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.