6 Summary, significance and future goals

In the previous chapters, I have shown that localized gene-trap mutagenesis can be achieved by regional trapping and that the gene-trap mutations generated can be made homozygous by inducible mitotic recombination. A genetic screen has been carried out on the isolated homozygous mutant clones using an ES cell *in vitro* differentiation assay. Clones that show abnormal morphological and gene expression changes during the differentiation process were identified. Other experiments were carried out to confirm these findings. Therefore, I have demonstrated that I can use this strategy to generate homozygous mutant clones in a given region of the mouse genome and use these clones for an *in vitro* recessive genetic screen. In principle, this strategy can be applied to other chromosomes in the mouse genome to create genome-wide homozygous mutant ES cells. This will be a valuable resource for *in vitro* recessive genetic screens.

Before I discuss the potential application of this strategy, I would like to describe some of the latest advancements in mutagenesis techniques, because no single mutagenesis method can completely replace the other methods, and mouse genetics will depend on a combination of these methods as a whole.

6.1 Chemical mutagenesis

Regional and genome-wide ENU mutagenesis in the mouse is a powerful way to generate dominant and recessive mutations for phenotype-driven genetic screens. Such screens can provide a large amount of information about a phenotype of interest or even a certain genetic pathway in a relatively short period of time.

A recent development in this field is to generate ENU- or EMS-induced alleles in mouse ES cells (Chen, Yee et al. 2000; Munroe, Bergstrom et al. 2000). Conventional germ cell mutagenesis with ENU is compromised by the inability to easily determine the mutation rate, strain and interlocus variation in mutation induction, as well as the extensive mouse husbandry requirements (Munroe, Bergstrom et al. 2000). Genome-wide recessive mutations

transmitted by ENU treated males can only be rendered homozygous after three generations of breeding, at which time phenotype screens can be performed. Chen *et al*. (2000) and Munroe *et al*. (2000) have both used the mouse *Hprt* locus to determine that the mutation rate in ES cell is comparable to the mutation rate in spermatogonia in adult male mice. By using ENU mutated ES cells, one generation can be eliminated from the complicated breeding strategy. Also storing ES cells is more convenient than cryopreserving sperm.

ENU/EMS mutagenesis in ES cells can be used for two different purposes, to screen for an allelic series of mutations of a target gene *in vitro* (Vivian, Chen et al. 2002; Greber, Lehrach et al. 2005) or to perform genome-wide recessive genetic screens *in vivo* (Munroe, Ackerman et al. 2004). Vivian *et al*. (2002) has used an RT-PCR based high throughput mutation detection technology to identify mutations in *Smad2* and *Smad4*, which are both embryonic lethal when the genes are knocked out. Of the five non-silent mutations that were transmitted through the germline and bred to homozygosity, one was a severe hypomorph, one was a dominant-negative allele, and the other three did not show any phenotype (Vivian, Chen et al. 2002). Munroe *et al*. (2004) have demonstrated the feasibility of performing genome-wide mutation screens with only two generations of breeding. This strategy was possible because chimeras derived from a single EMS treated ES cell clone transmit variations of the same mutagenized diploid genome, whereas ENU-treated males transmit numerous unrelated genomes (Munroe, Ackerman et al. 2004).

ENU mutagenesis has also been used to generate bi-allelic mutations in ES cells deficient in the Bloom's syndrome gene (*Blm*) (Yusa, Horie et al. 2004). Yusa *et al*. (2004) used a combination of ENU mutagenesis and transient loss of *Blm* expression to generate an ES cell library with genome-wide homozygous mutations. This library was evaluated by screening for mutants in a known pathway, glycosylphosphatidylinositol (GPI)-anchor biosynthesis. Mutants in12 out of 23 known genes involved in this pathway have been obtained, and two unknown mutants were also isolated (Yusa, Horie et al. 2004). Though ENU mutagenesis is proved to be an efficient tool to generate

mutants in ES cells, it is still a difficult task to identify the mutated gene. In cases when little is known about the pathway, this can only be achieved by expression cloning.

