *Msh2* or *Msh6*. These mutations were not identified by 200 kb resolution array CGH probably due to their small size. It is anticipated that this hypothesis will be confirmed by analysis of *Msh2* and *Msh6* transcripts.

Monosomies and trisomies were found in these mutants. For example, clone D6 appears to have a monosomy of chromosomes 2, 9, 13 and 16 (Figure 7-22), while clone G9 appears to have trisomies of chromosome 6 and 18 (Figure 7-20). In the clones C5, F9 and F16, some small homozygous deletions (single BAC) deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are associated with chromosome Y loss. This also confirmed the assumption that these BACs should be annotated to chromosome Y. The

clone D6 was selected to show chromosome copy number changes (Figure 7-22). This clone contains a 5.8 Mb heterozygous deletion including *Msh2* and *Msh6* genes on chromosome 17. A deletion on chromosome 2 may be homozygous. In addition, there seem to be monosomies on chromosomes 2, 9, 13 and 16 and trisomies on chromosome 1 and 14. Chromosome 11 appears to contain a segmental deletion at the centromere end. Several facts indicate that this clone might be a mosaic population or a tetraploid cell or both. Firstly, the Log_2(ratio) of chromosome Y is near -0.3. The ES cells I used are XY cells. If chromosome Y is lost, the Log_2(ratio) of chromosome Y on the array CGH should be like those of homozygous deletions. If one chromosome Y remains in a tetraploid cell, the Log_2(ratio) of it should be close to that of heterozygous deletions. However, the Log_2(ratio) of chromosome Y in clone D6 is near 0.3, suggesting a impure clone. Secondly, monosomies are very rare in ES cells. Thus, the clone might be a mosaic population of normal ploidy cell with tetraploid cells with one copy of chromosome loss (Log_2(¾)= -0.42). Thirdly, the Log_2(ratio) of chromosome 1 and 14 is less than 0.32, the Log_2 of a pentaploid chromosome in a tetraploid cell (Log_2(5/4)), indicating this clone is mosaic.

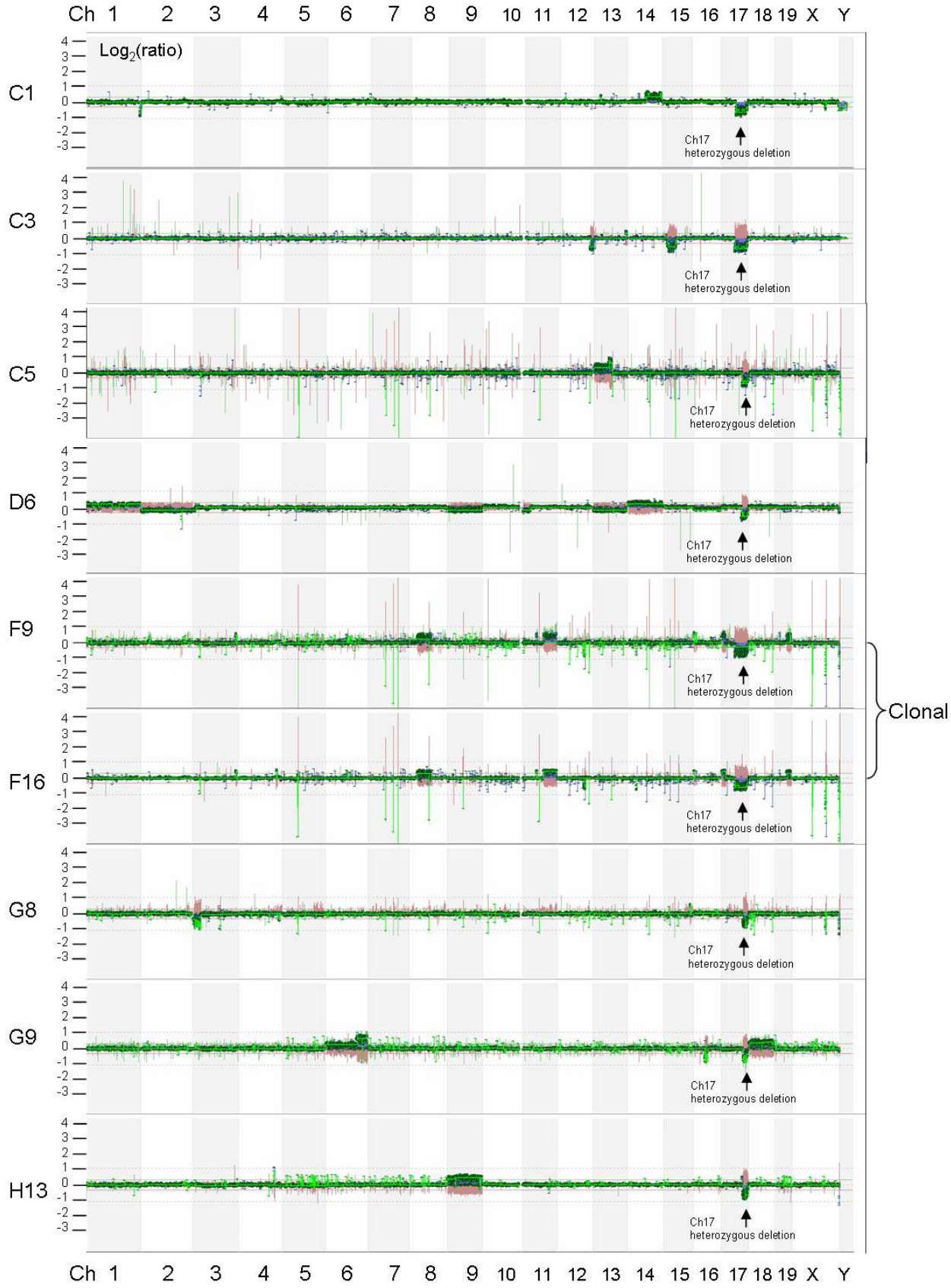


Figure 7-20 ***Msh2*** **and** ***Msh6*** **heterozygous mutants**

Figure 7-20 *Msh2* and *Msh6* heterozygous mutants

Array CGH analysis of nine mutant clones which show a heterozygous deletion covering *Msh2* and *Msh6* genes on chromosome 17. Chromosomes are also placed horizontally in ascending order followed by chromosomes X and Y. The Y axis is the Log_2(Mutant DNA signal/Reference DNA signal) of the BAC probes. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of the BAC clone probes. The red line indicates the Log_2(Reference DNA signal/Mutant DNA signal) of the BAC clone probes from the reciprocal hybridization experiment. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Besides the heterozygous deletions covering *Msh2* and *Msh6*, there are other genomic DNA copy number changes. For example, D6 appears to have a monosomy on chromosomes 2, 9, 13 and 16. Some small homozygous deletions (single BAC) deletions on chromosomes 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X are associated with chromosome Y loss. This further confirms the assumption that these BACs should be annotated to chromosome Y. The array profiles of clone F9 and F16 are highly similar. As a clonal relationship has been established by Southern blot between these two mutants, this is an expected result. This result further confirms the clonal relationships of other mutants built by Southern blot and also verifies the reliability of the array CGH analysis.

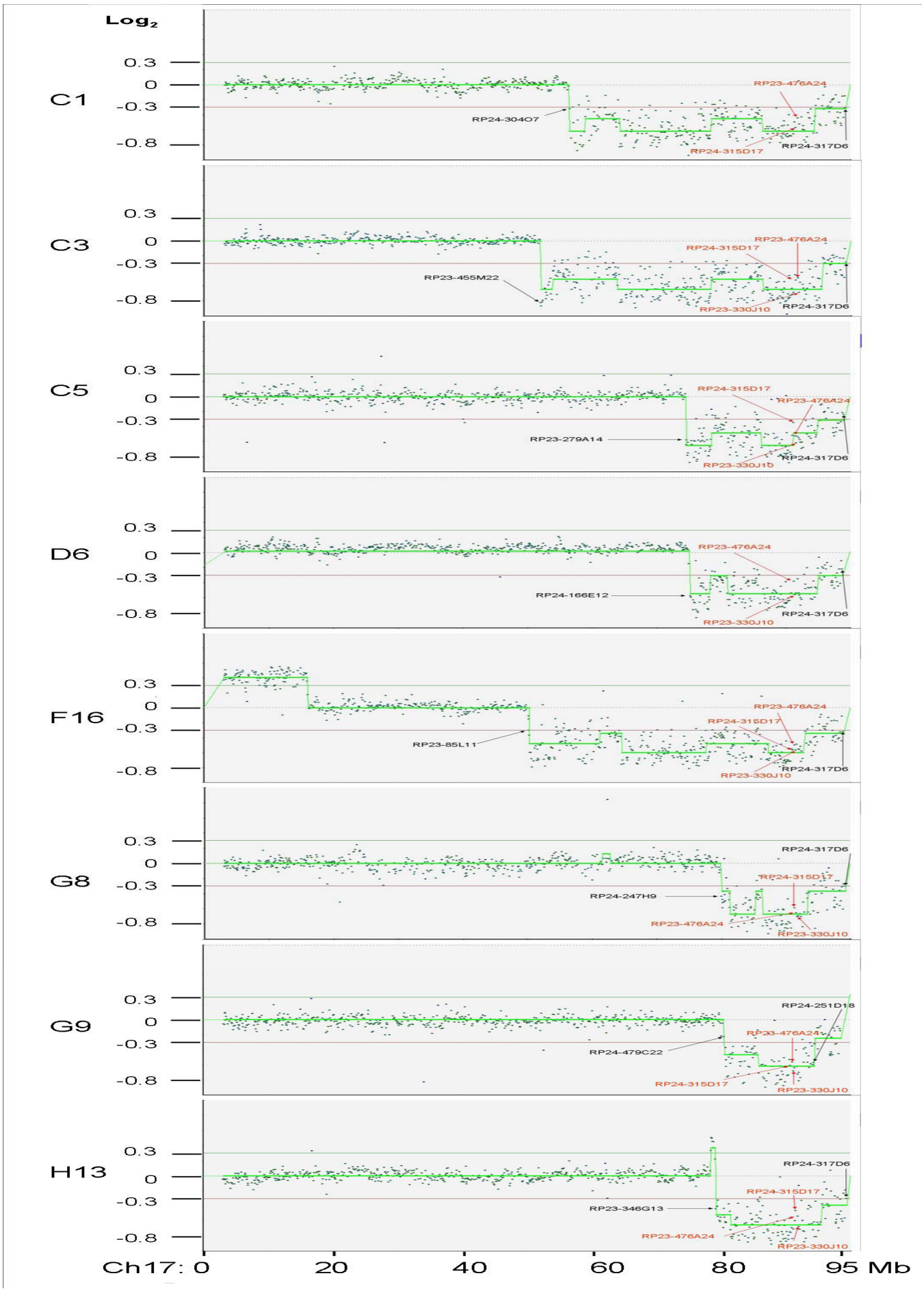


Figure 7-21 ***Msh2*** **and** ***Msh6*** **heterozygous mutants – chromosome 17**

Figure 7-21 *Msh2* and *Msh6* heterozygous mutants – chromosome 17

The array CGH profiles on chromosome 17 of eight clones (F9 and F16 is clonal thus F9 is not shown) with heterozygous deletions covering the *Msh2* and *Msh6* genes are shown. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. BAC RP24-315D17 (red font) covers the *Msh2* gene, while BACs RP23-476A24 and RP23-330J10 (red font) represent a region covering the *Msh6* gene. Log_2(ratio) of these three BACs is above -1 and below zero, therefore the two genes are regarded as heterozygous. The starting BAC probe and the ending BAC probe of each deleted region are shown in a black font. BAC RP24-317D6 is the most distal BAC of chromosome 17 on the 200 kb resolution array CGH.

Table 7-5 **Summary of heterozygous deletions covering *Msh2* and *Msh6***

Clone	Deleted region starting BAC (Mb)	Deleted region ending BAC (Mb)	Deletion length (Mb)	Average Log_2(ratio) over deleted region
C1	RP24-304O7 (55.4)	RP24-317D6 (95.1)	39.7	-0.51
C3	RP23-455M22 (51.3)	RP24-317D6 (95.1)	43.8	-0.59
C5	RP23-279A14 (72.2)	RP24-317D6 (95.1)	22.9	-0.52
D6	RP24-166E12 (72.9)	RP24-317D6 (95.1)	22.2	-0.43
F9	RP23-85L11 (49.1)	RP24-317D6 (95.1)	46.0	-0.51
F16	RP23-85L11 (49.1)	RP24-317D6 (95.1)	46.0	-0.47
G8	RP24-247H9 (77.3)	RP24-317D6 (95.1)	17.8	-0.49
G9	RP24-479C22 (78.2)	RP24-251D18 (91.5)	13.3	-0.55
H13	RP23-346G13 (76.3)	RP24-317D6 (95.1)	18.8	-0.52

The starting and ending BACs deleted on chromosome 17 including the *Msh2* and *Msh6* genes in nine mutant clones. The rough lengths of the deletions including these two genes are between 13.3 Mb and 46 Mb. The average deletion length is 28 Mb. The average Log_2(mutant DNA/*Blm*-deficient cell DNA) of the deleted regions is between -0.59 and -0.43. Genomic positions refer to Ensembl release 48 (NCBI m37).

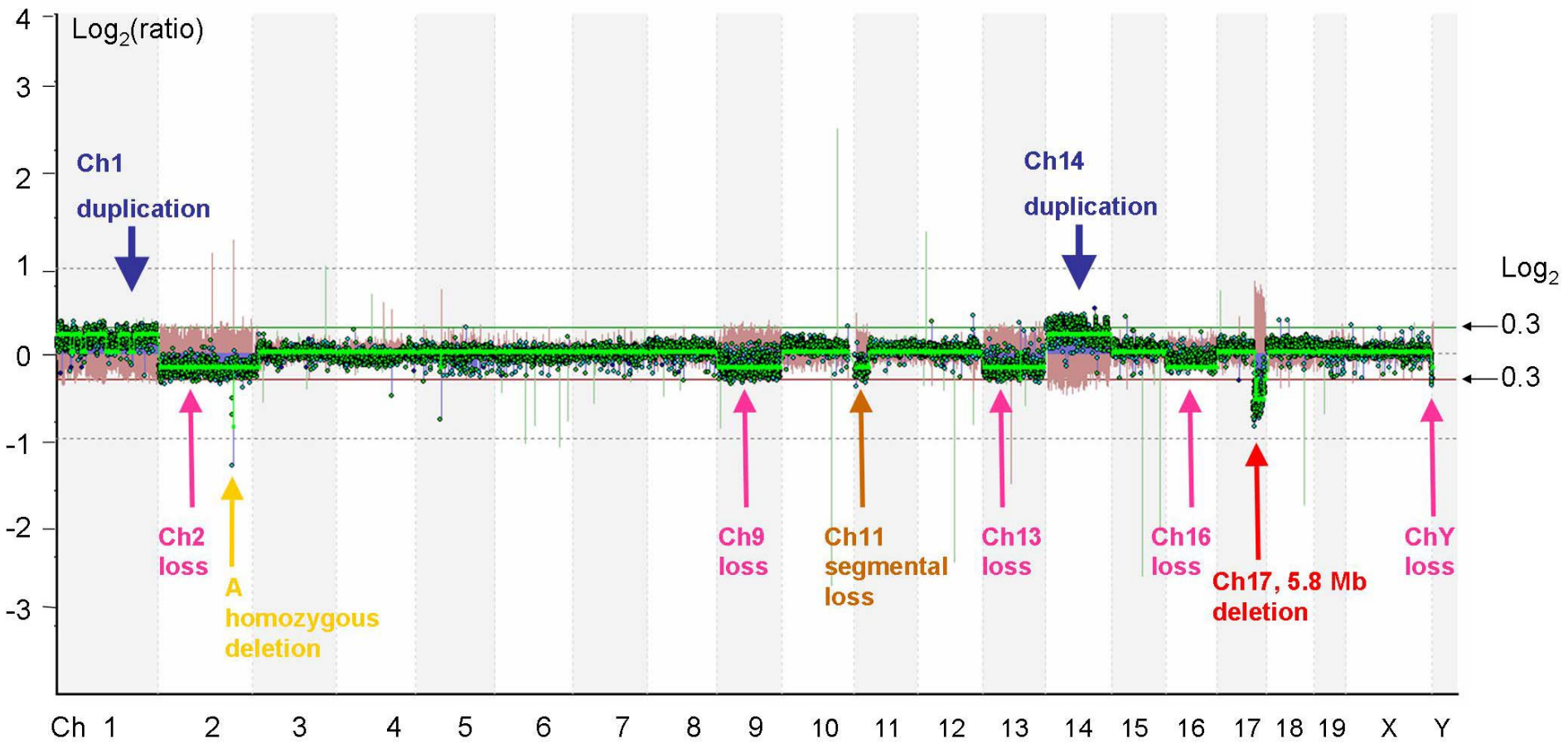


Figure 7-22 **Array CGH profile of mutant D6**

Figure 7-22 Array CGH profile of mutant D6

BACs were aligned on the X axis in ascending order according to their genomic position in every chromosome. Chromosomes were also placed in ascending order followed by X and Y. A dot indicates the middle point position of a given BAC. The Y axis lists the value of the Log_2(Mutant DNA signal/Reference DNA signal) of a given BAC probe. The Log_2(ratio) indicates an amplification when it is above +0.29 or a deletion when it is below -0.29. Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of the BAC probes from the reciprocal hybridization experiment. A red arrow indicates a 5.8 Mb heterozygous deletion including *Msh2* and *Msh6* genes on chromosome 17. There seem to be monosomies on chromosomes 2, 9, 13 and 16 (pink arrow). The Log_2(ratio) of chromosome Y is near -0.3, indicating this clone might be a mosaic population (pink arrow). A deletion on chromosome 2 appears to be homozygous (yellow arrow). Chromosome 11 appears to contain a segmental deletion at the centromere end (brown arrow). There seem to be trisomies on chromosome 1 and 14 (blue arrow). Some duplications or deletions observed only in one array but not in the reciprocal array. These duplications and deletions might be experimental variants, thus are excluded from the final results. The grey horizontal dotted lines are for Log_2(ratio) references of either +1 or -1.

7.2.6.3 *Msh6* mutant G10

The mutant clone G10, strongly resistant to 6TG, has a 3.2 Mb long homozygous deletion on chromosome 17 (Figure 7-24). In this clone, the *Msh2* gene is not deleted because the Log_2(ratio) of its representing BAC, RP24-315D17, is zero. The Log_2(ratio) of the two BACs representing *Msh6* (Figure 7-23) is different. BAC RP23-476A24 covers the whole *Msh6* locus. It may be partially deleted because its Log_2(ratio) is close to -0.3 (Figure 7-24). While BAC RP23-330J10 covers part of the *Msh6* gene and this BAC is homozygously deleted, with the Log_2(ratio) of ~-2. Therefore the *Msh6* gene is likely to be homozygously deleted. More experiments, such as array CGH analysis with higher resolution, FISH (fluorescent in situ hybridization) or expression array analysis may be of help to clarify this. Unfortunately, this clone was contaminated thus was not able to be further analysed.

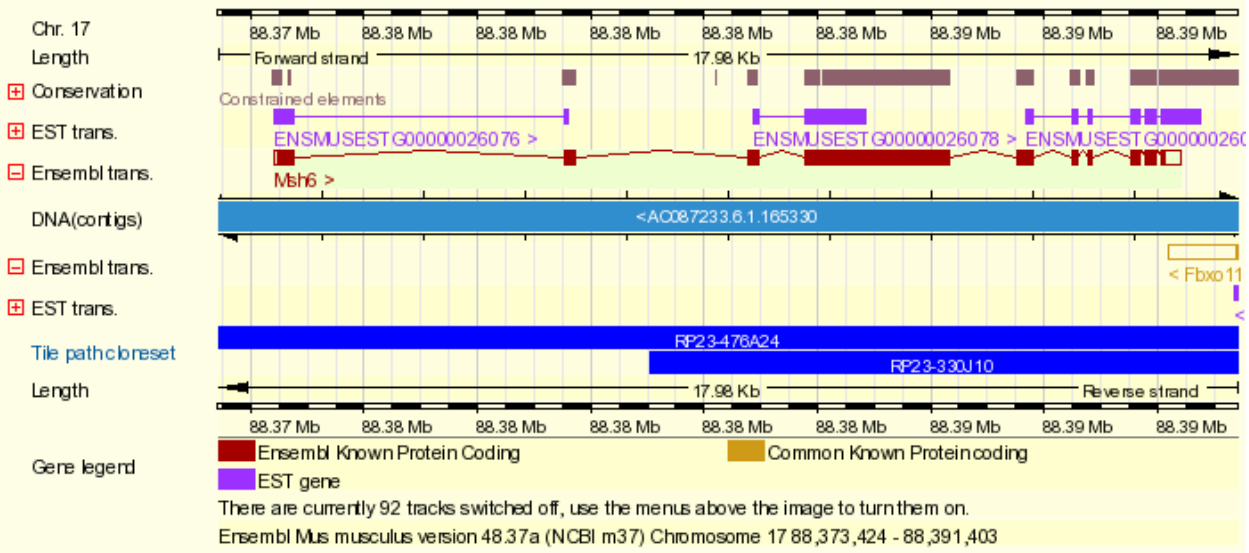


Figure 7-23 Tile path BACs at *Msh6* locus

A snapshot of the *Msh6* locus on mouse chromosome 17 (NCBI m37, Ensembl release 48). The *Msh6* transcript is shown in red boxes connected by red lines. Blue boxes indicate BACs used on the tile path array CGH (200 kb resolution). BAC RP23-476A24 covers the whole *Msh6* gene while RP23-330J10 covers part of the *Msh6* gene.

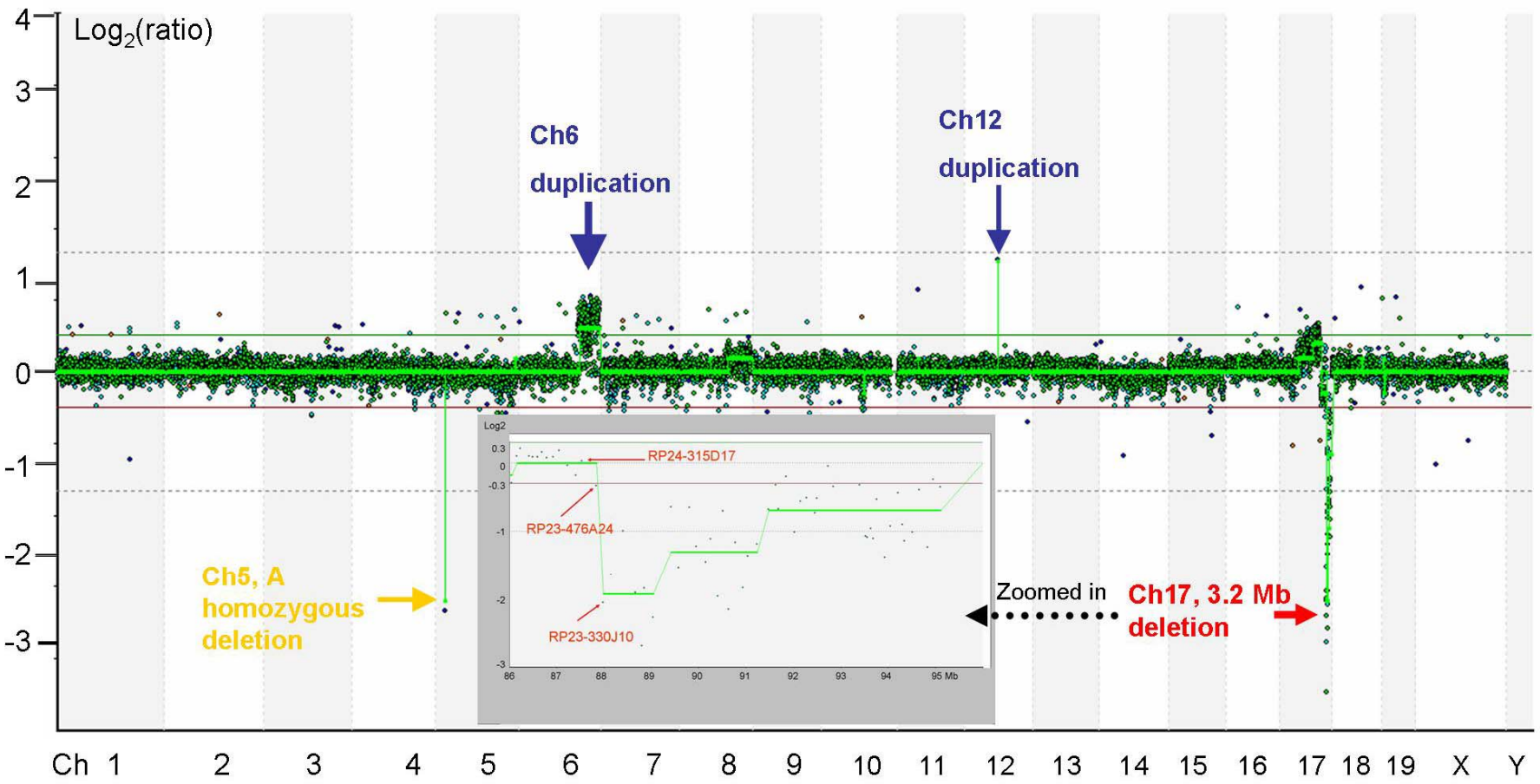


Figure 7-24 **Array CGH profile of mutant G10**

Figure 7-24 Array CGH profile of mutant G10

The array CGH profile of the mutant clone G10. A light green line indicates the average Log_2(Mutant DNA signal/Reference DNA signal) of regions. A 3.2 Mb long homozygous deletion is on chromosome 17 (red arrow). A zoomed-in figure of this deleted region is shown. BAC RP24-315D17 (red font) covers the *Msh2* gene and its Log_2(ratio) is zero, indicating the copy number of this BAC remains unchanged. BACs RP23-476A24 (red font) covers the whole *Msh6* gene region (Figure 7-23) while its Log_2(ratio) is about -0.3. RP23-330J10 (red font) represent part of the *Msh6* gene and its Log_2(ratio) is close to -2. In addition to this homozygous deletion, there is a homozygous deletion on chromosome 5 (single BAC, yellow arrow) and two duplications (blue arrows) on chromosome 6 (Mb) and 12 (single BAC). However, change of one BAC needs to be further examined because it can also be experimental variations.

7.2.6.4 A common deletion on chromosome 14

It was believed that 6TG resistance associated genes can be mutated for multiple times in the mutation pools and can be isolated. The homozygous and heterozygous mutations in the *Msh2* and *Msh6* genes described previously confirmed this assumption. Thus, frequently appeared mutations need to be investigated. It was found that eight clones, including B2, E1, E2, E4, E5, F2, G4 and H6 contain a common heterozygously deleted region on chromosome 14 (Figure 7-25). Five of them (E1, E4, E5, F2 and H6) are strongly resistant to 6TG, while the other three (B2, E2 and G4) are weakly resistant to 6TG. This common deleted region is between 107.9 Mb (RP24-115C5) and 119.6 Mb (RP23-407B7) on chromosome 14, indicated by the starting and the ending BAC probes.

The genes within this region were extracted from the Ensembl database (Table 7-6). Some clues were found to establish links between genes in this list and DNA repair function. The gene of ATP-binding cassette, sub-family C (CFTR/MRP), member 4 (*Abcc4*), the mouse homologue of human multidrug resistance-associated protein 4 gene (*MRP4*), is in this region. Mice deficient for this gene are more vulnerable to the nucleotide analogue, 9'-(2'-phosphonylmethoxyethyl)-adenine (PMEA), suggesting the protective function of ABCC4 in defence of DNA damage (Belinsky *et al.* 2007). Therefore, the *Abcc4* gene is worth further investigating.

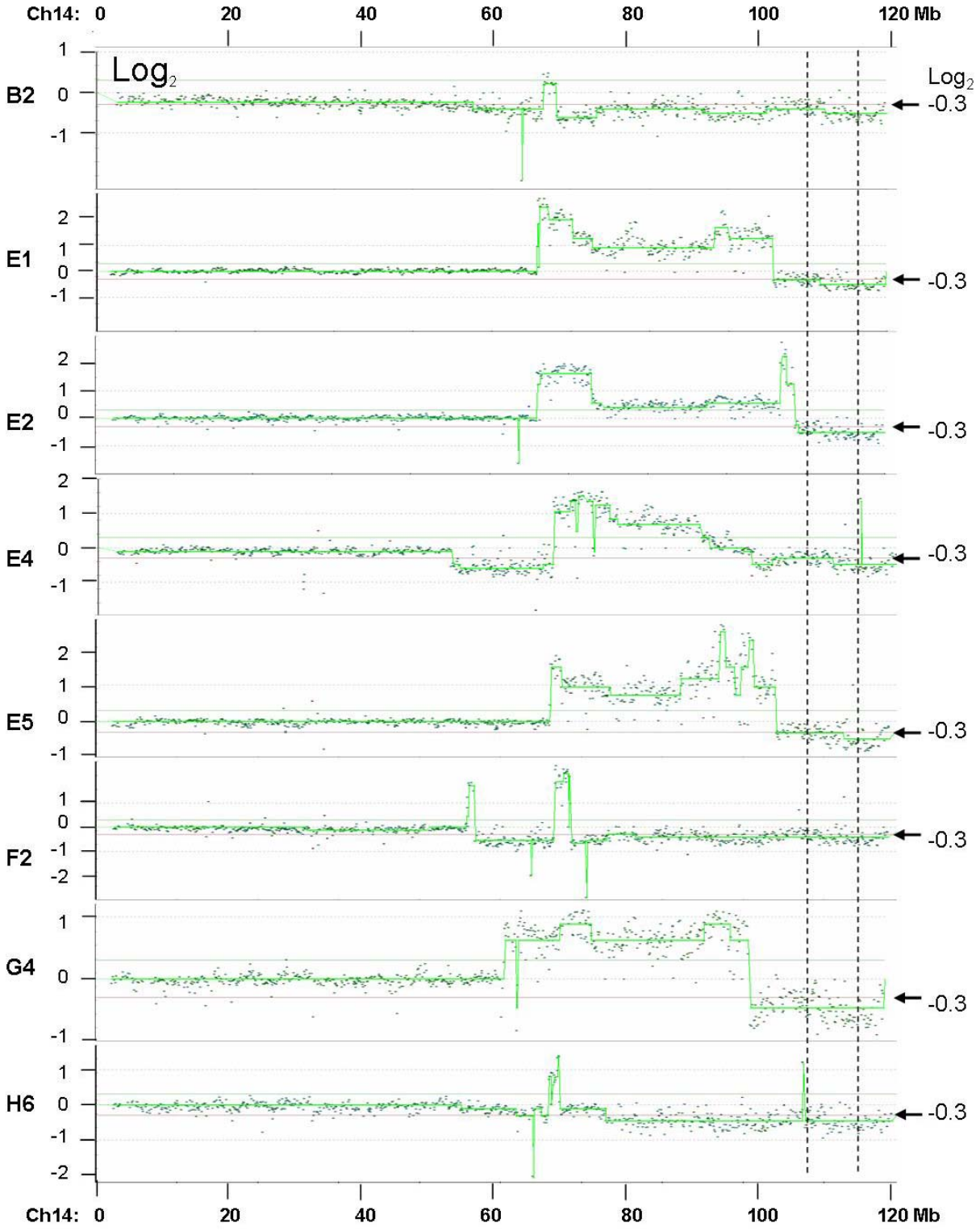


Figure 7-25 A common heterozygous deletion on chromosome 14

The array CGH profiles on chromosome 14 of the mutant clones B2, E1, E2, E4, E5, F2, G4 and H6. Two dotted lines indicate the common heterozygous deletion in the eight clones. The Log_2(ratio) of these deletions is below -0.3 and above -1.

Table 7-6 Genes in the common heterozygous deletion on chromosome 14

MGI symbol	Description
Slitrk1	SLIT and NTRK-like family, member 1 [MGI:2679446]
Slitrk6	SLIT and NTRK-like family, member 6 [MGI:2443198]
Slitrk5	SLIT and NTRK-like family, member 5 [MGI:2679448]
Tpm3	tropomyosin 3, gamma [MGI:1890149]
Q78E13_MOUSE	RNA for type IIB intracisternal A-particle (IAP) element encoding integrase, clone 111. (Fragment). Source: Uniprot/SPTREMBL Q78E13
Gpc5	glypican 5 [MGI:1194894]
Gpc6	glypican 6 [MGI:1346322]
Dct	dopachrome tautomerase [MGI:102563]
Tgds	TDP-glucose 4,6-dehydratase [MGI:1923605]
Gpr180	G protein-coupled receptor 180 [MGI:1930949]
Sox21	SRY-box containing gene 21 [MGI:2654070]
Abcc4	ATP-binding cassette, sub-family C (CFTR/MRP), member 4 [MGI:2443111]
A830021K08Rik	RIKEN cDNA A830021K08 gene [MGI:2443998]
Q8C3Q2_MOUSE	13 days embryo lung cDNA, RIKEN full-length enriched library, clone:D430043M02 product:hypothetical protein, full insert sequence. Source: Uniprot/SPTREMBL Q8C3Q2
Cldn10	claudin 10 [MGI:1913101]
Dnajc3	DnaJ (Hsp40) homolog, subfamily C, member 3 [MGI:107373]
Dzip1	DAZ interacting protein 1 [MGI:1914311]
Ugcgl2	UDP-glucose ceramide glucosyltransferase-like 2 [MGI:1913685]
Hs6st3	heparan sulfate 6-O-sulfotransferase 3 [MGI:1354960]

These genes locate between 107.9 and 119.6 Mb on chromosome 14 and are sorted in the direction from centromere to telomere.

7.2.6.5 The Other mutant clones

Of the thirty-four mutant clones analysed using array CGH, twenty-four have been described previously (summarized in Table 7-7). Although homozygous deletion covering the MMR genes of *Msh2* and *Msh6* were previously identified, other MMR genes, such as *Mlh1* and *Pms2*, failed to be identified in the other ten 6TG-isolated mutant clones (B3, B7, E3, E9, F1, F6, G6, H1, H3 and H5). One possible reason is that the adjacent genes of *Mlh1* and *Pms2* are essential to ES cell viability. Loss of function of these essential genes can result in cell lethality. Based on the data described previously, the homozygous deletions can be three to four million base pairs in size; and the adjacent regions of *Mlh1* and *Pms2* genes are gene-dense, therefore the probability to delete adjacent essential genes together with the MMR genes is high. The deficiency of some genes adjacent to *Mlh1* or *Pms2*, for instance *Tdgf1* (Ding *et al.* 1998), *Ubp1* (Parekh *et al.* 2004), *Rac1* (Sugihara *et al.* 1998) and *Trrap* (Herceg *et al.* 2001) causes embryonic lethality in knockout mice. These evidence implies that the deficiency of these genes may lead to cell lethality. Moreover, *Tdgf1* is 300 kb upstream of *Mlh1* and *Ubp1* is 2.7 Mb downstream of *Mlh1*; *Rac1* is 400 kb upstream of *Pms2* and *Trrap* is 800 kb downstream of *Pms2*. Within such short distances, homozygous deletions are able to delete these vital genes together with *Mlh1* or *Pms2,* thus *Mlh1* or *Pms2* deficient mutants can not be isolated.

The array CGH profiles of the other ten mutant clones are shown in Figure 7-26–35. Among them, the mutant clone B3 is weakly resistant to 6TG while the other nine clones are strongly resistant to 6TG (Table 7-2, C). The loss of chromosome Y was observed in seven out of ten mutants (B7, E3, E9, F1, F6, H1 and H3). Some single BAC homozygous-like deletions were seen on chromosome 3 (two BACs), 5, 7, 8, 11, 12, 13, 14, 15, 17, 18 and X. These deletions were found to be associated with chromosome Y loss. Thus, they were assumed to belong to chromosome Y but were incorrectly annotated to other chromosomes. Besides these single BAC deletions, some other single BAC homozygous deletions were observed, for example those on chromosomes 3, 15 and 17 of the clone G6, and that on chromosome 6 of the clone H5. However, single BAC changes are not convincing data. One needs to be cautious with them. Due to the complexity of the duplications and deletions in these clones, it is difficult to conclude which gene is responsible for the 6TG resistance. In an attempt to answer this question with information at transcriptional level, expression array analysis using an Illumina® platform was conducted (Chapter 8).

Table 7-7 Summary of the mutants analysed using array CGH

Types of mutation	ID of mutant clones	Number of mutant clones
Homozygous deletion of *Msh2* and *Msh6*	B6, D4, D8, F4, H14	5
Homozygous deletion of *Msh6*	D1	1
Heterozygous deletion of *Msh2* and *Msh6*	C1, C3, C5, D6, F9, F16, G8, G9, H13	9
A potential homozygous deletion of *Msh6*	G10	1
A common heterozygous deletion on ch14	B2, E1, E2, E4, E5, F2, G4, H6	8
Others	B3, B7, E3, E9, F1, F6, G6, H1, H3, H5	10
Total number		34

Types of the thirty-four mutants analysed using array CGH.

