A CHROMOSOME-SPECIFIC RECESSIVE GENETIC SCREEN FOR GENES INVOLVED IN *IN VITRO* DIFFERENTIATION IN MOUSE EMBRYONIC STEM CELLS

A dissertation submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

by Wei Wang

The Wellcome Trust Sanger Institute
Robinson College, University of Cambridge

August, 2005

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

Wei Wang

ACKNOWLEDGEMENT

I would like to thank my PhD supervisor, Dr. Allan Bradley, for the guidance, support and patience during the past four years. It is a privilege to have Allan as a mentor. I thank him for the inspiration he has given to me when I felt frustrated and also the freedom he has given to me in pursuing my interest.

I would like to thank Bradley lab members, past and present, for their kindness and generosity. Though I have not met some of them in person, they are always happy to help whenever Allan referred me to them. Special thanks to Dr. Ge Guo for her help in ES cell culture and virus work, and the instructive discussions at various stages of my project. I would express my gratitude to Dr. Pentao Liu for sharing with me some of his “magic tools” and for the instructive discussions. I thank Dr. Xiaozhong (Alec) Wang for the discussions and encouragement. I would like to thank Dr. Madhuri Warren for the histopathology analysis of the teratocarcinoma and for reading the last two chapters of my thesis. I thank Dr. Qin Si for her help in the embryo work. I am grateful to Francis Law and Alastair Beasley for the tissue culture support. I am also grateful to Evelyn Grau and Tina Hamilton for the blastocyst injection. I thank Dr. Yue Huang for his help in BAC preparation. I thank Dr. Haydn Prosser for providing some useful plasmids, Dr. Antony Rodriguez and Dr. Ilona Zvetkova for providing F1 hybrid mice.

I would like to thank Dr. Xiaolin Liu for her collaboration in identifying the mutant ES cell lines. I owe my thanks to Dr. Jos Jonkers for his help and the instructive discussions. I thank Dr. Dongrong Chen Murchie for her extensive work on *in situ* hybridization, and her husband, Dr. Alistair Murchie for reading the first chapter of my thesis. I thank Dr. Bill Skarnes for the instructive discussions. I think Dr. Lu Yu and Dr. Mercede Pardo for help with Western blotting. Thanks go to Dr. Christina Hedberg-Delouka for her kindly help in my application and also in the four years of my PhD study.

Finally, I would like to thank my parents, Ziyuan Wang and Huiyin Zhang, and my sister, Qin Wang for their unconditional love and support.

ABSTRACT

Mouse embryonic stem (ES) cells are pluripotent cells that retain unlimited self-renewal potential. ES cells can also differentiate into all the three germ layers and many specified cell types *in vitro*. The *in vitro* differentiation of ES cells recapitulates early embryogenesis, thus serves as a valuable *in vitro* model for developmental studies. Little is known about the genes which are important in the ES cell differentiation process, partly due to the difficulty of generating homozygous mutations in ES cells. In this study, we have developed a method, which explored induced mitotic recombination and regional trapping mutagenesis method to accumulate large numbers of homozygous gene-trap mutations in a genomic region of interest in mouse embryonic stem cells. A cell line was engineered to undergo mitotic recombination and to capture the subset of gene-trap mutations that were generated on chromosome 11. A large panel of genome-wide gene-traps were generated in this cell line and those that are located on chromosome 11 were specifically selected via Cre/*loxP* mediated inversions. The inversions were then made homozygous by induced mitotic recombination. Using this system, 66 independent homozygous gene-traps on chromosome 11 have been isolated from a library of about 10,000 gene-trap clones. These homozygous clones have been assessed for their developmental potential by an *in vitro* differentiation assay. The differentiation of each of these lines has been assessed by RT-PCR using a panel of markers that are characteristic for the three germ layers and various differentiated cell types. Clones that show abnormal expression of one or more markers are verified using microarrays, western blotting and *in situ* hybridization. A homozygous mutation of ATP-citrate lyase (*Acly*) gene was found to block the differentiation of ES cells *in vitro*. Under retinoic acid induction, the embryoid bodies derived from the mutated cell line were still mostly composed of undifferentiated embryonic stem cells. This was confirmed by the high expression of the epiblast makers, such as *Oct3/4*, *Nodal* and *Nanog*. BAC rescue experiment has been carried out to reverse the phenotype of this cell line and make a causal link between the expression level of *Acly* and the

phenotype. Therefore, we have shown that we can create random homozygous gene-trap mutagenesis in a candidate region of the genome and use these clones for functional genomics study.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENT

1 Introduction

1.1 Genetic screen in model organisms

1.1.1 Genetic screens in yeast

It is hard to imagine that so much of our knowledge of cell biology comes from the simple unicellular fungi, the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe*. Budding yeast and fission yeast are very different in their biology and are used to study different issues of cell biology. *S. cerevisiae* is the ideal model for signal transduction, cell cycle control, as well as chromosome structure (Forsburg 2001). On the other hand, *S. pombe* is a favourite for studies of cell-cycle control, mitosis and meiosis, DNA repair and recombination, and the checkpoint controls important for genome integrity (Wood, Gwilliam et al. 2002). In spite of their numerous differences, these organisms share one thing in common, namely the ease of genetic manipulation. This has resulted in the widespread use of these organisms as model organism to understand the biology of more complex systems.

S. cerevisiae has some milestones in biology. It was the first eukaryote to be transformed by plasmids, the first eukaryote for which gene-targeting became possible, the first eukaryote to be completely sequenced (Goffeau, Barrell et al. 1996). But surprisingly, the function of many of the 6000 or so genes still remains unknown. *S. pombe* has a relatively shorter history and a smaller research community. It contains 4,824 genes in its 13.8 Mb genome, which is the smallest number of protein-coding genes yet recorded for an eukaryote (Wood, Gwilliam et al. 2002). It diverged from budding yeast approximately 330 million years ago, around 4000 of the 4824 genes (83%) of *S. pombe* have homologues in *S. cerevisiae*, but they share no conserved synteny. Only 681 genes (14%) seem to be unique to *S. pombe*.

Both yeast species have a life cycle that is ideally suited for classical genetic analysis. They both can grow and divide as haploids and thus the phenotype of recessive mutations can be easily discovered. On the other hand, they also have a diploid sex cycle that allows the maintenance of lethal mutations and further characterisation of these. With the sequences of both genomes

finished (Goffeau, Barrell et al. 1996; Wood, Gwilliam et al. 2002), yeast geneticists are now equipped for genomic approaches, as well as their traditional genetic tools to answer fundamental questions: what is the function of each of the several thousand genes in the genome and how they work together to make the single cell function?

Because of the high efficiency of homologous recombination in yeast, it is relatively easy to design a construct with a nutritional or drug selection marker to disrupt a specific gene. In fact, budding yeast is the only eukaryotic organism in which every open reading frame has been knocked out or trapped (Kumar and Snyder 2001). These resources have made it possible to carry out large-scale screening in an efficient way.

But a complete loss-of-function mutation of an essential gene always leads to a lethal phenotype that impedes further functional analysis. So a partial loss-of-function mutation, which is functional only under a permissive condition, is a favourite of yeast geneticists. A classical example is the temperature sensitive mutation, which can either be a thermosensitive (*ts*) or a cold-sensitive (*cs*) mutation. The defective protein only works at low temperatures or high temperatures, respectively. Temperature sensitive mutants can be easily identified by replica plating and culturing at different temperatures. This simple method has directly led to the finding of genes involved in cell-division-cycle (*cdc*) machinery (Hartwell, Culotti et al. 1970; Nurse 1975). In 2001, Leland H. Hartwell, Tim Hunt and Sir Paul Nurse were awarded the Nobel Prize "for their discoveries of key regulators of the cell cycle".

Temperature sensitive mutants are not just used to find the genes affected in certain processes, they also serve as a starting point for screening for genes in the same or parallel pathways. Even at permissive temperature, the activity of the mutated allele is often attenuated, though the cell might look perfectly normal. An additional mutation at another locus can sometimes cause lethality, even at the permissive temperature. This is called “synthetic lethality”.

Another broad approach to identify other members in the network is to use suppression analysis. The basic logic behind this approach is that if one mutation will cause the strain to die under a non-permissive condition, and another mutation can rescue the phenotype, then the second mutation is likely to be in a gene that is functionally linked to the first gene. Though the concept itself sounds straightforward, the mechanisms that result in the rescue can be very different, for example a mutation in a direct interaction partner, a mutation that activates an alternative pathway, or even a mutation of the tRNA molecule that recognizes the mutated codon and translates it to the right amino acid.