6.2 Transposon mutagenesis

Retroviral and plasmid-based vectors are the two main approaches for insertional mutagenesis. Mutagenesis rates for these vectors are improved by ensuring that vector insertions coupled with actuation of a selectable marker, a concept known as a "gene trap". Different gene-trap vector designs are needed to achieve broad genome coverage in large-scale genetic screens. The synthetic *Sleeping Beauty* (SB) transposon system provides a promising alternative delivery method for gene-trap vectors (Ivics, Hackett et al. 1997).

Sleeping Beauty (SB) belongs to the *Tc1/mariner* superfamily of transposons. Ivics *et al*. (1997) reconstructed the transposon and transposase, *SB10*, from endogenous transposons inactivated by mutations accumulated in evolution. Both the reconstructed transposon and the transposase were shown to be active in mouse and human cell lines (Ivics, Hackett et al. 1997). It is composed of the SB transposon element and the separately expressed transposase. The SB transposon element contains two terminal inverted repeats (IR). The excision and re-insertion of the SB transposon element into the host genome occurs by a cut-and-paste process mediated by the transposase which 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 was first used as an insertional mutagen in mouse ES cells (Luo, Ivics et al. 1998). But in ES cells, the transposition efficiency is quite low (3.5×10^{-5} events/ cell per generation). Though there is still room to improve the efficiency of SB system *in vitro*, this system does not appear to be suitable for a genome-wide mutagenesis effort in ES cells. However, efficient transposition has been observed in the mouse germline, either by crossing males doubly transgenic for *SB10* transposase and a gene-trap transposon to wild-type females (Dupuy, Fritz et al. 2001), or by injecting transposon vectors

and SB10 mRNA together into one-cell mouse embryos (Dupuy, Clark et al. 2002). In these studies, on average, 1.5 to 2 transposon insertion were found in each of the offspring.

To determine sequence preferences and mutagenicity of SB-mediated transposition, Carlson *et al*. (2003) have cloned and analyzed 44 gene-trap transposon insertion sites from a panel of 30 mice. 19 of the 44 mapped transposon insertion sites were mapped to chromosome 9 where the transposon concatomer was located. The remaining insertion occurred on other chromosomes without obvious preference for chromosome or region. The local transposition interval appears to be between 5 to 15 Mb. Analysis of the transposon/host flanking sequence has shown that transposition sites are AT-rich and the favoured sequence is “ANNTANNT”. 27% transposon insertions were in transcription units. Of the 6 insertions in heterozygous animals which were bred in attempts to generate homozygous mice for the insertions, two were found to be homozygously lethal (Carlson, Dupuy et al. 2003). The transposition and gene insertion frequencies mean that *Sleeping Beauty* is still not efficient enough for a genome-wide mutagenesis screen.

The transposon and a transposase-expression vector can be electroporated into host cells where they co-exist episomally for a short period of time during which transposition is catalysed from the vector to the genome. Although this episomal method is very efficient in cultured somatic cells and in somatic cells *in vivo*, the transposition efficiency in mouse ES cells is very low (Luo, Ivics et al. 1998). Therefore it is not currently efficient enough for genome-wide mutagenesis in ES cells without a significant improvement of its efficiency in ES cells.

6.3 RNA interference

RNA interference (RNAi) was first noticed in *C.elegans* as a response to exogenous double strand RNA (dsRNA), which induce sequence specific knockdown of an endogenous gene’s function. Double strand RNA mediated gene inactivation is a highly conserved process. The basic mechanism of RNAi includes three major steps: first, a double strand RNA is cleaved by

Dicer protein into 21-25 nucleotides (nt) double strand RNAs; second, these small interfering RNAs (siRNA) associate with a complex (RISC, RNA-induced silencing complex) which has RNA nuclease activity; third, RISC unwinds siRNA and uses it as the template to capture and destroy endogenous transcript (Hannon 2002).

The RNAi phenomenon was quickly adopted for large-scale genome-wide genetic screens in *C. elegan*s. In *C. elegans*, this form of post-transcriptional gene silencing (PTGS) only requires a few molecules of double strand RNA in one cell to initiate the process. It can spread to all the cells in the body of the worm and pass through the germ line for several generations with almost complete penetrance (Kamath, Fraser et al. 2003). The delivery of dsRNA in *C. elegans* is also very simple, it can be achieved either by soaking the worms in dsRNA solution or feeding the worm with dsRNA-expressing *E. coli*.