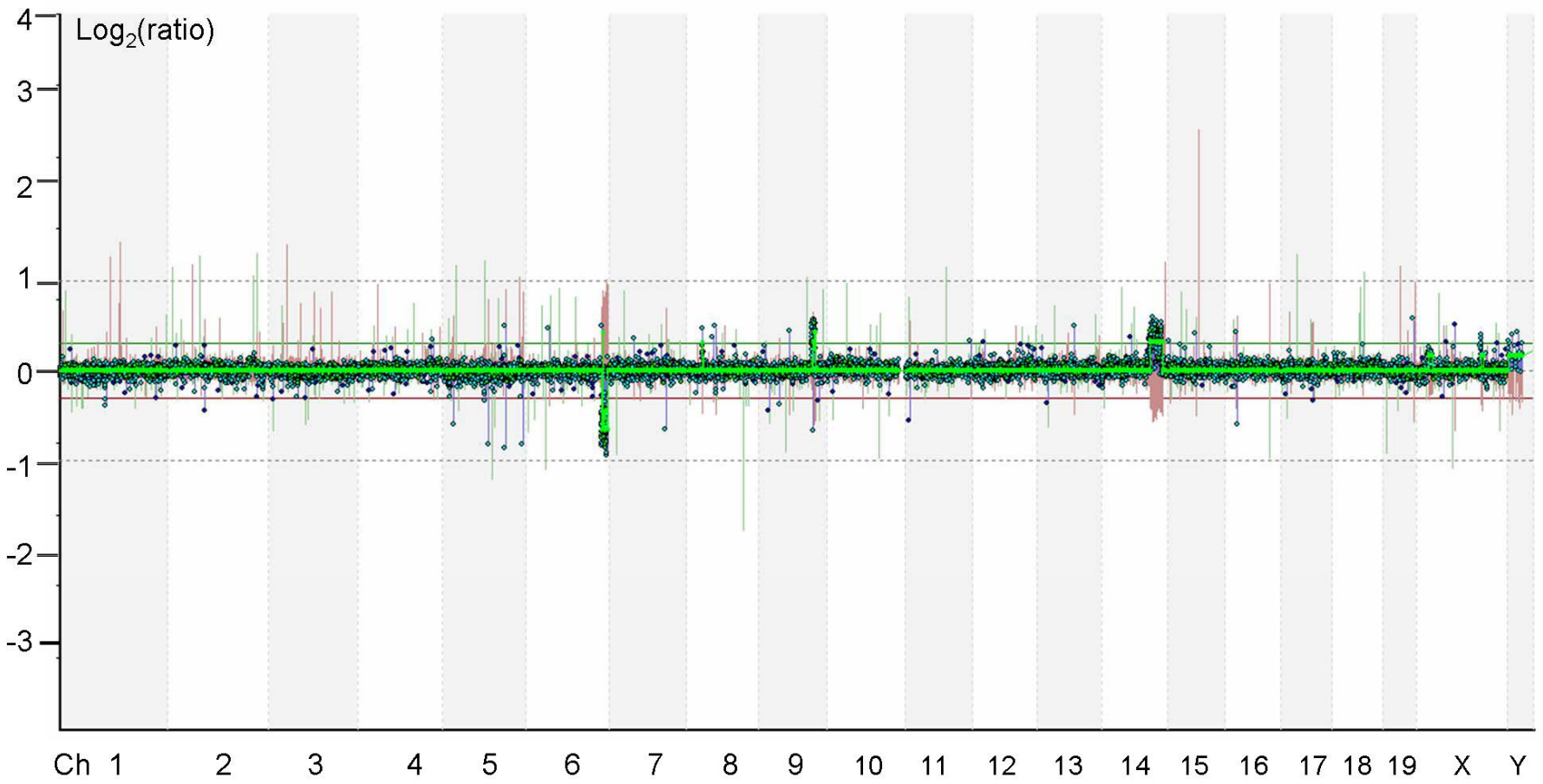


Figure 7-26 **Array CGH profile of mutant B3**

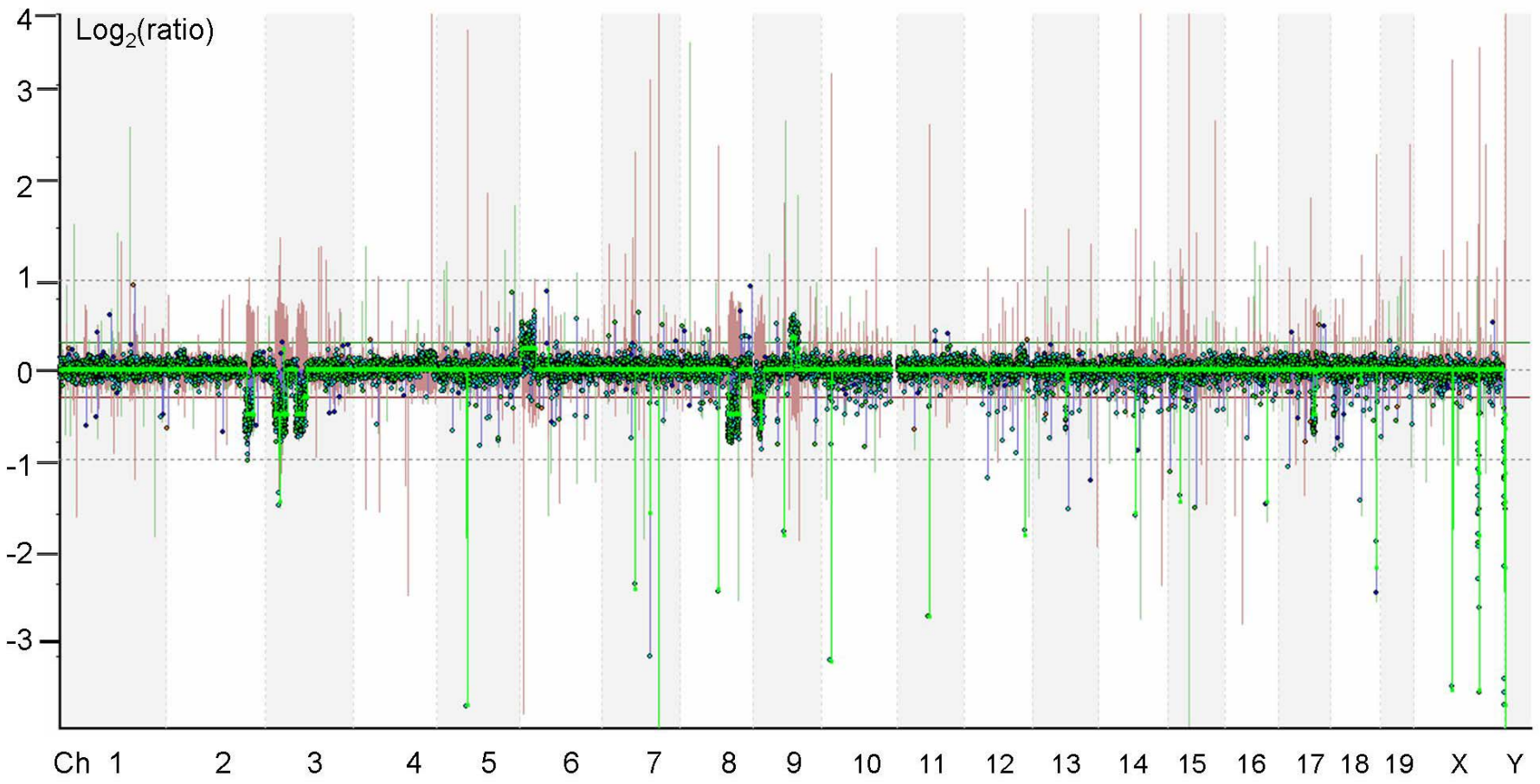


Figure 7-27 **Array CGH profile of mutant B7**

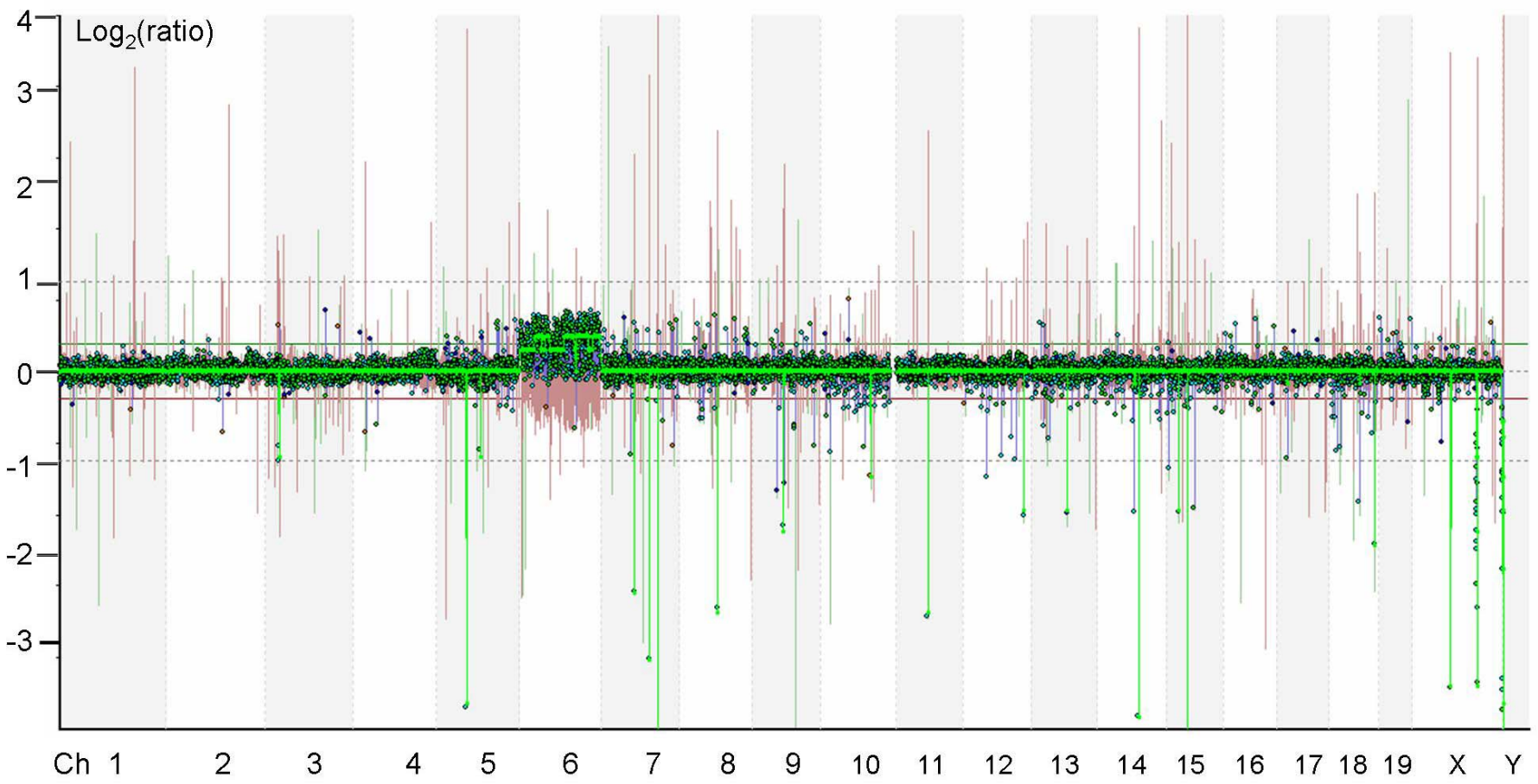


Figure 7-28 **Array CGH profile of mutant E3**

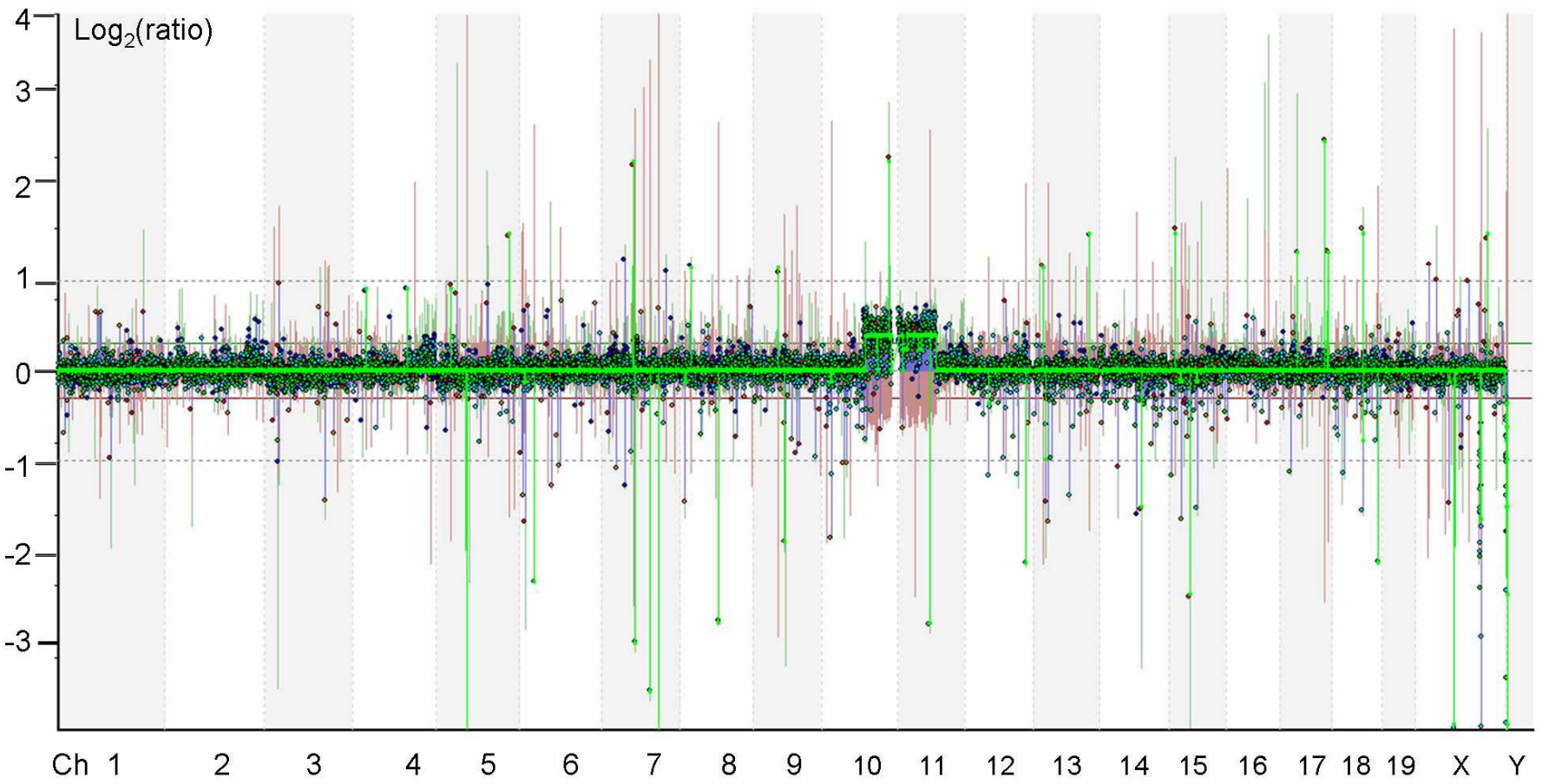


Figure 7-29 **Array CGH profile of mutant E9**

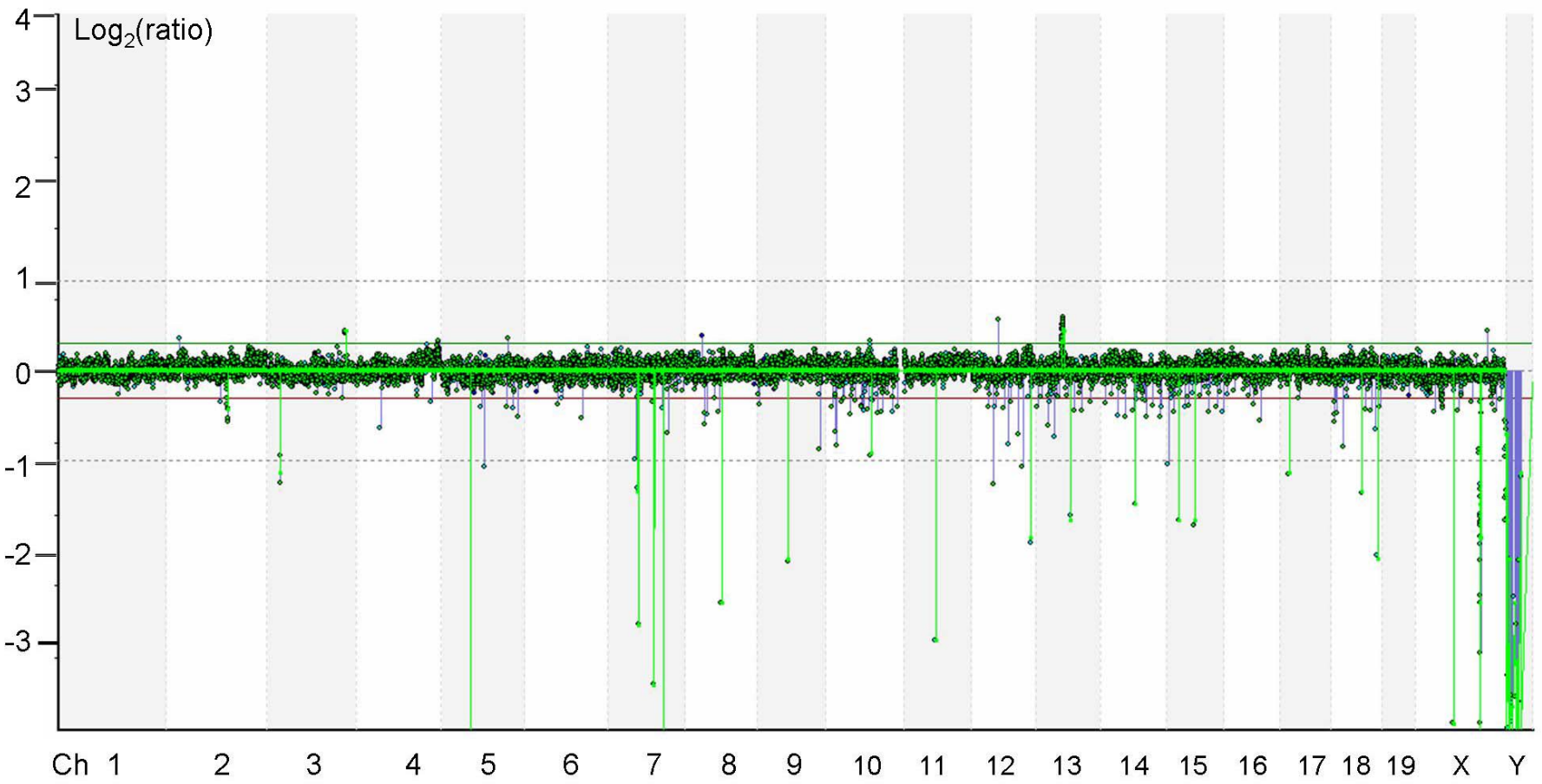


Figure 7-30 Array CGH profile of mutant F1

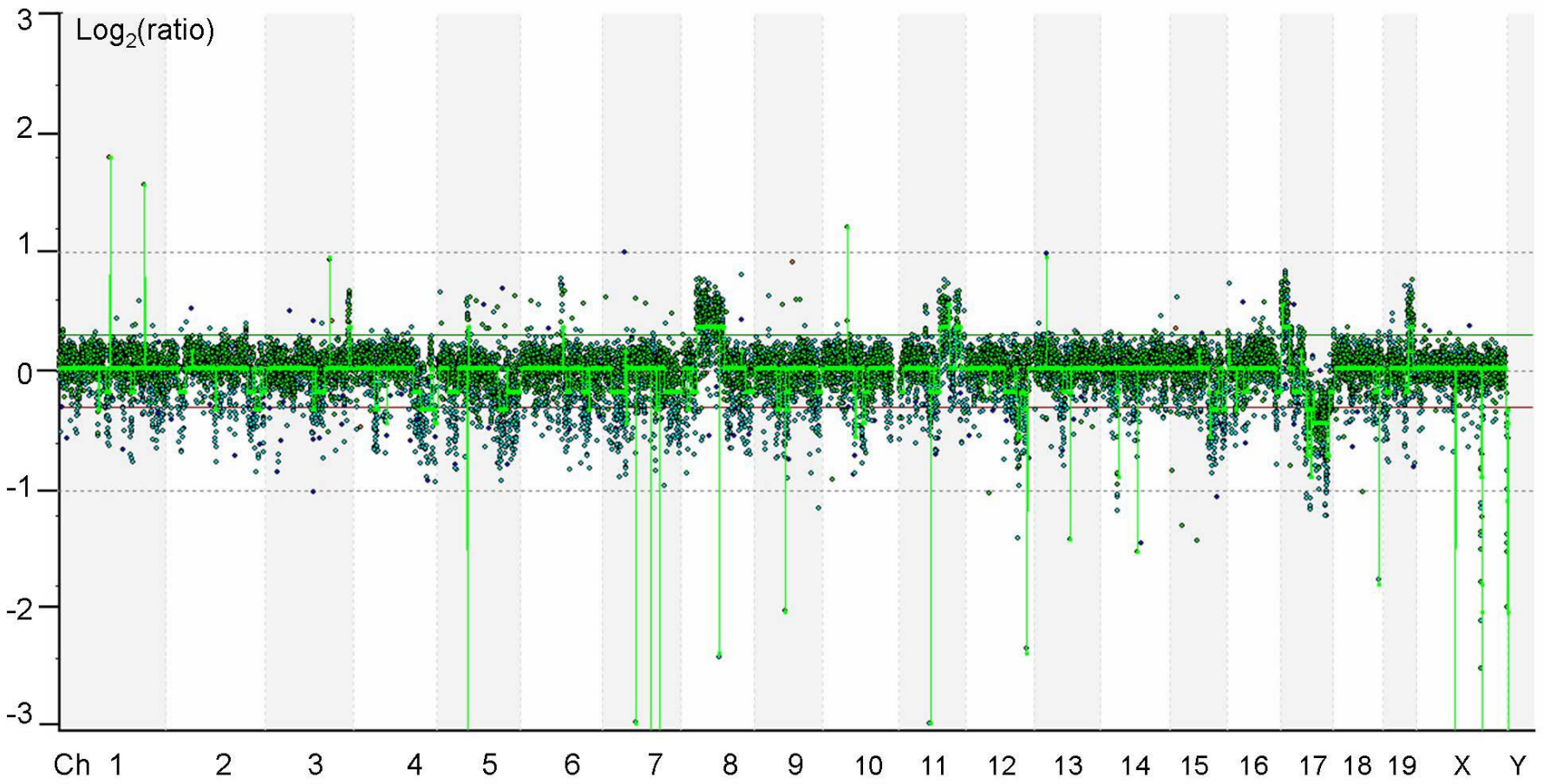


Figure 7-31 **Array CGH profile of mutant F6**

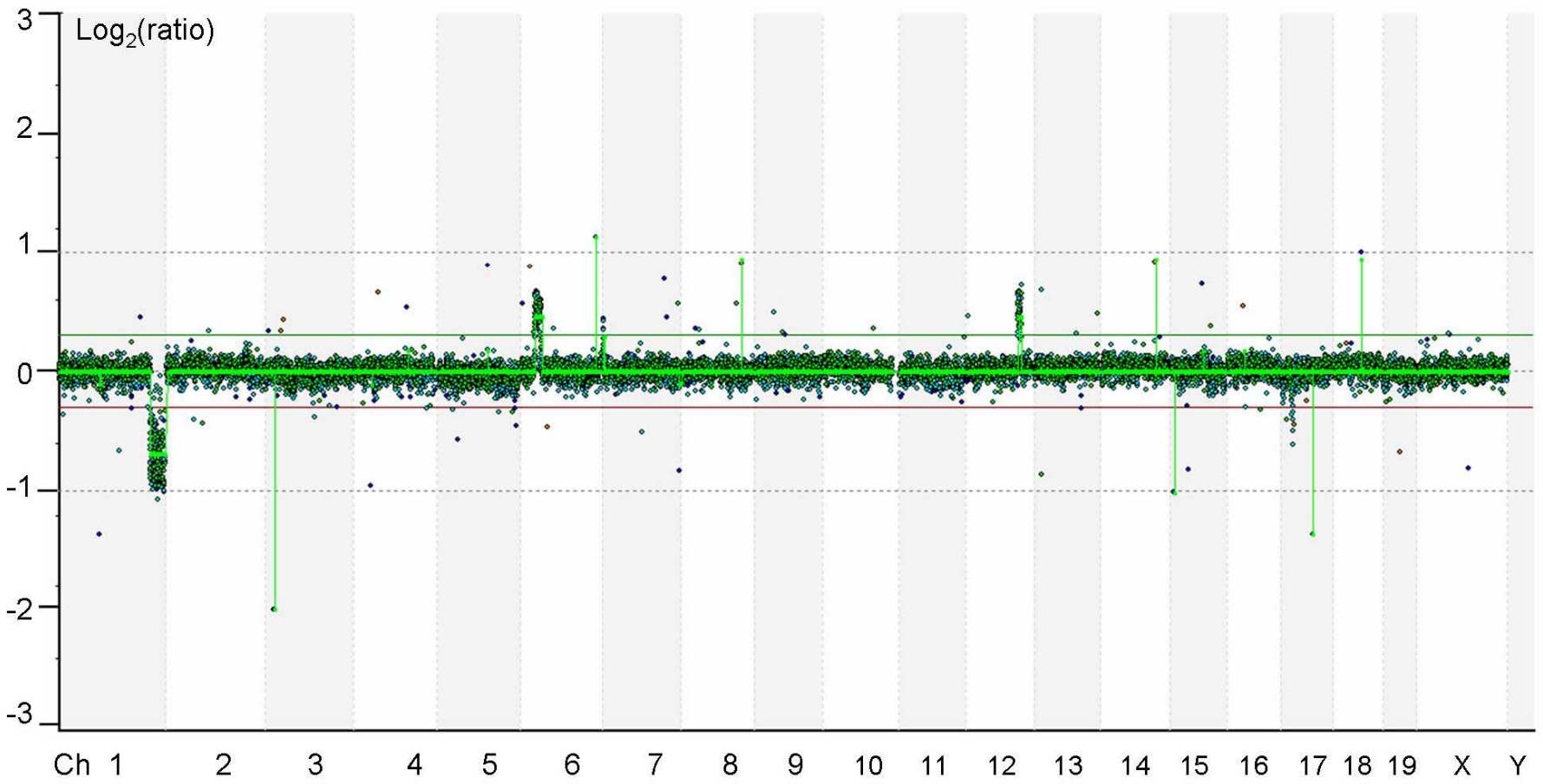


Figure 7-32 **Array CGH profile of mutant G6**

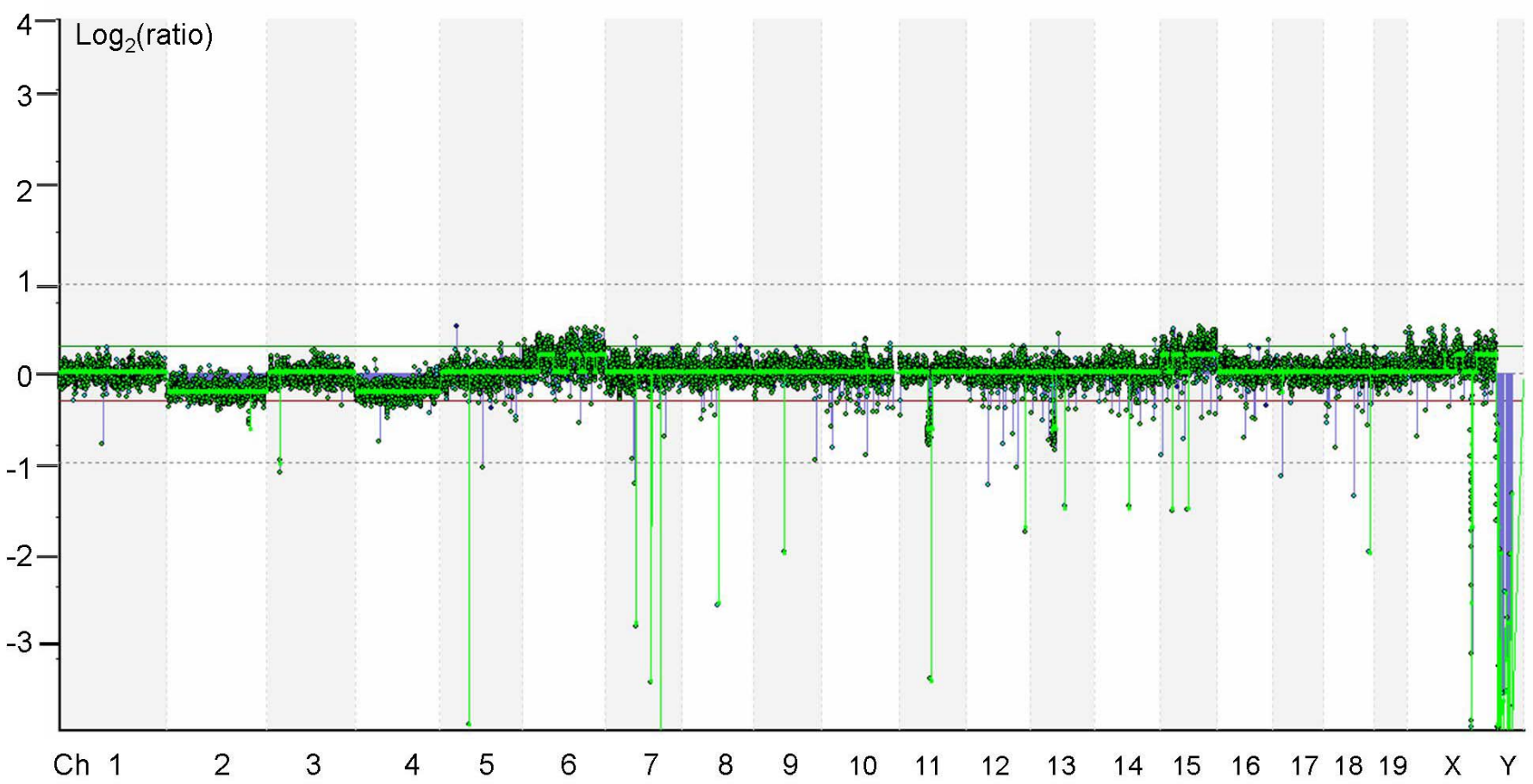


Figure 7-33 Array CGH profile of mutant H1

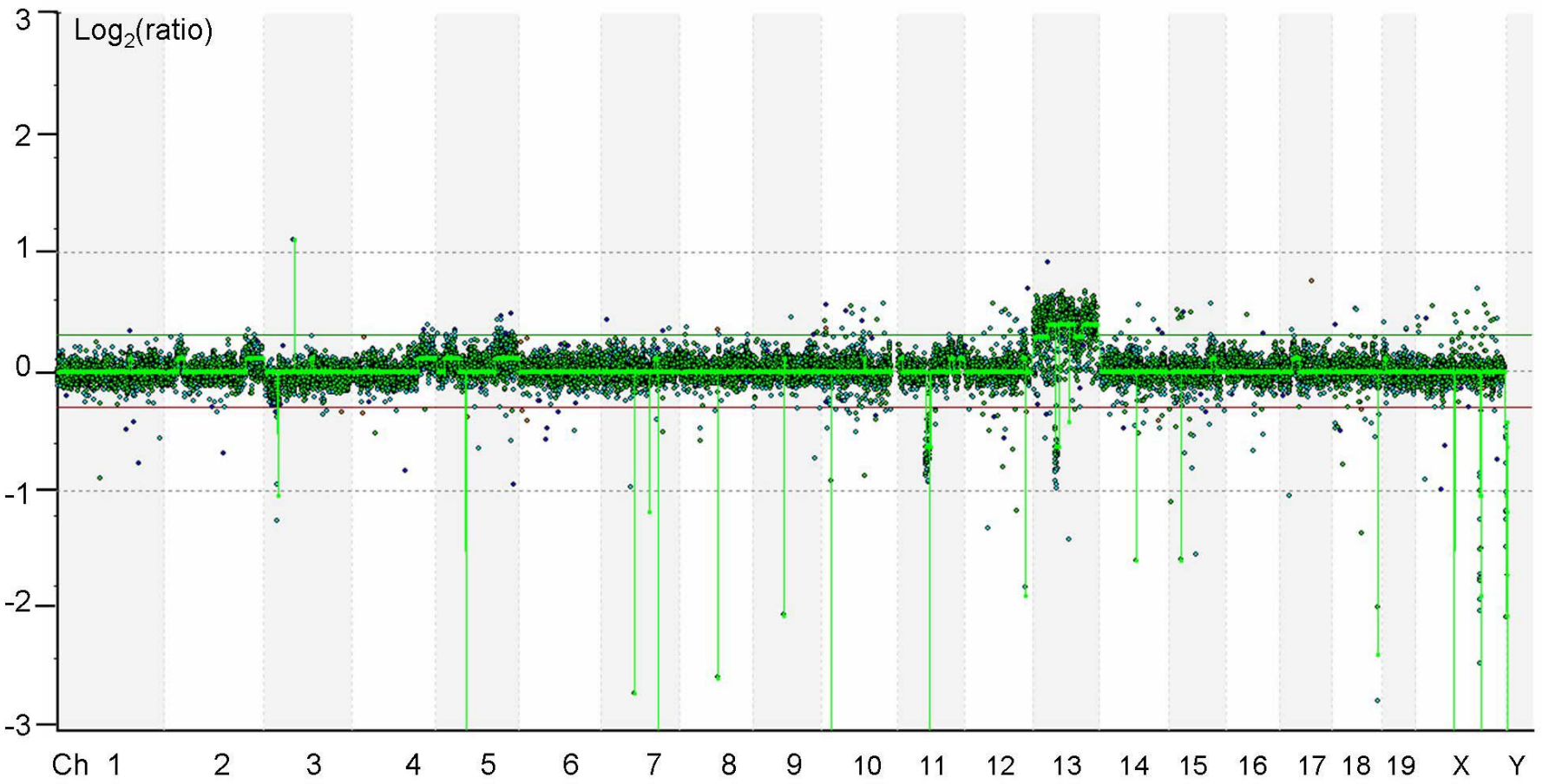


Figure 7-34 **Array CGH profile of mutant H3**

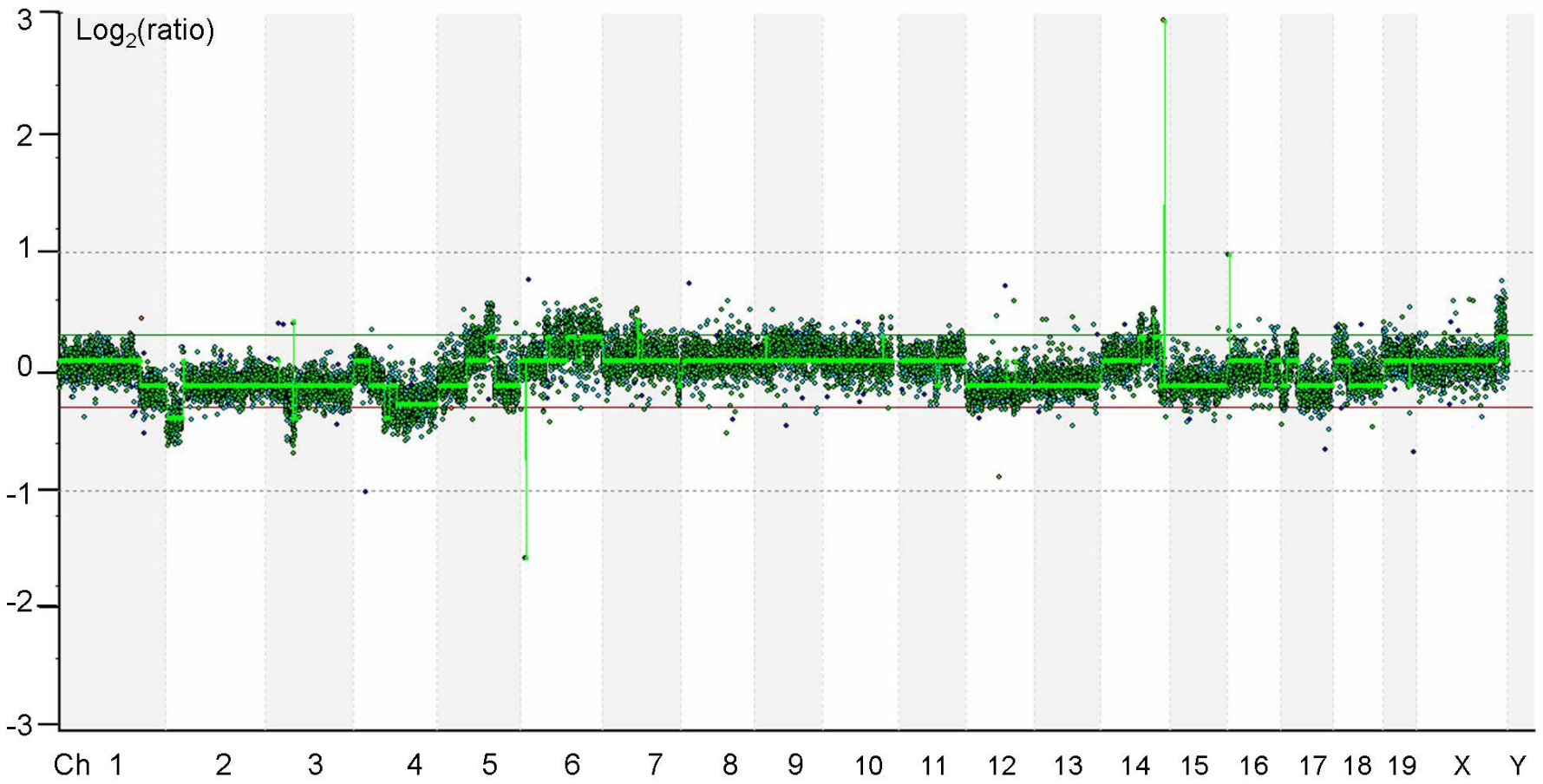


Figure 7-35 **Array CGH profile of mutant H5**

7.3 Conclusion and discussion

7.3.1 Advantages of establishing clonal relationships

Mutations generated by insertional mutagenesis methods are uniquely tagged and can be identified by molecular methods. However, mutations generated by irradiation cannot be so easily identified. The analysis methods for irradiation-induced mutations are genome-wide array CGH and expression arrays. These approaches are expensive. To prevent redundant experiments, molecular tags were introduced to distinguish clonal relationships between mutants and identify independent clones. A retroviral vector was incorporated into the cells before they were irradiated. By controlling the infection efficiency, it was simple to ensure each cell had a unique tag. In order to guarantee each cell contains one tag after irradiation, the complexity of each mutation pool was limited to approximate one thousand cells per pool, one sixth of the total number of tags before irradiation. Among 49 mutants with strong 6TG resistance, twenty-seven were identified as unique clones after analysing clonal relationships. Thus, only 27 unique clones needed to be analysed on arrays.