The suppressor assay is more attractive to yeast geneticists, because the rescue of the mutant phenotype can be selected, for example; survival under the original non-permissive conditions. In comparison, the synthetic lethality screen requires screening every clone under various conditions, which can be laborious and time-consuming work. Nevertheless, both methods are important to define genetic networks in yeast.

Yeast genetics and genomics studies continue to provide insights into the molecular mechanisms of eukaryotic cell biology. The small genome size and the limited homology with human genes can sometimes even be an advantage. The highly conserved genes are often the most fundamental components of the pathway. In the process of evolution, more regulators and modifiers are gradually added to pathways which allow more versatile and accurate control. If the phenotype of a mutation of a yeast gene can be rescued by its counterpart in a more complex eukaryotic organism, the two genes might have a similar if not the same function.

The arrival of the “genomics era” has rejuvenated studies in yeast. Interestingly, this organism serves as a platform to test high-throughput genomics techniques, such as genome sequencing and genome wide deletion libraries. All these techniques, once developed in yeast, have been quickly transferred to other organisms. Genetic screens in yeast have

provided and continue to provide informative insight into how to carry out similar experiments in more complex cells and organisms.

1.1.2 Genetic screens in fruitfly

Thomas Hunt Morgan's work in the fruitfly *Drosophila melanogaster* has been widely regarded as the beginning of the modern genetics. Considering the fact that he built his Nobel Prize winning chromosome theory of heredity on the spontaneous mutations isolated in a relatively short time (Rubin and Lewis 2000), the fruitfly is no doubt one of the most tractable multicellular organisms for genetic studies. Since then, generations of scientists have added numerous powerful tools to the fruitfly research, balancer chromosomes, deletion chromosomes, induced mitotic recombination, just to name a few.

Though flies and vertebrates diverged from a common ancestor about 700 million years ago, a lot of fundamental developmental processes are still essentially the same. *Drosophila melanogaster* has about 13,600 genes in its 180 Mb genome, much fewer than *Caenorhabditis elegans* (Adams, Celniker et al. 2000). But when both genomes are compared to human, the fruitfly has twice as many genes that have homologs in humans (Friedman and Hughes 2001). The fruitfly also has many homologs of human disease genes, which when mutated can provide insights of the molecular mechanisms of those human diseases. All these features have made *Drosophila melanogaster* an ideal system for genetic screens.

The Nobel Prize winning work by Christiane Nüsslein-Volhard and Eric Wieschaus (Nusslein-Volhard and Wieschaus 1980) changed the landscape of genetic studies completely. For the first time, a genome-wide mutagenesis was carried out in a multicellular organism to identify the mutations that would disrupt a given process, namely embryogenesis. Also for the first time, embryos rather than the adults were used for a genetic screen and many of the mutations found are actually the fundamental regulators of the whole development process (St Johnston 2002).

The famous Heidelberg screen has shown the tremendous power of a genetic screen. But like other successful screens in history, it also has its limitations. These limitations were either caused by the nature of the chosen organism, the design of the screen or the characteristics of the signaling pathway members. Since Christiane Nüsslein-Volhard and Eric Wieschaus had utilized this method to its logical extreme, most genetic screens in *Drosophila* after this were aimed at hunting for the mutations missed by the Heidelberg screen.

One efficient way to recover the missing parts of the genetic jigsaw is to find the mutations that are linked one way or another to known mutations. So enhancer and suppressor screens in a sensitized background, in which one component's function has already been partially disrupted, have become the favourite methods to expand knowledge of certain pathways and to fill the holes and the gaps. For most genes in the genome, the expression of one wild-type allele is enough for normal development. But if a second gene in the same pathway is also mutated, sometimes the level of expression of the first gene might no longer be enough to keep the signalling pathway running normally. In this way, dominant enhancers or suppressor of the first gene can be identified.

Compared to traditional screens, enhancer and suppressor screens have several important advantages. First of all, these can be an F1 screen rather than an F3 screen for recessive homozygous mutations, thus there is no need for a complicated breeding plan to make the mutations homozygous. Second, the function of an embryonic lethal mutation in later developmental stages can be studied. The most prominent example of a modifier screen was the identification of the components of the *Sevenless* (Sev) pathway, which controls the fate choice of the R7 photoreceptor cells in the eye (Simon 1994).

Another way to bypass the embryonic lethality limitation of a traditional screen is to perform a "mosaic screen". In this screen, induced mitotic recombination can be used to create homozygous mutations in a heterozygous background. When the recognition sites of the Flp recombinase, *FRT*s, are located at identical positions on homozygous chromosomes, FLP can mediate site-

specific recombination (Golic and Lindquist 1989) and make a mutation which lies distal to the *FRT* sites homozygous. The efficiency of this approach is surprisingly high, partly due to the fact that homologous chromosomes are paired in mitotic cells in *Drosophila.*

Interestingly, mosaic screens in *Drosophila* have identified not only those mutations that affect development, but also genes that can produce tumor outgrowth, which are almost impossible to find in a traditional screen (Xu, Wang et al. 1995). Because of the numerous tissue-specific FLP-expressing lines, *Drosophila* has become the only multicellular organism in which any cell type or developmental stages can be targeted for a genetic screen.

Even with all these powerful tools at hand, a lot of genes never show up in any of these loss-of-function screens. The challenge now is to annotate the remaining genes. Gain-of-function screens, which cause either over-expression of a gene in the right place or its mis-expression in the wrong tissue, can be used to identify the function of these genes (Rorth 1996; Rorth, Szabo et al. 1998).

The introduction of RNA Interference and the completion of the *Drosophila* genome sequence have now provided an unprecedented opportunity to carry out forward genetic screens (Carthew 2001). Nevertheless, in the foreseeable future, reverse genetic screens will still remain the favourite for the *Drosophila* community.

1.1.3 Genetic screens in nematode

Simplicity is one of the most important reasons why Sydney Brenner chose *Caenorhabditis elegans* as a new experimental organism to study the nervous system and embryonic development in 1963 (Ankeny 2001). But this small organism is not as simple as it looks. Surprisingly, this small creature with only 959 somatic nuclei (adult) has over 19,000 genes in its 97-Mb genome (1998), compared to the 23,000 genes of human and mouse. About one third of these genes have their counterparts in mammals.

This new model organism rapidly became quite a popular experimental system because of its unique hermaphroditic lifestyle, its rapid generation time, its simplicity and ease of manipulation. The nematode is the first multicellular model organism whose cell fate map has been fully described (Sulston 1976; Sulston and Horvitz 1977; Sulston, Schierenberg et al. 1983). From the fertilized egg to the 959-cell adult, every cell that appears or dies in the development process can be traced back to its origin. Thus, even a small change in cell number of a specific organ, like the vulva, can be used for genetic screens. The mutations recovered then build a bridge between one gene, its relevant development process and the aberrant anatomical structure caused by the mutation of this gene.

Though the worm was first picked with the view that this would enable studies of nervous system, many genetic screens have been carried out to elucidate various genetic pathways, such as apoptosis, RAS signalling, *Notch* signalling and sex determination. The early screens were mostly simple forward recessive screens that identified mutants with visible phenotypes. A typical example is Sydney Brenner's screen published in 1974 (Brenner 1974). Additional screens for the same phenotypes or modifier screens in the mutant lines have helped to identify more members of the same pathway.

A variety of mutagens have been used in *C. elegans* over the years. Ethyl methane sulphonate (EMS) is the commonly used mutagen. Mapping the point mutations caused by EMS largely depends on the available markers. *Mariner* elements from *Drosophila melanogaster* have also been used (Bessereau, Wright et al. 2001). The sequencing of the nematode genome (1998), which was the first multicellular organism to be completely sequenced, has greatly accelerated the identification of mutants. It has also enabled genome-wide RNAi screens (Kamath, Fraser et al. 2003).