Naturally, the success of RNAi technology in *C. elegans* inspired many to apply it to more complex mammalian systems. However at the beginning, this technology has encountered some problems. First, dsRNA becomes diluted in subsequent cell divisions, and the silencing phenotype can not be inherited unless a dsRNA-expressing construct is stably integrated in the genome. Second, dsRNA triggers a non-specific global translation inhibition by activating the RNA-dependent protein kinase (PKR) pathway (Hannon 2002). A way to bypass this problem is to express short hairpin RNA (shRNA) in mammalian cells

Elbashir *et al*. (2001) showed that 21 or 22 nucleotides double strand RNA could strongly induce gene-specific inactivation without eliciting the non-specific translation inhibition effect observed with longer dsRNAs (Elbashir, Harborth et al. 2001). However, the shRNA mediated RNAi effect in mammalian cells is not inherited nor can it spread to adjacent cells. Brummelkamp *et al*. (2002) developed a mammalian expression vector to synthesize short hairpin-structured RNA transcripts (shRNA) *in vivo*. The shRNA can be recognized and cleaved by the endogenous PTGS machinery and can trigger the RNAi process. With these developments, shRNA

technology has become a practical tool to study gene function in mammalian cells.

Recently, two groups have reported the construction and initial application of shRNA expressing libraries targeting human and mouse genes (Berns, Hijmans et al. 2004; Paddison, Silva et al. 2004). Berns *et al*. (2004) constructed a library of 23,472 distinct shRNAs targeting 7,914 human genes. They obtained on average 70% inhibition of expression for approximately 70% of the genes in the library. A screen using this library has successfully identified one known and five unknown modulators of the p53-dependent proliferation arrest (Berns, Hijmans et al. 2004). Paddison *et al*. (2004) targeted 9,610 human genes and over 5,563 mouse genes in their library. One quarter of this library was used to screen for shRNAs that interfere with 26S proteasome function. Nearly half of the shRNA clones that were expected to target proteosomal proteins were recovered as positive in the screen (Paddison, Silva et al. 2004). These experiments have shown that RNAi has become a practical tool for recessive genetic screens in mammalian cells in culture.

RNAi technology still has some limitations. First, it can only knockdown the expression of a gene. Incomplete inhibition will cause a hypomorphic phenotype in many cases. If the residual expression of the target gene is still enough for its normal function, it will be missed in large-scale genetic screens. An example of this is illustrated by a systematic function analysis of the *C. elegans* genome using RNAi. Although this screen targeted about 86% of the 19,427 predicted genes, mutant phenotypes were only identified for 1,722 genes (Kamath, Fraser et al. 2003). Another example of this limitation is that just 22 out of 55 shRNAs targeting 26S proteasome components were identified as positive in the screen. Another 14 shRNAs scored above background in the second focused assay in the same study (Paddison, Silva et al. 2004). Second, the design of an shRNA-expressing construct requires prior knowledge of its target, which is greatly limited by the annotation of the mouse genome. That means a genetic screen using this technology is always going to be a forward genetics screen. Any genes not in the library will never

be identified in the screen. So although shRNA screens are potentially powerful, they lack the coverage of a screen performed with a random mutagen like ENU.

6.4 Forward genetics versus reverse genetics

Forward genetics refers to the techniques used to identify mutations that produce a certain phenotype. A mutagen is often used to accelerate this process. Once mutants have been isolated, the mutated gene can be molecularly identified. Reverse genetics refers to the method to determine the phenotype that results from mutating a given gene, usually by deleting the gene of interest.

Historically, forward genetic screens have been the main method for gene function discovery in various model organisms. But in the mouse, the development of mouse gene knockout technology has made reverse genetics the most powerful and widely used functional genomics tool. The distinction between these two approaches is no longer so clear. For example, gene-trap insertional mutagenesis is a typical forward genetics approach that has been widely used in *in vitro* and *in vivo* forward genetic screens. But the development of 5' RACE technology has made the identification of the insertion site much easier than before, so a large number of mutant clones can be generated and identified in a high-throughput way (Skarnes, von Melchner et al. 2004), and reverse genetic screens can be carried out on these ES cell clones or the mice derived from them.