7.3.2 Mutations were identified in *Msh2* and *Msh6* MMR genes

Within 27 unique 6TG fully resistant mutants (Table 7-3), 22% (6/27) had homozygous deletions at MMR gene loci, including *Msh2* and *Msh6*. These six mutants, B6, D1, D4, D8, F4, and H14, were from four mutation pools. These *Msh2* and *Msh6* deletions ranged from 2.6 to 4.4 Mb. These samples provide size information about homozygous deletions generated by irradiation and the LOH events of the *Blm*-deficient cells. The average size of these six homozygous deletions is 3.6 Mb. It is adequate for 200 kb resolution array CGH to detect such deletions.

The mutant of other MMR genes, for instance *Mlh1* or *Pms2,* was not isolated. This may be due to presence of genes vital to cell viability around these two genes. *Tdgf1* is 300 kb upstream of *Mlh1* and *Ubp1* is 2.7 Mb downstream of *Mlh1*; *Rac1* is 400 kb upstream of *Pms2* and *Trrap* is 800 kb downstream of *Pms2*. The deficiency of these four genes causes embryonic lethality in knockout mice (Ding *et al.* 1998; Herceg *et al.* 2001; Parekh *et al.* 2004; Sugihara *et al.* 1998), implying that the deficiency of these genes may lead to cell lethality. It is known that homozygous deletions can be three to four million base pairs in size (Table 7-4), thus homozygous deletions are able to delete these vital genes together with *Mlh1* or *Pms2.*

In addition to the six clones with homozygous deletions, nine clones (C1, C3, C5, D6, F9,

F16, G8, G9 and H13) were identified which had heterozygous deletions of the same region on chromosome 17. The Log_2(ratio) of these deletions is between -0.5 and -0.7 (Figure 7-20). These nine deletions range from 13.3 Mb to 46 Mb in size (Table 7-5), much larger than the sizes of homozygous deletions covering the *Msh2* and *Msh6* genes (Table 7-4). The average length of heterozygous deletion is 28 Mb.

The region, including the *Msh2* and *Msh6* genes, is located at the 88 Mb position of the 95 Mb long chromosome 17 thus is near to telomere. It was believed that such a region is easy to be lost through homologous recombination. It seems all heterozygous deletions (Figure 7-21), except that in the mutant G9, in the eight clones extend to the last BAC probe on chromosome 17 (RP24-317D6), suggesting that these eight clones lost the whole distal part of chromosome 17. Although the heterozygous mutant clones still have one copy of genes *Msh2* and *Msh6*, six of them (C1, C5, F16, G8, G9 and H13) have shown strong 6TG resistance. It is assumed that the undeleted allele of either *Msh2* or *Msh6* has a small mutation, which results in a loss-of-function of either *Msh2* or *Msh6.* It is anticipated that expression array analysis will confirm that the transcripts of *Msh2* and *Msh6* are affected. Loss of distal part of chromosome 17 was also identified in the six mutants with homozygous *Msh2* and *Msh6* deletions (Figure 7-18).

Monosomies and trisomies were found in the heterozygous *Msh2* and *Msh6* mutants. For example, clone D6 appears to have monosomies of chromosomes 2, 9, 13 and 16 and trisomies on chromosome 1 and 14 (Figure 7-22). Chromosome 11 appears to contain a segmental deletion at the centromere end. However, the Log_2(ratio) of these changes is nearer to zero than that of other heterozygous deletions and duplications, indicating that this clone might be a mosaic population or a tetraploid cell or both. There are some evidence to support this hypothesis: firstly, the Log_2(ratio) of chromosome Y is near -0.3, higher than that of the normal heterozygous deletions. The ES cells are XY cells. If chromosome Y is lost, the Log_2(ratio) of chromosome Y on the array CGH should be like those of homozygous deletions. If one chromosome Y remains in a tetraploid cell, the Log_2(ratio) of it should be close to that of heterozygous deletions. However, the Log_2(ratio) of chromosome Y in clone D6 is near 0.3, suggesting that D6 is an impure clone. Secondly, monosomies are very rare in ES cells. Thus, the clone might be a mosaic population of normal ploidy cell with tetraploid cells with one copy of chromosome loss (Log_2(¾)= -0.42). Thirdly, the Log_2(ratio) of chromosome 1 and 14 is less than 0.32, the Log_2(ratio) of a pentaploid chromosome in a tetraploid cell (Log_2(5/4)), indicating this clone is mosaic.

7.3.3 Clues of potential MMR genes from a common deletion

The six homozygous and nine heterozygous mutations in the *Msh2* and *Msh6* genes demonstrate that the mutation of an MMR gene can be isolated in multiple clones. This kind of frequently appeared deletion needs to be investigated because 6TG resistance associated genes may locate in such regions. In the 6TG-isolated mutants, a common heterozygously deleted region on chromosome 14 was identified in eight mutant clones (B2, E1, E2, E4, E5, F2, G4 and H6) (Figure 7-25). This common deleted region is between 107.9 Mb (RP24-115C5) and 119.6 Mb (RP23-407B7) on chromosome 14. *Abcc4*, one of the genes in this region, was found to be associated with nucleotide analogue resistance, which is probably related to MMR. The *Abcc4* gene encodes ATP-binding cassette, sub-family C (CFTR/MRP), member 4 (*Abcc4*). Mice deficient for this gene exhibit increased lethality when treated with a nucleotide analogue, 9'-(2'-phosphonylmethoxyethyl)-adenine (PMEA) (Belinsky *et al.* 2007), suggesting that the *Abcc4* gene may play a role in the toxicity pathway of 6TG. Therefore, the deletion in a common region provides a clue to investigate potential MMR-related genes.

CHAPTER 8. Expression analysis of the 6TG^R mutants

8.1 Introduction

Some heterozygous deletions of the *Msh2* and *Msh6* genes are strongly resistant to 6TG. It was believed that small mutations exist in the undeleted alleles of these genes, which lead to the 6TG resistance. To test this assumption, expression array analysis was used to examine the presence of their transcripts. If the transcriptional level of heterozygously deleted genes is similar to that of homozygously deleted genes, the presence of such mutations is indirectly confirmed. Moreover, the transcriptional loss of homozygous deletions can be used to calibrate expression arrays. They can serve as negative controls of signals from non-expressed genes.

In addition, expression micro-arrays generate information not only about the presence of a given transcript, but also the expression level of a given gene. When probes on an array represent all of the genes in a genome, expression array can be used to screen transcriptional variations of the mutants. An Illumina® Mouse-6 Expression BeadChip was used to identify whole genome expression variations in the 6TG-isolated mutants.

8.2 Results

8.2.1 Expression analysis

Twenty-five unique clones (Table 7-3, without clone F16 and G10) with strong 6TG resistance were analysed not only by array CGH, but also by expression array. Homozygous deletions at known MMR genes, for instance *Msh2* and *Msh6*, were identified by array CGH analysis. As for the MMR genes, which are not homozygously deleted, absence of transcript in any of them can be expected to be identified through expression array analysis. This may be caused by small mutations (shorter than 100 kb) or nonsense-mediated mRNA decay in mutant clones.

Before initiating the expression array analysis on 6TG^R mutants, experiments were conducted to measure the sensitivity and accuracy of the arrays. A comparison between wild type cells (including both AB1 and AB2.2 cells) and *Blm*-deficient cells (cell line NGG5-3) was conducted to determine if there were any expression changes in the *Blm*-deficient cells. Individual MMR-deficient mutants were compared with the *Blm*-deficient cells to identify expression changes in the mutants. Downstream regulated gene expression

changes were identified by comparing and clustering mutants.

Biological replicates were used in the expression array analysis. These were RNA samples extracted from the same passage of a particular clone. Twenty-five 6TG-resistant mutant clones (three replicates each), six weakly resistant mutant clones (three replicates each), four wild type cell replicates (two in AB1 and two in AB2.2), three *Blm*-deficient cell replicates (NGG5-3) and three *Dnmt1* knockout cell replicates (ww56+*Hprt*) were used to conducted the analysis on an Illumina® platform. As described in the Methods, the ww56+*Hprt* cell line was derived from the AB2.2 cell line and has a targeted deletion of exons 2–4 in both alleles of *Dnmt1*. In addition, this cell line has an *Hprt* mini-gene, targeted at one allele of the *Gdf9* locus, providing the activity of hypoxanthine guanine phosphoribosyl transferase.

Each Illumina® array has approximately 30 features per probe and 19,400 genes on array. A total 7645 genes have more than one probe. A detection P-value was calculated for probes using BeadStudio (version 2.3.41) software. Probes with certain level of variations between its features are excluded from analysis. A detection P-value provides a measure of the probability of a measured signal being due to hybridization with Cy3 labelled complementary RNA rather than background non-specific binding. Each array experiment generated one signal value for a given feature of a probe. The expression fold change was calculated by comparing signals of one probe in replicates of one cell line and signals of the same probe in replicates of another cell line. The adjusted P-value was calculated per probe to indicate the confidence of expression fold change. The lower the adjusted P-value is, the more likely the expression fold change data is real. Five thousand probes with the lowest adjusted P-values in each pair of comparisons were selected for analysis. The highest P-value of five thousand probes in the arrays is between 10^{-4}–10^{-2}, thus gives a good confidence to the analysis. The other probes with higher adjusted P-values, even if they have a greater magnitude of signals, were excluded from the comparison analysis.

8.2.1.1 Controls of wild type cell lines – AB1 and AB2.2

AB1 and AB2.2 cell lines are both wild type and isolated from the same subline of 129SvEv mice (Simpson *et al.* 1997). The AB2.2 cell line was isolated from a mouse line that was still being backcrossed. In the AB2.2 cell line, the X-linked *Hprt* locus was inactivated by insertion of a myeloproliferative sarcoma virus (MPSV) (King *et al.* 1985; Kuehn *et al.* 1987; Lobel *et al.* 1985). MPSV is a variant of the Moloney strain of murine leukemia virus (MMuLV). By comparing probe signals between AB1 and AB2.2 cells, I could expect to

detect expression variations of both the *Hprt* gene and the MPSV or MMuLV associated elements.

Five thousand probes with the lowest adjusted P-values in the comparison were selected for analysis. The majority of genes are expressing at the same level between the AB1 and AB2.2 cell lines. There was no probe with a high fold change and high adjusted P-value. Thirteen probes (representing eleven genes) have twice the expression in AB1 than in AB2.2 (Table 8-1). Two out of the three *Hprt1* gene probes were among them, with 2.75 and 2.73 times the expression level in AB1 than in AB2.2 (Figure 8-2). Another *Hprt1* gene probe, scl0002995.1_2-S, has 1.98 times the expression level in AB1 than in AB2.2, close to the threshold of 2 times. This result is consistent with the fact that the *Hprt1* gene was inactivated in AB2.2 cells (Kuehn *et al.* 1987). Six probes were found to express at least 2 times more in AB2.2 than in AB1. These probes detect gene expression of the mouse retroviruses. Their sequences are aligned (Figure 8-1). A 23-nucleotide region is found to be identical among them and a Blast analysis (NCBI) of these sequences indicates that they are conserved with the v-mos gene, at 3' LTR of the MPSV proviral sequence. They are also conserved with LTR of other MMuLV viruses. The expression of these sequences in AB2.2 cells is consistent with the knowledge of this cell line, in which the *Hprt1* gene was silenced by a recombinant MPSV (King *et al.* 1985; Kuehn *et al.* 1987; Lobel *et al.* 1985).

The other ten genes with ~2 times the expression in the AB1 cell line than in the AB2.2 cell line are not expected. These transcriptional variants could be due to genetic or epigenetic alterations between the two cell lines. From this comparison analysis between two wild type cell lines, the Illumina® expression array platform is confirmed to be a stable platform to measure transcriptional variants between cell lines. It has demonstrated sensitivity to exogenous and endogenous gene expression, and it exhibits that two times the expression level can serve as a threshold to distinguish transcribed from non-transcribed genes. However, it should be borne in mind that a large expression fold change will result from a knock-down of an expressed gene.

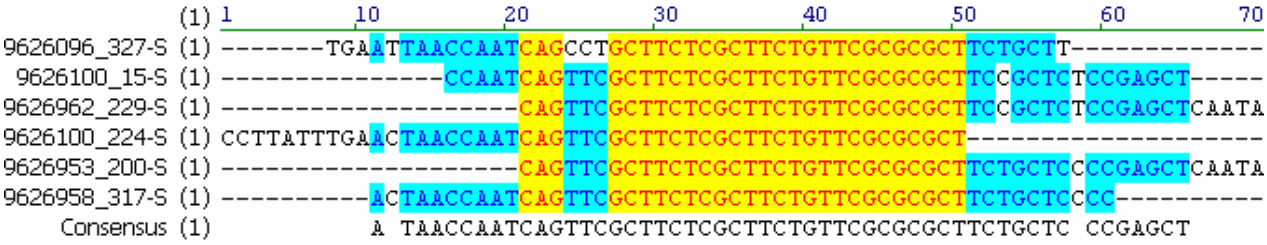


Figure 8-1 Alignment of probe sequences of MPSV

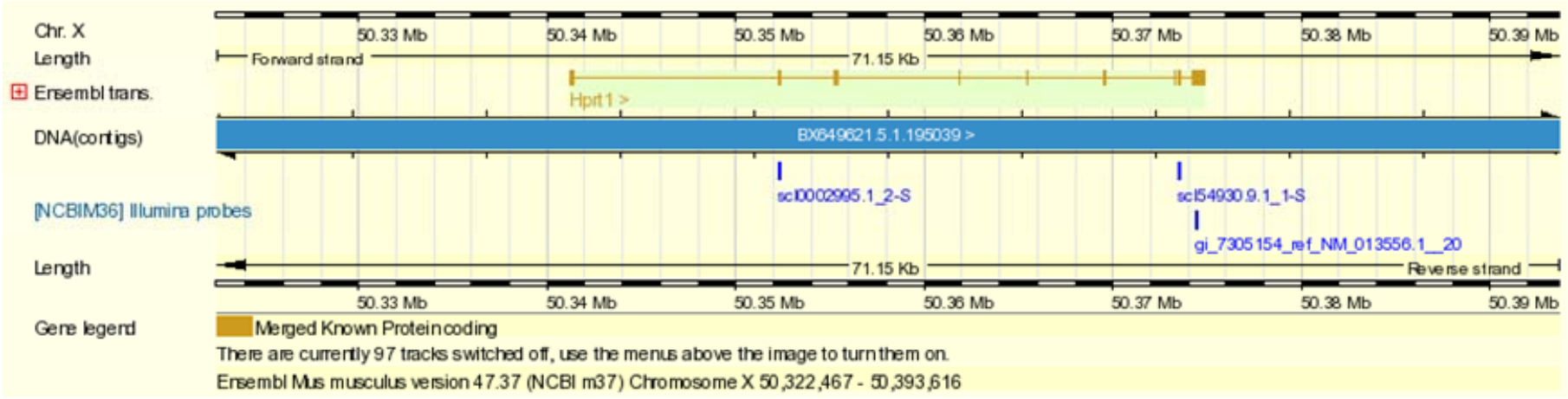
Probe names are listed on the left. These sequences detect transcripts which are expressed twice as much in the AB2.2 compared with the AB1 cell line. Their sequences are aligned and identical nucleotides are shown in a red font on a yellow background. Conserved nucleotides are shown in a dark blue font on a light blue background. Their consensus sequence is shown at the bottom.

Table 8-1 Main transcriptional variants between the AB1 and AB2.2 cell lines

Probe	Gene symbol	Expression fold change	Comparison P-value	Gene description
Expression AB1/AB2.2 ≥ 2.0		**AB1/AB2.2**		
scl30469.7_1-S	*Igf2*	3.89	1.90×10^{-8}	Mus musculus insulin-like growth factor 2 (Igf2), mRNA.
scl34188.7.1_72-S	*Acta1*	3.78	5.60×10^{-9}	Mus musculus actin, alpha 1, skeletal muscle (Acta1), mRNA.
scl54930.9.1_1-S	*Hprt1*	2.75	1.04×10^{-10}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
gi_7305154_ref_NM_013556.1__205-S	*Hprt1*	2.73	1.88×10^{-10}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
scl0068377.2_268-S		2.71	1.27×10^{-9}	Mus musculus RIKEN cDNA 0610041G09 gene (0610041G09Rik), mRNA.
scl24990.1.20_266-S	*Pou3f1*	2.51	3.08×10^{-13}	Mus musculus POU domain, class 3, transcription factor 1 (Pou3f1), mRNA.
scl54328.4.1_36-S	*Rhox9*	2.43	8.40×10^{-10}	Mus musculus placenta specific homeobox 2 (Psx2), mRNA.
scl16803.9_203-S	*Slc40a1*	2.39	1.57×10^{-9}	Mus musculus solute carrier family 40 (iron-regulated transporter), member 1 (Slc40a1), mRNA.
scl30493.4.19_120-S	*Sct*	2.20	4.52×10^{-8}	Mus musculus secretin (Sct), mRNA.
scl014468.10_101-S	*Gbp1*	2.16	8.60×10^{-11}	Mus musculus guanylate nucleotide binding protein 1 (Gbp1), mRNA.
scl50828.7.1_115-S	*Psmb8*	2.13	7.77×10^{-17}	Mus musculus proteosome (prosome, macropain) subunit, beta type 8 (large multifunctional protease 7) (Psmb8), mRNA.
scl52209.6.1_16-S	*Ttr*	2.08	2.54×10^{-6}	Mus musculus transthyretin (Ttr), mRNA.
scl018715.6_137-S	*Pim2*	2.04	2.70×10^{-13}	Mus musculus proviral integration site 2 (Pim2), mRNA.
scl0002995.1_2-S *	*Hprt1*	1.98	7.99×10^{-33}	Mus musculus hypoxanthine guanine phosphoribosyl transferase (Hprt), mRNA.
Expression AB2.2/AB1 ≥ 2.0		**AB2.2/AB1**		
9626962_229-S #	-	1.92	8.60×10^{-8}	Moloney murine sarcoma virus
9626096_327-S	-	2.01	1.09×10^{-8}	Friend murine leukemia virus
9626953_200-S	-	2.04	1.13×10^{-8}	Abelson murine leukemia virus
9626958_317-S	-	2.04	1.45×10^{-8}	Murine leukemia virus
9626100_15-S	-	2.10	1.07×10^{-10}	Murine sarcoma virus
9626100_224-S	-	2.30	1.46×10^{-11}	Murine sarcoma virus

*: This probe has 1.98 times the signal strength in AB1 than in AB2.2. #: This probe has 1.92 times the signal strength in AB2.2 than in AB2.2.

A



B

Illumina array probe on *Hprt1*	Genomic position on chromosome X (bp)	Within exon	Expression fold change AB1 vs AB2.2	Comparison P-value
scl0002995.1_2-S	49243733~49243782	2	1.98	7.99×10^{-33}
scl54930.9.1_1-S	49264970~49265019	8	2.75	1.04×10^{-10}
gi_7305154_ref_NM_013556.1__205-S	49265893~49265942	9	2.73	1.88×10^{-10}

Figure 8-2 **Mouse *Hprt1* gene expression variation between AB1 and AB2.2 cells**

A. Mouse *Hprt1* gene structure on NCBI m36, Ensembl release 46 with three Illumina® array probes annotated in this gene (blue text). **B**. Summary of expression fold change of these probes between the AB1 and AB2.2 cell lines. Probe coordinates are shown. Results indicate that *Hprt1* transcripts in the AB1 cell line are about 2 times more than those in the AB2.2 cell line. Comparison P-values (adjusted P-values) are near zero (10^{-33}–10^{-10}), indicating that these expression fold changes have a very low probability of being wrong.

8.2.1.2 Comparison between *Blm*-deficient and wild type cells

Since the *Blm*-deficient ES cells are used as the basis of comparisons with the mutant clones, it is important to examine their expression profile against the AB1 and AB2.2 ES cells. Using a comparable analysis strategy, transcriptional variations were analysed between *Blm*-deficient cells (NGG5-3) and wild type cells (AB1 and AB2.2). Data from AB1 and AB2.2 cells, two biological replicates each, was grouped together as the data of wild type cells. There were 111 genes expressing in wild type cells whose expression signals were two times more than those in the *Blm*-deficient cells (Table 8-2). The maximum expression fold change was seen for *Spink3* (serine peptidase inhibitor, Kazal type 3) with 7.4 times the expression in the wild type cells than in *Blm*-deficient cells. On the other hand, there were 67 genes expressing in *Blm*-deficient cells whose expression levels were more than twice of those in wild type cells (Table 8-3). The maximum expression fold change was of the AF067061 gene (2-cell-stage, variable group, member 3), which has 5.0 times more expression in the *Blm*-deficient cells than in wild type cells. The fact that so many genes express differently between the wild type and the *Blm*-deficient cells may be because of two reasons: 1. the loss of the genome surveillance function of the BLM protein. 2. homozygous deletion of transcription factors. All of these genes are distributed throughout the whole genome without a preference for a specific chromosome, which suggests that these changes are not due to a specific structural chromosomal alteration in the cells used for this analysis. There were two targeting experiments and two Cre recombination events conducted on a *Blm*-deficient cell line to generate the cell line of NGG5-3 (Guo *et al.* 2004), which was used in this project. These four steps and the long cell culture history may accumulate mutations and these mutations are likely to lose heterozygosity. When such a homozygous mutation occurs to a transcription factor, the gene expression of a batch of genes can be altered. In the Table 8-2, the expression of the *Tcf7l2* (transcription factor 7-like 2) gene is down-regulated to 19% of that in the wild type cells, indicating this gene does not express. Thus, *Tcf7l2* may contribute to such a non-chromosome-specific gene expression variation.

One change that is expected is in the *Blm* gene itself. The *Blm* gene expression level is detected by the probes scl00067.1_13-S in exon 2 and scl31124.24.1_96-S in exon 22, the last exon (Figure 8-3). As described in the Methods, the *Blm*-deficient cell line NGG5-3 has two different alleles: one has a duplicated exon 3 (allele tm3Brd) and the other one lost exon 2 (allele tm4Brd). Due to a reading frame shift caused by the extra exon 3 in allele tm3Brd, the first premature termination codon (PTC) can be generated in the second exon 3

and following exons. As for allele tm4Brd, a new start codon may be used in exon 3 because the original start codon is in exon 2 and exon 2 is deleted. If transcription starts from the new start codon, the reading frame will also be destroyed and an early premature termination codon can be generated within 30 bases' distance of the new start codon. These PTCs can initialize nonsense-mediated mRNA decay (NMD). The signal of the 3' probe scl31124.24.1_96-S is 1.6 times higher in wild type cells than in *Blm*-deficient cells, which is consistent with the genotypes of these cell lines. This expression fold change is less than two times because the wild type *Blm* gene is either not highly expressed or because NMD in *Blm*-deficient cells is not complete. As for the 5' probe scl00067.1_13-S, no obvious expression difference was observed between the wild type cell lines and the *Blm*-deficient cell line. One reason for this may be due to the fact that the *Blm* transcript is quite long (4491 bp), thus the 5' sequence might not be reverse transcribed very efficiently when using oligo d(T) primers to prepare the samples for expression array.

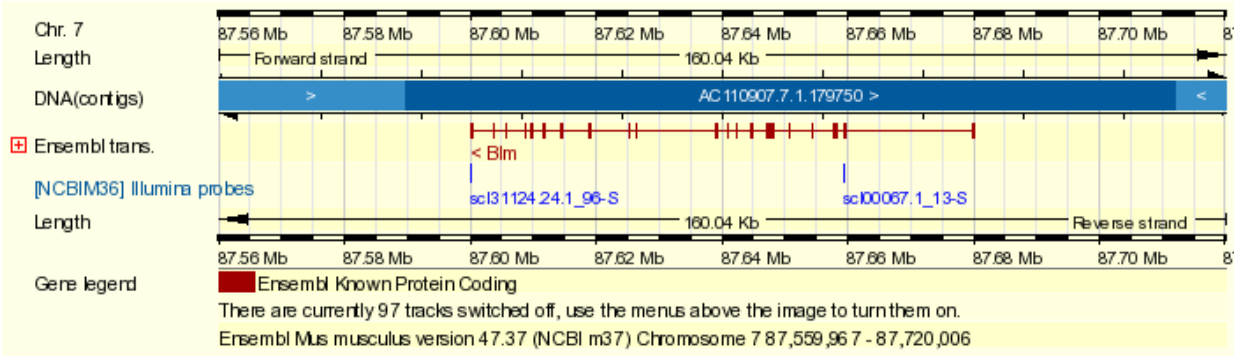


Figure 8-3 Expression array probes in the *Blm* gene

The *Blm* gene is transcribed from the telomere to the centromere direction on the long arm of mouse chromosome 7. There are two probes on the Illumina® array to detect its expression, scl31124.24.1_96-S (80328732–80328781 bp, reverse strand) and scl00067.1_13-S (80388138–80388187 bp, reverse strand).

Table 8-2 **Genes expressed in wild type cells twice as much as in *Blm*-deficient cells**

MGI symbol	Description	% wild type signal	Expression fold change (WT/*Blm*)	Comparison P-value	Chromosome Name
1200002N14Rik	RIKEN cDNA 1200002N14 gene [MGI:1918962]	33%	3.01	4.28×10^{-27}	10
1810015C04Rik	RIKEN cDNA 1810015C04 gene [MGI:1913520]	38%	2.60	4.83×10^{-16}	15
2610042L04Rik	predicted gene, ENSMUSG00000068790 [MGI:3704420]	34%	2.95	2.05×10^{-8}	14
3110043J09Rik	RIKEN cDNA 3110043J09 gene [MGI:1920417]	44%	2.25	1.58×10^{-14}	15
4631426J05Rik	RIKEN cDNA 4631426J05 gene [MGI:1924840]	43%	2.31	2.74×10^{-24}	7
5730469M10Rik	RIKEN cDNA 5730469M10 gene [MGI:1917814]	38%	2.66	7.48×10^{-16}	14
9930013L23Rik	RIKEN cDNA 9930013L23 gene [MGI:2443629]	47%	2.13	1.18×10^{-15}	7
Adam19	a disintegrin and metallopeptidase domain 19 (meltrin beta) [MGI:105377]	49%	2.04	1.27×10^{-19}	11
Amn	amnionless [MGI:1934943]	49%	2.04	9.24×10^{-19}	12
Arhgap17	Rho GTPase activating protein 17 [MGI:1917747]	23%	4.38	4.72×10^{-20}	7
Atp1b1	ATPase, Na+/K+ transporting, beta 1 polypeptide [MGI:88108]	40%	2.48	2.39×10^{-20}	1
Atp6v0a1	ATPase, H+ transporting, lysosomal V0 subunit A1 [MGI:103286]	45%	2.23	3.51×10^{-17}	11
Bach1	BTB and CNC homology 1 [MGI:894680]	49%	2.03	7.17×10^{-18}	16
Bat5	HLA-B associated transcript 5 [MGI:99476]	47%	2.14	5.45×10^{-9}	17
Car12	carbonic anyhydrase 12 [MGI:1923709]	47%	2.14	1.99×10^{-20}	9
Car4	carbonic anhydrase 4 [MGI:1096574]	24%	4.14	1.73×10^{-21}	11
Ccnt2	cyclin T2 [MGI:1920199]	42%	2.38	3.61×10^{-10}	1
Cited2	Cbp/p300-interacting transactivator, with Glu/Asp-rich carboxy-terminal domain, 2 [MGI:1306784]	50%	2.01	3.59×10^{-6}	10
Clic6	chloride intracellular channel 6 [MGI:2146607]	33%	3.01	2.42×10^{-19}	16
Clu	clusterin [MGI:88423]	22%	4.50	1.84×10^{-27}	14
Col4a1	procollagen, type IV, alpha 1 [MGI:88454]	40%	2.53	1.72×10^{-17}	8
Col4a2	procollagen, type IV, alpha 2 [MGI:88455]	24%	4.20	2.10×10^{-26}	8
Csf2ra	colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage) [MGI:1339754]	25%	3.94	3.12×10^{-29}	19
Ctgf	connective tissue growth factor [MGI:95537]	37%	2.68	3.12×10^{-29}	10
Ctns	cystinosis, nephropathic [MGI:1932872]	44%	2.27	4.74×10^{-12}	11
Ctsc	cathepsin C [MGI:109553]	49%	2.03	1.03×10^{-19}	7
Ctsh	cathepsin H [MGI:107285]	37%	2.73	4.85×10^{-21}	9
Cubn	cubilin (intrinsic factor-cobalamin receptor) [MGI:1931256]	26%	3.86	2.20×10^{-30}	2

Dab2	disabled homolog 2 (*Drosophila*) [MGI:109175]	44%	2.25	2.27×10^{-16}	15
Dnmt3b	DNA methyltransferase 3B [MGI:1261819]	20%	4.99	1.87×10^{-24}	2
Drp2	dystrophin related protein 2 [MGI:107432]	48%	2.07	6.64×10^{-25}	X
EG245190	predicted gene, EG245190 [MGI:3646425]	34%	2.97	3.54×10^{-11}	7
F3	coagulation factor III [MGI:88381]	35%	2.83	1.84×10^{-27}	3
Fabp3	fatty acid binding protein 3, muscle and heart [MGI:95476]	39%	2.58	1.69×10^{-22}	4
Fhl1	four and a half LIM domains 1 [MGI:1298387]	44%	2.25	7.17×10^{-18}	X
Flrt3	fibronectin leucine rich transmembrane protein 3 [MGI:1918686]	23%	4.26	2.59×10^{-21}	2
Foxq1	forkhead box Q1 [MGI:1298228]	42%	2.36	6.52×10^{-12}	13
Fst	follistatin [MGI:95586]	42%	2.36	2.73×10^{-16}	13
Fv1	Friend virus susceptibility 1 [MGI:95595]	31%	3.25	1.87×10^{-26}	4
Fxyd3	FXYD domain-containing ion transport regulator 3 [MGI:107497]	20%	4.89	1.43×10^{-25}	7
Gata4	GATA binding protein 4 [MGI:95664]	43%	2.33	8.03×10^{-16}	14
Gata6	GATA binding protein 6 [MGI:107516]	42%	2.39	5.91×10^{-12}	18
Gbp1	guanylate nucleotide binding protein 1 [MGI:95666]	32%	3.16	1.01×10^{-29}	3
Gcnt1	glucosaminyl (N-acetyl) transferase 1, core 2 [MGI:95676]	36%	2.81	3.68×10^{-26}	19
Gdpd5	glycerophosphodiester phosphodiesterase domain containing 5 [MGI:2686926]	21%	4.82	4.65×10^{-25}	7
Glipr1	GLI pathogenesis-related 1 (glioma) [MGI:1920940]	44%	2.27	7.78×10^{-11}	10
Gpc3	glypican 3 [MGI:104903]	32%	3.10	9.24×10^{-24}	X
Gpx3	glutathione peroxidase 3 [MGI:105102]	50%	2.01	3.03×10^{-19}	11
Grb10	growth factor receptor bound protein 10 [MGI:103232]	34%	2.93	4.60×10^{-19}	11
Gtl2	GTL2, imprinted maternally expressed untranslated mRNA [MGI:1202886]	35%	2.87	1.45×10^{-18}	12
H19	H19 fetal liver mRNA [MGI:95891]	38%	2.60	1.11×10^{-13}	7
Hkdc1	hexokinase domain containing 1 [MGI:2384910]	26%	3.78	7.73×10^{-30}	10
Hs3st1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1 [MGI:1201606]	25%	3.94	5.00×10^{-18}	5
Kdelr3	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 3 [MGI:2145953]	22%	4.50	5.66×10^{-26}	15
Kit	kit oncogene [MGI:96677]	26%	3.92	2.16×10^{-35}	5
Klb	klotho beta [MGI:1932466]	15%	6.68	9.50×10^{-31}	5
Krt18	keratin 18 [MGI:96692]	30%	3.32	1.39×10^{-26}	15
Krt8	keratin 8 [MGI:96705]	41%	2.46	2.43×10^{-16}	15
Lamb1-1	laminin B1 subunit 1 [MGI:96743]	41%	2.43	5.72×10^{-20}	12
Lamc1	laminin, gamma 1 [MGI:99914]	34%	2.93	9.81×10^{-10}	1
Ldlrap1	low density lipoprotein receptor adaptor protein 1 [MGI:2140175]	46%	2.19	1.86×10^{-6}	4

Lgmn	legumain [MGI:1330838]	41%	2.46	8.76×10^{-13}	12
Loxl2	lysyl oxidase-like 2 [MGI:2137913]	24%	4.23	5.22×10^{-35}	14
Lrp2	low density lipoprotein receptor-related protein 2 [MGI:95794]	38%	2.64	8.19×10^{-26}	2
Lrpap1	low density lipoprotein receptor-related protein associated protein 1 [MGI:96829]	27%	3.73	3.68×10^{-26}	5
Luc7l2	LUC7-like 2 (S. cerevisiae) [MGI:2183260]	34%	2.97	7.65×10^{-25}	6
Mod1	malic enzyme, supernatant [MGI:97043]	25%	4.03	3.17×10^{-19}	9
Myl3	myosin, light polypeptide 3 [MGI:97268]	48%	2.07	1.54×10^{-11}	9
Nav1	neuron navigator 1 [MGI:2183683]	46%	2.16	4.14×10^{-6}	1
Nostrin	nitric oxide synthase trafficker [MGI:3606242]	35%	2.87	3.99×10^{-16}	2
Nrk	Nik related kinase [MGI:1351326]	27%	3.73	9.31×10^{-22}	X
Nxf7	nuclear RNA export factor 7 [MGI:2159343]	48%	2.10	4.75×10^{-17}	X
Palld	palladin, cytoskeletal associated protein [MGI:1919583]	32%	3.10	1.97×10^{-17}	8
Pcgf5	polycomb group ring finger 5 [MGI:1923505]	45%	2.22	5.24×10^{-16}	19
Pdcd4	programmed cell death 4 [MGI:107490]	26%	3.84	1.01×10^{-29}	19
Pdpn	podoplanin [MGI:103098]	45%	2.22	1.03×10^{-18}	4
Peg3	paternally expressed 3 [MGI:104748]	46%	2.16	2.10×10^{-17}	7
Phf17	PHD finger protein 17 [MGI:1925835]	47%	2.14	8.03×10^{-14}	3
Pla2g12b	phospholipase A2, group XIIB [MGI:1917086]	49%	2.06	7.11×10^{-16}	10
Plod2	procollagen lysine, 2-oxoglutarate 5-dioxygenase 2 [MGI:1347007]	46%	2.17	3.44×10^{-12}	9
Podxl	podocalyxin-like [MGI:1351317]	42%	2.36	2.84×10^{-19}	6
Pthr1	parathyroid hormone receptor 1 [MGI:97801]	39%	2.57	1.67×10^{-26}	1
Pthr1	parathyroid hormone receptor 1 [MGI:97801]	44%	2.28	1.39×10^{-13}	9
Ptk6	Ptk6 protein tyrosine kinase 6 [MGI:99683]	27%	3.71	3.59×10^{-21}	2
Rab15	RAB15, member RAS oncogene family [MGI:1916865]	41%	2.43	1.06×10^{-19}	12
Ramp2	receptor (calcitonin) activity modifying protein 2 [MGI:1859650]	48%	2.08	5.03×10^{-25}	11
Rasd1	RAS, dexamethasone-induced 1 [MGI:1270848]	40%	2.53	6.37×10^{-16}	11
Rassf3	Ras association (RalGDS/AF-6) domain family 3 [MGI:2179722]	42%	2.36	6.27×10^{-19}	10
Rhox9	reproductive homeobox 9 [MGI:1890128]	39%	2.57	2.66×10^{-27}	X
S100a1	S100 calcium binding protein A1 [MGI:1338917]	41%	2.43	1.78×10^{-19}	3
Serpinh1	serine (or cysteine) peptidase inhibitor, clade H, member 1 [MGI:88283]	46%	2.17	9.00×10^{-19}	7
Sfmbt2	Scm-like with four mbt domains 2 [MGI:2447794]	45%	2.23	2.31×10^{-8}	2
Slc38a2	solute carrier family 38, member 2 [MGI:1915010]	44%	2.28	1.25×10^{-8}	15
Slc40a1	solute carrier family 40 (iron-regulated transporter), member 1 [MGI:1315204]	45%	2.20	1.47×10^{-13}	1
Slco2a1	solute carrier organic anion transporter family, member 2a1 [MGI:1346021]	33%	3.01	2.48×10^{-28}	9