The phenotype of the RNAi "knockdown" of a specific gene depend heavily on the timing and delivery method of dsRNA, the various characteristics of the target (protein stability and homology of the target gene to its family members) and the design of the dsRNA construct (Maine 2001). An example of the

limitations of RNAi is that the first genome-wide RNAi screen in *C. elegans* only identified mutant phenotypes of 1,722 genes out of 16,757 genes (86% of the 19,427 predicted genes) being targeted (Kamath, Fraser et al. 2003). So RNAi only serves as a complement, instead of a complete replacement to the traditional genetic "knockout" methods (Maine 2001). The lack of a reliable gene-targeting method to create null alleles for any given gene has limited the power of the worm genetics (Jorgensen and Mango 2002). A random deletion library and a standard method to recover the desired mutations have partly solved this problem (Jansen, Hazendonk et al. 1997; Liu, Spoerke et al. 1999).

Of the 19,000 to 20,000 *C. elegans* genes, it is estimated that only 6,000 will have a visible, lethal or sterile phenotype when mutated, and only a small portion of these have already been hit by various mutation methods (Jorgensen and Mango 2002). So the challenge now is to annotate the rest of the worm genome. Forward genetic screens are still the most powerful way of doing this, but new strategies need to be designed to screen for redundant genes. Combined with other genomics tools, such as cDNA microarrays and RNAi, genetic screens are now helping to discover the remaining secrets in the worm genome.

1.1.4 Genetic screen in mouse

Of all commonly used model organisms, the mouse is the closest relative to human. The conserved gene structures, sequence and the extensive comparative genetic linkage map make the mouse the best model to identify human gene function and provide models of human disease (Justice 2000).

Compared to the other model organisms, the history of large-scale genetic screens in mouse is relatively short. So it is no surprise that many genetic tools currently used in the mouse have already been applied to other model organisms a long time ago. Though the tools are old, the screens in mouse reveal the functions of many genes which are unique to mammals.

Interestingly, the mouse was the first multi-cellular organism in which gene targeting by homologous recombination became possible, owing to the development of mouse embryonic stem (ES) cell technology (Evans and Kaufman 1981; Bradley, Evans et al. 1984). Probably, it will also become the first multi-cellular organism in which all genes are systematically knocked-out, with the goal of creating a public resource that contains mutated alleles of every mouse genes (Austin, Battey et al. 2004; Auwerx, Avner et al. 2004).

In a model organism in which reverse genetics has been the major technology used for identifying gene function, it is important to notice the advantage of forward genetics and its role in future functional studies. In mouse, *N*-ethyl-*N*-nitrosourea (ENU) is the most commonly used mutagen for phenotype-driven screens. The point mutations induced by ENU can generate a wide range of alleles of a given gene, ranging from a complete loss-of-function null allele, a partial attenuated hypomorph to a gain-of-function allele. An allelic series provides researchers with a unique opportunity to dissect a gene's function.

It is no surprise that the first two large-scale genetic screens in the mouse were dominant screens for viable and visible phenotypes (Hrabe de Angelis, Flaswinkel et al. 2000; Nolan, Peters et al. 2000). Both screens identified a lot of visible mutations, including hair and skin, pigmentation, skeletal morphology and eye defects. Because the two different groups focused on different specialized phenotypes (the UK group on neurological phenotypes and the German group on haematological phenotypes), they identified a number of phenotypes directly linked to human disease. These two screens showed the efficacy of creating novel mutations of unknown genes by ENU mutagenesis. But they also revealed the bottleneck of this approach, the confirmation of the mutations and the subsequent identification of the point mutations (Justice 2000). Although the sequencing of the mouse genome and the creation of the mouse single nucleotide polymorphism (SNP) map have made mapping much easier, identification of the mutation is still a very laborious and time-consuming effort.

Most mutations in genes are recessive, which means a phenotype can only be observed when both alleles of a gene are disrupted. Dominant screens are thus limited to only a small subset of the 23,000 mouse genes. But recessive screens in mouse were quite difficult before the chromosome engineering techniques were developed (Ramirez-Solis, Liu et al. 1995; Zheng, Sage et al. 1999). A recent published recessive screen on mouse chromosome 11 (Kile, Hentges et al. 2003) has capitalized on a 24-cM inversion between the *Trp53* and *Wnt3* genes to isolate recessive mutations in this interval. The inversion served not only to suppress the recombination in this region, but also to simplify the genotyping by carrying a dominant *Agouti* coat colour marker.

This recessive screen has shown the tremendous power of a non-biased, phenotype-driven genetic screen that has already been proven in other model organisms. Though the cost and the time of mouse breeding has greatly limited the scale of the screens that can be carried out, the striking similarity between mice and humans has made every new mutation identified a human disease gene candidate. The fact that a single screen has tripled the number of the mutations in an already well characterised region confirms that ENU mutagenesis will still be a major player in the annotation of the mouse genome (Kile, Hentges et al. 2003).

1.1.5 Genetic screens in mammalian cells

About 40 years ago, extensive tests were carried out to test the suitable growth media for growing mammalian cells *in vitro* (Grimm 2004). This work has laid the foundation of modern cell biology. Cultured mammalian cells soon become a favourite tool for geneticists because of the difficulty to perform genetic screens at the organism level in mammals. Even today, when the development of mouse genetics and genomics tools has made large-scale genetic screens in mouse feasible, mammalian cells of various origins are still of fundamental importance to utilize the vast quantities of data generated by the genome projects.

As a model system for genetic studies, cultured mammalian cells are very similar to yeast. They can both be grown on defined media, which makes the

growth conditions precisely controllable. They can both grow quickly under “ideal” conditions, which make it easy to accumulate a lot of experimental material in a short time. They both have well-documented origins and a uniform make-up, which make the identification of mutations relatively easy. Last but not least, they both can be easily manipulated genetically, which facilitates all sorts of genetic screen designs.

Yeast was the first model used for studying eukaryotic gene functions. Genetic screens in yeast shed light on many different molecular mechanisms from signal transduction to cell-cycle control, chromosome structure to secretion (Forsburg 2001). But bioinformatic analysis of the yeast genome sequence has shown that many human genes do not have homologs in yeast (Goffeau, Barrell et al. 1996). For example, yeast does not have any real counterparts of certain cellular processes ranging from apoptosis, tissue specific differentiation to oncogenic transformation (Grimm 2004).

Similar limitations also exist in other model systems. Because of the phylogenetic distance between these model organisms and mammals, extra care needs to be taken when researchers try to interpret the exact mechanisms in human according to the results obtained from these systems. Even for the mouse, genetic screens cannot be easily carried out for some cellular alterations because of the complexity of the intrinsic and extrinsic environments. Thus, the cultured mammalian cells become the ideal substitute for the model organisms.

Because of efficient mutagenesis protocols, the ease with which genetic material can be introduced into mammalian cells, the uniform genetic make-up and the defined growth conditions, cultured mammalian cells are one of the most widely used biological systems (Grimm 2004). The basic logic of a genetic screen in cultured mammalian cells is essentially the same as a screen in other genetic systems, but it also has some clear differences. First, in contrast to the whole organisms, which constitute a wide range of different cell types, mammalian cell lines, no matter what their tissue origins are, comprise a phenotypically and genetically uniform population. Second,

mammalian cell lines are also stable under the proper culture conditions. For most cell lines, they will not enter a developmental pathway without induction. This is a very important characteristic which can be used for screens for the determinants of the developmental fate decision of a cell line. Third, mammalian cell lines are kept under defined conditions, which make the variation in external environments negligible. In more complicated systems, a simple genetic change can sometimes cause various phenotypes, and a lot of different cell types might be involved, which makes the phenotype description and dissection difficult. Thus, mammalian cells provide a reductionist model that eliminates external variation.

Cultured mammalian cells provide a simple model system to study gene function in a complex organism. This is a big advantage for studying some basic biological processes, such as cell cycle and apoptosis. The knowledge obtained from this simplified model system might not exactly reflect what has happened *in vivo*, however it does provide a good start point to extrapolate the possible functions *in vivo*.

1.2 Mouse as a genetic tool

1.2.1 Introduction

The laboratory mouse, *Mus Musculus*, has been used to study human disease throughout the last century. For a long time, the study was limited to a few visible spontaneous mutations such as *agouti*, *reeler* and *obese* (Austin, Battey et al. 2004). The work on these spontaneous mutations has provided important insights into the molecular mechanisms of the relevant human diseases. However spontaneous mutations in mice do not provide enough different mutants for genetic studies. Many different methods have been developed to generate mutants in mouse at a higher rate, including gene-trapping, ENU mutagenesis and gene targeting.