The completion of the mouse and human genome has provided an unprecented opportunity for both forward and reverse genetics studies. For forward genetics, it is now much easier to map and identify the causative genetic change. For reverse genetics, the availability of the sequence information for each mouse gene has made it possible to knockout any gene in the mouse genome by gene-targeting or it can be knocked down by RNAi.

Though reverse genetics is more straightforward, and the phenotype can be quickly linked to the mutation, forward genetics has its own advantages. First,

it is quick to generate a lot of mutations for phenotype analysis. Second, it is an unbiased, phenotype-driven approach and no previous knowledge of the pathway involved is needed. It is not surprising that even a screen for a well-characterized pathway can still identify unknown components. Third, a variety of allelic mutations can be generated and they might affect a gene's function in different ways. So forward genetics will play an increasingly important role in mouse functional genomics.

6.5 Selection versus screening

Most of the genetic screens performed in mammalian cells are in fact selections. The distinction between a selection and a screen depends on the method used to detect the phenotype of the mutants. A selection requires a strategy to distinguish those mutant cells that show a given phenotype from the rest of the cell population. This can be achieved by two ways, either by accumulating the cells that carry the desired mutations, or more often, by selectively killing the rest of the cells that do not carry the relevant mutations (Grimm 2004).

On the other hand, in a screen, mutants must be examined one by one to determine whether and to what extent they have the desired phenotype. So for a selection or a screen conducted on the same scale, a screen will require much more time and labour. Geneticists always prefer to perform a selection whenever it is possible. But screens are particularly useful when a broad dynamic range of gene activity is examined (Shuman and Silhavy 2003), for example the mutations that affect ES cell *in vitro* differentiation in our study.

The development of FACS technology has made it possible to turn a screen into a selection by selectively accumulating the mutants that show a certain phenotype. For example, if we want to carry out a screen on ES cell differentiation into mesodermal lineages, mutant ES cells can first be differentiated on collagen IV coated dishes, and *Flk1*$^{+}$ cells derived from embryonic stem cells can then be sorted by FACS (Yamashita, Itoh et al. 2000), while the undifferentiated mutant ES cells can be sorted by ES cell specific markers, such as SSEA-1. If a cell lineage-specific cell surface

marker is not available, a fluorescence reporter can be used to tag an intra-cellular lineage-specific gene. Examples for this strategy is the use of *Sox1*-GFP knock-in to track the differentiation of ES cells into neuroectodermal precursors (Ying, Stavridis et al. 2003) and the use of a *Gsc*-GFP reporter to investigate the differentiation course of mesendodermal cells (Tada, Era et al. 2005). Random mutations can then generated in this modified cell line. The mutant cells are induced to differentiate under optimized conditions, and the cells that do not express the reporter can be sorted out by FACS and further analyzed. Fluorescent cells can also be screened in a high-throughput anner using live cell imaging machines.

6.6 The future of genetic screens in mouse ES cells

As I discussed before, mouse ES cells are a unique experimental system that not only has the potential to be a model for mouse early embryogenesis but also sheds the light on how to manipulate their human counterparts to treat human diseases. However the factors and the pathways that direct their differentiation are still not well understood. So genetic screens for discrete differentiation steps can provide an immense amount of data and information to elucidate the regulation of pathways underlying this process (Grimm 2004).

The biggest obstacle for a genetic screen in ES cells is the generation of recessive mutations. We have demonstrated that we can use a strategy which combines regional trapping and inducible mitotic recombination to generate recessive mutations in a region of interest. A genetic screen using these homozygous clones has identified genes that are involved in ES cell *in vitro* differentiation. Thus we have shown that a genetic screen of a complex pathway like *in vitro* differentiation is feasible in ES cells.

Other mutagenesis methods in ES cells can also be combined with inducible mitotic recombination to generate homozygous mutations, such as ENU, irradiation, transposons and gene targeting. RNAi can also be used to perform recessive genetic screens *in vitro*. Because of the limitations of every existing mutagenesis method, it is likely that a combination of different methods is needed to saturate the mouse genome.

