Part I: A system for recessive screens

7.1 *Blm*-deficient cells

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

7.2 Gene trap mutagenesis

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

7.3 Broad applications of the Blm-deficiency in recessive screens

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

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

7.4 Transponson-mediated mutagenesis

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

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

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

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

7.5 Combination of deletional mutations with *Blm*-deficiency

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

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

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

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

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

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

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

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

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

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

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

Part II: DNA methylation and mismatch repair surveillance

7.7 *Dnmt1*, a MMR surveillance gene

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

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

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

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

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

7.8 Clinical implications

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