Since the gene targeting technology became a reality in ES cells in the late 1980s (Thomas and Capecchi 1987; Capecchi 1989), the mouse has played a prominent role in functional genetics and genomics studies. Compared to other model organisms, *Mus Musculus* has some unique advantages for

studying human biology and disease. As a mammal, its development, body plan, physiology, behaviour and even its diseases can be very similar to human. It is also one of the model organisms that have the highest homology to the human.

1.2.2 Similarity between human and mouse

With the completion of the human genome (Lander, Linton et al. 2001; Venter, Adams et al. 2001), the biggest challenge now is to annotate the 2.9 billion nucleotides and decode all the information. Mouse is undoubtedly the key player in the process. After about 75 million years of divergence, the genomes of mouse and human have been altered so much by evolution that there is nearly one substitution for every two nucleotides, as well as deletions, insertions, translocations and inversions (Waterston, Lindblad-Toh et al. 2002). In spite of the divergence rate, systematic genome comparisons can still identify the highly conserved regions between these genomes, which indicate functional importance. Comparative genomics also help to identify the key differences between these two organisms and elucidate the driving force shaping their genomes.

The mouse genome is 14% smaller than the human genome (2.5 Gb compared to 2.9 Gb). Over 90% of the mouse and human genomes can be partitioned into corresponding regions of conserved synteny. Approximately 40% of the human genome can be aligned to the mouse genome, which represents most of the orthologous sequences that remain in both lineages from a common ancestor. The mouse and human genomes each seem to contain about 23,000 protein-coding genes. Approximately 80% of mouse genes have at least one identifiable orthologue in the human genome. Less than 1% of the mouse genes do not have any homologue currently detectable in the human genome (Waterston, Lindblad-Toh et al. 2002).

1.2.3 Tools available for mouse genetics and genomic studies

The widespread use of the mouse for biomedical research is largely due to the development of many genetic and genomic tools. One of the landmarks in mouse genetics was the isolation of pluripotent mouse embryonic stem (ES)

cells from mouse blastocysts (Evans and Kaufman 1981) and the subsequent demonstration that cultured ES cells can transmit through the mouse germline when reintroduced into host blastocysts (Bradley, Evans et al. 1984). Importantly, cultured ES cells maintain their pluripotency after modification of their genome which allows these modifications to be established in mice. Initially, the targets for modification were random or limited to a couple of mouse genes whose disruption could be selected by drugs, such as the *Hprt* gene on the hemizygous X chromosome (Kuehn, Bradley et al. 1987; Thomas and Capecchi 1987). A more general technology was needed to allow the disruption of the genes that could not be selected *in vitro* (Goldstein 2001).

Then came the second important breakthrough. Several groups independently demonstrated that targeted mutations could be introduced into ES cells by homologous recombination (Zijlstra, Li et al. 1989; Koller, Marrack et al. 1990; McMahon and Bradley 1990; Schwartzberg, Robertson et al. 1990). This technique allowed the precise disruption of any of the 23,000 mouse genes.

This pioneering work has established a new era in mouse genetics. Precisely engineered loss- or gain-of-function mutations can be established in the mouse through *in vitro* manipulation of ES cells. These approaches, together with the transgenic technique of zygote injection, are all classified as reverse genetics. Interestingly, the laboratory mouse is the first multi-cellular animal model organism in which gene targeting by homologous recombination became possible. Reverse genetics has become the main approach to identify gene function in mouse. This situation is partly due to the ease of genetic manipulation of mouse ES cells. Another important reason is that the cost of mouse breeding makes forward genetic screens a lot more expensive using mice compared to other model organisms.

Many new genetic and genomic tools have since been developed to help to decipher the information encoded in the mouse genome. Some of these, for example Cre/*lox*P technology, chromosome engineering and induced mitotic recombination, will be reviewed in the following chapters.

1.3 Mouse embryonic stem cells as a genetic tool

1.3.1 A brief history

The foundation of the mouse embryonic stem cell (ES) technology can be traced back to the observations made on teratocarcinomas and the embryonal carcinoma (EC) cells derived from them in 1970s (Chambers and Smith 2004). Teratomas, which arise spontaneously from independent germ cells in mouse testis, contain different types of tissue derived from all the three germ layers. In some cases, the tumours also contain undifferentiated stem cells, and these malignant teratomas are thus named teratocarcinomas.

The “stemness” of the teratocarcinomas can be demonstrated by the ability of these cells to form secondary teratocarcinomas after transplantation. Undifferentiated teratocarcinoma cells can also be maintained *in vitro*. Moreover, if the EC cells are injected into blastocysts, they can sometimes be incorporated into the embryos and contribute to various cell types (Chambers and Smith 2004).

Because of their tumour origin, EC cells are mostly aneuploid, which greatly limits their ability to differentiate *in vitro*. But the unique characteristics of these pluripotent cells *in vitro* and *in vivo* raised an important question: Do they have a pluripotent counterpart in normal blastocysts that acts similarly? This presumed similarity eventually led to the isolation of the pluripotent mouse embryonic stem (ES) cells (Evans and Kaufman 1981). Initially, these cells were called “teratocarcinoma stem cells” because they share so much similarity with EC cells, such as their appearance, the culture conditions, their unlimited self-renewal and their ability to differentiate *in vivo* and *in vitro*. However, these embryo-derived cells were shown to be more stable and thus more controllable in genetic terms.

The pioneering work in ES cell biology has made it possible to target any mouse gene precisely and then transmit the mutation to the germline. The ease of manipulation of mouse ES cells has made them the “workhorse” of mouse genetics.

1.3.2 Mouse embryonic stem cells compared with the mouse

Compared to the other model organisms, the laboratory mouse can provide more accurate disease models for the human. But no model organism is perfect, and the mouse is no exception. Logistical and cost considerations that come with breeding large numbers of mice greatly limit the scale of mouse genetic studies. Even when the proposed International Knockout Mouse Project (Austin, Battey et al. 2004) and the European Mouse Genome Mutagenesis Program (Auwerx, Avner et al. 2004) becomes reality, the scale of the genetic screens will still be limited by the mouse lines that can be accommodated in a given institution. Different targeting strategies, ES cell origins and mouse strain background will all affect the phenotype observed even when a same gene is disrupted.

Another complication of a global knockout project is the fact that the disruption of many of the 23,000 mouse genes will cause embryonic lethality. Mouse embryos are covered by many layers of maternal tissues in the uterus. So observations can not easily be made without killing the pregnant female. The identification of the cause of embryonic lethality thus requires a lot of experimental analysis.

Mouse ES cells have become a key tool for mouse genetics. But their unique characteristics also mean that they can be used as an independent experimental system for studying early embryonic development. ES cells retain their unlimited self-renewal and differentiation capacity under appropriate culture conditions. They can also differentiate, both *in vitro* and *in vivo*, into almost all specialized cell types and their *in vitro* differentiation recapitulates the early embryogenesis (Wobus 2001).

The *in vitro* differentiation of ES cells has been widely studied to define the parallels with early embryonic development. Using a panel of markers representative of the early germ layers and late cell lineages, progressive differentiation of embryoid bodies (EB) has been correlated with early embryogenesis of mouse embryos (Leahy, Xiong et al. 1999). It is interesting to note that the temporal and spatial expression pattern of these markers is

strikingly similar between the EBs and the embryos. Thus *in vitro* differentiation of ES cells can serve as a system to study early lineage determination and organogenesis in mouse. These markers can be used to screen for mutations in ES cells that disrupt these processes.

1.3.3 Mouse embryonic stem cells compared with other cultured mammalian cells

Compared with other mammalian culture cell lines, mouse ES cells offer some unique advantages in cell-based screens.

First, many mammalian cell lines that are in use now, for example the human cervical carcinoma “Hela” cells and the mouse monkey embryonic kidney “Cos-7” cells, were either transformed *in vitro* or obtained from tumours to get immortalized lines. Most of these cell lines are aneuploid (Grimm 2004) and thus care needs to be taken when the data obtained from these cells is used to interpret what really happens in normal cells. In contrast, although ES cells exhibit unlimited growth in culture, they are stable in their genomic structure and they remain undifferentiated with a stable phenotype under the appropriate culture conditions. It is reasonable to argue that the biology of the ES cells more precisely reflects a normal biological and physiological status than that displayed by highly aneuploid transformed cell lines.