To use mouse ES cell *in vitro* differentiation in a genetic screen, a lot of fundamental work still needs to be done. For example, it would be an advantage to know how the expression of each mouse gene changes during the whole differentiation process. This will not only provide a background control for mutant phenotyping, it will also provide a set of markers for each of the differentiation steps and cell lineages, which will be more reliable than just monitoring a few markers.

The limiting factor for a high throughput genetic assay in mammalian cells is always the read-out, or the detection of the cellular changes (Grimm 2004). The use of cDNA and oligonucleotide microarrays is one of the solutions. FACS sorting based on different cell lineage specific markers is another promising way to determine ES cell *in vitro* differentiation potential. Or florescence reporters can be knocked into cell lineage marker genes and these can be used to monitor the expression of these markers in the differentiation process.

The International mouse knockout project has already proposed to systematically knockout every mouse gene (Austin, Battey et al. 2004; Auwerx, Avner et al. 2004). Known or predicted human disease genes will likely be high priority candidates. But how to decide the priority of other genes, especially those genes that no biological function has ever been attributed, will be a challenge for the organizers of this international program. *In vitro* data can provide some useful information about the function of these unknown genes. For example, it will be helpful for the researchers to decide which targeting strategy to use (for example, conventional or conditional knockout) and even which phenotypes to expect. So an ES cell *in vitro* differentiation screen can serve as a pre-screen for the analysis of gene function in whole animals in a large-scale knockout project.

To make such a genetic screen possible, it is necessary to make a library of homozygous mutant ES cells. It can be achieve by generating a library of mutants of a mixture of different genotypes (Guo, Wang et al. 2004; Yusa,

Horie et al. 2004). The advantage of this strategy is that the library is easy to make and maintain. However, this strategy has limited the application of the library to genetic screens in which mutants are identified by their resistance to a specific mutagen. It is impossible to select for mutants that are sensitive to the same mutagen which can be equally important to elucidate a complicated genetic pathway. On the other hand, a genetic screen can also be performed on an array of homozygous ES cells mutants. These homozygous mutants, which can be maintained in a format convenient for high-throughput screens, can be exposed to a range of different concentrations of a specific mutagen, which can not only identify mutants that are sensitive or resistant to this mutagen, but also determine the levels of resistance or sensitivity of these mutants, which can be informative to their role in the interested genetic pathway. Pure homozygous mutant ES cell clones are particularly important for genetic screens on ES cell differentiation because mutants are difficult to be identified by drug selection. Homozygous mutant ES cell clones can be exposed to different differentiation inducers to analysis their differentiation into a variety of cell lineages.

In this study, we have demonstrated that inducible mitotic recombination can be used to generate homozygous gene-trap mutations in mouse embryonic stem cells in a high-throughput way. Homozygous mutant ES cells lines produced by this strategy can be used for genetic screens. However, the genetic instability of ES cells in culture and the epigenetic changes caused by induced mitotic recombination might interfere with the phenotype-driven screens. Care need be taken to choose appropriate positive and negative control cell lines to keep the background of the screens to a reasonable level. On the other hand, genetic and epigenetic instabilities also exist in the other existing high-throughput method to generate homozygous mutant ES cells using *Blm*-deficient ES cells. *Blm*-deficient ES cells have already been successfully used for phenotype-driven screens (Guo, Wang et al. 2004; Yusa, Horie et al. 2004), so it is reasonable to predict these background interferences can be controlled by a good experimental design.

Inducible mitotic recombination is also compatible with other mutagenesis methods, including ENU (Chen, Yee et al. 2000; Munroe, Bergstrom et al. 2000), transposon mutagenesis (Ivics, Hackett et al. 1997; Luo, Ivics et al. 1998) and gene targeting (Thomas and Capecchi 1987). RNAi is another way to knock down gene expression for recessive screens in ES cells (Berns, Hijmans et al. 2004; Paddison, Silva et al. 2004). The limitations of the existing mutagenesis methods suggest that the most effective way to saturate the genome with recessive mutations is to use a combination of these methods. Recessive genetic screens in mouse ES cells will accelerate functional studies of genes in the mouse, as well as provide a foundation for applied research to differentiate human ES cells into cell types that can be potentially used to treat the human diseases.