Soat1	sterol O-acyltransferase 1 [MGI:104665]	40%	2.51	2.87×10^{-18}	1
Sox17	SRY-box containing gene 17 [MGI:107543]	42%	2.36	1.70×10^{-12}	1
Sox7	SRY-box containing gene 7 [MGI:98369]	41%	2.41	1.45×10^{-9}	14
Spink3	serine peptidase inhibitor, Kazal type 3 [MGI:106202]	29%	3.41	2.98×10^{-21}	18
Srgn	serglycin [MGI:97756]	47%	2.14	7.46×10^{-16}	10
Stard8	START domain containing 8 [MGI:2448556]	16%	6.15	5.05×10^{-29}	X
Syvn1	synovial apoptosis inhibitor 1, synoviolin [MGI:1921376]	32%	3.14	2.65×10^{-27}	19
Tax1bp3	Tax1 (human T-cell leukemia virus type I) binding protein 3 [MGI:1923531]	13%	7.41	4.40×10^{-20}	11
Tcf7l2	transcription factor 7-like 2, T-cell specific, HMG-box [MGI:1202879]	19%	5.31	3.51×10^{-17}	19
Tfpi	tissue factor pathway inhibitor [MGI:1095418]	37%	2.71	5.45×10^{-21}	2
Tinagl	tubulointerstitial nephritis antigen-like [MGI:2137617]	47%	2.11	3.31×10^{-22}	4
Tmem166	transmembrane protein 166 [MGI:2385247]	39%	2.55	7.23×10^{-25}	6
Tnk1	tyrosine kinase, non-receptor, 1 [MGI:1930958]	42%	2.38	1.85×10^{-24}	11
Tspan2	tetraspanin 2 [MGI:1917997]	43%	2.33	2.32×10^{-9}	3
Ttr	transthyretin [MGI:98865]	37%	2.73	5.40×10^{-16}	18
Txndc12	thioredoxin domain containing 12 (endoplasmic reticulum) [MGI:1913323]	39%	2.58	2.98×10^{-21}	4

Table 8-3 **Genes expressed in *Blm*-deficient cells twice as much as in wild type cells**

MGI symbol	Description	Expression fold change (*Blm*/WT)	Comparison P-value	Chromosome Name
2200001I15Rik	RIKEN cDNA 2200001I15 gene [MGI:1916384]	2.31	4.51×10^{-17}	14
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	2.97	1.58×10^{-11}	13
2310039H08Rik	RIKEN cDNA 2310039H08 gene [MGI:1914351]	2.08	4.88×10^{-15}	17
2410116G06Rik	RIKEN cDNA 2410116G06 gene [MGI:1915486]	2.30	3.07×10^{-11}	2
6330406I15Rik	RIKEN cDNA 6330406I15 gene [MGI:1917967]	2.99	2.98×10^{-9}	5
Aebp1	AE binding protein 1 [MGI:1197012]	2.04	1.60×10^{-5}	11
AF067061	2-cell-stage, variable group, member 3 [MGI:2675349]	4.99	5.74×10^{-16}	NT_053651
Antxr2	anthrax toxin receptor 2 [MGI:1919164]	2.13	6.00×10^{-10}	5
Aqp1	aquaporin 1 [MGI:103201]	2.20	2.65×10^{-8}	6
Axl	AXL receptor tyrosine kinase [MGI:1347244]	2.55	2.97×10^{-10}	7
Bgn	biglycan [MGI:88158]	2.23	1.49×10^{-9}	X
Cav1	caveolin, caveolae protein 1 [MGI:102709]	2.23	3.50×10^{-9}	6
Cd44	CD44 antigen [MGI:88338]	2.10	2.93×10^{-9}	2
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	2.43	6.22×10^{-9}	4
Col4a5	procollagen, type IV, alpha 5 [MGI:88456]	2.35	2.18×10^{-11}	X
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	2.62	1.36×10^{-8}	10
Cpe	carboxypeptidase E [MGI:101932]	2.14	7.82×10^{-8}	8
Crlf1	cytokine receptor-like factor 1 [MGI:1340030]	2.19	5.29×10^{-12}	8
Cul1	cullin 1 [MGI:1349658]	2.08	3.44×10^{-12}	6
Cxcl1	chemokine (C-X-C motif) ligand 1 [MGI:108068]	2.03	3.35×10^{-10}	5
D12Ertd647e	ISG12a protein. [Source:Uniprot/SPTREMBL;Acc:Q70LN0]	2.36	2.89×10^{-9}	12
Dcn	decorin [MGI:94872]	3.27	3.56×10^{-8}	10
Dlk1	delta-like 1 homolog (*Drosophila*) [MGI:94900]	2.16	1.64×10^{-4}	12
Ecm1	extracellular matrix protein 1 [MGI:103060]	2.27	8.73×10^{-12}	3
Emp3	epithelial membrane protein 3 [MGI:1098729]	2.46	6.29×10^{-9}	7
Eno2	enolase 2, gamma neuronal [MGI:95394]	2.10	1.15×10^{-10}	6
Fbln2	fibulin 2 [MGI:95488]	2.10	3.08×10^{-7}	6
Fez1	fasciculation and elongation protein zeta 1 (zygin I) [MGI:2670976]	2.07	1.41×10^{-16}	9
Fxyd5	FXYD domain-containing ion transport regulator 5 [MGI:1201785]	2.20	1.97×10^{-11}	7
Gbx2	gastrulation brain homeobox 2 [MGI:95668]	2.73	2.75×10^{-12}	1

H2-D1	histocompatibility 2, D region [MGI:95912]	3.07	2.06×10^{-13}	17
Ifitm3	interferon induced transmembrane protein 3 [MGI:1913391]	2.50	2.97×10^{-15}	7
Igfbp7	insulin-like growth factor binding protein 7 [MGI:1352480]	2.16	2.07×10^{-10}	5
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	2.66	8.63×10^{-8}	1
Jam2	junction adhesion molecule 2 [MGI:1933820]	2.57	5.06×10^{-23}	16
Lgals1	lectin, galactose binding, soluble 1 [MGI:96777]	2.83	9.08×10^{-15}	15
Lmna	lamin A [MGI:96794]	2.20	4.23×10^{-10}	3
Lox	lysyl oxidase [MGI:96817]	2.11	1.94×10^{-9}	18
Loxl1	lysyl oxidase-like 1 [MGI:106096]	2.62	4.70×10^{-14}	9
Ltbp2	latent transforming growth factor beta binding protein 2 [MGI:99502]	2.53	6.86×10^{-10}	12
Ltbp3	latent transforming growth factor beta binding protein 3 [MGI:1101355]	2.71	4.64×10^{-11}	19
Ly6a	lymphocyte antigen 6 complex, locus A [MGI:107527]	2.45	1.71×10^{-9}	15
Masp1	mannan-binding lectin serine peptidase 1 [MGI:88492]	2.08	1.75×10^{-8}	16
Mmd	monocyte to macrophage differentiation-associated [MGI:1914718]	2.31	8.32×10^{-11}	11
Mmp14	matrix metallopeptidase 14 (membrane-inserted) [MGI:101900]	2.14	1.07×10^{-11}	14
Msn	moesin [MGI:97167]	2.00	8.45×10^{-12}	X
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	3.51	4.15×10^{-18}	7
Myo1f	myosin IF [MGI:107711]	2.30	4.40×10^{-20}	17
Notch4	Notch gene homolog 4 (*Drosophila*) [MGI:107471]	2.01	2.22×10^{-13}	17
Nrp1	neuropilin 1 [MGI:106206]	2.03	3.14×10^{-10}	8
Nupr1	nuclear protein 1 [MGI:1891834]	2.13	8.46×10^{-18}	7
Plau	plasminogen activator, urokinase [MGI:97611]	2.69	7.03×10^{-9}	14
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	2.31	8.04×10^{9}	5
Ptn	pleiotrophin [MGI:97804]	2.25	3.01×10^{-9}	6
Rbp1	retinol binding protein 1, cellular [MGI:97876]	2.13	4.20×10^{-11}	9
S100a4	S100 calcium binding protein A4 [MGI:1330282]	2.48	3.40×10^{-9}	3
Sdpr	serum deprivation response [MGI:99513]	2.01	4.37×10^{-11}	1
Serpinf1	serine (or cysteine) peptidase inhibitor, clade F, member 1 [MGI:108080]	2.27	4.38×10^{-7}	11
Spon2	spondin 2, extracellular matrix protein [MGI:1923724]	2.10	1.98×10^{-8}	5
Tcf15	transcription factor 15 [MGI:104664]	2.25	4.44×10^{-14}	2
Tcstv3	2-cell-stage, variable group, member 3 [MGI:2675349]	3.32	1.71×10^{-13}	NT_053651
Tnfrsf22	tumour necrosis factor receptor superfamily, member 22 [MGI:1930270]	2.41	3.83×10^{-12}	7
Tnnt2	troponin T2, cardiac [MGI:104597]	2.31	9.78×10^{-12}	1
Tspo	translocator protein [MGI:88222]	2.08	4.03×10^{-11}	15

Ube1l2	ubiquitin-activating enzyme E1-like 2 [MGI:1913894]	2.60	5.81×10^{-12}	5
Ugt1a6b	UDP glucuronosyltransferase 1 family, polypeptide A7C [MGI:3032636]	2.14	1.20×10^{-7}	1
Vcam1	vascular cell adhesion molecule 1 [MGI:98926]	2.27	3.37×10^{-10}	3

8.2.1.3 Comparison between *Dnmt1*-deficient and AB2.2 wild type cells

The *Dnmt1* gene is involved in DNA MMR (Guo *et al.* 2004). Its deficiency leads to 6TG resistance and microsatellite instability. One expected change from this comparison is the expression change in the *Dnmt1* gene. This gene has one probe (scl36157.39.1_18-S) on the expression array. However, the expression signal of it is 76% of that in the AB2.2 wild type cells. This *Dnmt1*-deficient cell line was generated by deleting the exons 2–4 thourgh gene targeting and a frame shift mutation was generated in this gene. Loss of function of DNMT1 protein is resulted from this frame shift mutation while its gene transcripts can still exist if nonsense-mediated decay does not degrade its transcripts efficiently. The scl36157.39.1_18-S probe locates in the exon 39, the last exon of the *Dnmt1* gene (Figure 8-11), thus its signal may not be affected.

To explore the downstream regulated genes in *Dnmt1*-deficient cells, the expression profile of the ww56+*Hprt* cell line was compared with that of the AB2.2 cell line, from which ww56+*Hprt* was derived. It is assumed that some genes regulated by *Dnmt1* may also be regulated by other MMR components. Thus, the comparison data between the ww56+*Hprt* and AB2.2 cell lines provides a reference for identifying MMR proteins regulated genes in 6TG-resistant mutants.

The expression signals of ten genes in the *Dnmt1*-deficient cell line were two or more times greater than those of the AB2.2 wild type cell line (Table 8-4). The maximum expression fold difference is 3.5 fold for the *Fv1* gene. In addition, the signal of the *neomycin* gene probe is 16 times higher in the *Dnmt1*-deficient cell line ww56+*Hprt*. This is consistent with the selection marker used in the ww56+*Hprt* cell line. Three genes (*2410116G06Rik*, *Chac1* and *Mybl2*) on chromosome 2 and three genes (*Rhox2*, *Scml2* and *Xlr3a*) on chromosome X show altered expression, however, these genes are far from each other. Twenty-two genes are expressed more than two times higher in AB2.2 wild type ES cells than in *Dnmt1*-deficient cells (Table 8-5). The maximum expression fold difference was 9.2 fold for the *Eif2s3y* gene. There is only one probe on the Illumina® array for the *Dnmt1* gene. This probe, scl36157.39.1_18-S, is the same as the genomic sequence between 20657646 and 20657695 bp. It is in the 3' end of the gene on the reverse strand of chromosome 9. The signal of this probe is 1.3 times higher in the AB2.2 than in the ww56+*Hprt* cell line, similar to the *Blm* gene probe in the *Blm*-deficient cells, this modest expression difference may be due to uncompleted NMD.

Table 8-4 **Genes expressed more than two fold higher in the *Dnmt1*-deficient cells compared with the AB2.2 wild type cells**

MGI symbol	Description	Expression fold change (Dnmt1/AB2.2)	Comparison P-value	Chromosome Name	Gene Start (bp)	Gene End (bp)
2410116G06Rik	RIKEN cDNA 2410116G06 gene [MGI:1915486]	2.14	3.66×10^{-11}	2	162912770	162913591
Chac1	ChaC, cation transport regulator-like 1 (E. coli) [MGI:1916315]	2.57	3.70×10^{-9}	2	119176992	119180117
Fv1	Friend virus susceptibility 1 [MGI:95595]	3.51	1.73×10^{-15}	4	147243088	147244467
Gbx2	gastrulation brain homeobox 2 [MGI:95668]	2.14	8.24×10^{-6}	1	91824559	91827751
Mybl2	myeloblastosis oncogene-like 2 [MGI:101785]	2.03	4.74×10^{-10}	2	162880371	162910423
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	2.75	2.47×10^{-10}	7	134355122	134357801
Rhox2	reproductive homeobox 2 [MGI:1922449]	2.06	2.23×10^{-20}	X	34784987	34789685
Scml2	sex comb on midleg-like 2 (*Drosophila*) [MGI:1340042]	2.01	2.02×10^{-12}	X	157555125	157696145
Sgip1	SH3-domain GRB2-like (endophilin) interacting protein 1 [MGI:1920344]	2.07	7.32×10^{-09}	4	102413902	102644468
Xlr3a	X-linked lymphocyte-regulated 3A [MGI:109506]	2.17	3.84×10^{-12}	X	70331634	70342380

These ten genes increased their expression in the *Dnmt1*-deficient cells to 2–3.5 times of those in the parental cell line, AB2.2.

Table 8-5 **Genes expressed more than two fold higher in the AB2.2 wild type cells compared with the *Dnmt1*-deficient cells**

MGI symbol	Description	% wild type signal	Expression fold change (AB2.2/*Dnmt1*)	Comparison P-value	Chromosome Name
Amn	amnionless [MGI:1934943]	46%	2.16	2.77×10^{-6}	12
Car12	carbonic anyhydrase 12 [MGI:1923709]	49%	2.04	1.14×10^{-6}	9
Cdkn1c	cyclin-dependent kinase inhibitor 1C (P57) [MGI:104564]	44%	2.25	5.73×10^{-9}	7
Clic6	chloride intracellular channel 6 [MGI:2146607]	45%	2.22	1.16×10^{-11}	16
Clu	clusterin [MGI:88423]	30%	3.39	3.89×10^{-18}	14
Col4a2	procollagen, type IV, alpha 2 [MGI:88455]	49%	2.03	2.42×10^{-11}	8
Ctsh	cathepsin H [MGI:107285]	40%	2.51	3.63×10^{-17}	9
EG245190	predicted gene, EG245190 [MGI:3646425]	45%	2.20	1.57×10^{-20}	7
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	11%	9.19	1.63×10^{-33}	Y
F3	coagulation factor III [MGI:88381]	43%	2.31	3.76×10^{-8}	3
Fst	follistatin [MGI:95586]	47%	2.14	3.81×10^{-10}	13
Fxyd3	FXYD domain-containing ion transport regulator 3 [MGI:107497]	44%	2.27	2.53×10^{-17}	7
Grb10	growth factor receptor bound protein 10 [MGI:103232]	23%	4.29	2.58×10^{-27}	11
Hkdc1	hexokinase domain containing 1 [MGI:2384910]	40%	2.53	3.33×10^{-23}	10
Hs3st1	heparan sulfate (glucosamine) 3-O-sulfotransferase 1 [MGI:1201606]	44%	2.30	4.22×10^{-10}	5
Igf2	insulin-like growth factor 2 [MGI:96434]	14%	6.96	5.88×10^{-16}	7
Lamc1	laminin, gamma 1 [MGI:99914]	47%	2.13	2.19×10^{-16}	1
Phlda2	pleckstrin homology-like domain, family A, member 2 [MGI:1202307]	38%	2.62	2.43×10^{-12}	7
Podxl	podocalyxin-like [MGI:1351317]	46%	2.16	2.72×10^{-8}	6
Slco2a1	solute carrier organic anion transporter family, member 2a1 [MGI:1346021]	49%	2.04	4.34×10^{-16}	9
Spink3	serine peptidase inhibitor, 0Kazal type 3 [MGI:106202]	39%	2.55	2.69×10^{-5}	18

The above genes have decreased their expression to 10–50% of that in the AB2.2 wild type cells.

8.2.1.4 *Msh2* and *Msh6* homozygous mutants

There are five mutants (B6, D4, D8, F4 and H14), in which both *Msh2* and *Msh6* genes are homozygously deleted. All of the mutant clones with *Msh2* and *Msh6* deletions were analysed by expression arrays to validate the array for deletions of the homozygously deleted genes and to obtain a downstream regulated gene profile of these MMR mutants. The downstream genes are possibly regulated by MMR defects, thus consistent patterns may be observed in other MMR mutants. A gene regulatory profile can also be used as a reference to judge if another mutant has an *Msh2* and *Msh6* deficiency. In the analysis, I expected to see that the *Msh2* and *Msh6* genes were not expressed.

On the Illumina® mouse expression array, *Msh2* has three probes: ri|8030473G08|PX00650P10|AK078, ri|D230024H20|PX00188O22|AK051 and scl50397.16.1_74-S (Figure 8-4). Among them, probe scl50397.16.1_74-S gave the most reliable data since its expression was more than four times higher in the *Blm*-deficient ES cells than in the mutants. The expression signal of probe ri|8030473G08|PX00650P10|AK078 was only 1.38 times higher in the *Blm*-deficient cell line NGG5-3 than in the mutants, which may be regarded as non-complementary cross-hybridization, as this level of expression difference was also observed between wild type cells. The signal of ri|D230024H20|PX00188O22|AK051 was similar to background, thus it was not counted.

Msh6 has two probes on the array: scl50396.10.5_23-S and scl0017688.1_95-S positioed in exons 9 and 10, respectively (Figure 8-5). The expression signal of probe scl0017688.1_95-S was 30–47 times higher in the *Blm*-deficient cells than in the five clones which have *Msh2* and *Msh6* deleted. The signal of the other probe scl50396.10.5_23-S was only 1.4–1.5 times higher in the *Blm*-deficient cells compared with the mutants, indicating that the signal detected by this probe may come from non-complementary hybridization.

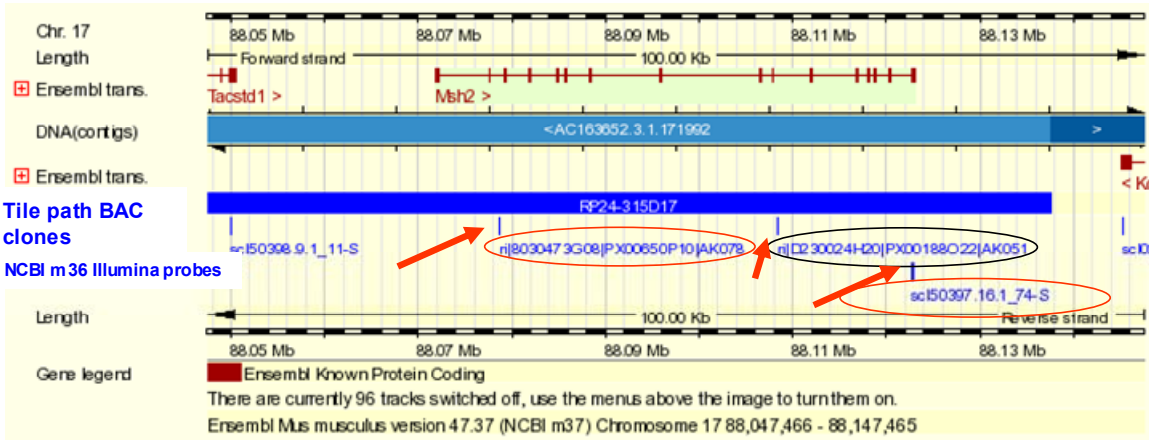
These analysis indicated that *Msh2* and *Msh6* are not expressed as expected. However, this analysis also illustrated that many of the probes on the Illumina array were not that reliable. This can also be observed from the previous data that the probes of *Blm*, *Dnmt1* and *Hprt* genes only show 1–2 times signal difference or no difference between the wild type cells and the gene deficient cells. The genes which show a change in expression are listed in

Table 8-6. The array profiles of these mutants provide an opportunity to determine if there are any other consistent transcriptional differences between the mutants. By comparing all the mutants, 15 genes were found to be down-regulated more than 50% in all of the *Msh2* and *Msh6* homozygous mutants. These genes are distributed throughout in the mouse genome thus are not linked with the homozygous deletion of chromosome 17, though *Msh2* and *Msh6* are both detected by this analysis. Although no evidence has been found for any of these genes to be involved in DNA mismatch repair until now, it can be suggested that these genes are downstream indirectly associated with either *Msh2* function or *Msh6* function. However, other possibilities should be borne in mind, such as that these expression variations could be permanently introduced by gamma irradiation. The genes in the common homozygous deleted region were expected to be found in the Table 8-6 but they were not. These genes have probes on the expression array. Their signals might be excluded from the analysis due to relatively big variations between replicates.

By comparing all the mutants to detect upregulated genes, only one probe scl0012865.1_86-S was detected which showed 2.0–3.5 times increased expression. This probe was not annotated correctly in the mouse genome, but Blast analysis identified this as part of the *Cox7a1* transcript (Figure 8-6). Interestingly, the expression signal of this gene in the *Dnmt1* knockout cells (ww56+*Hprt*) is 1.48 times more than that in the AB2.2 wild type cells. The *Cox7a1* gene is the mouse homologue of the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1*) in humans. COX7A1 is a nuclear-encoded subunit, which may play a regulatory role in the fuction of cytochrome c oxidase (OMIM). The cytochrome c oxidase is the last component of the mitochondrial respiratory chain. This oxidase catalyzes the transfer of electrons from reduced cytochrome c to molecular oxygen. No relation between this gene and MMR was found in literature. It could be related to MMR-deficiency. To elucidate this, more expression analysis on targeted or gene trapped $Msh2^{-/-}$ or $Msh6^{-/-}$ cells with normal BLM function is required.

Based on the above analysis, the expression signals of individual probes were consistent in these clones. This helps to validate the performance of the probes on the other arrays. If the high quality data from these probes can be extended to all of the probes, whole genome expression profiles can be compared confidently.

A



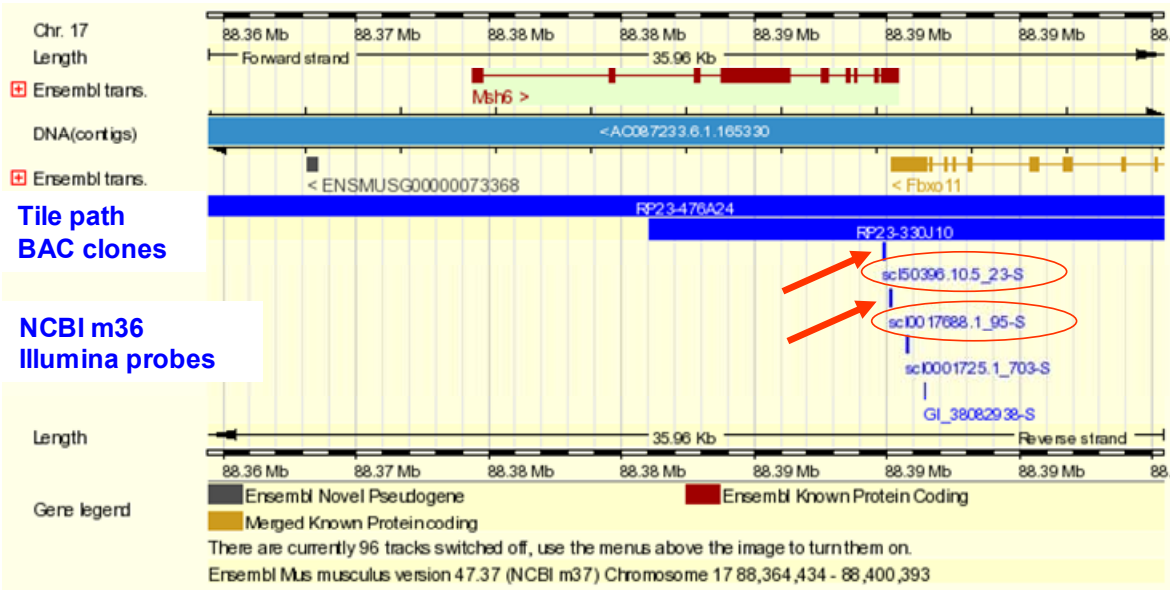
B

Clone	Expression fold change (*Blm*/mutant)		Expression change (mutant/*Blm*)	
	ri|8030473G08|PX00650P10|AK078	scl50397.16.1_74-S	ri|8030473G08|PX00650P10|AK078	scl50397.16.1_74-S
B6	1.38	4.32	72%	23%
D4	1.36	4.63	74%	22%
D8	1.38	4.79	72%	21%
F4	1.39	4.47	72%	22%
H14	1.39	4.56	72%	22%

Figure 8-4 Expression fold change of the *Msh2* gene in homozygously deleted mutants

A. The expression array probes are annotated (small blue bars) on the *Msh2* transcript (NCBI m37, Ensembl release 47). The *Msh2* gene has three probes (circled): ri|8030473G08|PX00650P10|AK078, ri|D230024H20|PX00188O22|AK051 and scl50397.16.1_74-S. Their genomic positions are indicated by red arrows. The genomic DNA of BAC RP24-315D17 is shown in a large blue bar. This BAC was used in the 200 kb resolution tile path array CGH. **B**. The expression level of *Msh2* probes compared to the *Blm*-deficient cells. The expression signal of probe scl50397.16.1_74-S (red circled) in the *Blm*-deficient cell line is around four times more than that in the mutants. The expression signal of probe ri|8030473G08|PX00650P10|AK078 (red circled) in the *Blm*-deficient cells is around 1.38 times more than that in the mutants, suggesting that the signals are probably due to non-complementary cross-hybridization. The signal of probe ri|D230024H20|PX00188O22|AK051 (black circled) was not consistent among its features on the arrays, thus it was not counted.

A



B

Clone	Expression fold change (*Blm*/mutant)		Expression change (mutant/*Blm*)	
	scl50396.10.5_23-S	scl0017688.1_95-S	scl50396.10.5_23-S	scl0017688.1_95-S
B6	1.48	30.7	68%	3%
D1	1.48	42.8	68%	2%
D4	1.49	43.4	67%	2%
D8	1.46	38.9	68%	3%
F4	1.51	36.3	66%	3%
H14	1.49	46.9	67%	2%

Figure 8-5 Expression fold change of the *Msh6* gene in homozygously deleted mutants

A. The expression array probes are annotated (small blue bars) on the *Msh6* transcript (NCBI m37, Ensembl release 47). The *Msh6* gene has two probes (red circled, their positions are indicated by red arrows): scl50396.10.5_23-S (on exon 9) and scl0017688.1_95-S (on exon 10, the last exon). The genomic DNA of BAC RP23-476A24 and RP23-330J10 is shown in large blue bars. These BACs were used in the 200 kb resolution tile path array CGH. **B**. The expression level of *Msh6* probes compared to the *Blm*-deficient cells. The expression signal of probe scl0017688.1_95-S in the *Blm*-deficient cells is around 30–47 times more than that in the mutants. The signal of the other probe, scl50396.10.5_23-S, in the *Blm*-deficient cells is around 1.4–1.5 times more than that in the mutants, indicating that the signal detected by this probe may come from non-complementary hybridization.

Table 8-6 **Down-regulated genes in all five *Msh2* and *Msh6* homozygously deleted mutants compared with the *Blm*-deficient NGG5-3 cell line**

MGI symbol	Description	% $Blm^{-/-}$ signal					Average % $Blm^{-/-}$ signal ± SD	Chromosome Name	Gene Start (bp)	Gene End (bp)
		B6	D4	D8	F4	H14				
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	44%	28%	31%	37%	31%	34±6%	13	113832971	113836596
Ank	progressive ankylosis [MGI:3045421]	41%	35%	42%	38%	39%	39±3%	15	27396432	27524660
Bst2	bone marrow stromal cell antigen 2 [MGI:1916800]	19%	27%	42%	34%	34%	31±9%	8	74058163	74061336
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	38%	40%	46%	35%	41±5%	4	129725129	129776519
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	50%	42%	44%	45%	29%	42±8%	10	76171537	76188789
Cpe	carboxypeptidase E [MGI:101932]	50%	40%	40%	46%	38%	43±5%	8	67071355	67171851
Dcn	decorin [MGI:94872]	33%	22%	27%	25%	28%	27±4%	10	96942245	96980777
Ecm1	Extracellular matrix protein 1 [MGI:103060]	39%	47%	48%	44%	43%	44±4%	3	95538073	95543494
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	12%	12%	12%	11%	13%	12±1%	Y	347055	365035
H2-D1	histocompatibility 2, D region [MGI:95912]	41%	39%	47%	41%	31%	40±6%	17	35400039	35404440
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	43%	36%	33%	40%	32%	37±5%	1	40496621	40522241
Lmna	lamin A [MGI:96794]	35%	35%	45%	38%	33%	37±5%	3	88285071	88307229
Ltbp3	latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	43%	41%	37%	39%	39±3%	19	5740904	5758532
Msh2	mutS homolog 2 (*E. coli*) [MGI:101816]	23%	22%	21%	22%	22%	22±1%	17	88071880	88123052
Msh6	mutS homolog 6 (*E. coli*) [MGI:1343961]	3%	2%	3%	3%	2%	3±0%	17	88374424	88390403
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	48%	41%	48%	41%	42%	44±4%	5	99360287	99466098
Tnfrsf22	tumour necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	35%	37%	30%	37%	35±3%	7	150820717	150835557

A

	B6	D4	D8	F4	H14
Signal in mutant/ signal in *Blm*$^{-/-}$ cells	3.05	3.46	2.02	2.20	2.33
Adjusted P-value	2.77×10^{-20}	1.97×10^{-22}	$3.07\times^{-10}$	4.98×10^{-14}	2.10×10^{-14}

B

scl0012865.1_86-S probe sequence

GGCAGAGAAGCAGAAGCTCTTCCAGGCCGACAATGACCTCCCAGTACACT

1 23 26 50

C

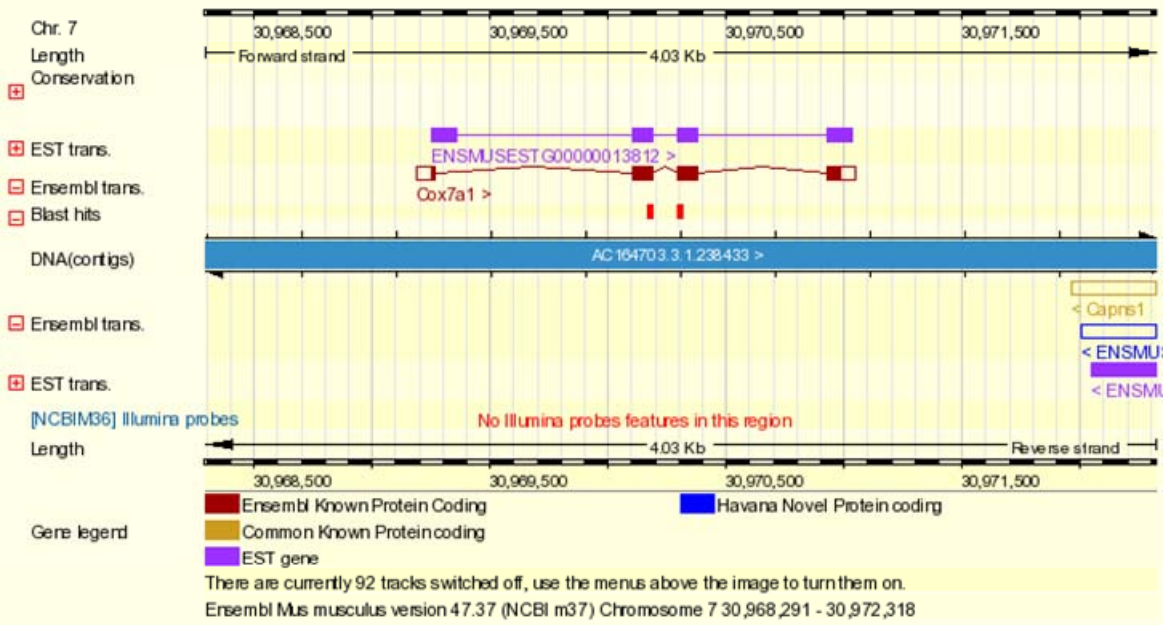


Figure 8-6 *Cox7a1* is up-regulated in all five *Msh2* and *Msh6* homozygously deleted mutants compared with the *Blm*-deficient NGG5-3 cell line

A. Comparison of the signals of probe scl0012865.1_86-S, the only probe for gene *Cox7a1*. The signal of this probe in any mutant is compared with its signal in the NGG5-3 *Blm*-deficient cell line. Fold changes are indicated. The adjusted P-value is used to indicate the possibility that this fold change is incorrect. **B**. Sequence of probe scl0012865.1_86-S. The red font will be explained in C. **C**. Probe scl0012865.1_86-S annotated in the mouse genome (NCBI m37). An Ensembl snapshot is shown here. The Illumina DAS source has not correctly annotated this probe in the genome, therefore Blast analysis was conducted. Two fragments of the probe sequence are 100% identical to the gene's sequence in exons 2 and 3. These two fragments are shown in red boxes. The first fragment contains nucleotide 1–26 of the probe. The second fragment contains nucleotide 23–50 of the probe. The red characters in B. indicate the overlapped nucleotides between two fragments.

8.2.1.5 *Msh6* homozygous deletion in mutant D1

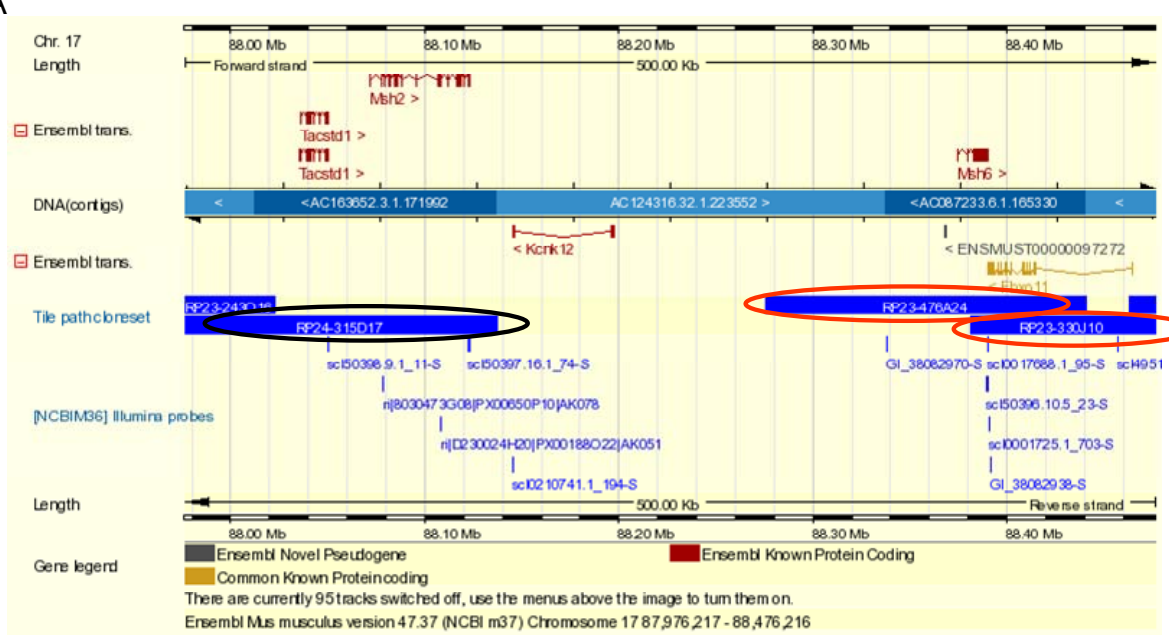
One mutant clone, D1, has lost both alleles of the *Msh6* gene. As *Msh2* is not affected in this clone, this clone was used to investigate downstream gene expression differences caused by *Msh6* loss and the differences compared with loss of both *Msh2* and *Msh6*. The D1 clone is strongly resistant to 6TG and has a 160 kb homozygous deletion from 87.8 Mb to 87.9 Mb on chromosome 17 with a Log_2(ratio) of -1.30 (Figure 8-7). By comparing probe signals on the expression array between the *Blm*-deficient cell line NGG5-3 and D1, a series of expression variations were identified. Most of the genes identified by comparing clones which had a deletion of both *Msh2* and *Msh6* (Table 8-6) were found to vary in this clone, too. The genes down-regulated by more than 50% in both D1 and the five *Msh2* and *Msh6* homozygous deletion mutants listed in Table 8-7.

Compared with the previous analysis (Table 8-6), three genes, *Bst2*, *Eif2s3y* and *Msh2,* were not altered in the D1 clone. The *Eif2s3y* gene is Y-linked, considering that the five mutants with *Msh2* and *Msh6* homozygous deletions had all lost the Y chromosome while the D1 mutant had not, it is reasonable that this gene is not altered. *Msh2* is not deleted in D1, thus it is also expected that this gene was found to be unaltered, too. The expression of *Bst2* (bone marrow stromal cell antigen 2) appears to be differentially affected by the activity of *Msh2*. Bone marrow stromal cells can regulate B-cell development through their surface molecules (Ishikawa *et al.* 1995).

The analysis of genes with increased expression in the D1 clone also identified the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1)* gene which doubled (2.08 times) its expression signal in the D1 mutant compared with the *Blm*-deficient cell line NGG5-3. This confirms that *Cox7a1* is transcriptionally upregulated in MMR-deficient ES cells. Similarly, the expression signal of this gene in the *Dnmt1* knockout cells (ww56+*Hprt*) is 1.48 times more than that in the AB2.2 wild type cells.

When comparing the down-regulated genes in the *Msh6* mutant D1 and the five mutant clones with the *Msh2* and *Msh6* homozygous deletions with the down-regulated genes in the *Dnmt1* knockout control cell line ww56+Hprt, no gene appears in both comparisons. This indicates that the gene expression changes in *Dnmt1* and *Msh2/Msh6* mutants are different.