Second, homologous recombination in ES cells is much more efficient than that in the other somatic cell lines with the possible exception of DT40 cells. The ease of genetic manipulation allows the introduction of virtually any kinds of gene-targeting construct and reporter cassettes into almost any locus in ES cells.

Third, because most mammalian cell lines were derived from a variety of differentiated tissues, their differentiation capacity is highly limited (Grimm 2004). To identify the functions of a gene in different tissue contexts, cell lines from various differentiation stages need to be used. It is more efficient to disrupt the gene in ES cells and differentiate them into different cell types *in vitro* and *in vivo*.

It is estimated that more than 10,000 genes are expressed in ES cells (Sharov, Piao et al. 2003). Most of these genes are either the structural components or the essential players of basic processes common to all cells, for example metabolism, signalling, cell division and DNA repair. The others are genes that govern the special properties of embryonic stem cells. The second category of genes are especially attractive because the knowledge of these will have implications not only for academic research, but also for the clinical application of ES cells in cell replacement therapy (Ramalho-Santos, Yoon et al. 2002).

Several genetic screens have recently been published in which ES cells have been used identify genes in different pathways. Chambers *et al.* (2003) have used a gain-of-function approach to isolate self-renewal determinants in mouse ES cells (Chambers, Colby et al. 2003). Expression cloning was used to identify a homeodomain protein, *Nanog*, which when over-expressed can drive ES cell self-renewal without LIF. Another two groups have exploited the high rate of mitotic recombination in *Bloom*-deficient ES cells to screen for recessive mutations related to the DNA mismatch repair pathway (Guo, Wang et al. 2004) and glycosylphosphatidylinasitol-anchor biosynthesis pathway (Yusa, Horie et al. 2004).

1.3.4 Genetic and epigenetic instability of mouse embryonic stem cells

Although mouse embryonic stem cells show great potential in cell-based genetic screens, cautions need to be taken in designing screens in ES cells and also in interpreting the results of the screens because of the genetic and epigenetic instability of ES cells maintained in culture.

It was noticed that there is clear clonal variance in the efficiency of germ line transmission of ES cell clones derived even from the same parental cell line. The germline transmission ability of ES cells also decreases when the passage number increases (Nagy, Rossant et al. 1993). A possible explanation is that genetic alterations, especially those which will provide the mutant ES cells with growth advantages, can accumulate in ES cells

cultivated *in vitro*. When the passage number increases, these mutant cells will dominate the cell population because of their growth advantage, and thus interfere with the germline transmission.

When the growth rate, karyotype and the efficiency of germ line transmission are examined, it was found that chromosomal abnormality occurred rather frequently in ES cells (Nichols, Evans et al. 1990; Liu, Wu et al. 1997). A number of chromosomes can be randomly duplicated in culture, especially trisomy 8 and trisomy 11, which are directly associated a growth advantage *in vitro* and the failure of ES cells to contribute to the germ line. It is reasonable to predict that the mutant cells with trisomy 8 or trisomy 11 will also have abnormality in *in vitro* differentiation, which might interfere with genetic screens using ES cell differentiation.

Although the mutant cells with trisomies can have dramatic abnormalities both *in vivo* and *in vitro*, these cells can be distinguished by their accelerated growth rate or by karotype analysis. Subcloning of the parental cell line is an easy way to get a normal population of ES cells for further analysis.

Epigenetic instability in cultured ES cells can also impact on genetic screens using ES cells. Mouse embryonic stem cells were isolated from inner cell mass of the blastocysts (Evans and Kaufman 1981). Theoretically, they should carry the same epigenetic information as their *in vivo* counterparts. However, epigenetic state of the ES cell genome may not be stable under *in vitro* culture conditions. Epigenetic variance was observed in different ES cell lines and even in those cells derived from ES cells of the same subclone (Humpherys, Eggan et al. 2001). Epigenetic alterations at one imprinted locus did not necessarily predict changes at other loci, which suggests that the epigenetic instability of ES cells is more likely to be caused by random local loss of imprinting, instead of global increase or decrease of the methylation level in the ES cell genome.

The epigenetic variability was even found in the placentas of cloned mice derived from the same cell line (Humpherys, Eggan et al. 2001). However,

epigenetic instability of murine ES cells does not interfere with their germline transmission efficiency. Mammalian development may be rather tolerant to local epigenetic abnormalities, and unless a global loss of imprinting happens, the pluripotency of the ES cells will not be compromised. However, differentiation may be biased by altered imprinting (Mann, Gadi et al. 1990).

In summary, the unique intrinsic characteristics of ES cells have already made them a promising system to address a wide range of basic cell biology and developmental questions. The potential clinical application of ES cell biology will attract more and more researchers to use different methods to understand how the ES cells maintain self-renewal and how they differentiate into other cell types. However, care needs to be taken to monitor and control for the genetic and epigenetic status of cultured ES cells.

1.4 Cre/*lox*P site specific recombination

1.4.1 A brief history

Site-specific recombination in multicellular organisms was first achieved in *Drosophila* (Golic and Lindquist 1989). Flp recombinase from the yeast 2μ plasmid can efficiently mediate site-specific recombination between *FRT* (Flp recombinase target) sites in the fruitfly. The Flp/*FRT* system has been widely used for creating deletions, duplications, inversions and genetic mosaics.

The most widely used site-specific recombination method in mouse is based on another recombinase, Cre, although recent work shows that Flp/*FRT* works as efficiently in mouse as it does in the fruitfly. The recombinase, Cre, from the P1 bacteriophage belongs to the integrase family of site-specific recombinases (Hamilton and Abremski 1984). Cre can catalyze the recombination between two *lox*P sites. The *lox*P site is a 34-bp consensus sequence, which includes two inverted 13-bp flanking sequences on both sides of an 8-bp core spacer sequence. The core spacer decides the orientation of the *lox*P site, but the flanking sequences are the actual binding site of Cre.

The Cre/*lox*P site-specific recombination system was first shown to work in mammalian cells in the late 1980s (Sauer and Henderson 1988). In the early 1990s, this system was shown to work in mouse when Cre was expressed *in vivo* (Lakso, Sauer et al. 1992; Orban, Chui et al. 1992; Gu, Zou et al. 1993). The Cre/*lox*P system has been widely used in mouse genetics, combined with the gene targeting and transgenic technology, the Cre/*loxP* system has made it easy for mouse geneticists to tailor the mouse genome almost without any limitations, from one-base-pair point mutations to mega-base-level deletions, inversions, duplications and translocations.

1.4.2 Basic Characteristics of Cre/*lox*P system

The 34-bp *lox*P site is short enough to be put into large introns without disrupting the transcription of the gene. It is also long enough to avoid the random occurrence of intrinsic *loxP* site in the mouse genome. With the completion of the sequencing of several major model organisms, searches reveal that no perfectly matched *lox*P site has even been found in any organisms other than the P1 bacteriophage. It has been noted that some pseudo recognition sites exist in the mouse genome but the efficiency of recombination between wild type *lox*P sites and these pseudo sites has not been thoroughly studied.

In vitro, Cre-mediated recombination is efficient enough to excise genomic regions as large as 400 kb, and recombinants can be identified without selection (Nagy 2000). The Cre recombinase is also very efficient *in vivo*. Numerous Cre transgenic lines have been established in the last decade to facilitate efficient Cre-mediated excision in a lot of different developmental stages and different cell types. One aim is to generate more Cre transgenic lines to cover all development processes and cell types (Nagy 2000). A resource like this will greatly help the study of gene function *in vivo*, especially for the genes that cause lethality at early stages when disrupted.

1.4.3 Application of Cre/*lox*P system in mouse genetics

1.4.3.1 Conditional gene knock-out

One of the common uses of the Cre/*lox*P site-specific recombination is for conditional gene knockouts. The logic behind this powerful tool is simple: two *loxP* sites in the same orientation are placed on both sides of the most important functional domain of the gene of interest. Since the *lox*P site is only 34 bp, usually it will not affect the gene transcription if it is placed in the non-conserved region of an intron. The targeted ES cells and the animals containing such an allele are perfectly normal compared to wild type animals. But when the animals are crossed to a Cre-expressing transgenic line, the progeny that carry both the Cre transgene and the *lox*P-flanked allele will excise the *lox*P-flanked portion of the gene in the cells that express Cre (Tsien, Chen et al. 1996).

Two main issues with the conditional knockout approach are the design of the conditional targeting construct and the specificity and efficiency of the Cre line. When a conditional targeting vector is designed, the region that is selected to be flanked by *lox*P sites needs to be important enough to disrupt the gene function completely when excised. The flanked region also needs to be small enough for the two *lox*P sites to be introduced into ES cells in one targeting step. The availability of restriction enzyme sites, the size of the genomic insert that a vector can incorporate and the subsequent genotyping strategy all limit the choice of the position of *lox*P sites. The development of *E. coli* recombineering technology recently has greatly simplified the method and allowed the flexibility of design of conditional targeting vector (Copeland, Jenkins et al. 2001).

Although many Cre transgenic lines have been generated in the last decade, they are still not enough to satisfy the increasing need (an incomplete list of available Cre excision lines can be found on the webpage of Dr. Andras Nagy's lab http://www.mshri.on.ca/nagy). Even for existing Cre lines, leaky expression of Cre in the wrong cell type and/or developmental stage, or incomplete excision in target cells makes the interpretation of the phenotypes difficult, or can result in no phenotype at all, for example in a mosaic tissue.

In spite of these problems, the conditional gene knockout technology, which combines the strength of gene targeting and site-specific recombination, is still a powerful tool for the mouse geneticist, which will play an increasingly important role in functional genetic studies.

1.4.3.2 Selectable marker removal and recycling

It has been reported that in some cases, the selection cassette used for gene targeting will affect the expression of genes nearby *in vivo*. An easy way to circumvent this problem is to flank the selection marker with two *lox*P sites and "pop out" the cassette either *in vivo* or *in vitro*. This has already become a routine procedure for both traditional and conditional targeting.

Another advantage for marker removal is that the same selection marker can be reused in the subsequent manipulations of the ES cells. For studying gene function *in vitro*, especially for genetic screens, multiple gene targeting events are usually required to disrupt a number of loci or introduce reporter cassettes. If the usable markers are exhausted, this will limit downstream analysis. Positive-negative selection marker flanked by *lox*P sites can be used to generate mouse ES cells that carry multiple targeted mutations but devoid of any exogenous markers (Abuin and Bradley 1996).

1.4.3.3 Subtle change and hypomorphic alleles

Another important application of the Cre/*lox*P system is to create subtle mutations. For most of the mouse knockout lines published so far, either an important domain of a gene or even the whole gene has been deleted. This approach is more likely to create a null allele, but considering the fact that many human hereditary diseases are caused by point mutations, small deletions and small insertions, null alleles in mice might not generate an ideal model for their relevant human disease. Also, hypomorphic alleles, which partially disrupt gene function, are sometimes more useful for genetic screens.

A typical way to introduce subtle genetic changes into a gene in mouse ES cells is to make the change in one of the homology arms of the targeting vector, and include a selection marker flanked by *lox*P sites in a non-coding region. After subtle mutation has been confirmed to have been incorporated, the selection marker can be popped out either *in vitro* or *in vivo*, leaving only the small change (Nagy, Moens et al. 1998).

1.4.3.4 Chromosome rearrangement

Chromosome rearrangements happen spontaneously in almost all the eukaryotic species. They play a very important role in evolution, but in humans, they are also one of the most common causes of foetal losses, developmental disorders and cancer (Yu and Bradley 2001). Thus, engineered chromosome rearrangements in mouse generated by long range Cre/*lox*P recombination can be used to model their human counterparts and investigate the molecular mechanisms underlying a variety of different human genetic disorders.

Chromosome deletions and inversions are also useful tools for performing recessive genetic screens. Large deletions can reduce the diploid genome to areas of segmental haploidy, which allows F1 screens for recessive mutations in the deletion region (Ramirez-Solis, Liu et al. 1995). On the other hand, inversions, which serve as balancer chromosomes, can be used to maintain lethal recessive mutations in the inversion interval (Zheng, Sage et al. 1999). Creating a resource of inversions and deletions throughout the mouse genome will be important for large-scale phenotype-driven mutagenesis programs.

1.5 Chromosome engineering

1.5.1 A brief history

Chromosome engineering has its origin in *Drosophila* genetics. Spontaneous chromosome rearrangements were found and mapped by observing fruitfly salivary gland polytene chromosomes under the microscope. These rearrangements are very useful tools in genetic studies. For example, inversions could be used to maintain lethal mutations without selection.

In 1927, Muller showed that ionizing radiation could induce different kinds of genetic damage, including chromosomal rearrangements (Rubin and Lewis 2000). He was awarded the Nobel Prize in 1946 for this finding. This method has been exploited to its logical extremes by D. L. Lindsley and his colleagues in 1972 to generate an ordered set of duplications and deletions spanning the major autosomes (Lindsley, Sandler et al. 1972). This effort leads to the collection of deletion lines that provide maximal coverage of the genome in a minimum number of stocks held in the Blooming Stock Center (St Johnston 2002). This resource provides ideal starting material for region-specific mutagenesis screens. It also provides a rapid way to map recessive mutations found in genetic screens.

Though X-rays are also very efficient in inducing genetic damage in other species, a similar genome-wide chromosomal rearrangement resource is not available in any other multi-cellular organisms. This is partly due to one unique characteristic of *Drosophila*, the salivary gland polytene chromosomes, which have made the physical mapping of chromosome rearrangements significantly easier than in other species. The lack of efficient methods to determine the endpoints of rearrangements has made it hard to replicate this genome-wide resource in other species.

1.5.2 Engineering mouse chromosome with Cre/*lox*P

Spontaneous chromosome arrangements are very rare in nature, and even if they do happen, there are practical difficulties in recovering them. The same problem also exists in the arrangements induced by radiation and other chemical mutagens (Yu and Bradley 2001). Clearly, induced arrangements with pre-determined end points will be more useful for genetic studies. The Cre/*lox*P site-specific recombination system is now the most commonly used method to generate these rearrangements in mouse.

1.5.2.1 Deletions, duplications and inversions

Deletions, duplications, translocations and inversions with pre-determined end points are useful not only for creating human disease models, but also for making genetic tools for functional genomics studies.

If large genomic regions are involved, an efficient selection strategy is needed to identify the ES clones that carry the desired genomic re-arrangement. A common way to achieve this is to put the *loxP* sites into two non-functional halves of the hypoxanthine phosphoribosyl transferase (*Hprt*) mini-gene. Then the two halves are sequentially targeted to two pre-determined end points of *Hprt*-deficient ES cells. Transient Cre expression induces the site-specific recombination and restores the activity of *Hprt* mini-gene. HAT selection can directly select the clones with the desired chromosomal rearrangements (Ramirez-Solis, Liu et al. 1995; Smith, De Sousa et al. 1995).

The bottleneck for making the targeting vectors used for chromosome engineering is to isolate the end-point genomic fragments, which normally requires laborious genomic library screening. A two-library system of pre-made targeting vectors has greatly simplified the procedure (Zheng, Mills et al. 1999). Also, by incorporating the coat colour markers into the vector backbone, the mice that carry the chromosome rearrangements can be genotyped easily by eye.

To evaluate the efficiency of Cre-mediated recombination over long distances, Zheng et al. (2000) created a series of deletions, duplication and inversions on mouse chromosome 11 and compared their relative efficiency *in vitro* (Zheng, Sage et al. 2000). It has been shown that, although the site-specific recombination efficiency decreases with increasing distances, rearrangements as large as three quarters of chromosome 11 can be achieved with a proper selection strategy. The only limitation seems to be the haploinsufficiency that comes with large deletions (Liu, Zhang et al. 1998; Zheng, Sage et al. 2000). Although the recombination can still occur over these distances, HAT resistant clones are only recovered if they duplicate the wild type chromosome to compensate the loss caused by the large deletion.

1.5.2.2 Nested deletions

For human genetic disorders that are caused by spontaneous deletions, a common way to map the disease gene is to identify the end points of the deletions in a collection of patients. Since the deletions are often of different sizes and with different end points, a minimum overlapping region can be defined and used to locate the disease gene(s). The availability of the genetic material is limited, thus it is not always possible to identify the relevant gene(s). Instead, a key region and several candidate genes will often be suggested for future studies. If the minimum overlapping region is relatively small and only a few genes are involved, it is possible to knockout these genes one by one and analyse the mouse phenotype to identify the disease gene. But if the range is too big, additional steps are needed to further reduce the size of the region.