A



B

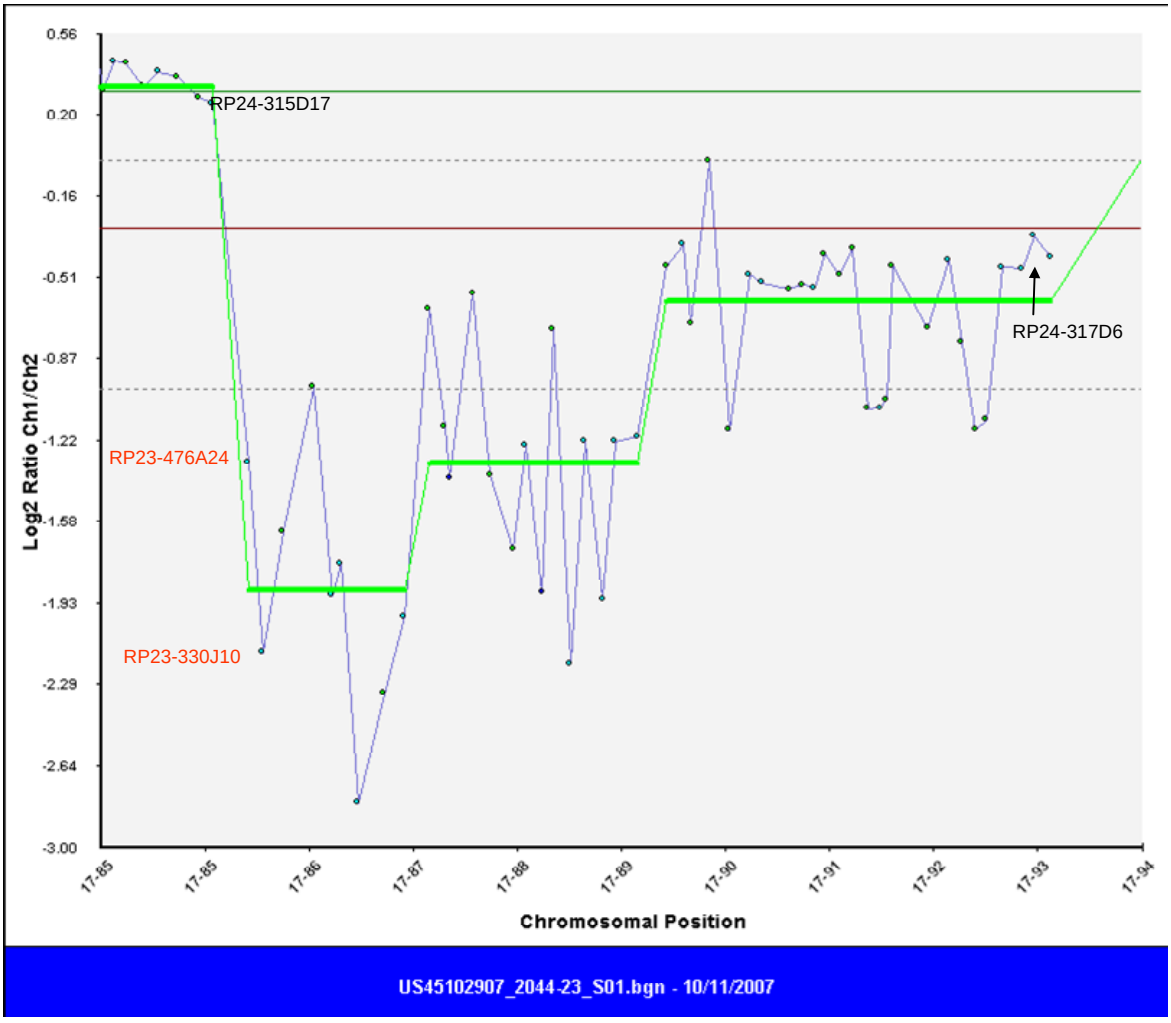


Figure 8-7 Analysis of the mutant D1, a clone with an *Msh6* homozygous deletion

Figure 8-7 Analysis of the mutant D1, a clone with *Msh6* homozygous deletion

A. An Ensembl snapshot of a 500 kb region showing the *Msh2* and *Msh6* genes. Ensembl annotated transcripts are shown in dark red. Tile path BACs for array CGH are shown as blue boxes. The black circled BAC, RP24-315D17, covers the *Msh2* gene. Two red circled BACs, RP23-476A24 and RP23-330J10, cover the *Msh6* gene. There is a gap in the tile path between RP24-315D17 and RP23-476A24. Gene-specific probes on the Illumina® array are shown as small blue bars. **B**. Zoomed-in array CGH profile of distal chromosome 17. BAC RP24-315D17 does not exhibit any copy number change compared with the *Blm*-deficient cell line NGG5-3. However, BAC RP23-476A24 and RP23-330J10 (red font) are within a homozygous deletion. BAC RP24-316D6 is the last tile path array BAC on chromosome 17. The Log_2(ratio) does not confirm the profile expected for a homozygous deletion. This clone is likely to have a complex deletion from BAC RP23-476A24 to the end of chromosome 17. One interpretation of this pattern is that one allele has deleted 160 kb, including *Msh6,* while the other has suffered a much larger deletion. Thus, only the 160 kb region shows a homozygous deletion.

Table 8-7 Genes down-regulated in both D1 (*Msh6*$^{-/-}$ deletion) and the five *Msh2* and *Msh6* deficient mutants compared to *Blm*-deficient ES cells

MGI symbol	Description	% *Blm*$^{-/-}$ signal						Average % *Blm*-/- signal ± SD	Chromosome Name
		B6	D1	D4	D8	F4	H14		
2310016C16Rik	RIKEN cDNA 2310016C16 gene [MGI:1916840]	44%	27%	28%	31%	37%	31%	33±6%	13
Ank	progressive ankylosis [MGI:3045421]	41%	37%	35%	42%	38%	39%	39±3%	15
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	41%	38%	40%	46%	35%	41±5%	4
Col6a1	procollagen, type VI, alpha 1 [MGI:88459]	50%	39%	42%	44%	45%	29%	42±7%	10
Cpe	carboxypeptidase E [MGI:101932]	50%	36%	40%	40%	46%	38%	42±5%	8
Dcn	decorin [MGI:94872]	33%	20%	22%	27%	25%	28%	26±5%	10
Ecm1	extracellular matrix protein 1 [MGI:103060]	39%	45%	47%	48%	44%	43%	44±3%	3
H2-D1	histocompatibility 2, D region [MGI:95912]	41%	40%	39%	47%	41%	31%	40±5%	17
Il1rl1	interleukin 1 receptor-like 1 [MGI:98427]	43%	32%	36%	33%	40%	32%	36±5%	1
Lmna	Lamin A [MGI:96794]	35%	36%	35%	45%	38%	33%	37±4%	3
Ltbp3	Latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	40%	43%	41%	37%	39%	39±3%	19
Msh6	mutS homolog 6 (*E. coli*) [MGI:1343961]	3%	2%	2%	3%	3%	2%	3±0%	17
Prkg2	protein kinase, cGMP-dependent, type II [MGI:108173]	48%	43%	41%	48%	41%	42%	44±3%	5
Tnfrsf22	Tumor necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	34%	35%	37%	30%	37%	34±3%	7

The above 14 genes are down-regulated more than 50% in the mutants of B6, D4, D8, F4, H14 and D1, comparing to the expression levels of all genes in the *Blm*-deficient control ES cell line NGG5-3. The individual and the average expression levels of the mutants compared to the *Blm*-deficient cells are shown with standard deviation (SD).

8.2.1.6 Expression status of *Msh2* and *Msh6* in heterozygous deletions

There were nine mutants (C1, C3, C5, D6, F9, F16, G8, G9 and H13) which had heterozygous deletions covering both *Msh2* and *Msh6* genes (see array CGH analysis). Six of them (C1, C5, F16, G8, G9 and H13) are strongly resistant to 6TG. Three (C3, D6 and F9) are weakly resistant to 6TG. F9 and F16 are clonal thus expression array analysis was performed on one of them, F9. As heterozygous mutation of *Msh2* does not lead to resistance to 6TG (Abuin *et al.* 2000; de Wind *et al.* 1995) and one copy of *Msh2* and *Msh6* should be able to produce functional proteins, either *Msh2* or *Msh6* are expected to have a mutation in the other allele, which may result in a loss of gene expression caused by changes in the reading frame and generation of premature termination codons (PTCs). Using these PTCs as signals, the nonsense-mediated decay (NMD) mechanism can selectively degrade mutant mRNA.

Expression array analysis identified a number of mutants which had the same magnitude of signal reduction for *Msh2* or *Msh6* as in the homozygous deletions (Table 8-8). The *Msh2* probe scl50397.16.1_74-S and the *Msh6* probe scl0017688.1_95-S are reliable probes as these two probes show 2.5– 25% expression signals of that in the *Blm*-deficient control cells, while the other probes of these two genes exhibit 65–75% expression signals or inconsistent signals (probe ri|D230024H20|PX00188O22|AK051) in the homozygous deletions compared with the *Blm*-deficient control cells (Figure 8-4 and Figure 8-5). The expression level of the *Msh2* probe scl50397.16.1_74-S in the mutant C5 and H13 is 25–26% of that in the *Blm*-deficient control cells. This expression level is similar to those in the homozygous deletions (21–22%). Additionally, the expression level of the *Msh6* probe scl0017688.1_95-S in the mutant G8 is 2% of that in the *Blm*-deficient control cells, the same signal level of that in the homozygously deleted mutants. These data expain why the mutant clones C5, H13 and G8 are strongly resistant to 6TG. These results indicate that these clones carry mutations in the other allele of *Msh2* or *Msh6*. However, one clone (G9) shows a reduced magnitude of reduction of the *Msh6* signal (nine fold) compared to that seen in homozygous mutations and no reduction in the *Msh2* signal. This clone is believed to be an *Msh6* mutant, too. This explains the strong resistance to 6TG of the clone G9. In a summary, six homozygous mutants for the *Msh2* and *Msh6* genes have lost the expression signals of the both genes. Two heterozygous mutants (C5 and H13) lost the transcripts of *Msh2* and two heterozygous mutants (G8 and G9) lost or significantly reduced the expression of *Msh6*. These four mutants may contain compound mutations of deletion/small mutation in either *Msh2* or *Msh6*, thus are strongly resistant to 6TG. While the cause of the strong 6TG resistance of the mutant C1 remains unclear. But the weak 6TG resistance of

Table 8-8 Expression signal decrease of *Msh2* and *Msh6* probes in homozygously and heterozygously deleted mutants

A

	Clone	6TG^R	*Msh2* probe scl50397.16.1_74-S	
			Expression fold change (NGG5-3/mutant)	% *Blm*$^{-/-}$ signal
Homozygous deletion	B6	Strong	4.32	23%
	D4	Strong	4.63	22%
	D8	Strong	4.79	21%
	F4	Strong	4.47	22%
	H14	Strong	4.56	22%
Heterozygous deletion	C1	Strong	1.97	51%
	C3	Weak	2.01	50%
	C5	Strong	3.81	26%
	D6	Weak	2.25	44%
	F9	Weak	NA	NA
	G8	Strong	1.78	56%
	G9	Strong	NCD	NCD
	H13	Strong	4.06	25%

B

	Clone	6TG^R	*Msh6* probe scl0017688.1_95-S	
			Expression fold change (NGG5-3/mutant)	% *Blm*$^{-/-}$ signal
Homozygous deletion	B6	Strong	30.7	3%
	D1	Strong	42.8	2%
	D4	Strong	43.4	2%
	D8	Strong	38.9	3%
	F4	Strong	36.3	3%
	H14	Strong	46.9	2%
Heterozygous deletion	C1	Strong	NCD	NCD
	C3	Weak	NCD	NCD
	C5	Strong	1.67	60%
	D6	Weak	1.38	72%
	F9	Weak	NA	NA
	G8	Strong	49.5	2%
	G9	Strong	9.32	11%
	H13	Strong	1.5	67%

A. Signal decrease of an *Msh2* probe in mutants with *Msh2* and *Msh6* homozygous and heterozygous deletions. The phenotype of 6TG resistance is shown. NA: data not available. NCD: no convincing data (comparison P-value is high). **B**. Signal decrease of an *Msh6* probe in mutants with *Msh2* and *Msh6* heterozygous deletions.

the mutant clones C3, D6 and F9 may result from the hypomorphic mutation of the *Msh2* and *Msh6* genes.

By comparing all the mutants, five genes were found to be down-regulated to 20–50% in all of the *Msh2* and/or *Msh6* homozygous and compound mutants (Table 8-9). These genes are *Col16a1* (procollagen, type XVI, alpha 1), *Dcn* (decorin), *Lmna* (Lamin A), *Ltbp3* (Latent transforming growth factor beta binding protein 3) and *Tnfrsf22* (Tumor necrosis factor receptor superfamily, member 22). A common feature of these genes is that their products may locate on or close to cell membrane and nuclear membrane (Lamin A). The down-regulation of these genes may be secondary effects caused by MMR defects.

It is known that the cytochrome c oxidase, subunit VIIa 1 (*Cox7a1)* gene probe scl0012865.1_86-S was detected which showed 2.0–3.5 times increased expression in the six *Msh2/Msh6* homozygous mutants (B6, D1, D4, D8, F4 and H14). When comparing the compound heterozygous mutants (C5, H13, G8 and G9) which lost the function of *Msh2* or *Msh6* to detect upregulated genes, the *Cox7a1* gene is one of the ten genes up-regulated in the C5, G8 and G9 mutants. The expression levels of this gene in C5, G8 and G9 are 2.3–3.7 times of that in the *Blm*-deficient control cells. While the probe performance of *Cox7a1* was not so consistent in the mutant H13 because the comparison P-value of this probe is higher than the lowest five thousand probes; therefore, this probe was not included in these comparisons. But it is possible that *Cox7a1* is also over expressed in the H13 clone. In a summary, the *Cox7a1* gene showed 2.0–3.7 times increased expression in the nine *Msh2/Msh6* homozygous or compound mutants (B6, C5, D1, D4, D8, F4, G8, G9 and H14).

Table 8-9 Genes down-regulated in ten *Msh2/Msh6* deficient mutants compared to *Blm*-deficient ES cells

MGI symbol	Description	% *Blm*$^{-/-}$ signal										Average % *Blm*-/- signal ± SD	Chromosome Name
		B6	C5	D1	D4	D8	F4	G8	G9	H13	H14		
Col16a1	procollagen, type XVI, alpha 1 [MGI:1095396]	47%	47%	41%	38%	40%	46%	30%	38%	46%	35%	41±6%	4
Dcn	decorin [MGI:94872]	33%	35%	20%	22%	27%	25%	21%	21%	43%	28%	27±7%	10
Lmna	Lamin A [MGI:96794]	35%	37%	36%	35%	45%	38%	26%	27%	33%	33%	34±5%	3
Ltbp3	Latent transforming growth factor beta binding protein 3 [MGI:1101355]	35%	46%	40%	43%	41%	37%	31%	38%	46%	39%	40±5%	19
Tnfrsf22	Tumor necrosis factor receptor superfamily, member 22 [MGI:1930270]	35%	36%	34%	35%	37%	30%	24%	32%	49%	37%	35±6%	7

The above five genes are down-regulated to 20–50% in the mutants of B6, C5, D1, D4, D8, F4, G8, G9, H13, and H14, comparing to the expression levels in the *Blm*-deficient control ES cell line NGG5-3. The individual and the average expression levels of the mutants compared to the *Blm*-deficient cells are shown with standard deviation (SD).

8.2.2 Combined analysis of array CGH and expression array results

8.2.2.1 Positive relationship between copy number and expression variation

To test the usefulness of a combined analysis of array CGH and expression array data, a small-scale analysis was conducted to examine the relationship between a given gene's transcription level and its copy number. The samples in this small-scale experiment included mutant clones B2 (weak $6TG^R$), E1 (strong $6TG^R$), E2 (weak $6TG^R$), F1 (strong $6TG^R$) and G4 (weak $6TG^R$). These clones were selected because these mutant clones had all deleted one allele of a 12 Mb genomic region on chromosome 14 (107.9 Mb–119.6 Mb, Figure 7-25). It was expected that genes homozygously deleted in the mutants should show reduced expression while amplified genes should express at higher levels. The array signals of genes in the amplified and deleted regions in the five mutants were compared with array signals of *Blm*-deficient cells. The Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) data from the expression array is used to indicate the expression of a given gene compared with the *Blm*-deficient control cell line NGG5-3 and the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) data from the array CGH is used to indicate the copy number change of a given gene. By plotting these two values in a figure, the relationship between copy number and expression level can be discovered.

From this analysis (Figure 8-8), it was found that genes in amplified regions express more mRNA than in the control cell line; and most of the genes express less mRNA than the control cell line does when the region is deleted. Thus, there is a positive relationship between copy number and expression level in most of the cases. It is noticeable that one gene is homozygously deleted in four out of five mutant clones (except E1) with the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) being less than minus one. This gene is *Eif2s3y*, eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked. This kind of gene could be the best MMR candidate resulting from the screen. However, this gene is on the Y chromosome and the Y chromosome is lost in all four of these clones. Clone E1 has an intact Y chromosome thus does not show this phenomenon.

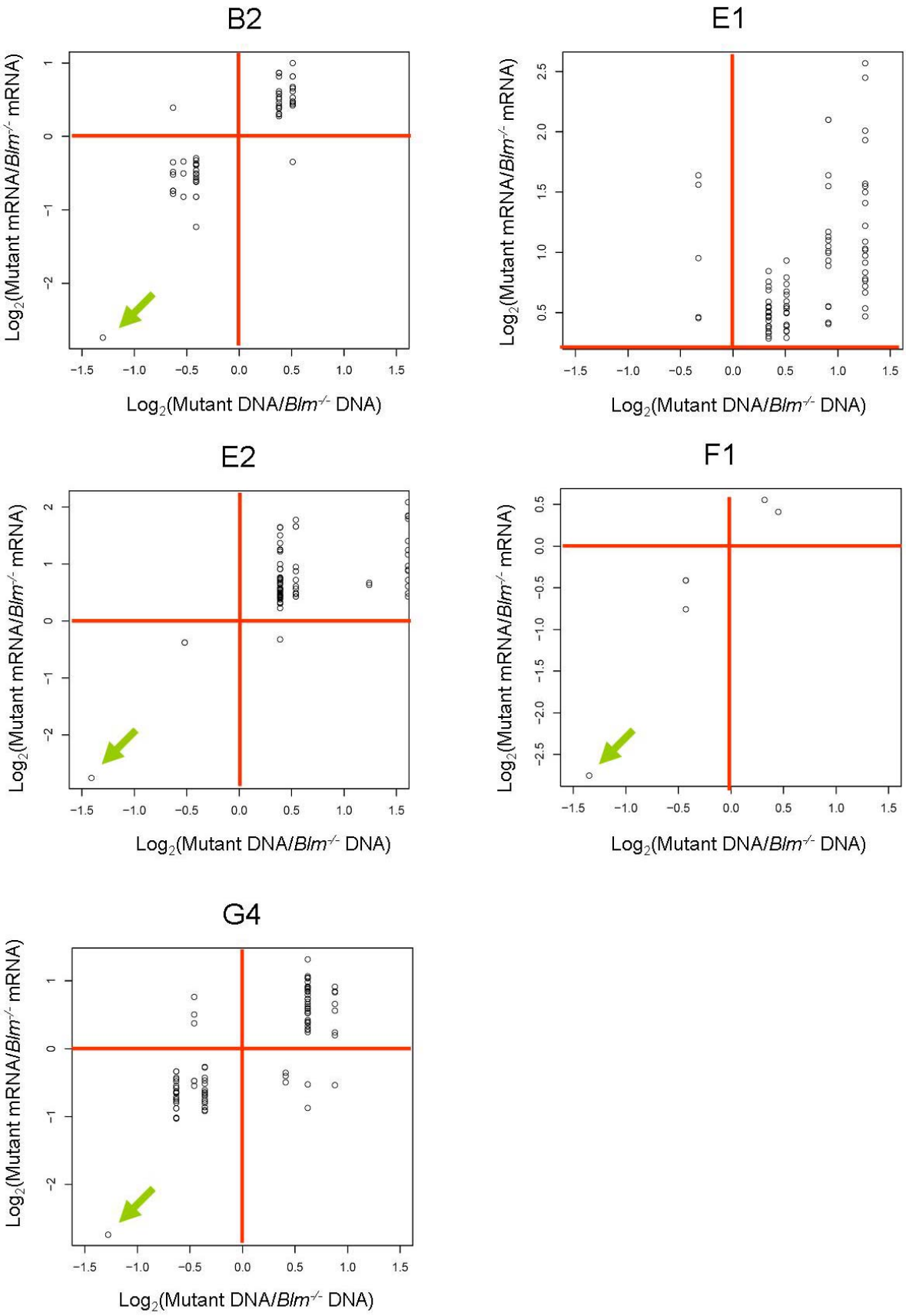


Figure 8-8 **Positive relationship between copy number and expression level**

Figure 8-8 Positive relationship between copy number and expression level

In mutants B2, E1, E2, F1 and G4, data of expression change of amplified and deleted genes is plotted against data of their copy number changes. Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is used to indicate the expression variations of mutant genes compared with the expression of their counterparts in the *Blm*-deficient control cell line NGG5-3. The Y axis measures Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA). A given mutant gene is overexpressed when the Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is greater than zero; and it is down-regulated when the Log_2(Mutant mRNA/*Blm*$^{-/-}$ mRNA) is less than zero. The X axis measures Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA). These data are from the array CGH analysis. Similarly, a given gene is amplified when the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) is greater than zero; and it is deleted when the Log_2(Mutant DNA/*Blm*$^{-/-}$ DNA) is less than zero. The red lines indicate zero on both the X and Y axes. Most of the genes fall into either the (Plus, Plus) region or the (Minus, Minus) region, indicating that in general, when a gene gains more copies in the genome, it expresses more; when a gene loses one or both copies of its sequence, its transcription is also decreased. The green arrows indicate that in four out of five mutant clones (except E1), gene *Eif2s3y* is deleted and its expression is shut down.

8.2.2.2 Genes affected by MMR defects

Using a similar strategy to the previous analysis section, the expression level of genes in deleted regions can be examined within thirty-four $6TG^R$ mutant clones, which were analysed using array CGH. It was assumed that if a gene was homozygously deleted and its transcription was lost in different mutant clones, it would be likely that the deficiency of this gene would be associated with 6TG resistance and MMR. However, a loss-of-function mutation may be caused by a large heterozygous deletion combined with a small mutation at the relevant locus. This type of combination of mutations cannot be shown as a homozygous deletion. Thus, in this analysis, any gene was selected to examine the decrease of its expression if they were homozygously deleted in one mutant clone. The transcription levels of these genes in the mutants were separately compared with the NGG5-3 *Blm*-deficient cells.

By combining data from both array CGH analysis and expression array analysis, 34 genes were found to be homozygously deleted in at least one mutant and their transcripts can be observed in the *Blm*-deficient ES cells, but the expression level of these genes in the $6TG^R$ mutants was less than 50% of that in the *Blm*-deficient cells (Table 8-10). The genes in this list may be indirectly affected by MMR defects or 6TG metabolism pathway. The frequency of homozygous deletion of genes and the number of $6TG^R$ mutants in which each gene is down-regulated are shown in the Table 8-10. The *Msh2* and *Msh6* genes can be found in Table 8-10, both being down-regulated in nine $6TG^R$ mutants. Some genes on the list are on chromosome 17, which could be associated with the *Msh2* and *Msh6* homozygous deletions. Examples are the *Cox7a2l*, *Fbxo11*, *Ppm1b* and *Socs5* genes, which are down-regulated in ten to twenty-one $6TG^R$ mutants and are within a short distance of the *Msh2* and *Msh6* genes. All genes down-regulated on chromosome 17 are within a 32 Mb region, ranging from 56.4 Mb (*Uhrf1*) to 88.4 Mb (*Fbxo11*). These genes are down-regulated probably because they were deleted together with the *Msh2* and *Msh6* genes. However, this does not rule out the possibility that these genes and other genes on chromosome 17 may be involved in MMR, such as genes *Bcat1* (Ch6, reduced expression in 10 clones), *Eps8* (Ch6, reduced expression in 6 clones), *Eif5* (Ch12, reduced expression in 7 clones), *Ank* (Ch15, reduced expression in 22 clones) and *Dap* (Ch15, reduced expression in 8 clones) are not on chromosome 17 and were down-regulated in between 6 (*Eps8*) and 22 (*Ank*) mutant clones. This provides some evidence that these genes affected by DNA mismatch repair defects. The *Zfp459* (Ch13, reduced expression in 8 clones) gene seems to be homozygously deleted for seven times. However, this gene is located on a BAC,

which should belong to chromosome Y instead of chromosome 13. Its decreased expression may be associated with loss of chromosome Y.

In addition, over-expression of the human homologue *EPS8* was observed in pancreatic cancer, that can facilitate colon cancer cell growth (Maa *et al.* 2007; Welsch *et al.* 2007). The EPS8 protein may function through the expression of the focal adhesion kinase (*Fak*) gene (Maa *et al.* 2007). Among these six genes, only *Dap* has been shown to be associated with MMR in the literature. The promoter region of death-associated protein (DAP) kinase gene (*DAPK-1*) was observed to be hypermethylated in early-onset sporadic gastric carcinoma and B-cell lymphoma cell lines (Katzenellenbogen *et al.* 1999; Kim *et al.* 2005; Kissil *et al.* 1997). The expressional down-regulation of *DAPK-1* due to hypermethylation indicates that this gene might play a role in tumour suppression.

Because the NGG5-3 *Blm*-deficient cells have been modified by multiple steps of gene targeting, and *Blm* deficiency causes genome instability to the cells, these cells may contain unknown mutations. Therefore, the above analysis was also conducted on wild type cells (both AB1 and AB2.2 cell lines, two replicates each). The genes, which were homozygously deleted in at least one $6TG^R$ mutant, were compared at a transcriptional level between any single mutant and the wild type cells. The number of mutants in which gene expression is down-regulated compared to the wild type cells is shown. These genes (42) are listed in Table 8-11. Several genes were identified from this comparison which had not been identified in the comparison with the *Blm*-deficient line NGG5-3. For example, the *Amn* and *1810015C04Rik* genes show reduced transcription in 26 (*Amn*) or 21 (*1810015C04Rik*) of thirty-one $6TG^R$ mutants analysed. The reason for such a high frequency and difference between the *Blm*-deficient cells and the wild type cells is due to their lower expression in the *Blm*-deficient reference cells and the *Blm*-deficient mutant cells. Then expression differences may be spontaneously introduced by *Blm* deficiency during multiple targeting procedures.

Twenty-six of the fourty-two genes in Table 8-11 overlapped with the genes in Table 8-10. The 26 overlapping genes are shown in Table 8-12, in which genes were homozygously deleted in at least one $6TG^R$ mutant and the gene expression in some mutants was reduced to less than 50% of that in both the *Blm*-deficient cells and the wild type cells. Again, genes *Msh2*, *Msh6* and other genes on chromosome 17 are in Table 8-12. Five genes on chromosome 15 and sixteen genes on chromosome 17 are found in these 26 genes and are illustrated in Figure 8-9. *Bcat1, Eps8* and *Dap* are not in Table 8-11 and Table 8-12. An

explanation may be that the expression level of these three genes has a large range of variation. The *Eif5* and *Ank* genes are in Table 8-12 with a relatively higher frequency of being down-regulated in the $6TG^R$ mutants. This indicates that they are more likely to be involved in or affected by MMR deficiency.

Table 8-10 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with the *Blm*-deficient cells

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency in deletion	Frequency of down-regulation
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	22
Socs5	suppressor of cytokine signaling 5 [MGI:2385459]	17	87507019	87536923	5	21
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	10
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	10
Ppm1b	protein phosphatase 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	10
Bcat1	branched chain aminotransferase 1, cytosolic [MGI:104861]	6	144947678	145024704	1	10
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	9
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	9
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	8
Dap	death-associated protein [MGI:1918190]	15	31154147	31204095	1	8
Eif5	eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	7
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	6
Eps8	epidermal growth factor receptor pathway substrate 8 [MGI:104684]	6	137425766	137597806	1	6
Ppp4r1	protein phosphatase 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	4
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	4
Bmp4	bone morphogenetic protein 4 [MGI:88180]	14	47003195	47010274	1	3
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	3
Gemin6	gem (nuclear organelle) associated protein 6 [MGI:1914492]	17	80623791	80627810	2	2
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	2
Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	2
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	2
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	2
Ttc27	tetratricopeptide repeat domain 27 [MGI:1921446]	17	75117090	75262910	1	2
Clpp	caseinolytic peptidase, ATP-dependent, proteolytic subunit homolog (E. coli) [MGI:1858213]	17	57129725	57135610	3	1

Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	1
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	1
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1
Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	1

Table 8-11 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with wild type cells (AB1 and AB2.2)

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency of deletion vs *Blm*$^{-/-}$ cells	Frequency of down-regulation vs wild type cells
Amn	amnionless [MGI:1934943]	12	112509322	112514637	0	26
1810015C04Rik	RIKEN cDNA 1810015C04 gene [MGI:1913520]	15	25773019	25903442	0	21
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	11
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	11
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	10
Ppp4r1	protein phosphatase 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	10
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	9
Ppm1b	protein phosphatase 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	8
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	8
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	7
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	7
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	6
Crip2	cysteine rich protein 2 [MGI:1915587]	12	114378742	114383715	0	5
Rcor1	REST corepressor 1 [MGI:106340]	12	112277825	112351595	0	5
Wars	tryptophanyl-tRNA synthetase [MGI:104630]	12	110098244	110132355	0	5
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	5
Eif5	eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	4
Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	4
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	4
1700019H03Rik	RIKEN cDNA 1700019H03 gene [MGI:1919107]	2	180459968	180476985	0	3
AW555464	expressed sequence AW555464 [MGI:2145043]	12	113960445	113984802	0	3
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	3
Ppp2r5c	protein phosphatase 2, regulatory subunit B (B56), gamma isoform [MGI:1349475]	12	111685518	111821271	0	3
Eml4	echinoderm microtubule associated protein like 4 [MGI:1926048]	17	83750322	83879697	0	3
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	3
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	3
Rfx2	Regulatory factor X, 2 (influences HLA class II expression) [MGI:106583]	17	56915323	56970436	0	3
Zfyve21	zinc finger, FYVE domain containing 21 [MGI:1915770]	12	113052381	113066596	0	2
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	2
M6prbp1	mannose-6-phosphate receptor binding protein 1 [MGI:1914155]	17	56416899	56429934	0	2

Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2
Ythdf1	YTH domain family 1 [MGI:1917431]	2	180639084	180655641	0	1
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	1
Bmpr1a	bone morphogenetic protein receptor, type 1A [MGI:1338938]	14	35224255	35315732	0	1
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1
Hrsp12	heat-responsive protein 12 [MGI:1095401]	15	34413776	34424935	0	1
Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	1
Sfrs7	splicing factor, arginine/serine-rich 7 [MGI:1926232]	17	80599426	80606643	0	1

Table 8-12 Genes which are deleted in one mutant and whose transcripts are down-regulated in mutants compared with wild type cells and the *Blm*-deficient cells

MGI symbol	Description	Chromosome Name	Gene Start (bp)	Gene End (bp)	Frequency of deletions vs *Blm*$^{-/-}$ cells	Frequency of down-regulation vs wild type cells	Frequency of down-regulation vs Blm$^{-/-}$ cells
Ank	progressive ankylosis [MGI:3045421]	15	27396432	27524660	1	2	22
Cox7a2l	cytochrome c oxidase subunit VIIa polypeptide 2-like [MGI:106015]	17	83901270	83913670	2	4	10
Fbxo11	F-box protein 11 [MGI:2147134]	17	88390199	88464625	12	7	10
Ppm1b	protein phosphatise 1B, magnesium dependent, beta isoform [MGI:101841]	17	85357341	85423074	2	8	10
Msh2	mutS homolog 2 (E. coli) [MGI:101816]	17	88071880	88123052	6	11	9
Msh6	mutS homolog 6 (E. coli) [MGI:1343961]	17	88374424	88390403	12	10	9
Zfp459	zinc finger protein 459 [MGI:3040701]	13	67506658	67514893	7	7	8
Eif5	Eukaryotic translation initiation factor 5 [MGI:95309]	12	112776312	112784822	1	4	7
Fez2	fasciculation and elongation protein zeta 2 (zygin II) [MGI:2675856]	17	78777225	78817471	2	11	6
Ppp4r1	protein phosphatise 4, regulatory subunit 1 [MGI:1917601]	17	66132738	66191187	2	10	4
Vapa	vesicle-associated membrane protein, associated protein A [MGI:1353561]	17	65929396	65962895	1	8	4
1110012J17Rik	RIKEN cDNA 1110012J17 gene [MGI:1915867]	17	66686326	66799087	1	5	3
Ldhb	lactate dehydrogenase B [MGI:96763]	6	142438772	142456477	1	1	2
Hsp90aa1	heat shock protein 90kDa alpha (cytosolic), class A member 1 [MGI:96250]	12	111929238	111934605	1	3	2
Grhl2	grainyhead-like 2 (*Drosophila*) [MGI:2182543]	15	37163002	37293317	1	9	2
Map4k3	mitogen-activated protein kinase kinase kinase kinase 3 [MGI:2154405]	17	80979852	81127433	2	1	2
Rab31	RAB31, member RAS oncogene family [MGI:1914603]	17	66001070	66122092	2	2	2
Uhrf1	ubiquitin-like, containing PHD and RING finger domains, 1 [MGI:1338889]	17	56442823	56462909	2	2	2
Entpd4	ectonucleoside triphosphate diphosphohydrolase 4 [MGI:1914714]	14	69955231	69985327	1	1	1
Atp6v1c1	ATPase, H+ transporting, lysosomal V1 subunit C1 [MGI:1913585]	15	38591699	38622197	1	1	1

Mtdh	Metadherin [MGI:1914404]	15	34012474	34072140	1	1	1
Pabpc1	poly A binding protein, cytoplasmic 1 [MGI:1349722]	15	36525416	36538728	1	6	1
Arl6ip2	ADP-ribosylation factor-like 6 interacting protein 2 [MGI:1929492]	17	80247732	80295463	2	4	1
Epb4.1l3	erythrocyte protein band 4.1-like 3 [MGI:103008]	17	69506150	69639325	2	1	1
Mta3	metastasis associated 3 [MGI:2151172]	17	84105503	84223055	2	3	1
Mylc2b	myosin light chain, regulatory B [MGI:107494]	17	71323282	71339897	2	3	1

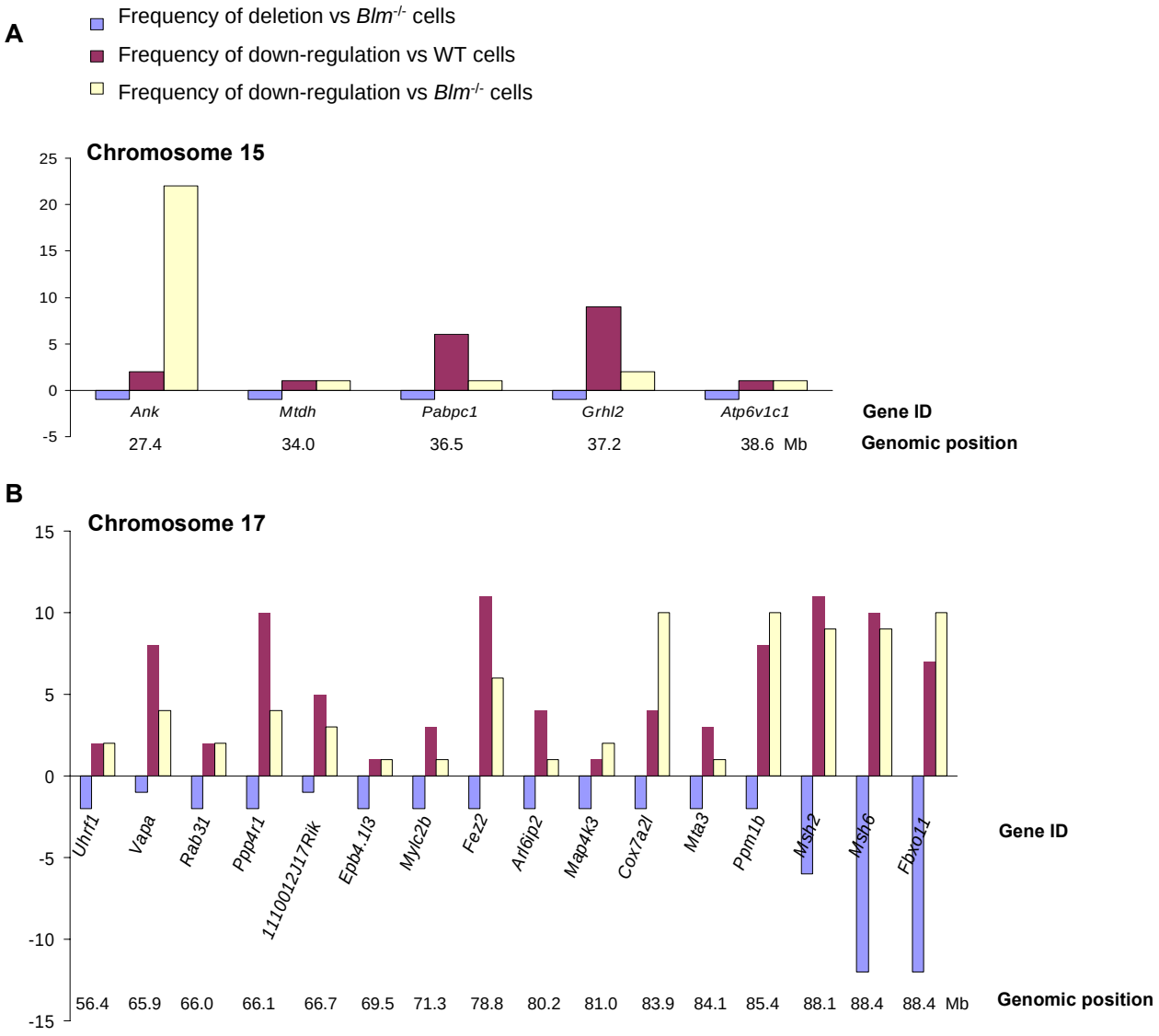


Figure 8-9 Genes on chromosome 15 and 17 which was deleted in one mutant and down-regulated in others

According to Table 8-12, five genes on chromosome 15 and sixteen genes on chromosome 17 were deleted in more than one mutant. And these genes are also down-regulated in more than one mutant compared with the wild type cells (AB1 and AB2.2) and the *Blm*-deficient cells (NGG5-3). These genes are plotted with the frequency of either deletion or down-regulation stated in Table 8-12. These genes are sorted in an ascending order of their genomic position. **A**. Chromosome 15; **B**. Chromosome 17.