The conserved synteny between human and mouse can be used to define the region corresponding to the deletion in the human genome. Nested deletions can then be constructed to map the disease gene (Yu and Bradley 2001). First, a 5' *hprt*-*lox*P cassette was targeted to a predetermined locus to serve as an anchor point. The *lox*P-3' *hprt* cassette was then randomly integrated into the ES cell genome by retroviral infection. ES cells with nested deletions can be isolated by transient expression of Cre and subsequent HAT selection (Su, Wang et al. 2000).

In this strategy, the introduction of a second *lox*P site is random, thus only a small subset of the viral integrations will occur on the same chromosome as the fist anchor *lox*P site. The strategy to recover these rare events from the pool of random insertions in the ES genome is based on two key observations of Cre/*lox*P recombination efficiency. First, Cre efficiency does not change appreciably over the range of several megabases (Zheng, Sage et al. 2000). So this predicts that small deletions will not be generated any more efficiently than large ones. Second, Cre-mediated recombination within a few megabases on the same chromosome (*cis*) is two-three orders of magnitude more efficient than recombination between two *lox*P sites located on

homologous chromosomes (*trans*) (Ramirez-Solis, Liu et al. 1995; Liu, Zhang et al. 1998). The recombination efficiency between *lox*P sites on non-homologous chromosomes is even lower (Zheng, Sage et al. 2000).

Taking these two observations together, it is reasonable to predict that in from a pool of random insertions, the HAT-resistant clones recovered after exposure to Cre should have deletions within several megabases of the anchor point. This prediction has been proved to be true, a nested set of deletions were generated on the mouse chromosome X and chromosome 11 (Su, Wang et al. 2000). Most deletions recovered in this report were mapped within 1 cM distal or proximal to *Hprt* or E_2DH. It was also noticed that although the deletions proximal to *Hsd17b1*, the chromosome 11 anchor point, could be transmitted through the germline and maintained as heterozygotes, the deletions distal to *Hsd17b1* could not be transmitted, which might be due to the severe haploinsufficiency reported in that region (Liu, Zhang et al. 1998).

The nested deletion strategy has been successfully used to map *Tbx1*, the gene responsible for the DiGeorge Syndrome (Lindsay, Vitelli et al. 2001). The key deletion region in human encompasses around 1 Mb, and contains at least 15 genes. Since human deletion map was not helpful to determine the causal gene, several deletions were made in mouse ES cells by both traditional chromosome engineering and nested deletion methods. Some of the deletion mouse lines showed the typical clinical phenotype, while others did not. By comparing these phenotypes with the deletion regions, the critical region was reduced to 200 kb between *T10* and *Cdcrel1* genes. PAC transgenesis was then used to rescue the phenotype in these deletion lines and finally identify the disease gene as *Tbx1*.

1.5.2.3 Regional trapping

The nested deletion strategy described by Su et al. (2000) was shown to be an efficient way to create a series of deficiencies in a region of interest. But like other deletion strategies, a large portion of the deletions generated cannot be transmitted into the germline, greatly limiting the use of this strategy as a

convenient way to mutate genes (Liu, Zhang et al. 1998; Su, Wang et al. 2000; Zheng, Sage et al. 2000).

On the other hand, the success of the nested deletion strategy demonstrated that retrovirus integrations in a small region around an anchor point could be recovered using an appropriate selection procedure *in vitro*. If retroviral integrations occur in the coding region of a gene, then a variation of this strategy could be utilized to accumulate mutations around an anchor point. This has led to the idea of “regional trapping”, which use a gene-trapping virus to disrupt genes and then use site-specific recombination to select for trapping events that occur in the desired region (Wentland, unpublished data).

Cre recombination occurs efficiently between *lox*P sites on the same chromosome (*cis*) (Liu, Zhang et al. 1998; Zheng, Sage et al. 2000). For a given physical distance, the efficiency of recovery of inversions is higher than deletions because inversions do not result in the loss of any genetic material and thus ES cells with inversions do not suffer growth disadvantages compared with cell lines with deletions (Zheng, Sage et al. 2000). Theoretically, an inversion should only disrupt the genes close to the two end points. If the phenotype of the anchor point in the regional trapping is already known, any new phenotype of the inversion is most likely related to the other breakpoint, although it is possible that an inversion can sometimes disrupt long-range regulatory elements.

A strategy similar to that described for making nested deletions has been used for regional trapping. In this strategy, the anchor point was introduced into ES cells by targeting the 5’ *Hprt-lox*P cassette into the anchor point locus, *Hsd17b1*. The *loxP*-3’*Hprt* cassette was then randomly integrated into the ES cell genome by retroviral infection using a gene-trapping retroviral vector. Gene-trap insertions on chromosome 11 can be selected by transient expression of Cre which induces an inversion, cells with recombination events can be selected in HAT (Wentland, unpublished data).

In this strategy, gene-trap insertions along the whole length of the chromosome have been recovered. As predicted, about 86% of the gene-traps were concentrated on the distal part of chromosome 11, and fall within a 43 Mb region surrounding the *Hsd17b1* locus. The largest confirmed inversion was 82 Mb (Wentland, unpublished data). The range over which this strategy is effective was much larger than the range of deletions achieved by Su et al. (2000). Since the same anchor point was used for both experiments, the data shows that cells with inversions are more likely to be viable than cells with deletions of similar size.

It is interesting to note that in this study, a large portion of trapped loci were neither predicted by *in-silico* gene prediction, nor supported by any EST sequences. Some of these genes were verified to express in the embryo or adult tissues by RT-PCR, but few of them were expressed in ES cells. This data demonstrates the use of 3' trapping to mutate genes that do not normally express in undifferentiated ES cells (Wentland, unpublished data).

The "genes" which do not appear to be expressed might have been recovered with the 3' gene-trap selection because of pseudo splice acceptors and polyadenylation signals scattered in the genome. Some of the 3' RACE products matched repeats such as retrotransposons and SINEs. Though approximately 37.5% of the mouse genome is comprised of these repeat elements, this did not cause any serious background problems.

Regional trapping disrupts a trapped gene by moving the part of the gene upstream of the proviral integration site away from the downstream part. So this method should be more mutagenic than traditional trapping. But inversions have the potential complication of disrupting genes nearby whose coding region or transcriptional regulatory regions overlap or fall on either end of the trapped gene. Thus expression of the genes around the breakpoint needs to be checked to avoid misinterpretation of the phenotype (Wentland, unpublished data).

The regional trapping strategy is potentially useful for finding disease genes in a given region. Because of the conserved synteny between the human and mouse genomes, it is easy to find a mouse genomic region which corresponds to a human disease candidate region. By regional trapping, it is possible to mutate a large portion of the genes in that region in a quick and efficient way, facilitating regional screening of disease genes in the mouse genome (Wentland, unpublished data).

1.6 Induced mitotic recombination

1.6.1 A brief history

Chimaeras are individuals that are formed from cells of different origins, and genetic mosaics refer to the individuals that contains cells of different genotype but of the same origin (Rossant and Spence 1998). Mosaics are particularly important for the study of cell lineage, cell fate determination and cell-cell interactions. Some human diseases are caused by somatic loss of heterozygousity (LOH) caused by mitotic recombination or chromosome loss and re-duplication. Genetic mosaics can serve as models for these types of human diseases.

If mitotic recombination occurs during the G2 phase of the cell cycle and the recombinant chromatids segregate to different daughter cells (G2-X), the daughter cells will have either two copies of the paternal chromosome or two copies of the maternal chromosome. If the recombinant chromatids segregate to the same daughter cells, the daughter cells will still have one copy of maternal chromosome and one copy paternal chromosome (G2-Z). If the mitotic recombination happens during G1 phase of the cell cycle, this is genetically neutral as all the daughter cells will be the same as the parental cell (Golic 1991).

Historically, mosaics have been widely used to address many different developmental questions (Xu and Harrison 1994). Since the spontaneous mitotic recombination rate is too low for any practical use in genetic studies, various methods have been developed to induce mitotic recombination. X-ray irradiation is the most frequently used method to achieve this purpose. But the

shortcomings of this method greatly limit its usage as a genetic tool. First, the dosage required to induce mitotic recombination limits experimental use. High dosage of ionizing irradiation causes excessive cell death, whilst a low dosage is inefficient at inducing mitotic recombination. Second, the mitotic recombination induced by X-rays occurs randomly in the genome, so unless a genetic marker which is visible at the cellular level (such as the eye colour markers in the fruitfly and the coat colour markers in the mouse) is located near the allele of interest, it is impossible to distinguish the mosaic clones from their background.