8.2.2.3 *Dnmt1* mutant B7

According to array CGH data, gene *Dnmt1* was heterozygously deleted in mutant B7. A 400 kb region (20.4–20.8 Mb) on chromosome 9 is deleted with a Log_2(ratio) of -0.51 (Figure 8-10, a zoomed-in view shown in Figure 8-11). This mutant clone B7 demonstrated strong 6TG resistance, the phenotype of *Dnmt1*-deficiency. It was assumed that there is a small mutation in the undeleted allele of this gene. Therefore, it is worth examining the *Dnmt1* gene at the transcription level to test this assumption. When the expression array data of B7 was compared with the data from the *Blm*-deficient control cells, the transcription signal of the *Dnmt1* probe (scl36157.39.1_18-S) in B7 is 24% of that in the NGG5-3 cell line. This expression decrease can be regarded as a transcript loss. Thus, it can be assumed that in the non-deleted allele there must be some kind of mutation which destroys the *Dnmt1* transcription but which cannot be detected by 200 kb resolution array CGH. The absence of functional *Dnmt1* mRNA will lead to the full resistance to 6TG which was previously observed by Guo (Guo *et al.* 2004).

By comparing the gene expression regulation of the *Dnmt1* mutant B7 with the *Blm*-deficient cells, and comparing that of the *Dnmt1* knockout cell line (ww56+*Hprt*) with its parental wild type cell line AB2.2, six genes (*Fv1*, *Mylpf*, *Rhox2*, *Scml2*, *Sgip1* and *Xlr3a*) were found to have increased expression in the B7 clone and the *Dnmt1* knockout cell line (Table 8-13). Three of them are located on chromosome X and two of them are located on chromosome 4, though their genomic locations are far from each other. While two genes, *Eif2s3y* (eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked) and *Igf2* (insulin-like growth factor 2) were expressed at 10–20% of that in the reference cells. The down regulation of *Igf2* was also seen in the comparison between AB1 and AB2.2 wild type cells (Table 8-1), thus its probe might have high variations. Another explanation is that *Igf2* is expressed in high level in AB1 and it deceased to ~25% in AB2.2 and to ~3% in the two *Dnmt1*-deficient cell lines.

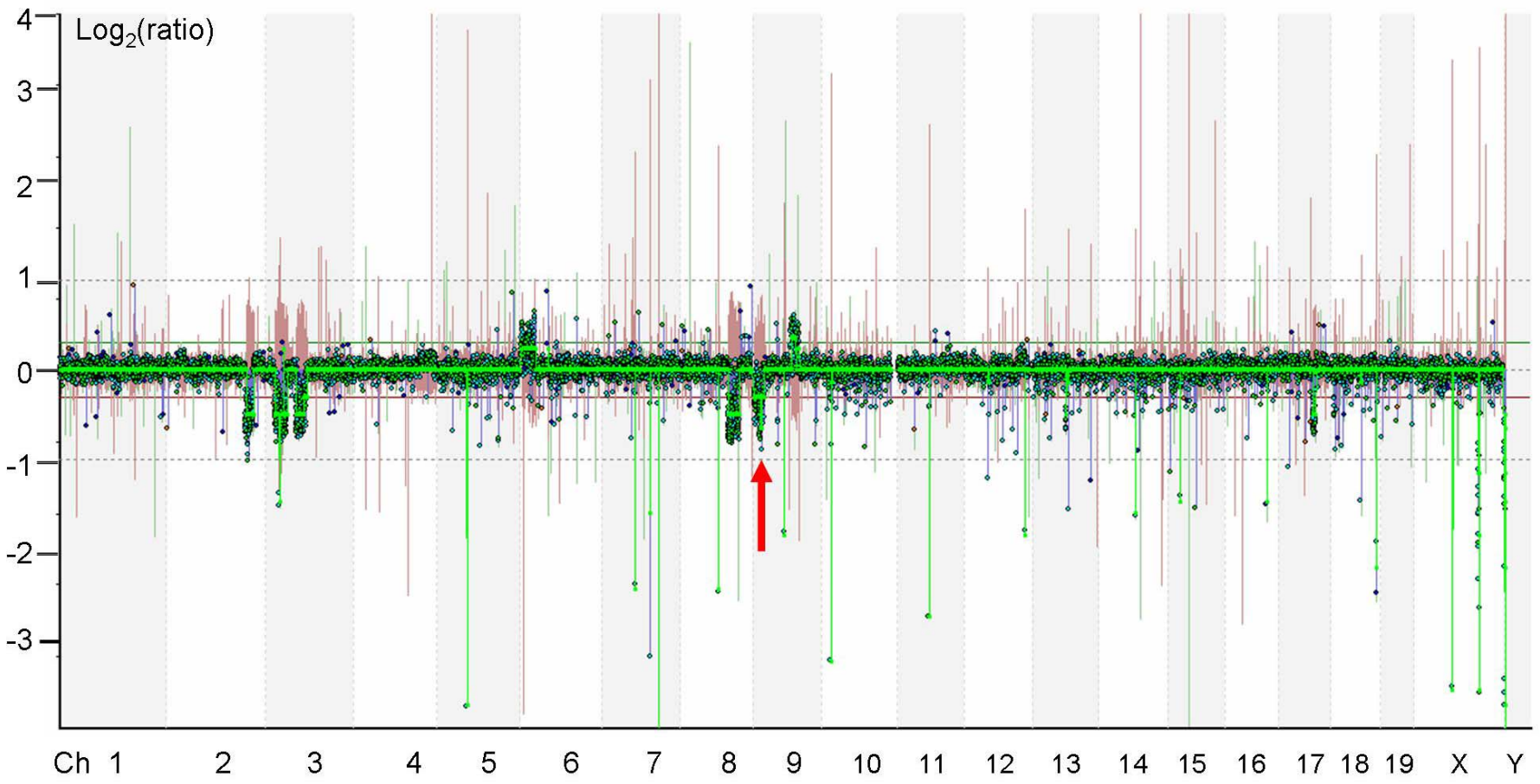
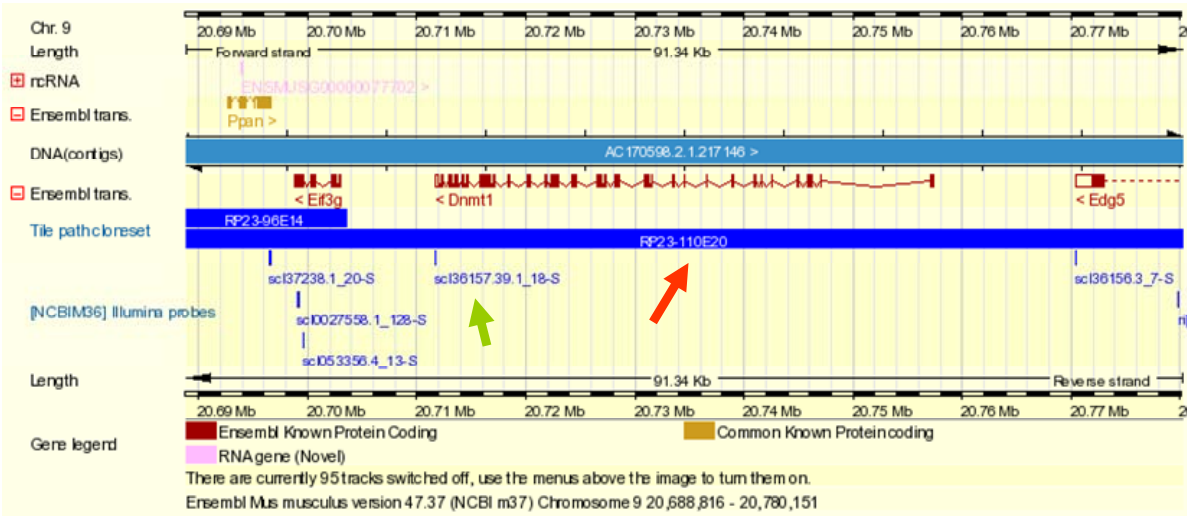


Figure 8-10 **Array CGH profile of mutant B7**

Figure 8-10 Array CGH profile of mutant B7

Green lines were smoothly connected between the values of Log_2(Mutant DNA signal/Reference DNA signal) of BAC clone probes. The red line indicates Log_2(Reference DNA signal/Mutant DNA signal) of BAC clone probes from the reciprocal hybridization experiment. The grey dotted lines are for Log_2(ratio) references of either +1 or -1. One deletion (indicated by a red arrow) on chromosome 9 includes the *Dnmt1* locus. The next figure shows a zoomed-in region of the *Dnmt1* locus. There are some single BAC homozygous deletions on chromosomes 3, 5, 7, 8, 9, 11, 12, 13, 17, 18, X and Y. Like this clone, these deletions were found in many other experiments and associated with chromosome Y loss, too. Thus, they were assumed to belong to chromosome Y but incorrectly annotated to other chromosomes. In addition, there are some heterozygous deletions on chromosomes 2, 3 and 8.

A



B

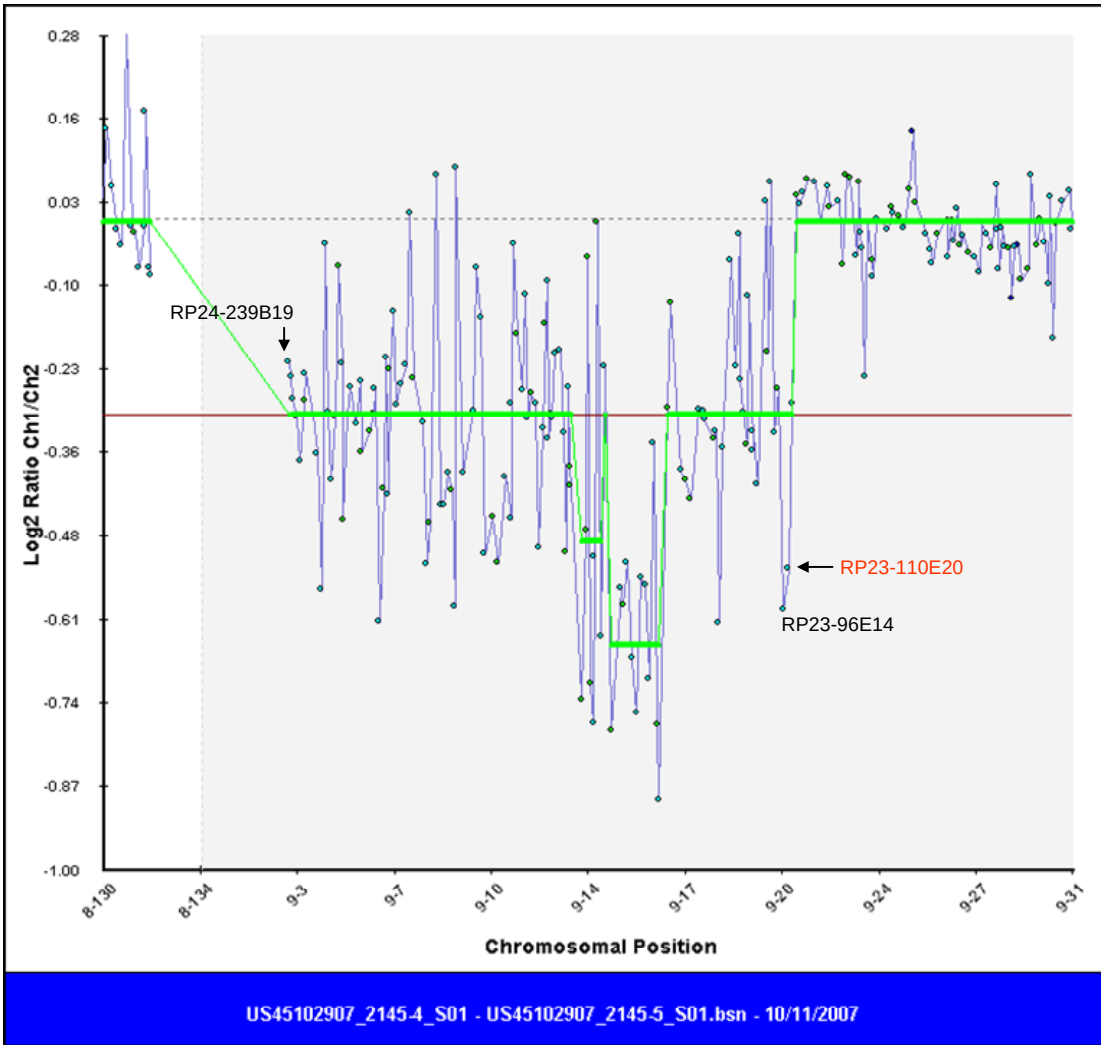


Figure 8-11 **Array CGH analysis of clone B7 at a zoomed-in region of the *Dnmt1* locus**

Figure 8-11 Array CGH analysis of clone B7 at a zoomed-in region of *Dnmt1* locus

A. An Ensembl snapshot of a 91 kb region with the gene *Dnmt1*. Ensembl annotated transcripts are shown in dark red. Tile path BACs for array CGH are shown in blue boxes. BAC RP23-110E20 (indicated by a red arrow) covers the gene *Dnmt1*. Gene-specific probes on the Illumina® array are shown in small blue bars. A probe for the gene *Dnmt1* is indicated by a green arrow. This probe is in exon 39, the last exon of *Dnmt1.* **B**. A zoomed-in array CGH profile of the chromosome 9. BAC RP23-110E20 (red font), covering the gene *Dnmt1,* is within a heterozygous deletion. BAC RP24-239B19 is the starting tile path array BAC on chromosome 9. It is noticeable that the Log_2(Mutant DNA signal/Reference DNA signal) is not the same within a deletion. This can vary depending on the different BAC probes.

Table 8-13 Genes regulated in both mutant clone B7 and *Dnmt1* knockout cells

A

MGI symbol	Description	Expression fold change mutant/reference		Chromosome Name	Gene Start (bp)	Gene End (bp)
		B7	*Dnmt1*$^{-/-}$ knockout			
Fv1	Friend virus susceptibility 1 [MGI:95595]	20.1	3.51	4	147243088	147244467
Mylpf	myosin light chain, phosphorylatable, fast skeletal muscle [MGI:97273]	2.89	2.75	7	134355122	134357801
Rhox2	reproductive homeobox 2 [MGI:1922449]	5.43	2.06	X	34784987	34789685
Scml2	sex comb on midleg-like 2 (Drosophila) [MGI:1340042]	4.14	2.01	X	157555125	157696145
Sgip1	SH3-domain GRB2-like (endophilin) interacting protein 1 [MGI:1920344]	6.73	2.07	4	102413902	102644468
Xlr3a	X-linked lymphocyte-regulated 3A [MGI:109506]	2.95	2.17	X	70331634	70342380

B

MGI symbol	Description	% reference signal		Chromosome Name	Gene Start (bp)	Gene End (bp)
		B7	*Dnmt1*$^{-/-}$ knockout			
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked [MGI:1349430]	12%	11%	Y	347055	365035
Igf2	insulin-like growth factor 2 [MGI:96434]	22%	14%	7	149836671	149852721

A. The genes over expressed in the mutant clone B7 and the *Dnmt*$^{-/-}$ knockout cell line (ww56+*Hprt*). **B**. The genes with decreased expression in the mutant clone B7 and the *Dnmt1*-deficient knockout cell line (ww56+*Hprt*). The reference cell of the clone B7 is the *Blm*-deficient cell line (NGG5-3); the reference cell of the *Dnmt1*-deficient knockout cell line is AB2.2 wild type cell line.

8.3 Conclusion and discussion

8.3.1 Confirmation of *Msh2, Msh6* and *Dnmt1* deficiency in mutants

The six mutants, B6, D1, D4, D8, F4, and H14, had homozygous deletions at MMR gene loci, including *Msh2* and *Msh6*. Analysis of these mutant clones with expression arrays confirmed that the transcripts of *Msh2* and *Msh6* genes were significantly reduced. Probe signals were 2.5–25% of levels in the *Blm*-deficient control cells than in the mutants.

Another eight unique mutants (C1, C3, C5, D6, F9, G8, G9 and H13) had lost one allele of the *Msh2* and *Msh6* MMR genes based on array CGH analysis results. Five of these heterozygous mutants were fully resistant to 6TG. Expression array analysis revealed that Four of the five strongly 6TG resistant mutants had lost either *Msh2* (C5 and H13) or *Msh6* (G8 and G9) transcripts (Table 8-8). Their expression array probe signals were 3.8–49 times higher in control cells than in the mutants. This level of expression change was consistent with that observed in the *Msh2* and *Msh6* homozygous deleted mutants and consistent with the strong 6TG resistance phenotype of these four clones. Mutant C3 and D6 were weakly resistant to 6TG and believed to be hypomorphic because they show reduced expression of *Msh2*. In these clones, *Msh2* signals were 40–50% of the level in the controls.

However, the phenotype of heterozygous mutant C1 is not consistent with *Msh2* and *Msh6* gene expression variation. C1 is fully resistant to 6TG, but the transcription levels of both *Msh2* and *Msh6* genes were not significantly reduced. One possible explanation is that a small mutation in the undeleted allele caused amino acid alteration without interfering the normal transcription.

The *Dnmt1* gene was detected as heterozygously deleted in the B7 mutant clone and its transcript was absent, leading to loss-of-function of the Dnmt1 protein. As reported previously (Guo *et al.* 2004), *Dnmt1* deficiency resulted in microsatellite instability and full resistance to 6TG.

By using genomic and expression arrays, a total of eleven mutants were found to have lost-of-function mutations at known MMR genes, including five *Msh2/Msh6* homozygous deletions (B6, D4, D8, F4 and H14), one *Msh6* homozygous deletion (D1), two compound *Msh2* mutations (C5 and H13), two compounds *Msh6* mutations (G8 and G9) and one compound *Dnmt1* mutation (B7). Two out of three mutants (C3, D6 and F9)

with heterozygous *Msh2* and *Msh6* deletions and weak 6TG resistance were examined by expression array analysis. The mutant C3 and D6 showed hypomorphic expression on the *Msh2* gene. One mutant (C1) has a heterozygous deletion at *Msh2* and *Msh6* loci but did not show transcription decrease of these two genes.

8.3.2 Genes regulated by *Msh2* and *Msh6*

The expression array signals of the *Blm*-deficient control cell line NGG5-3 were compared separately with the signals of individual MMR-deficient clones. The expression level of a set of genes (13 genes excluding *Msh6*) were down-regulated to less than 50% in all of the five mutant clones with the *Msh2* and *Msh6* homozygous deletions and in the mutant clone D1 with homozygous deleted *Msh6*. Considering the significance of these data, it is strong evidence that these 13 genes are downstream regulated by MSH6. The normal function of MSH6 is associated with normal transcription of these genes and loss-of-function mutation of MSH6 may play a role in the silencing effect of these genes' expression. Besides, gene *Bst2* was down-regulated in all five *Msh2* and *Msh6* deficient mutants but not in the Msh6 deleted mutant D1. The transcription of this gene is probably downstream of the normal function of MSH2.

In the ten mutants, which lost the transcription of *Msh2* and/or *Msh6* genes, five genes were identified to have 20–50% expression levels of that in the *Blm*-deficient control cells (Table 8-9). These genes (*Col16a1*, *Dcn*, *Lmna*, *Ltbp3* and *Tnfrsf22*) are similar because gene products are extracellular, transmembrane or nuclear membrance associated. The down-regulation of these genes in the MMR mutants may be secondary effects caused by MMR defects.

8.3.3 Potential genes involved in the MMR pathway

By analysing the data from array CGH and expression array, 26 genes (Table 8-12) were found to be homozygously deleted in at least one mutant; at the same time, their transcripts can be observed in the *Blm*-deficient ES cells and the AB1 and AB2.2 wild type cells, but their transcripts were lost in more than two $6TG^R$ mutants. As expected, genes *Msh2* and *Msh6* were among them. Some non-chromosome 17-linked genes such as *Ank, Zfp459* and *Eif5* are in the Table 8-12 with relatively high frequencies of being down-regulated in the $6TG^R$ mutants. This indicates that they are more likely to be involved in the MMR defects. However, the expression decrease of these genes may also be secondary effects of DNA mismatch repair defects.

The selection process in this analysis is stringent as it required genes to be homozygously deleted in one mutant. Considering that compound mutations are common in the previously described *Msh2/Msh6* mutants (Chapter 7), heterozygous deletions may contain MMR genes. Thus, the genes in heterozygously deleted regions also need be investigated in the future.

CHAPTER 9. GENERAL DISCUSSION

9.1 *Blm*-deficient ES cells for genetic screens

In the year 2000, a viable *Blm*-deficient mouse line was reported in our laboratory (Luo *et al.* 2000). One important feature of this mouse was the high rate of loss of heterozygosity observed in both ES cells and somatic cells in mice, suggesting the possibility of generating homozygous mutations through LOH. The LOH rate in *Blm*-deficient ES cells is 4.2×10^{-4} events per cell per generation at the *Gdf9* locus, an 18-fold increase of that in the same locus of the wild type ES cells. However, this rate will vary at different genomic locations. For instance, the LOH rate is highest close to telomere and lowest close to the centromere. Using this LOH rate, it is estimated that 12 doublings of a heterozygous mutant cell is sufficient to segregate homozygous mutants. This time course can be longer if the heterozygous locus is nearer to centromere and shorter if it is nearer to telomere. Therefore in practice, 15 cell doublings or more are used to generate a homozygous mutation library. The LOH rate in wild type cells is 2.3×10^{-5} events per cell per generation at the *Gdf9* locus given this rate at least. Fifty-thousand cells are needed to segregate a homozygous mutant in a wild type background. This number of cells is too high to practically generate a mutation library with 10^4 independent heterozygous mutants, as 5×10^8 cells will be needed for screening. However, in a *Blm*-deficient background only 2500 cells are needed to segregate one homozygous mutant thus only 2.5×10^7 cells are required to efficiently screen recessive mutations from a heterozygous mutation library of similar complexity. In principle, the *Blm*-deficient ES cells are very useful for genome-wide recessive screens.

Three reports have shown that *Blm*-deficient mouse ES cells are a reliable platform for genome-wide recessive genetic screens. Guo and Yusa independently demonstrated that cells with heterozygous point mutations could segregate homozygous mutations (Guo, PhD thesis, Yusa, *et al*. 2004). These reports described mutants in both DNA MMR genes and glycosylphosphatidylinositol-anchor biosynthesis. Yusa *et al*. used a chemical mutagen, while Guo *et al*. used an insertional mutagen. Guo *et al*. identified a new MMR-associated gene *Dnmt1* by screening for resistance to the DNA methylating agent 6TG. In addition to 6TG resistance, loss of functional Dnmt1 protein resulted in elevated microsatellite instability (MSI), a hallmark of human hereditary non-polyposis colorectal carcinoma (HNPCC). The library generated by Guo *et al*. has also be screened by Wei Wang (Wang and Bradley 2007) to isolate mutants that prevent MMuLV infection.

The screen performed by Ge Guo although successful was believed to be incomplete. The number of MMR-deficient mutants were imbalanced as six *Msh6* mutants were identified while only two mutants were *Dnmt1*–deficient. Moreover, this screen failed to recover mutants of other MMR genes, for instance *Mlh1* and *Pms2*. The reason of this is probably due to the non-random integration of retroviruses. Thus, for my research I selected to use irradiation to generate mutations in *Blm*-deficient ES cells. Irradiation mutagenesis does not have a sequence recognition process thus is a random mutagen. It was expected that a MMR screen on a mutation library induced by irradiation could recover all known MMR genes and novel MMR components. However, it was unknown how well this mutagen would work in *Blm*-deficient ES cells.

A mutant ES cell library in *Blm*-deficient ES cells was constructed through gamma irradiation, and a genome-wide recessive screen was conducted to recover 6TG-resistant mutants. Approximately 7000 individual mutants were generated through irradiation (Table 7-1). Considering the average size of deleted region is ~10 Mb per cell (Table 5-7), a twenty-fold genome coverage was achieved. Mutants of the MMR genes including *Msh2*, *Msh6*, *Mlh1*, *Pms2* and *Dnmt1* were expected to be recovered from this library.

Clones with homozygous biallelic mutations of the MMR genes *Msh2*, *Msh6* and *Dnmt1* were identified by array CGH and gene expression analysis. A total of eleven mutants were found to have loss-of-function mutations at known MMR genes, including five *Msh2/Msh6* homozygous deletions, one *Msh6* homozygous deletion, two compound *Msh2* mutations, two compound *Msh6* mutations and one compound *Dnmt1* mutation. However, mutations were not recovered at *Pms2* and *Mlh1*. As irradiation-induced mutations do not favour particular mutation loci, the lack of deletions at *Mlh1* and *Pms2* suggest that genes deleted with these MMR genes may be required for ES cell viability. Deletions generated by irradiation can be several million base pairs in length. When a gene important to cell viability is homozygously deleted together with an MMR gene, this cell will die. Consistent with this hypothesis, the *Mlh1* and *Pms2* genes are in gene-dense regions. There are approximate 70 genes in the 5 Mb surrounding *Mlh1* and *Pms2*, which makes a gene density of 14 genes/Mb, higher than the average gene density (~10 genes/Mb) of the mouse genome. Moreover, some genes in the 0.3–2.7 Mb surrounding *Mlh1* and *Pms2* were found to cause early embryonic lethality as early as embryonic day 3.5 (Ding *et al.* 1998; Herceg *et al.* 2001; Parekh *et al.* 2004; Sugihara *et al.* 1998). The early embryonic lethality of these genes implies that ES cells deficient for their proteins may not survive. Since deletion sizes are relatively large after irradiation, these genes are likely to be deleted

together with *Mlh1* and *Pms2*.

Although MMR genes, such as *Msh2* and *Msh6*, were homozygously deleted in six mutants, the distal parts of two homologous chromosomes in these mutants are not the same (Figure 7-18). The so-called “homozygous” deletion of *Msh2* and *Msh6* described in my thesis are all compound heterozygous mutations, which seems to be caused by a larger heterozygous deletion on one homologous chromosome and a small deletion in the other. The pure of homozygous deletions generated by increased LOH events in the *Blm*-deficient cells and described by Guo *et al*. (2004) and Wang *et al*. (2007) were not observed. This might also be due to selection against homozygous deletions induced by irradiation.

Since irradiation can generate high genome coverage mutation libraries, fewer cells are needed for screening. In principle, three hundred mutant clones are sufficient to saturate the mouse genome. As discussed previously, 50,000 wild type cells can segregate one homozygous mutant. Thus fifteen million wild type cells would be enough for genome-wide screening, which means this screen could be performed on wild type ES cells. One advantage to do so is that wild type cells are more genetically stable than the *Blm*-deficient cells. In addition, wild type cells do not contain genetic modifications thus are more accessible for native gene functions.

9.2 Selecting for deletions with *HSV-tk*

My initial plans were to use negative selection to recover deletions using FIAU selection. To model this, the *PuroΔtk* cassette was targeted into a gene rare region on chromosome 6, 0.01–1 per cent of the targeted cells were still resistant to FIAU though all cells were resistant to puromycin; thus the designed selection scheme for homozygous deletions at this locus was not achieved. Southern blot analysis confirmed that 100% (22/22) of the puromycin and FIAU double resistant clones contained intact *PuroΔtk* cassette. Random integration events of the targeting vector was used to explore this phenomenon further and a similar FIAU-resistant background (10^{-5}–10^{-2}) was observed in the *Blm*-deficient cells but not in the AB1 wild type cells, suggesting the appearance of puro-resistant plus FIAU-resistant clone is a genetic event associated with the *Blm*-deficient ES cell line. This type of event occurred during cell culture and must be due to LOH, which segregated homozygous mutations in a gene which gives FIAU resistance to a cell expressing *PuroΔtk*. As two steps of *Gdf9* targeting experiments and two steps of Cre-mediated recombination experiments were performed after the *Blm* gene was double targeted to generate the NGG5-3 cell line (Guo *et al.* 2004), spontaneous mutations can be anticipated in this cell line. One possible

explanation is that a spontaneous mutation has occurred in a gene critical for FIAU metabolism and that this mutation segregates homozygous mutations in both the *PuroΔtk* targeted and the *PuroΔtk* randomly integrated *Blm*$^{-/-}$ cells. The gene responsible for this FIAU resistance could be identified in a screen analogue to the one I have performed.

9.3 Gamma irradiation as an efficient mutagen

Ionizing radiation has been utilized to produce nested deletions in mouse ES cells (Kushi *et al.* 1998; Schimenti *et al.* 2000; Thomas *et al.* 1998; You *et al.* 1997). Generating mutations by irradiation is a rapid approach when compared to other methods of mutagenesis. Irradiation creates random DNA breakages which are repaired but often result in deletions. From this point of view, it is suitable for genome-wide mutagenesis. However, a methodology for scanning the whole genome to quickly identify deleted loci had not available until recently. The development of high resolution comparative genomic hybridization (CGH) arrays has solved this problem (Cai *et al.* 2002; Chung *et al.* 2004). Using array CGH, regions of the genome with a change of copy number can be identified. In *Blm*-deficient ES cells, loss-of-function mutations were expected to be caused by homozygous deletions caused by LOH. Thus, homozygous deletions were anticipated in 6TG-resistant irradiation-induced mutants. However, the homozygously deleted regions in the 6TG^{R} mutants detected on the CGH arrays are actually compound heterozygous deletions, thus the homologous chromosomes are not homozygous. Within twenty-seven unique mutants, which are all strongly resistant to 6TG, seven mutants had compound mutations at *Msh2*, eight at *Msh6* and one at *Dnmt1*.

Although radiation has been used for decades, the relationship between deletion length and irradiation dosage is not clearly defined. Kushi has observed that X-rays make deletions ranging from 200 to 700 kb around the *Hprt* locus on the X chromosome of mouse E14 ES cell (Kushi *et al.* 1998). Schimenti observed irradiation-induced deletions up to approximate 70 Mb on chromosome 5. Schimenti used a targeted *HSV-tk* gene to negatively select cells with deletions (Schimenti *et al.* 2000). However, these approaches have been limited to specific chromosomes. I have used array CGH to reveal copy number changes induced by irradiation across the genome which provides a wider view of the effects of irradiation. There was no obvious correlation between the irradiation dose and the number of deletions per cell or the total length of deleted regions per cell. Two-thirds of *Blm*-deficient cells surviving 10 Gray irradiation contain deletions (Table 5-6) with an average total length per cell of 10 Mb (Table 5-7). Therefore, approximately five hundred mutants can saturate the mouse genome. This number of cells is much lower than required for other mutagenesis

systems, which makes it possible to screen recessive mutations in wild type cells.

However, the estimate of the size of the deleted regions per *Blm*-deficient cell in the non-selected clones was probably under estimated in Table 5-7 due to low sample numbers. The heterozygous deletions in eight unique *Msh2* and *Msh6* mutants range from 13 to 46 Mb (Table 7-5). The average length per deletion of chromosome 17 is 28 Mb in the eight mutants. It is noticed that seven of these eight deletions appear to have lost distal chromosome 17, which requires only one double-strand break instead of two DSBs for an interstitial deletion.

I have shown that irradiation can be used to generate deletions in mouse ES cells ranging in length from several hundred kilo bases to a few million base pairs. Mutations induced by irradiation show no sequence preference, being essentially random, and thus have no loci preference, which is an important advantage when screening for new mutations.

Although irradiation-induced mutations do not exhibit loci preference, the lack of any homozygous and heterozygous deletions at MMR genes, such as *Mlh1* and *Pms2* indicated that there are limitations of irradiation as a mutagen in the *Blm*-deficient ES cells. As I relied on LOH events to generate homozygous deletions and the deletions generated by irradiation can be millions of base pairs in length, large homozygous deletions may contain not only the MMR genes but also genes that are essential for cell viability. When one of these essential genes is located in close proximity to one of the MMR genes and is homozygously deleted together with an MMR gene, the ES cell cannot survive and contribute to the library of selected mutants.

Irradiation also generates smaller deletions, which are below the minimum detection limit for the array CGH used in this study (200 kb). If a small genomic change mutates a gene causing a loss of transcript, it can be detected using expression arrays. MMR gene mutants with one small mutation and a deletion of the other allele (compound heterozygote) have been isolated in my screen. For example, mutant clone B7 has a heterozygous deletion at *Dnmt1* but the transcript of *Dnmt1* is absent according to expression array analysis. Overall, I have detected five clones with heterozygous deletions of one allele and loss of transcript on the other allele.

The combination of CGH and expression arrays is a powerful detection system, thus irradiation mutagenesis system could be used together with a chemical mutagen to overcome the limitation of *Blm* in irradiated cells. One strategy would be to use irradiation to

generate large heterozygous deletions in pools then use a chemical mutagen, such as ENU, to introduce point mutations. If irradiation at 10 Gray deletes 20 Mb of DNA per cell and each cell contain such deletions, 100–150 surviving cells will be sufficient to saturate the mouse genome with heterozygous deletions. Then ENU mutagenesis has a 1% possibility of generating loss-of-function mutations per cell. Five thousand cells need to be generated if ENU generates 1 mutation per 500 kb, 1% of the mouse genome codes genes and 10% of mutations generated by ENU cause loss-of-function mutations.

9.4 Mutation detection methods

Using array CGH it is possible to identify deletions and duplications in a whole genome relative to a reference genome. However, there are some limitations when applying this technology. Firstly, the resolution of an array CGH determines the resolution of the analysis. In this project, the BAC array has a resolution of about 200 kb (1 BAC). But small duplications and deletions can not be detected. Thus, data from array CGH may not report a gene as deleted, but the transcript may be absent (false negatives). The use of a higher resolution array CGH may help to overcome this to some degree. In the analysis reported here, the genomic location of a BAC probe is recorded as the position of its mid-point. We are unable to determine how much of the sequence within a BAC is hybridizing or the location of the hybridizing sequence within the BAC. Moreover, 50 kb of homologous sequence in a sample may generate enough signal to interpret a particular 200 kb BAC probe as positive. This means that, in this example, we may be unable to detect a 150 kb deletion. When comparing deletions between samples, the start and end points of deletions encompassing the same set of BAC probes will be quite different (up to 150 kb).

Array CGH is too expensive to be used in very large scale projects. Because one is trying to detect relatively small differences (loss of one copy vs two in the reference), reciprocal and repeated experiments must be performed, which raise the cost. Also, array CGH cannot detect chromosome translocations and inversions, both of which can result in truncated transcripts. I screened for deletions generated by irradiation by array CGH and followed by expression array analysis. One array can only be used to analyse one candidate DNA or RNA. The process of hybridization, washing and scanning arrays limits the throughput to less than 20 samples per week. Compared with the PCR-based mutation mapping strategies used during insertional mutagenesis approaches, the mapping of irradiation-induced mutations using array CGH is inefficient.

In this project, I used Illumina® expression arrays to measure expression variations

experimentally. This proved to be a very stable platform. This type of array analyses the expression of 19,400 mouse genes. As an example of the sensitivity of the method, only 14 genes were detected whose expression differed significantly between the AB1 and AB2.2 ES cells. Considering that AB2.2 is highly related to AB1 but differs due to inactivation of expression of the *Hprt* gene by viral integration, this degree of expression similarity indicates the reliability of the Illumina® expression array platform. Additionally, this analysis detected known variations, such as the change of *Hprt* expression and the viral elements expression in these two cell lines.

Several mutations were detected in known MMR genes by expression array analysis. These mutations were either homozygous or heterozygous deletions of *Msh2*, *Msh6* and *Dnmt1*, which were also detected by array CGH. The transcripts of *Msh2*, *Msh6* and *Dnmt1* were usually down regulated to 2–25% of the wild type levels. Thus, data from expression array analysis matches that of array CGH analysis.

Secondary expression changes of MMR defects were identified using expression arrays. Five genes (*Col16a1*, *Dcn*, *Lmna*, *Ltbp3* and *Tnfrsf22*) were down-regulated to 20–50% of wild type levels in the ten *Msh2*/*Msh6* mutants (Table 8-9). A common feature of these genes is that gene products are extracellular, transmembrane or nuclear membrance associated. These products are not easily linked to MMR defects and may be secondary. Additionally, the cytochrome c oxidase, subunit VIIa 1 gene (*Cox7a1*) was over-expressed in nine out of ten *Msh2/Msh6* mutants. The decreased expression of the five genes and the increased expression of *Cox7a1* may be secondary effects caused by MMR defects or related to deletion of gene(s) on chromosome 17.

Although the expression array analysis is sensitive and reliable, it cannot be used alone to detect mutations. For example in my analysis without the CGH data the list of genes extend to a hundred or more which had expressions. This list is too long to identify which gene is responsible for the $6TG^R$ phenotype without clues from array CGH analysis.

Both array CGH data and expression array data can be analysed in a variety of dimensions. One of the approaches to identify new MMR genes is to find statistically significant associations between 6TG resistance, genomic deletions and absence of specific transcripts. For example, I have identified just the subset of genes (Table 8-12) deleted in one or more 6TG-resistant mutants which fulfilled two additional requirements that their transcripts were absent in greater than two $6TG^R$ mutant clones and their transcripts were present in the *Blm*-deficient control ES cells and the wild type control cells. This set of 26

genes are highly likely to be involved in the DNA mismatch repair pathway; thus they are worth investigating further.

9.5 Future analysis of the MMR mutants

DNA mismatch repair is an essential DNA surveillance system to prevent spontaneous or induced replication errors in somatic cells and gametes. However, knowledge about MMR in mammals is not complete. For instance, the processes of moving MMR-components to a site of damage, strand discrimination and strand excision between new and old DNA are not clear. The present screen has generated $6TG^R$ MMR mutants and a list of potential MMR genes, both of which ought to be analysed for other phenotypes of mismatch repair deficiency. To test the deficiency of potential MMR genes in the mouse ES cells, some other characteristics, such as an increased accumulation of point mutations and frame shift mutations, hyperrecombination, microsatellite instability (MSI) and inability of mismatch repair (de Wind *et al.* 1995) need to be investigated.

For example, the mutation rate in the mutants can be compared with that in the wild type cells in several ways. The simplest is to sequence certain genes in both cell types to measure the frequency of spontaneous point mutations and frame shift mutations. This method may need to examine a large number of samples. To accelerate the speed of mutation accumulation, nucleotide analogues or methylating agents, such as MNNG (N-methyl-N'-nitro-N-nitrosoguanidine) or MNU (N-methyl-N-nitrosourea) can be used to induce mutations. A critical lesion induced by methylating agents is O^6-mehtylguanine (O^6-MeG), which causes mismatches when this methylated site is replicated by DNA polymerase. Measurements can be performed at certain times after the addition of methylating agents. To facilitate measuring the frequency of mutations, reporter genes, such as the green fluorescent protein (*GFP*) gene, can be introduced into genomic loci; and one only needs to analyse number of green and normal-colored cells. Alternatively, the whole genome or certain chromosomes of methylating agent treated or un-treated cells could be sequenced using new generation sequencing systems, such as Solexa platform, to measure the frequency of mutations. Sex chromosomes, which can be isolated using chromosome sorting machines, are monoploid thus are convenient for sequencing point mutations.

Similarly, the frequency of enhanced recombination can be measured in the mutants. It is shown that sequence alterations in homologous sequences will suppress homologous recombination in wild type cells, while *Msh2*-deficient cells have lost its ability to suppress homologous recombination between poorly matched sequences (te Riele *et al.* 1992;

Waldman *et al.* 1988). Moreover, the occurrence of MSI can be measured using a slippage assay, such as that described by Guo *et al*. in the *Dnmt1* mutants (Guo *et al.* 2004). A repair assay of heteroduplex DNA can also be performed to examine the efficiency of mismatch repair (Wildenberg *et al.* 1975).

Another way to confirm which specific gene are responsible for 6TG resistance and DNA mismatch repair is to restore their normal functions in mutants. This can be achieved by transforming cDNA or genomic loci, for instance BACs, into the mutant cells to recover their normal functions.

The mutant clones I have identified carry many genetic changes. Thus to examine if the expression effects in the mutants result from MMR-deficiency or are secondary changes, expression array analysis can be repeated in targeted cells to check if the same expression changes are present. However, when these expression changes need to be analysed in a series of time course, conditional targeted cells will be generated so that genes can be mutated within desired time frames.

The $6TG^R$ mutants I have isolated are also *Blm*-deficient. Since the Blm protein is involved in DNA repair (Davalos *et al.* 2003; Wang *et al.* 2000), the mutant phenotype may be enhanced or suppressed by the *Blm*-deficiency; thus knockout experiments should be conducted in wild type cells to examine these potential MMR genes in more detail.

9.6 Other mutagenesis systems

Considering the limitations of irradiation as a mutagen and the mutation mapping procedures, a more efficient mutagen is needed for genome-wide mutagenesis studies. RNA interference (RNAi) and transposon-mediated mutagenesis are alternatives to the irradiation mutagenesis system.

The concept of RNAi was developed by Fire *et al*. (1998). He observed that sense and anti-sense RNA and also double-strand RNA could significantly reduce gene expression (Fire *et al.* 1998). This phenomenon was soon employed and developed as a powerful genetic tool for loss-of-function mutagenesis screens. It is an extremely convenient tool for *C.elegans* genetics as double-strand RNA can be delivered into worms by feeding them with bacteria that express double-strand RNA. RNAi has been successfully used in genetic screens in cultured human cells and new components of the p53-dependent profileration arrest process has been identified (Berns *et al.* 2004). One advantage of RNAi is that it does not

require a mutation in both alleles of a gene to achieve a loss of function. In this case, experiments can be performed in any cells; thus, the *Blm*-associated high LOH rate is not needed. However, the expression suppressing effect of RNAi in mammalian cells is not as clean as expected. Nearly 50% of the expected targets were not inhibited in one experiment to suppress expression of the components of 26S proteosome (Paddison *et al.* 2004). In many cases, there are off-target effects and the knockdown of gene expression is incomplete. Moreover, it is logistically very complicated to deal with a library covering the whole genome of 25,000 genes and 3 RNAi vector per gene (75,000 clones).

One promising alternative, which can be used with the *Blm*-deficient cells are transposable elements, which are known to efficiently integrate foreign DNA into mammalian cells as well as moving positions within cells through a “cut-and-paste” mechanism (Kaufman *et al.* 1992). Genetically modified transposon systems are emerging as promising tools for mutagenesis, gene discovery and therapeutic gene delivery in mammals. Transposable elements have been shown to be successful for transgenesis and insertional mutagenesis in a number of host genomes.

One very useful transposon is *Sleeping Beauty* (*SB*), which is a Tc1 like transposon which has been constructed from transpositionally inactive transposon sequences in Teleost fish by eliminating the inactivating mutations accumulated during evolution (Ivics *et al.* 1997). It has been shown that *SB* can be mobilized 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 with long-term transgene expression. When using *SB* as a mutagenesis system, exogenous DNA, which can be a selection marker, is placed between two terminal inverted repeats. *SB* has been used to identify cancer associated genes in mice (Collier *et al.* 2005; Dupuy *et al.* 2005). One limitation of this system is the strong propensity for “local hopping” events. Three quarters of transpositions are found to be within the same chromosome in mice (Horie *et al.* 2003), which has limited its application as a genome-wide mutagenesis system.

Piggybac (*PB*) elements may solve this problem. *PB* can carry transgenes up to 9.1 kb without decreasing transposition efficiency (Ding *et al.* 2005). There was no obvious *PB* insertional preference and local hopping has not been observed in any chromosomes. Although transposon systems demonstrate a small but significant bias toward genes, transcriptional start sites and gene upstream regulatory sequences, their genome-wide transposition and mouse germ line transpositional ability indicates their promising future as genetic analysis tools in model genomes.

REFERENCE

Aaltonen, L. A., P. Peltomaki, et al. (1993). "Clues to the pathogenesis of familial colorectal cancer." Science 260(5109): 812-6.

Ababou, M., S. Dutertre, et al. (2000). "ATM-dependent phosphorylation and accumulation of endogenous BLM protein in response to ionizing radiation." Oncogene 19(52): 5955-63.

Abuin, A. and A. Bradley (1996). "Recycling selectable markers in mouse embryonic stem cells." Mol Cell Biol 16(4): 1851-6.

Abuin, A., H. Zhang, et al. (2000). "Genetic analysis of mouse embryonic stem cells bearing Msh3 and Msh2 single and compound mutations." Mol Cell Biol 20(1): 149-57.

Adam, M. A. and A. D. Miller (1988). "Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions." J Virol 62(10): 3802-6.

Adams, D. J., P. J. Biggs, et al. (2004). "Mutagenic insertion and chromosome engineering resource (MICER)." Nat Genet 36(8): 867-71.

Aebi, S., D. Fink, et al. (1997). "Resistance to cytotoxic drugs in DNA mismatch repair-deficient cells." Clin Cancer Res 3(10): 1763-7.

Alani, E., R. A. Reenan, et al. (1994). "Interaction between mismatch repair and genetic recombination in Saccharomyces cerevisiae." Genetics 137(1): 19-39.

Alani, E., T. Sokolsky, et al. (1997). "Genetic and biochemical analysis of Msh2p-Msh6p: Role of ATP hydrolysis and Msh2p-Msh6p subunit interactions in mismatch base pair recognition." Molecular and Cellular Biology 17(5): 2436-2447.

Allen, N. D., D. G. Cran, et al. (1988). "Transgenes as probes for active chromosomal domains in mouse development." Nature 333(6176): 852-5.

Amin, N. S., M. N. Nguyen, et al. (2001). "exo1-Dependent mutator mutations: model system for studying functional interactions in mismatch repair." Mol Cell Biol 21(15): 5142-55.

Antoch, M. P., E. J. Song, et al. (1997). "Functional identification of the mouse circadian Clock gene by transgenic BAC rescue." Cell 89(4): 655-67.

Asakawa, J., N. Nakamura, et al. (2007). "Estimation of mutation induction rates in AT-rich sequences using a genome scanning approach after X irradiation of mouse spermatogonia." Radiat Res 168(2): 158-67.

Bader, S., M. Walker, et al. (1999). "Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency." Oncogene 18(56): 8044-7.

Baer, A. and J. Bode (2001). "Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes." Curr Opin Biotechnol 12(5): 473-80.

Baker, S. M., C. E. Bronner, et al. (1995). "Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis." Cell 82(2): 309-19.

Baker, S. M., A. W. Plug, et al. (1996). "Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over." Nat Genet 13(3): 336-42.

Ban, C. and W. Yang (1998). "Structural basis for MutH activation in E.coli mismatch repair and relationship of MutH to restriction endonucleases." Embo J 17(5): 1526-34.

Belinsky, M. G., P. Guo, et al. (2007). "Multidrug resistance protein 4 protects bone marrow, thymus, spleen, and intestine from nucleotide analogue-induced damage." Cancer Res 67(1): 262-8.

Bellacosa, A., L. Cicchillitti, et al. (1999). "MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1." Proc Natl Acad Sci U S A 96(7): 3969-74.

Bergstrom, R. A., Y. You, et al. (1998). "Deletion mapping of the head tilt (het) gene in mice: a vestibular mutation causing specific absence of otoliths." Genetics 150(2): 815-22.

Berns, K., E. M. Hijmans, et al. (2004). "A large-scale RNAi screen in human cells identifies new components of the p53 pathway." Nature 428(6981): 431-7.

Billet, S., S. Bardin, et al. (2007). "Gain-of-function mutant of angiotensin II receptor, type 1A, causes hypertension and cardiovascular fibrosis in mice." J Clin Invest 117(7): 1914-25.

Bingham, P. M., M. G. Kidwell, et al. (1982). "The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family." Cell 29(3): 995-1004.

Bjelland, S. and E. Seeberg (2003). "Mutagenicity, toxicity and repair of DNA base damage induced by oxidation." Mutat Res 531(1-2): 37-80.

Bloom, D. (1954). "Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs; probably a syndrome entity." AMA Am J Dis Child 88(6): 754-8.

Bocker, T., A. Barusevicius, et al. (1999). "hMSH5: a human MutS homologue that forms a novel heterodimer with hMSH4 and is expressed during spermatogenesis." Cancer Res 59(4): 816-22.

Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived

teratocarcinoma cell lines." Nature 309(5965): 255-6.
Breen, A. P. and J. A. Murphy (1995). "Reactions of oxyl radicals with DNA." Free Radic Biol Med 18(6): 1033-77.
Brewen, J. G., R. J. Preston, et al. (1973). "Genetic hazards of ionizing radiations: cytogenetic extrapolations from mouse to man." Mutat Res 17(2): 245-54.
Bronson, S. K., E. G. Plaehn, et al. (1996). "Single-copy transgenic mice with chosen-site integration." Proc Natl Acad Sci U S A 93(17): 9067-72.
Buermeyer, A. B., S. M. Deschenes, et al. (1999). "Mammalian DNA mismatch repair." Annu Rev Genet 33: 533-64.
Cai, W. W., J. H. Mao, et al. (2002). "Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays." Nat Biotechnol 20(4): 393-6.
Calin, G., V. Herlea, et al. (1998). "The coding region of the Bloom syndrome BLM gene and of the CBL proto-oncogene is mutated in genetically unstable sporadic gastrointestinal tumors." Cancer Res 58(17): 3777-81.
Capecchi, M. R. (1989). "Altering the genome by homologous recombination." Science 244(4910): 1288-92.
Carlson, C. M., A. J. Dupuy, et al. (2003). "Transposon mutagenesis of the mouse germline." Genetics 165(1): 243-56.
Carmeliet, P., V. Ferreira, et al. (1996). "Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele." Nature 380(6573): 435-9.
Carraway, M. and M. G. Marinus (1993). "Repair of heteroduplex DNA molecules with multibase loops in Escherichia coli." J Bacteriol 175(13): 3972-80.
Carter, D. M. and C. M. Radding (1971). "The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease." J Biol Chem 246(8): 2502-12.
Cary, L. C., M. Goebel, et al. (1989). "Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses." Virology 172(1): 156-69.
Casselton, L. and M. Zolan (2002). "The art and design of genetic screens: filamentous fungi." Nat Rev Genet 3(9): 683-97.
Cattanach, B. M. and J. Jones (1994). "Genetic imprinting in the mouse: implications for gene regulation." J Inherit Metab Dis 17(4): 403-20.
Cejka, P., L. Stojic, et al. (2003). "Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1." Embo J 22(9): 2245-54.
Celeste, A., S. Petersen, et al. (2002). "Genomic instability in mice lacking histone H2AX." Science 296(5569): 922-7.
Chandler, K. J., R. L. Chandler, et al. (2007). "Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression." Mamm Genome 18(10): 693-708.
Chang, Y. F., J. S. Imam, et al. (2007). "The nonsense-mediated decay RNA surveillance pathway." Annu Rev Biochem 76: 51-74.
Chao, H. H., S. E. Mentzer, et al. (2003). "Overlapping deletions define novel embryonic lethal loci in the mouse t complex." Genesis 35(2): 133-42.
Chen, W. and S. Jinks-Robertson (1999). "The role of the mismatch repair machinery in regulating mitotic and meiotic recombination between diverged sequences in yeast." Genetics 151(4): 1299-313.
Chen, Y., D. Yee, et al. (2000). "Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells." Nat Genet 24(3): 314-7.
Chen, Y. T. and A. Bradley (2000). "A new positive/negative selectable marker, puDeltatk, for use in embryonic stem cells." Genesis 28(1): 31-5.
Chester, N., F. Kuo, et al. (1998). "Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene." Genes Dev 12(21): 3382-93.
Chick, W. S., S. E. Mentzer, et al. (2005). "X-ray-induced deletion complexes in embryonic stem cells on mouse chromosome 15." Mamm Genome 16(9): 661-71.
Chung, Y. J., J. Jonkers, et al. (2004). "A whole-genome mouse BAC microarray with 1-Mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization." Genome Res 14(1): 188-96.
Coffin, J., Hughes, S., Varmus, H. (1997). Retroviruses, Cold Spring Harbor Laboratory Press.
Collier, L. S., C. M. Carlson, et al. (2005). "Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse." Nature 436(7048): 272-6.
Copeland, N. G., N. A. Jenkins, et al. (2001). "Recombineering: a powerful new tool for mouse functional genomics." Nat Rev Genet 2(10): 769-79.
Cordes, S. P. (2005). "N-ethyl-N-nitrosourea mutagenesis: boarding the mouse mutant express."

Microbiol Mol Biol Rev 69(3): 426-39.
Cortellino, S., D. Turner, et al. (2003). "The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity." Proc Natl Acad Sci U S A 100(25): 15071-6.
Cox, E. C. (1976). "Bacterial mutator genes and the control of spontaneous mutation." Annu Rev Genet 10: 135-56.
Crow, J. F. (1995). "Spontaneous mutation as a risk factor." Exp Clin Immunogenet 12(3): 121-8.
D'Atri, S., L. Tentori, et al. (1998). "Involvement of the mismatch repair system in temozolomide-induced apoptosis." Mol Pharmacol 54(2): 334-41.
Datta, A., M. Hendrix, et al. (1997). "Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast." Proc Natl Acad Sci U S A 94(18): 9757-62.
Davalos, A. R. and J. Campisi (2003). "Bloom syndrome cells undergo p53-dependent apoptosis and delayed assembly of BRCA1 and NBS1 repair complexes at stalled replication forks." J Cell Biol 162(7): 1197-209.
de Martinville, B., A. R. Wyman, et al. (1982). "Assignment of first random restriction fragment length polymorphism (RFLP) locus ((D14S1) to a region of human chromosome 14." Am J Hum Genet 34(2): 216-26.
de Vries, S. S., E. B. Baart, et al. (1999). "Mouse MutS-like protein Msh5 is required for proper chromosome synapsis in male and female meiosis." Genes Dev 13(5): 523-31.
de Wind, N., M. Dekker, et al. (1995). "Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer." Cell 82(2): 321-30.
de Wind, N., M. Dekker, et al. (1999). "HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions." Nat Genet 23(3): 359-62.
de Wind, N., M. Dekker, et al. (1998). "Mouse models for hereditary nonpolyposis colorectal cancer." Cancer Res 58(2): 248-55.
Dehal, P., P. Predki, et al. (2001). "Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution." Science 293(5527): 104-11.
Deng, H. K., D. Unutmaz, et al. (1997). "Expression cloning of new receptors used by simian and human immunodeficiency viruses." Nature 388(6639): 296-300.
di Pietro, M., G. Marra, et al. (2003). "Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent N-methyl-n'-nitro-N-nitrosoguanidine." Cancer Res 63(23): 8158-66.
Ding, J., L. Yang, et al. (1998). "Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo." Nature 395(6703): 702-7.
Ding, S., X. Wu, et al. (2005). "Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice." Cell 122(3): 473-83.
Dizdaroglu, M. (1992). "Oxidative damage to DNA in mammalian chromatin." Mutat Res 275(3-6): 331-42.
Doetschman, T., R. G. Gregg, et al. (1987). "Targetted correction of a mutant HPRT gene in mouse embryonic stem cells." Nature 330(6148): 576-8.
Dohet, C., R. Wagner, et al. (1986). "Methyl-directed repair of frameshift mutations in heteroduplex DNA." Proc Natl Acad Sci U S A 83(10): 3395-7.
Dong, J., D. F. Albertini, et al. (1996). "Growth differentiation factor-9 is required during early ovarian folliculogenesis." Nature 383(6600): 531-5.
Duckett, D. R., J. T. Drummond, et al. (1996). "Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct." Proc Natl Acad Sci U S A 93(13): 6443-7.
Dufort, D., L. Schwartz, et al. (1998). "The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis." Development 125(16): 3015-25.
Dupuy, A. J., K. Akagi, et al. (2005). "Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system." Nature 436(7048): 221-6.
Dupuy, A. J., K. Clark, et al. (2002). "Mammalian germ-line transgenesis by transposition." Proc Natl Acad Sci U S A 99(7): 4495-9.
Dupuy, A. J., S. Fritz, et al. (2001). "Transposition and gene disruption in the male germline of the mouse." Genesis 30(2): 82-8.
Dutertre, S., R. Sekhri, et al. (2002). "Dephosphorylation and subcellular compartment change of the mitotic Bloom's syndrome DNA helicase in response to ionizing radiation." J Biol Chem 277(8): 6280-6.
Edelmann, W., P. E. Cohen, et al. (1996). "Meiotic pachytene arrest in MLH1-deficient mice." Cell 85(7): 1125-34.
Edelmann, W., P. E. Cohen, et al. (1999). "Mammalian MutS homologue 5 is required for chromosome pairing in meiosis." Nat Genet 21(1): 123-7.

Edelmann, W., A. Umar, et al. (2000). "The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression." Cancer Res 60(4): 803-7.
Edelmann, W., K. Yang, et al. (1997). "Mutation in the mismatch repair gene Msh6 causes cancer susceptibility." Cell 91(4): 467-77.
Eizirik, E., W. J. Murphy, et al. (2001). "Molecular dating and biogeography of the early placental mammal radiation." J Hered 92(2): 212-9.
Ellis, N. A., S. Ciocci, et al. (1998). "The Ashkenazic Jewish Bloom syndrome mutation blmAsh is present in non-Jewish Americans of Spanish ancestry." Am J Hum Genet 63(6): 1685-93.
Ellis, N. A., J. Groden, et al. (1995). "The Bloom's syndrome gene product is homologous to RecQ helicases." Cell 83(4): 655-66.
Ellis, N. A., A. M. Roe, et al. (1994). "Linkage disequilibrium between the FES, D15S127, and BLM loci in Ashkenazi Jews with Bloom syndrome." Am J Hum Genet 55(3): 453-60.
Emmons, S. W., L. Yesner, et al. (1983). "Evidence for a transposon in Caenorhabditis elegans." Cell 32(1): 55-65.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature 292(5819): 154-6.
Evans, M. J., T. von Hahn, et al. (2007). "Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry." Nature 446(7137): 801-5.
Fairchild, C. R., J. Maybaum, et al. (1986). "Concurrent unilateral chromatid damage and DNA strand breakage in response to 6-thioguanine treatment." Biochem Pharmacol 35(20): 3533-41.
Fang, J. Y., R. Lu, et al. (2006). "Regulation of hMSH2 and hMLH1 expression in the human colon cancer cell line SW1116 by DNA methyltransferase 1." Cancer Lett 233(1): 124-30.
Favor, J., A. Neuhauser-Klaus, et al. (1991). "The induction of forward and reverse specific-locus mutations and dominant cataract mutations in spermatogonia of treated strain DBA/2 mice by ethylnitrosourea." Mutat Res 249(2): 293-300.
Feng, W. Y., E. H. Lee, et al. (1991). "Recombinagenic processing of UV-light photoproducts in nonreplicating phage DNA by the Escherichia coli methyl-directed mismatch repair system." Genetics 129(4): 1007-20.
Festing, M. F., E. M. Simpson, et al. (1999). "Revised nomenclature for strain 129 mice." Mamm Genome 10(8): 836.
Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature 391(6669): 806-11.
Fishel, R., M. K. Lescoe, et al. (1993). "The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer." Cell 75(5): 1027-38.
Flores-Rozas, H. and R. D. Kolodner (1998). "The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations." Proceedings of the National Academy of Sciences of the United States of America 95(21): 12404-12409.
Flores-Rozas, H. and R. D. Kolodner (1998). "The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations." Proc Natl Acad Sci U S A 95(21): 12404-9.
Folger, K. R., E. A. Wong, et al. (1982). "Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules." Mol Cell Biol 2(11): 1372-87.
Forsburg, S. L. (2001). "The art and design of genetic screens: yeast." Nat Rev Genet 2(9): 659-68.
Fox, J. G. (2007). The Mouse in Biomedical Research, Academic Press Inc.,U.S.
Frankenberg-Schwager, M. (1990). "Induction, repair and biological relevance of radiation-induced DNA lesions in eukaryotic cells." Radiat Environ Biophys 29(4): 273-92.
Fraser, M. J., L. Cary, et al. (1995). "Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA." Virology 211(2): 397-407.
Fraser, M. J., T. Ciszczon, et al. (1996). "Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera." Insect Mol Biol 5(2): 141-51.
Frengen, E., D. Weichenhan, et al. (1999). "A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites." Genomics 58(3): 250-3.
Friedberg, E. C., G. C. Walker, et al. (2005). DNA Repair and Mutagenesis, ASM Press, American Society for Microbiology.
Friedrich, G. and P. Soriano (1991). "Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice." Genes Dev 5(9): 1513-23.
Gangloff, S., J. P. McDonald, et al. (1994). "The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase." Mol Cell Biol 14(12): 8391-8.
Garkavtsev, I. V., N. Kley, et al. (2001). "The Bloom syndrome protein interacts and cooperates with p53 in regulation of transcription and cell growth control." Oncogene 20(57): 8276-80.
Genschel, J., L. R. Bazemore, et al. (2002). "Human exonuclease I is required for 5' and 3' mismatch

repair." J Biol Chem 277(15): 13302-11.
Genschel, J., S. J. Littman, et al. (1998). "Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha." J Biol Chem 273(31): 19895-901.
German, J. (1969). "Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients." Am J Hum Genet 21(2): 196-227.
German, J. (1995). "Bloom's syndrome." Dermatol Clin 13(1): 7-18.
German, J. (1997). "Bloom's syndrome. XX. The first 100 cancers." Cancer Genet Cytogenet 93(1): 100-6.
German, J., A. M. Roe, et al. (1994). "Bloom syndrome: an analysis of consanguineous families assigns the locus mutated to chromosome band 15q26.1." Proc Natl Acad Sci U S A 91(14): 6669-73.
German, J. L., 3rd and E. Passarge (1990). "Bloom's Syndrome Registry." Int J Dermatol 29(3): 233-4.
Glickman, B. W. (1979). "Spontaneous mutagenesis in Escherichia coli strains lacking 6-methyladenine residues in their DNA: an altered mutational spectrum in dam- mutants." Mutat Res 61(2): 153-62.
Goodhead, D. T. (1989). "The initial physical damage produced by ionizing radiations." Int J Radiat Biol 56(5): 623-34.
Goodwin, N. C., Y. Ishida, et al. (2001). "DelBank: a mouse ES-cell resource for generating deletions." Nat Genet 28(4): 310-1.
Goss, K. H., M. A. Risinger, et al. (2002). "Enhanced tumor formation in mice heterozygous for Blm mutation." Science 297(5589): 2051-3.
Gossler, A., A. L. Joyner, et al. (1989). "Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes." Science 244(4903): 463-5.
Goulaouic, H., T. Roulon, et al. (1999). "Purification and characterization of human DNA topoisomerase IIIalpha." Nucleic Acids Res 27(12): 2443-50.
Gray, S. L., K. J. Cummings, et al. (2001). "Targeted disruption of the pituitary adenylate cyclase-activating polypeptide gene results in early postnatal death associated with dysfunction of lipid and carbohydrate metabolism." Mol Endocrinol 15(10): 1739-47.
Green, E. L., Roderick, T. H. (1966). Radiation Genetics. Biology of the Laboratory Mouse. . New York, McGraw-Hill: 165-185.
Griffin, S., P. Branch, et al. (1994). "DNA mismatch binding and incision at modified guanine bases by extracts of mammalian cells: implications for tolerance to DNA methylation damage." Biochemistry 33(16): 4787-93.
Grimm, S. (2004). "The art and design of genetic screens: mammalian culture cells." Nat Rev Genet 5(3): 179-89.
Gruber, S. B., N. A. Ellis, et al. (2002). "BLM heterozygosity and the risk of colorectal cancer." Science 297(5589): 2013.
Gu, L., Y. Hong, et al. (1998). "ATP-dependent interaction of human mismatch repair proteins and dual role of PCNA in mismatch repair." Nucleic Acids Research 26(5): 1173-1178.
Guenet, J. L. (2004). "Chemical mutagenesis of the mouse genome: an overview." Genetica 122(1): 9-24.
Guerra, B., O. G. Issinger, et al. (2003). "Modulation of human checkpoint kinase Chk1 by the regulatory beta-subunit of protein kinase CK2." Oncogene 22(32): 4933-42.
Guo, G. (2004). Recessive genetic screen for mismatch repair components in BLM-deficient ES cells. The Wellcome Trust Sanger Institute, University of Cambridge.
Guo, G., W. Wang, et al. (2004). "Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells." Nature 429(6994): 891-5.
Gurin, C. C., M. G. Federici, et al. (1999). "Causes and consequences of microsatellite instability in endometrial carcinoma." Cancer Res 59(2): 462-6.
Habraken, Y., L. Prakash, et al. (1998). "ATP-dependent assembly of a ternary complex consisting of a DNA mismatch and the yeast MSH2-MSH6 and MLH1-PMS1 protein complexes." Journal of Biological Chemistry 273(16): 9837-9841.
Habraken, Y., P. Sung, et al. (1996). "Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3." Current Biology 6(9): 1185-1187.
Habraken, Y., P. Sung, et al. (1997). "Enhancement of MSH2-MSH3-mediated mismatch recognition by the yeast MLH1-PMS1 complex." Current Biology 7(10): 790-793.
Haldane, J. (1915). "Reduplication in mice." Journal Genetics 5: 133-135.
Handler, A. M., S. P. Gomez, et al. (1993). "A functional analysis of the P-element gene-transfer vector in insects." Arch Insect Biochem Physiol 22(3-4): 373-84.
Hansen, J., T. Floss, et al. (2003). "A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome." Proc Natl Acad Sci U S A 100(17): 9918-22.
Harfe, B. D. and S. Jinks-Robertson (2000). "Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae." Genetics 156(2): 571-8.
Hawn, M. T., A. Umar, et al. (1995). "Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint." Cancer Res 55(17): 3721-5.
Hayakawa, T., K. Yusa, et al. (2006). "Bloom's syndrome gene-deficient phenotype in mouse primary cells

induced by a modified tetracycline-controlled trans-silencer." Gene 369: 80-9.
Heaney, J. D., A. N. Rettew, et al. (2004). "Tissue-specific expression of a BAC transgene targeted to the Hprt locus in mouse embryonic stem cells." Genomics 83(6): 1072-82.
Hedges, S. B., P. H. Parker, et al. (1996). "Continental breakup and the ordinal diversification of birds and mammals." Nature 381(6579): 226-9.
Heinz-Albert Becker, W.-E. L. (2001). Transpsons: Eukaryotic. Bencyclopedia of life sciences, NPG.
Herceg, Z., W. Hulla, et al. (2001). "Disruption of Trrap causes early embryonic lethality and defects in cell cycle progression." Nat Genet 29(2): 206-11.
Herman, J. G. and S. B. Baylin (2003). "Gene silencing in cancer in association with promoter hypermethylation." N Engl J Med 349(21): 2042-54.
Herman, J. G., J. R. Graff, et al. (1996). "Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands." Proc Natl Acad Sci U S A 93(18): 9821-6.
Herman, J. G., A. Umar, et al. (1998). "Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma." Proc Natl Acad Sci U S A 95(12): 6870-5.
Hill, S. A., M. M. Stahl, et al. (1997). "Single-strand DNA intermediates in phage lambda's Red recombination pathway." Proc Natl Acad Sci U S A 94(7): 2951-6.
Hoess, R. H., A. Wierzbicki, et al. (1986). "The role of the loxP spacer region in P1 site-specific recombination." Nucleic Acids Res 14(5): 2287-300.
Hollingsworth, N. M., L. Ponte, et al. (1995). "MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in Saccharomyces cerevisiae but not mismatch repair." Genes Dev 9(14): 1728-39.
Horie, K., K. Yusa, et al. (2003). "Characterization of Sleeping Beauty transposition and its application to genetic screening in mice." Mol Cell Biol 23(24): 9189-207.
Hrabe de Angelis, M. H., H. Flaswinkel, et al. (2000). "Genome-wide, large-scale production of mutant mice by ENU mutagenesis." Nat Genet 25(4): 444-7.
Hsieh, P. (2001). "Molecular mechanisms of DNA mismatch repair." Mutat Res 486(2): 71-87.
Hu, S., E. Otsubo, et al. (1975). "Electron microscope heteroduplex studies of sequence relations among bacterial plasmids: identification and mapping of the insertion sequences IS1 and IS2 in F and R plasmids." J Bacteriol 122(2): 764-75.
Huang, J., N. Papadopoulos, et al. (1996). "APC mutations in colorectal tumors with mismatch repair deficiency." Proc Natl Acad Sci U S A 93(17): 9049-54.
Hughes, M. J. and J. Jiricny (1992). "The purification of a human mismatch-binding protein and identification of its associated ATPase and helicase activities." Journal of Biological Chemistry 267(33): 23876-23882.
Hunter, N. and R. H. Borts (1997). "Mlh1 is unique among mismatch repair proteins in its ability to promote crossing-over during meiosis." Genes Dev 11(12): 1573-82.
Hutchinson, F. (1985). "Chemical changes induced in DNA by ionizing radiation." Prog Nucleic Acid Res Mol Biol 32: 115-54.
Iaccarino, I., G. Marra, et al. (1998). "hMSH2 and hMSH6 play distinct roles in mismatch binding and contribute differently to the ATPase activity of hMutS?" EMBO Journal 17(9): 2677-2686.
Ionov, Y., M. A. Peinado, et al. (1993). "Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis." Nature 363(6429): 558-61.
Ishida, Y. and P. Leder (1999). "RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells." Nucleic Acids Res 27(24): e35.
Ishikawa, J., T. Kaisho, et al. (1995). "Molecular cloning and chromosomal mapping of a bone marrow stromal cell surface gene, BST2, that may be involved in pre-B-cell growth." Genomics 26(3): 527-34.
Ivics, Z., P. B. Hackett, et al. (1997). "Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells." Cell 91(4): 501-10.
Iyer, R. R., A. Pluciennik, et al. (2006). "DNA mismatch repair: functions and mechanisms." Chem Rev 106(2): 302-23.
Izsvak, Z., Z. Ivics, et al. (2000). "Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates." J Mol Biol 302(1): 93-102.
Jacobson, J. W., M. M. Medhora, et al. (1986). "Molecular structure of a somatically unstable transposable element in Drosophila." Proc Natl Acad Sci U S A 83(22): 8684-8.
Jaenisch, R. (1976). "Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus." Proc Natl Acad Sci U S A 73(4): 1260-4.
Jaenisch, R., D. Jahner, et al. (1981). "Chromosomal position and activation of retroviral genomes inserted into the germ line of mice." Cell 24(2): 519-29.
Jiricny, J. (1998). "Replication errors: Cha(lle)nging the genome." EMBO Journal 17(22): 6427-6436.
Jones, M., R. Wagner, et al. (1987). "Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence." Genetics 115(4): 605-10.
Jones, P. A. and P. W. Laird (1999). "Cancer epigenetics comes of age." Nat Genet 21(2): 163-7.

Jorgensen, E. M. and S. E. Mango (2002). "The art and design of genetic screens: caenorhabditis elegans." Nat Rev Genet 3(5): 356-69.
Junop, M. S., G. Obmolova, et al. (2001). "Composite active site of an ABC ATPase: MutS uses ATP to verify mismatch recognition and authorize DNA repair." Mol Cell 7(1): 1-12.
Justice, M. J. and V. C. Bode (1986). "Induction of new mutations in a mouse t-haplotype using ethylnitrosourea mutagenesis." Genet Res 47(3): 187-92.
Justice, M. J., J. K. Noveroske, et al. (1999). "Mouse ENU mutagenesis." Hum Mol Genet 8(10): 1955-63.
Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science 258(5083): 818-21.
Karmakar, P., M. Seki, et al. (2006). "BLM is an early responder to DNA double-strand breaks." Biochem Biophys Res Commun 348(1): 62-9.
Karow, J. K., R. K. Chakraverty, et al. (1997). "The Bloom's syndrome gene product is a 3'-5' DNA helicase." J Biol Chem 272(49): 30611-4.
Karran, P. and M. Bignami (1992). "Self-destruction and tolerance in resistance of mammalian cells to alkylation damage." Nucleic Acids Res 20(12): 2933-40.
Karran, P. and R. Hampson (1996). "Genomic instability and tolerance to alkylating agents." Cancer Surveys 28: 69-85.
Karran, P. and M. G. Marinus (1982). "Mismatch correction at O6-methylguanine residues in E. coli DNA." Nature 296(5860): 868-9.
Karran, P., J. Offman, et al. (2003). "Human mismatch repair, drug-induced DNA damage, and secondary cancer." Biochimie 85(11): 1149-60.
Kasarskis, A., K. Manova, et al. (1998). "A phenotype-based screen for embryonic lethal mutations in the mouse." Proc Natl Acad Sci U S A 95(13): 7485-90.
Kat, A., W. G. Thilly, et al. (1993). "An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair." Proc Natl Acad Sci U S A 90(14): 6424-8.
Katzenellenbogen, R. A., S. B. Baylin, et al. (1999). "Hypermethylation of the DAP-kinase CpG island is a common alteration in B-cell malignancies." Blood 93(12): 4347-53.
Kaufman, P. D. and D. C. Rio (1992). "P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor." Cell 69(1): 27-39.
Kawate, H., K. Sakumi, et al. (1998). "Separation of killing and tumorigenic effects of an alkylating agent in mice defective in two of the DNA repair genes." Proc Natl Acad Sci U S A 95(9): 5116-20.
Kiko, H., E. Niggemann, et al. (1979). "Physical mapping of the restriction fragments obtained from bacteriophage T4 dC-DNA with the restriction endonucleases SmaI, KpnI and BglII." Mol Gen Genet 172(3): 303-12.
Kile, B. T. and D. J. Hilton (2005). "The art and design of genetic screens: mouse." Nat Rev Genet 6(7): 557-67.
Kim, H. C., J. C. Kim, et al. (2005). "Aberrant CpG island methylation in early-onset sporadic gastric carcinoma." J Cancer Res Clin Oncol 131(11): 733-40.
Kim, M., B. N. Trinh, et al. (2004). "Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells." Nucleic Acids Res 32(19): 5742-9.
King, W., M. D. Patel, et al. (1985). "Insertion mutagenesis of embryonal carcinoma cells by retroviruses." Science 228(4699): 554-8.
Kissil, J. L., E. Feinstein, et al. (1997). "DAP-kinase loss of expression in various carcinoma and B-cell lymphoma cell lines: possible implications for role as tumor suppressor gene." Oncogene 15(4): 403-7.
Kitao, S., A. Shimamoto, et al. (1999). "Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome." Nat Genet 22(1): 82-4.
Kneitz, B., P. E. Cohen, et al. (2000). "MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice." Genes Dev 14(9): 1085-97.
Koc, A., L. J. Wheeler, et al. (2003). "Replication-independent MCB gene induction and deoxyribonucleotide accumulation at G1/S in Saccharomyces cerevisiae." J Biol Chem 278(11): 9345-52.
Koi, M., A. Umar, et al. (1994). "Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation." Cancer Res 54(16): 4308-12.
Koller, B. H., L. J. Hagemann, et al. (1989). "Germ-line transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells." Proc Natl Acad Sci U S A 86(22): 8927-31.
Kolodner, R. (1996). "Biochemistry and genetics of eukaryotic mismatch repair." Genes and Development 10(12): 1433-1442.
Kolodner, R. D. and G. T. Marsischky (1999). "Eukaryotic DNA mismatch repair." Current Opinion in Genetics and Development 9(1): 89-96.
Korn, R., M. Schoor, et al. (1992). "Enhancer trap integrations in mouse embryonic stem cells give rise to

staining patterns in chimaeric embryos with a high frequency and detect endogenous genes." Mech Dev 39(1-2): 95-109.
Kothary, R., S. Clapoff, et al. (1988). "A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube." Nature 335(6189): 435-7.
Kramer, B., W. Kramer, et al. (1984). "Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli." Cell 38(3): 879-87.
Kuehn, M. R., A. Bradley, et al. (1987). "A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice." Nature 326(6110): 295-8.
Kuhn, E. M. and E. Therman (1986). "Cytogenetics of Bloom's syndrome." Cancer Genet Cytogenet 22(1): 1-18.
Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annu Rev Biochem 74: 681-710.
Kushi, A., K. Edamura, et al. (1998). "Generation of mutant mice with large chromosomal deletion by use of irradiated ES cells--analysis of large deletion around hprt locus of ES cell." Mamm Genome 9(4): 269-73.
Lahue, R. S., S. S. Su, et al. (1987). "Requirement for d(GATC) sequences in Escherichia coli mutHLS mismatch correction." Proc Natl Acad Sci U S A 84(6): 1482-6.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.
Langin, T., P. Capy, et al. (1995). "The transposable element impala, a fungal member of the Tc1-mariner superfamily." Mol Gen Genet 246(1): 19-28.
Langland, G., J. Elliott, et al. (2002). "The BLM helicase is necessary for normal DNA double-strand break repair." Cancer Res 62(10): 2766-70.
Learn, B. A. and R. H. Grafstrom (1989). "Methyl-directed repair of frameshift heteroduplexes in cell extracts from Escherichia coli." J Bacteriol 171(12): 6473-81.
Lee, E. C., D. Yu, et al. (2001). "A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA." Genomics 73(1): 56-65.
Lefebvre, L., N. Dionne, et al. (2001). "Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity." Nat Genet 27(3): 257-8.
Legerski, R. and C. Peterson (1992). "Expression cloning of a human DNA repair gene involved in xeroderma pigmentosum group C." Nature 359(6390): 70-3.
Lett, J. T. (1990). "Damage to DNA and chromatin structure from ionizing radiations, and the radiation sensitivities of mammalian cells." Prog Nucleic Acid Res Mol Biol 39: 305-52.
Li, G. M., H. Wang, et al. (1996). "Human MutSalpha specifically binds to DNA containing aminofluorene and acetylaminofluorene adducts." J Biol Chem 271(39): 24084-8.
Lieber, M. R., Y. Ma, et al. (2003). "Mechanism and regulation of human non-homologous DNA end-joining." Nat Rev Mol Cell Biol 4(9): 712-20.
Liebl, F. L., K. M. Werner, et al. (2006). "Genome-wide P-element screen for Drosophila synaptogenesis mutants." J Neurobiol 66(4): 332-47.
Lindblom, A., P. Tannergard, et al. (1993). "Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer." Nat Genet 5(3): 279-82.
Lindor, N. M., Y. Furuichi, et al. (2000). "Rothmund-Thomson syndrome due to RECQ4 helicase mutations: report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome." Am J Med Genet 90(3): 223-8.
Ling, Y. H., J. Y. Chan, et al. (1992). "Consequences of 6-thioguanine incorporation into DNA on polymerase, ligase, and endonuclease reactions." Mol Pharmacol 42(5): 802-7.
Lipkin, S. M., P. B. Moens, et al. (2002). "Meiotic arrest and aneuploidy in MLH3-deficient mice." Nat Genet 31(4): 385-90.
Lipkin, S. M., V. Wang, et al. (2000). "MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability." Nat Genet 24(1): 27-35.
Little, C. C., Bagg, H. J. (1924). "The occurrencce of four inheritable morphological variations in mice and their possible relation to treatment iwth X-rays." J. Exp. Zool. 41: 45-92.
Liu, K., S. Hipkens, et al. (2006). "Recombinase-mediated cassette exchange to rapidly and efficiently generate mice with human cardiac sodium channels." Genesis 44(11): 556-64.
Liu, P., N. A. Jenkins, et al. (2002). "Efficient Cre-loxP-induced mitotic recombination in mouse embryonic stem cells." Nat Genet 30(1): 66-72.
Liu, P., N. A. Jenkins, et al. (2003). "A highly efficient recombineering-based method for generating conditional knockout mutations." Genome Res 13(3): 476-84.
Lobel, L. I., M. Patel, et al. (1985). "Construction and recovery of viable retroviral genomes carrying a bacterial suppressor transfer RNA gene." Science 228(4697): 329-32.
Lobrich, M., P. K. Cooper, et al. (1996). "Non-random distribution of DNA double-strand breaks induced by particle irradiation." Int J Radiat Biol 70(5): 493-503.
Loots, G. G., R. M. Locksley, et al. (2000). "Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons." Science 288(5463): 136-40.

Lopez de Saro, F. J. and M. O'Donnell (2001). "Interaction of the beta sliding clamp with MutS, ligase, and DNA polymerase I." Proc Natl Acad Sci U S A 98(15): 8376-80.
Lu, A. L. (1987). "Influence of GATC sequences on Escherichia coli DNA mismatch repair in vitro." J Bacteriol 169(3): 1254-9.
Lu, A. L., S. Clark, et al. (1983). "Methyl-directed repair of DNA base-pair mismatches in vitro." Proc Natl Acad Sci U S A 80(15): 4639-43.
Lund, A. H., G. Turner, et al. (2002). "Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice." Nat Genet 32(1): 160-5.
Luo, G., Z. Ivics, et al. (1998). "Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells." Proc Natl Acad Sci U S A 95(18): 10769-73.
Luo, G., I. M. Santoro, et al. (2000). "Cancer predisposition caused by elevated mitotic recombination in Bloom mice." Nat Genet 26(4): 424-9.
Lynch, H. T. and A. de la Chapelle (1999). "Genetic susceptibility to non-polyposis colorectal cancer." J Med Genet 36(11): 801-18.
Lynch, H. T., T. C. Smyrk, et al. (1993). "Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review." Gastroenterology 104(5): 1535-49.
Lyons, S. M. and P. F. Schendel (1984). "Kinetics of methylation in Escherichia coli K-12." J Bacteriol 159(1): 421-3.
Maa, M. C., J. C. Lee, et al. (2007). "Eps8 facilitates cellular growth and motility of colon cancer cells by increasing the expression and activity of focal adhesion kinase." J Biol Chem 282(27): 19399-409.
Malling, H. V. and L. R. Valcovic (1977). "A biochemical specific locus mutation system in mice." Arch Toxicol 38(1-2): 45-51.
Mann, R., R. C. Mulligan, et al. (1983). "Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus." Cell 33(1): 153-9.
Markowitz, S., J. Wang, et al. (1995). "Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability." Science 268(5215): 1336-8.
Marsischky, G. T., N. Filosi, et al. (1996). "Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair." Genes and Development 10(4): 407-420.
Martin, R. H., A. Rademaker, et al. (1994). "Chromosomal breakage in human spermatozoa, a heterozygous effect of the Bloom syndrome mutation." Am J Hum Genet 55(6): 1242-6.
Matsuura, S., J. Komatsu, et al. (2001). "Real-time observation of a single DNA digestion by lambda exonuclease under a fluorescence microscope field." Nucleic Acids Res 29(16): E79.
Mayo, K. E., W. Vale, et al. (1983). "Expression-cloning and sequence of a cDNA encoding human growth hormone-releasing factor." Nature 306(5938): 86-8.
McClintock, B. (1948). Mutable loci in maize., Carnegie Institute of Washington Year Book. 47: 155-169.
McClintock, B. (1949). Mutable loci in maize., Carnegie Institute of Washington Year Book. 48: 142-154.
McClintock, B. (1950). "The origin and behavior of mutable loci in maize." Proc Natl Acad Sci U S A 36(6): 344-55.
McDaniel, L. D., N. Chester, et al. (2003). "Chromosome instability and tumor predisposition inversely correlate with BLM protein levels." DNA Repair (Amst) 2(12): 1387-404.
McDaniel, L. D. and R. A. Schultz (1992). "Elevated sister chromatid exchange phenotype of Bloom syndrome cells is complemented by human chromosome 15." Proc Natl Acad Sci U S A 89(17): 7968-72.
McKinnon, P. J. and K. W. Caldecott (2007). "DNA Strand Break Repair and Human Genetic Disease." Annu Rev Genomics Hum Genet 8: 37-55.
McMahon, A. P. and A. Bradley (1990). "The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain." Cell 62(6): 1073-85.
Menoyo, A., H. Alazzouzi, et al. (2001). "Somatic mutations in the DNA damage-response genes ATR and CHK1 in sporadic stomach tumors with microsatellite instability." Cancer Res 61(21): 7727-30.
Michael, B. D. and P. O'Neill (2000). "Molecular biology. A sting in the tail of electron tracks." Science 287(5458): 1603-4.
Mikkers, H., J. Allen, et al. (2002). "High-throughput retroviral tagging to identify components of specific signaling pathways in cancer." Nat Genet 32(1): 153-9.
Millar, C. B., J. Guy, et al. (2002). "Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice." Science 297(5580): 403-5.
Mitchell, R. S., B. F. Beitzel, et al. (2004). "Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences." PLoS Biol 2(8): E234.
Modrich, P. (1997). "Strand-specific mismatch repair in mammalian cells." J Biol Chem 272(40): 24727-30.
Modrich, P. and R. Lahue (1996). "Mismatch repair in replication fidelity, genetic recombination, and cancer biology." Annu Rev Biochem 65: 101-33.
Monk, R. J., F. G. Malik, et al. (1992). "Direct determination of the point mutation rate of a murine retrovirus." J Virol 66(6): 3683-9.
Mortensen, R. M., D. A. Conner, et al. (1992). "Production of homozygous mutant ES cells with a single

targeting construct." Mol Cell Biol 12(5): 2391-5.
Muller, H. J. (1927). "Artificial Transmutation of the Gene." Science 66(1699): 84-87.
Muniyappa, K. and C. M. Radding (1986). "The homologous recombination system of phage lambda. Pairing activities of beta protein." J Biol Chem 261(16): 7472-8.
Murphy, K. C. (1991). "Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme." J Bacteriol 173(18): 5808-21.
Muzumdar, M. D., L. Luo, et al. (2007). "Modeling sporadic loss of heterozygosity in mice by using mosaic analysis with double markers (MADM)." Proc Natl Acad Sci U S A 104(11): 4495-500.
Mythili, E., K. A. Kumar, et al. (1996). "Characterization of the DNA-binding domain of beta protein, a component of phage lambda red-pathway, by UV catalyzed cross-linking." Gene 182(1-2): 81-7.
Nackerdien, Z., G. Rao, et al. (1991). "Chemical nature of DNA-protein cross-links produced in mammalian chromatin by hydrogen peroxide in the presence of iron or copper ions." Biochemistry 30(20): 4873-9.
Nagy, A., J. Rossant, et al. (1993). "Derivation of completely cell culture-derived mice from early-passage embryonic stem cells." Proc Natl Acad Sci U S A 90(18): 8424-8.
Nakayama, H. (2002). "RecQ family helicases: roles as tumor suppressor proteins." Oncogene 21(58): 9008-21.
Ng, H. H. and A. Bird (1999). "DNA methylation and chromatin modification." Curr Opin Genet Dev 9(2): 158-63.
Ni, T. T., G. T. Marsischky, et al. (1999). "MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxo-guanine in S. cerevisiae." Mol Cell 4(3): 439-44.
Nicholson, A., M. Hendrix, et al. (2000). "Regulation of mitotic homeologous recombination in yeast. Functions of mismatch repair and nucleotide excision repair genes." Genetics 154(1): 133-46.
Nobrega, M. A., I. Ovcharenko, et al. (2003). "Scanning human gene deserts for long-range enhancers." Science 302(5644): 413.
Nobrega, M. A., Y. Zhu, et al. (2004). "Megabase deletions of gene deserts result in viable mice." Nature 431(7011): 988-93.
Nolan, P. M., J. Peters, et al. (2000). "A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse." Nat Genet 25(4): 440-3.
Nord, A. S., P. J. Chang, et al. (2006). "The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse." Nucleic Acids Res 34(Database issue): D642-8.
Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." Nature 287(5785): 795-801.
O'Brien, S. J., M. Menotti-Raymond, et al. (1999). "The promise of comparative genomics in mammals." Science 286(5439): 458-62, 479-81.
O'Hare, K. and G. M. Rubin (1983). "Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome." Cell 34(1): 25-35.
Oeltjen, J. C., T. M. Malley, et al. (1997). "Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains." Genome Res 7(4): 315-29.
Ohtoshi, A., A. Bradley, et al. (2006). "Generation and maintenance of Dmbx1 gene-targeted mutant alleles." Mamm Genome 17(7): 744-50.
Oostlander, A. E., G. A. Meijer, et al. (2004). "Microarray-based comparative genomic hybridization and its applications in human genetics." Clin Genet 66(6): 488-95.
Paddison, P. J., J. M. Silva, et al. (2004). "A resource for large-scale RNA-interference-based screens in mammals." Nature 428(6981): 427-31.
Page, D. R. and U. Grossniklaus (2002). "The art and design of genetic screens: Arabidopsis thaliana." Nat Rev Genet 3(2): 124-36.
Pan, B. F. and J. A. Nelson (1990). "Characterization of the DNA damage in 6-thioguanine-treated cells." Biochem Pharmacol 40(5): 1063-9.
Pan, B. F., T. S. Priebe, et al. (1992). "Mechanisms of resistance to 6-thioguanine in a murine pancreatic tumor." Cancer Chemother Pharmacol 29(6): 471-4.
Papadopoulos, N., N. C. Nicolaides, et al. (1994). "Mutation of a mutL homolog in hereditary colon cancer." Science 263(5153): 1625-9.
Parekh, V., A. McEwen, et al. (2004). "Defective extraembryonic angiogenesis in mice lacking LBP-1a, a member of the grainyhead family of transcription factors." Mol Cell Biol 24(16): 7113-29.
Parker, B. O. and M. G. Marinus (1992). "Repair of DNA heteroduplexes containing small heterologous sequences in Escherichia coli." Proc Natl Acad Sci U S A 89(5): 1730-4.
Parsons, B. L. (2003). "MED1: a central molecule for maintenance of genome integrity and response to DNA damage." Proc Natl Acad Sci U S A 100(25): 14601-2.
Parsons, R., G. M. Li, et al. (1993). "Hypermutability and mismatch repair deficiency in RER+ tumor cells." Cell 75(6): 1227-36.
Patton, E. E. and L. I. Zon (2001). "The art and design of genetic screens: zebrafish." Nat Rev Genet

2(12): 956-66.
Paull, T. T., E. P. Rogakou, et al. (2000). "A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage." Curr Biol 10(15): 886-95.
Pedrazzi, G., C. Perrera, et al. (2001). "Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1." Nucleic Acids Res 29(21): 4378-86.
Peltomaki, P., L. A. Aaltonen, et al. (1993). "Genetic mapping of a locus predisposing to human colorectal cancer." Science 260(5109): 810-2.
Pennacchio, L. A. and E. M. Rubin (2001). "Genomic strategies to identify mammalian regulatory sequences." Nat Rev Genet 2(2): 100-9.
Petit, M. A., J. Dimpfl, et al. (1991). "Control of large chromosomal duplications in Escherichia coli by the mismatch repair system." Genetics 129(2): 327-32.
Plasterk, R. H., Z. Izsvak, et al. (1999). "Resident aliens: the Tc1/mariner superfamily of transposable elements." Trends Genet 15(8): 326-32.
Prolla, T. A., S. M. Baker, et al. (1998). "Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair." Nat Genet 18(3): 276-9.
Prolla, T. A., D. M. Christie, et al. (1994). "Dual requirement in yeast DNA mismatch repair for MLH1 and PMS1, two homologs of the bacterial mutL gene." Mol Cell Biol 14(1): 407-15.
Prolla, T. A., Q. Pang, et al. (1994). "MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast." Science 265(5175): 1091-3.
Prosser, H. M. R., Agnieszka K.; Steel, Karen P.; Bradley, Allan (2007 unpublished). Mosaic complementation demonstrates a regulatory role for myosin VIIa in stereocilia actin dynamics.
Pukkila, P. J., J. Peterson, et al. (1983). "Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in Escherichia coli." Genetics 104(4): 571-82.
Puranam, K. L. and P. J. Blackshear (1994). "Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ." J Biol Chem 269(47): 29838-45.
Radding, C. M. and D. M. Carter (1971). "The role of exonuclease and beta protein of phage lambda in genetic recombination. 3. Binding to deoxyribonucleic acid." J Biol Chem 246(8): 2513-8.
Ramirez-Solis, R., A. C. Davis, et al. (1993). "Gene targeting in embryonic stem cells." Methods Enzymol 225: 855-78.
Rampino, N., H. Yamamoto, et al. (1997). "Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype." Science 275(5302): 967-9.
Rao, V. A., C. Conti, et al. (2007). "Endogenous gamma-H2AX-ATM-Chk2 checkpoint activation in Bloom's syndrome helicase deficient cells is related to DNA replication arrested forks." Mol Cancer Res 5(7): 713-24.
Rayssiguier, C., D. S. Thaler, et al. (1989). "The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants." Nature 342(6248): 396-401.
Reaume, A. G., P. A. de Sousa, et al. (1995). "Cardiac malformation in neonatal mice lacking connexin43." Science 267(5205): 1831-4.
Redon, R., S. Ishikawa, et al. (2006). "Global variation in copy number in the human genome." Nature 444(7118): 444-54.
Reitmair, A. H., R. Schmits, et al. (1995). "MSH2 deficient mice are viable and susceptible to lymphoid tumours." Nat Genet 11(1): 64-70.
Riccio, A., L. A. Aaltonen, et al. (1999). "The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability." Nat Genet 23(3): 266-8.
Riley, P. A. (1994). "Free radicals in biology: oxidative stress and the effects of ionizing radiation." Int J Radiat Biol 65(1): 27-33.
Rinchik, E. M. and D. A. Carpenter (1999). "N-ethyl-N-nitrosourea mutagenesis of a 6- to 11-cM subregion of the Fah-Hbb interval of mouse chromosome 7: Completed testing of 4557 gametes and deletion mapping and complementation analysis of 31 mutations." Genetics 152(1): 373-83.
Robert, M. F., S. Morin, et al. (2003). "DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells." Nat Genet 33(1): 61-5.
Robertson, E., A. Bradley, et al. (1986). "Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector." Nature 323(6087): 445-8.
Rogakou, E. P., D. R. Pilch, et al. (1998). "DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139." J Biol Chem 273(10): 5858-68.
Roos, W., M. Baumgartner, et al. (2004). "Apoptosis triggered by DNA damage O6-methylguanine in human lymphocytes requires DNA replication and is mediated by p53 and Fas/CD95/Apo-1." Oncogene 23(2): 359-67.
Ross-Macdonald, P. and G. S. Roeder (1994). "Mutation of a meiosis-specific MutS homolog decreases crossing over but not mismatch correction." Cell 79(6): 1069-80.
Rowley, P. T. (2005). "Inherited susceptibility to colorectal cancer." Annu Rev Med 56: 539-54.
Rubin, G. M., M. G. Kidwell, et al. (1982). "The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations." Cell 29(3): 987-94.

Russell, W. L., E. M. Kelly, et al. (1979). "Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse." Proc Natl Acad Sci U S A 76(11): 5818-9.
Sansom, O. J., J. Zabkiewicz, et al. (2003). "MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine." Oncogene 22(46): 7130-6.
Santucci-Darmanin, S., S. Neyton, et al. (2002). "The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination." Hum Mol Genet 11(15): 1697-706.
Schimenti, J. C., B. J. Libby, et al. (2000). "Interdigitated deletion complexes on mouse chromosome 5 induced by irradiation of embryonic stem cells." Genome Res 10(7): 1043-50.
Schwartzberg, P. L., S. P. Goff, et al. (1989). "Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells." Science 246(4931): 799-803.
Searle, A. G. and C. V. Beechey (1974). "Cytogenetic effects of X-rays and fission neutrons in female mice." Mutat Res 24(2): 171-86.
Seki, M., H. Miyazawa, et al. (1994). "Molecular cloning of cDNA encoding human DNA helicase Q1 which has homology to Escherichia coli Rec Q helicase and localization of the gene at chromosome 12p12." Nucleic Acids Res 22(22): 4566-73.
Seki, T., M. Seki, et al. (1998). "Cloning of cDNA encoding a novel mouse DNA topoisomerase III (Topo IIIbeta) possessing negatively supercoiled DNA relaxing activity, whose message is highly expressed in the testis." J Biol Chem 273(44): 28553-6.
Selva, E. M., A. B. Maderazo, et al. (1997). "Differential effects of the mismatch repair genes MSH2 and MSH3 on homeologous recombination in Saccharomyces cerevisiae." Mol Gen Genet 257(1): 71-82.
Selva, E. M., L. New, et al. (1995). "Mismatch correction acts as a barrier to homeologous recombination in Saccharomyces cerevisiae." Genetics 139(3): 1175-88.
Selzer, R. R., T. A. Richmond, et al. (2005). "Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH." Genes Chromosomes Cancer 44(3): 305-19.
Shedlovsky, A., T. R. King, et al. (1988). "Saturation germ line mutagenesis of the murine t region including a lethal allele at the quaking locus." Proc Natl Acad Sci U S A 85(1): 180-4.
Shedlovsky, A., J. D. McDonald, et al. (1993). "Mouse models of human phenylketonuria." Genetics 134(4): 1205-10.
Shuman, H. A. and T. J. Silhavy (2003). "The art and design of genetic screens: Escherichia coli." Nat Rev Genet 4(6): 419-31.
Siepka, S. M., S. H. Yoo, et al. (2007). "Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression." Cell 129(5): 1011-23.
Silver, L. M. (1995). Mouse Genetics, Oxford University Press.
Simpson, E. M., C. C. Linder, et al. (1997). "Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice." Nat Genet 16(1): 19-27.
Skvorak, K., B. Vissel, et al. (2006). "Production of conditional point mutant knockin mice." Genesis 44(7): 345-53.
Smits, R., N. Hofland, et al. (2000). "Somatic Apc mutations are selected upon their capacity to inactivate the beta-catenin downregulating activity." Genes Chromosomes Cancer 29(3): 229-39.
Sonntag, C. V. (1987). Chemical Basis of Radiation Biology, Taylor & Francis.
Sonoda, E., M. S. Sasaki, et al. (1999). "Sister chromatid exchanges are mediated by homologous recombination in vertebrate cells." Mol Cell Biol 19(7): 5166-9.
Soultanas, P. and D. B. Wigley (2001). "Unwinding the 'Gordian knot' of helicase action." Trends Biochem Sci 26(1): 47-54.
Spradling, A. C. and G. M. Rubin (1982). "Transposition of cloned P elements into Drosophila germ line chromosomes." Science 218(4570): 341-7.
St Johnston, D. (2002). "The art and design of genetic screens: Drosophila melanogaster." Nat Rev Genet 3(3): 176-88.
Stanford, W. L., J. B. Cohn, et al. (2001). "Gene-trap mutagenesis: past, present and beyond." Nat Rev Genet 2(10): 756-68.
Stojic, L., N. Mojas, et al. (2004). "Mismatch repair-dependent G2 checkpoint induced by low doses of SN1 type methylating agents requires the ATR kinase." Genes Dev 18(11): 1331-44.
Strand, M., T. A. Prolla, et al. (1993). "Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair." Nature 365(6443): 274-6.
Su, H., X. Wang, et al. (2000). "Nested chromosomal deletions induced with retroviral vectors in mice." Nat Genet 24(1): 92-5.
Sugawara, A., K. Goto, et al. (2006). "Current status of chromosomal abnormalities in mouse embryonic stem cell lines used in Japan." Comp Med 56(1): 31-4.
Sugihara, K., N. Nakatsuji, et al. (1998). "Rac1 is required for the formation of three germ layers during gastrulation." Oncogene 17(26): 3427-33.

Swann, P. F., T. R. Waters, et al. (1996). "Role of postreplicative DNA mismatch repair in the cytotoxic action of thioguanine." Science 273(5278): 1109-11.
te Riele, H., E. R. Maandag, et al. (1992). "Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs." Proc Natl Acad Sci U S A 89(11): 5128-32.
te Riele, H., E. R. Maandag, et al. (1990). "Consecutive inactivation of both alleles of the pim-1 proto-oncogene by homologous recombination in embryonic stem cells." Nature 348(6302): 649-51.
Thibodeau, S. N., G. Bren, et al. (1993). "Microsatellite instability in cancer of the proximal colon." Science 260(5109): 816-9.
Thomas, J. W., C. LaMantia, et al. (1998). "X-ray-induced mutations in mouse embryonic stem cells." Proc Natl Acad Sci U S A 95(3): 1114-9.
Thomas, K. R. and M. R. Capecchi (1987). "Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells." Cell 51(3): 503-12.
Thompson, L. H., K. W. Brookman, et al. (1994). "Molecular cloning of the human nucleotide-excision-repair gene ERCC4." Proc Natl Acad Sci U S A 91(15): 6855-9.
Thompson, S., A. R. Clarke, et al. (1989). "Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells." Cell 56(2): 313-21.
Threadgill, D. W., A. Matin, et al. (1997). "SSLPs to map genetic differences between the 129 inbred strains and closed-colony, random-bred CD-1 mice." Mamm Genome 8(6): 441-2.
Tishkoff, D. X., A. L. Boerger, et al. (1997). "Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2." Proc Natl Acad Sci U S A 94(14): 7487-92.
Tran, P. T., J. A. Simon, et al. (2001). "Interactions of Exo1p with components of MutLalpha in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A 98(17): 9760-5.
Traut, H. and W. Scheid (1971). "The production of monosomic-trisomic individuals in Drosophila melanogaster by x-irradiation of immature oocytes." Mutat Res 13(4): 429-32.
Trinh, B. N., T. I. Long, et al. (2002). "DNA methyltransferase deficiency modifies cancer susceptibility in mice lacking DNA mismatch repair." Mol Cell Biol 22(9): 2906-17.
Umar, A., M. Koi, et al. (1997). "Correction of hypermutability, N-methyl-N'-nitro-N-nitrosoguanidine resistance, and defective DNA mismatch repair by introducing chromosome 2 into human tumor cells with mutations in MSH2 and MSH6." Cancer Res 57(18): 3949-55.
Umar, A., J. I. Risinger, et al. (1998). "Functional overlap in mismatch repair by human MSH3 and MSH6." Genetics 148(4): 1637-46.
Umezu, K. and H. Nakayama (1993). "RecQ DNA helicase of Escherichia coli. Characterization of the helix-unwinding activity with emphasis on the effect of single-stranded DNA-binding protein." J Mol Biol 230(4): 1145-50.
Umezu, K., K. Nakayama, et al. (1990). "Escherichia coli RecQ protein is a DNA helicase." Proc Natl Acad Sci U S A 87(14): 5363-7.
Upcroft, P., B. Carter, et al. (1980). "Analysis of recombination in mammalian cells using SV40 genome segments having homologous overlapping termini." Nucleic Acids Res 8(12): 2725-36.
Upcroft, P., B. Carter, et al. (1980). "Mammalian cell function mediating recombination of genetic elements." Nucleic Acids Res 8(23): 5835-44.
van Beers, E. H. and P. M. Nederlof (2006). "Array-CGH and breast cancer." Breast Cancer Res 8(3): 210.
van der Weyden, L., D. J. Adams, et al. (2002). "Tools for targeted manipulation of the mouse genome." Physiol Genomics 11(3): 133-64.
van der Weyden, L., D. J. Adams, et al. (2005). "Null and conditional semaphorin 3B alleles using a flexible puroDeltatk loxP/FRT vector." Genesis 41(4): 171-8.
Van Gelder, R. N., M. E. von Zastrow, et al. (1990). "Amplified RNA synthesized from limited quantities of heterogeneous cDNA." Proc Natl Acad Sci U S A 87(5): 1663-7.
Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science 291(5507): 1304-51.
Vitaterna, M. H., D. P. King, et al. (1994). "Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior." Science 264(5159): 719-25.
von Melchner, H., J. V. DeGregori, et al. (1992). "Selective disruption of genes expressed in totipotent embryonal stem cells." Genes Dev 6(6): 919-27.
von Melchner, H. and H. E. Ruley (1989). "Identification of cellular promoters by using a retrovirus promoter trap." J Virol 63(8): 3227-33.
Waldman, A. S. and R. M. Liskay (1988). "Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology." Mol Cell Biol 8(12): 5350-7.
Wallace, H. A., F. Marques-Kranc, et al. (2007). "Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence." Cell 128(1): 197-209.
Walpita, D., A. W. Plug, et al. (1999). "Bloom's syndrome protein, BLM, colocalizes with replication protein A in meiotic prophase nuclei of mammalian spermatocytes." Proc Natl Acad Sci U S A 96(10):

5622-7.
Wang, K. Y. and C. K. James Shen (2004). "DNA methyltransferase Dnmt1 and mismatch repair." Oncogene 23(47): 7898-902.
Wang, W. and A. Bradley (2007). "A recessive genetic screen for host factors required for retroviral infection in a library of insertionally mutated Blm-/- ES cells." Genome Biol 8(4): R48.
Wang, W., M. Warren, et al. (2007). "Induced mitotic recombination of p53 in vivo." Proc Natl Acad Sci U S A 104(11): 4501-5.
Wang, Y., D. Cortez, et al. (2000). "BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures." Genes Dev 14(8): 927-39.
Wang, Y. and J. Qin (2003). "MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation." Proc Natl Acad Sci U S A 100(26): 15387-92.
Ward, I. M. and J. Chen (2001). "Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress." J Biol Chem 276(51): 47759-62.
Ward, J. F. (1985). "Biochemistry of DNA lesions." Radiat Res Suppl 8: S103-11.
Ward, J. F. (1988). "DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability." Prog Nucleic Acid Res Mol Biol 35: 95-125.
Ward, J. F. (1990). "The yield of DNA double-strand breaks produced intracellularly by ionizing radiation: a review." Int J Radiat Biol 57(6): 1141-50.
Waters, T. R. and P. F. Swann (1997). "Cytotoxic mechanism of 6-thioguanine: hMutSalpha, the human mismatch binding heterodimer, binds to DNA containing S6-methylthioguanine." Biochemistry 36(9): 2501-6.
Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature 420(6915): 520-62.
Watt, P. M., E. J. Louis, et al. (1995). "Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation." Cell 81(2): 253-60.
Weber, F., J. de Villiers, et al. (1984). "An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences." Cell 36(4): 983-92.
Wei, K., A. B. Clark, et al. (2003). "Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility." Genes Dev 17(5): 603-14.
Wei, K., R. Kucherlapati, et al. (2002). "Mouse models for human DNA mismatch-repair gene defects." Trends Mol Med 8(7): 346-53.
Weksberg, R., C. Smith, et al. (1988). "Bloom syndrome: a single complementation group defines patients of diverse ethnic origin." Am J Hum Genet 42(6): 816-24.
Welsch, T., K. Endlich, et al. (2007). "Eps8 is increased in pancreatic cancer and required for dynamic actin-based cell protrusions and intercellular cytoskeletal organization." Cancer Lett 255(2): 205-18.
West, S. C. (2003). "Molecular views of recombination proteins and their control." Nat Rev Mol Cell Biol 4(6): 435-45.
Wieacker, P., N. Horn, et al. (1983). "Menkes kinky hair disease: a search for closely linked restriction fragment length polymorphism." Hum Genet 64(2): 139-42.
Wilber, A., J. L. Linehan, et al. (2007). "Efficient and Stable Transgene Expression in Human Embryonic Stem Cells Using Transposon-Mediated Gene Transfer." Stem Cells.
Wildenberg, J. and M. Meselson (1975). "Mismatch repair in heteroduplex DNA." Proc Natl Acad Sci U S A 72(6): 2202-6.
Williams, R. L., D. J. Hilton, et al. (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature 336(6200): 684-7.
Wong, E., K. Yang, et al. (2002). "Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation." Proc Natl Acad Sci U S A 99(23): 14937-42.
Wong, M. H., J. R. Saam, et al. (2000). "Genetic mosaic analysis based on Cre recombinase and navigated laser capture microdissection." Proc Natl Acad Sci U S A 97(23): 12601-6.
Wu, L., S. L. Davies, et al. (2001). "Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51." J Biol Chem 276(22): 19375-81.
Wu, L., S. L. Davies, et al. (2000). "The Bloom's syndrome gene product interacts with topoisomerase III." J Biol Chem 275(13): 9636-44.
Wu, L. and I. D. Hickson (2002). "The Bloom's syndrome helicase stimulates the activity of human topoisomerase IIIalpha." Nucleic Acids Res 30(22): 4823-9.
Wurst, W., J. Rossant, et al. (1995). "A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice." Genetics 139(2): 889-99.
Xu, T. and G. M. Rubin (1993). "Analysis of genetic mosaics in developing and adult Drosophila tissues." Development 117(4): 1223-37.
Xu, X. and D. F. Stern (2003). "NFBD1/KIAA0170 is a chromatin-associated protein involved in DNA damage signaling pathways." J Biol Chem 278(10): 8795-803.
Yagi, T., Y. Ikawa, et al. (1990). "Homologous recombination at c-fyn locus of mouse embryonic stem cells

with use of diphtheria toxin A-fragment gene in negative selection." Proc Natl Acad Sci U S A 87(24): 9918-22.
Yamane, K. and T. J. Kinsella (2005). "Casein kinase 2 regulates both apoptosis and the cell cycle following DNA damage induced by 6-thioguanine." Clin Cancer Res 11(6): 2355-63.
Yamane, K., J. E. Schupp, et al. (2007). "BRCA1 activates a G2-M cell cycle checkpoint following 6-thioguanine-induced DNA mismatch damage." Cancer Res 67(13): 6286-92.
Yamane, K., K. Taylor, et al. (2004). "Mismatch repair-mediated G2/M arrest by 6-thioguanine involves the ATR-Chk1 pathway." Biochem Biophys Res Commun 318(1): 297-302.
Yan, T., S. E. Berry, et al. (2003). "DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells." Clin Cancer Res 9(6): 2327-34.
Yang, Q., R. Zhang, et al. (2004). "The mismatch DNA repair heterodimer, hMSH2/6, regulates BLM helicase." Oncogene 23(21): 3749-56.
Yang, Q., R. Zhang, et al. (2002). "The processing of Holliday junctions by BLM and WRN helicases is regulated by p53." J Biol Chem 277(35): 31980-7.
Yant, S. R., L. Meuse, et al. (2000). "Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system." Nat Genet 25(1): 35-41.
Yao, X., A. B. Buermeyer, et al. (1999). "Different mutator phenotypes in Mlh1- versus Pms2-deficient mice." Proc Natl Acad Sci U S A 96(12): 6850-5.
You, Y., R. Bergstrom, et al. (1997). "Chromosomal deletion complexes in mice by radiation of embryonic stem cells." Nat Genet 15(3): 285-8.
Yu, C. E., J. Oshima, et al. (1996). "Positional cloning of the Werner's syndrome gene." Science 272(5259): 258-62.
Yu, D., H. M. Ellis, et al. (2000). "An efficient recombination system for chromosome engineering in Escherichia coli." Proc Natl Acad Sci U S A 97(11): 5978-83.
Yusa, K., K. Horie, et al. (2004). "Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene." Nature 429(6994): 896-9.
Yusa, K., J. Takeda, et al. (2004). "Enhancement of Sleeping Beauty transposition by CpG methylation: possible role of heterochromatin formation." Mol Cell Biol 24(9): 4004-18.
Zambrowicz, B. P., G. A. Friedrich, et al. (1998). "Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells." Nature 392(6676): 608-11.
Zayed, H., Z. Izsvak, et al. (2004). "Development of hyperactive sleeping beauty transposon vectors by mutational analysis." Mol Ther 9(2): 292-304.
Zhang, Y., F. Buchholz, et al. (1998). "A new logic for DNA engineering using recombination in Escherichia coli." Nat Genet 20(2): 123-8.
Zheng, B., D. W. Larkin, et al. (1999). "The mPer2 gene encodes a functional component of the mammalian circadian clock." Nature 400(6740): 169-73.
Zheng, B., A. A. Mills, et al. (1999). "A system for rapid generation of coat color-tagged knockouts and defined chromosomal rearrangements in mice." Nucleic Acids Res 27(11): 2354-60.
Zheng, B., M. Sage, et al. (1999). "Engineering a mouse balancer chromosome." Nat Genet 22(4): 375-8.
Zheng, B., M. Sage, et al. (2000). "Engineering mouse chromosomes with Cre-loxP: range, efficiency, and somatic applications." Mol Cell Biol 20(2): 648-55.
Zhou, B. B. and S. J. Elledge (2000). "The DNA damage response: putting checkpoints in perspective." Nature 408(6811): 433-9.
Zijlstra, M., E. Li, et al. (1989). "Germ-line transmission of a disrupted beta 2-microglobulin gene produced by homologous recombination in embryonic stem cells." Nature 342(6248): 435-8.