1.6.2 Induced mitotic recombination in fruitfly

After the Flp/*FRT* system was shown to be able to mediate site-specific recombination in *Drosophila* (Golic and Lindquist 1989), this system was quickly applied to generate mitotic recombination clones by creating flies with transgenic FRT sites at the same position on homologous chromosomes (Golic 1991). Xu et al. (1993) have constructed a series of *FRT* transgenic lines to cover all the four *Drosophila* chromosomes. This resource has made it possible to create mosaic animals for 95% of *Drosophila* genes. Each of these lines carries not only an *FRT* site close to the centromere of one of the chromosomes, but also a cell-autonomous marker distal to the *FRT* site. When these *FRT* lines are crossed to a Flp recombinase transgenic line, the marker can be used to distinguish cells of different genotypes (Xu and Rubin 1993).

This method has been successfully used to screen for mutations that produce tumorous outgrowth in the imaginal discs (Xu, Wang et al. 1995). A tumour suppressor gene, *large tumor suppressor (lats)*, which encodes a protein kinase, has been discovered in this screen. This example clearly shows the usefulness of induced mitotic recombination. The *lats* gene was found to cause a wide range of defects throughout development. All the alleles of *lats* were also found to be lethal at different stages (Xu, Wang et al. 1995). So it would have been almost impossible for a traditional recessive screen to identify the tumor-suppressor function of *lats* in the adult.

Induced mitotic recombination is very efficient in *Drosophila*, largely due to its chromosome structural and mechanical characteristics (Liu, Jenkins et al. 2002). For example, in *Drosophila*, nearly two-thirds of the mitotic recombination events are G2-X, and G1 recombination accounts for the remaining one-third, whereas G2-Z events are very infrequent (Golic 1991; Xu, Wang et al. 1995). Also in *Drosophila*, the homologous chromosomes are paired in somatic cells, which makes mitotic recombination more likely to happen.

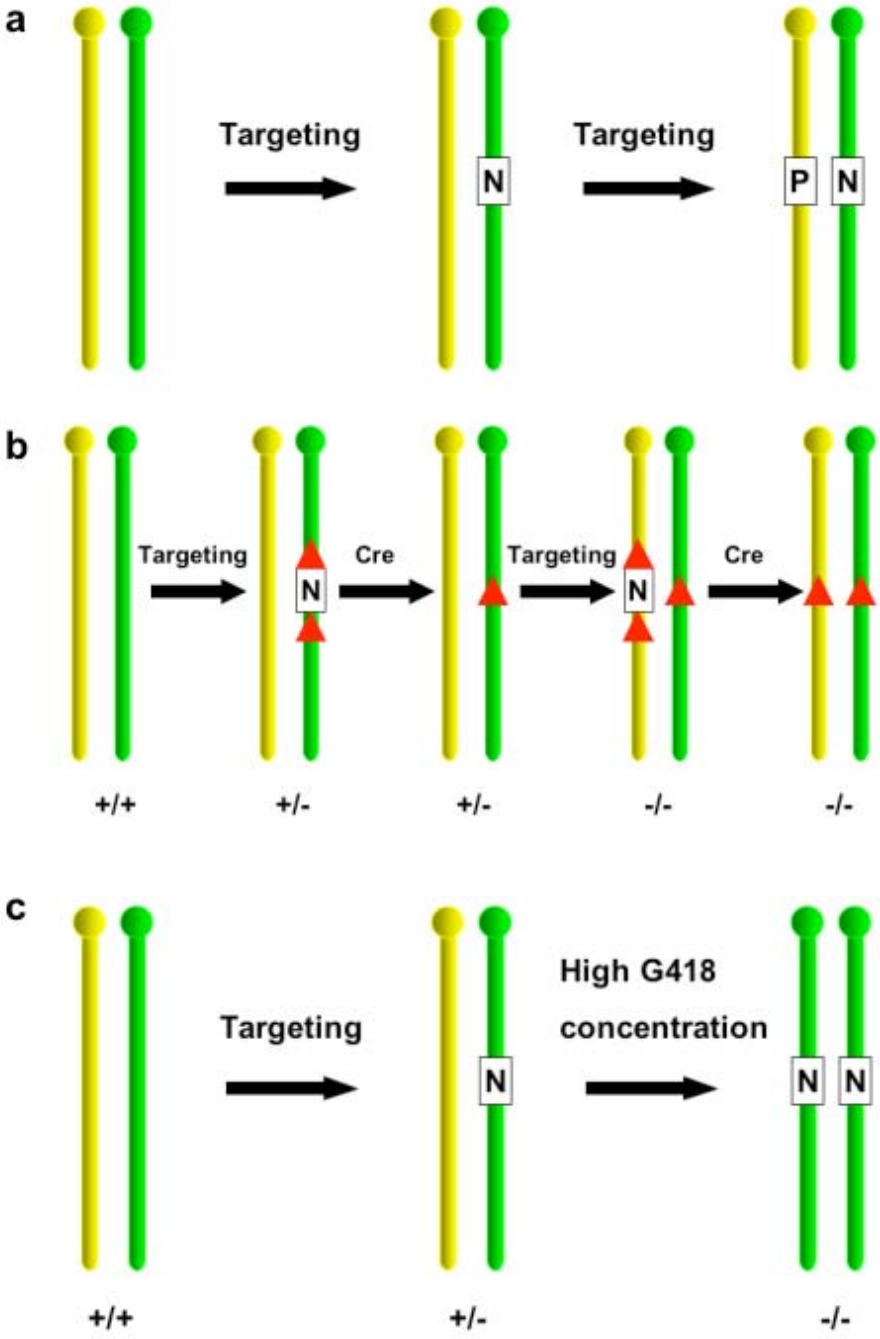
One disadvantage of the induced mitotic recombination approach is that it can only be used to screen for genes distal to the *FRT* site. Most *FRT* lines were created by random integration of a dominantly marked *FRT*-containing P element and FISH was then used to map the integration sites (Xu and Rubin 1993). Using this approach, it is hard to find integrations that are located very close to the centromeres. Also, only chromosome-specific screens can be carried out using this method. Separate screens for each of the five arms are required to cover the entire *Drosophila* genome (St Johnston 2002).

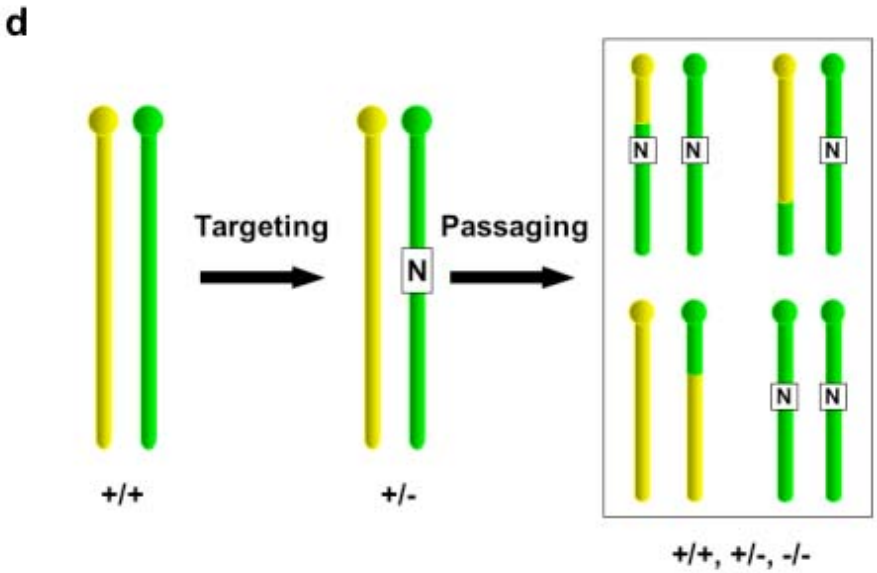
1.6.3 Traditional methods to generate homozygous mutations in mouse embryonic stem cells

Homozygously mutated ES cells are important resources to study gene function *in vitro*. To make a homozygous mutant ES cell line is a complicated process that normally requires several targeting steps and a screening strategy to identify the correctly targeted clones.

1.6.3.1 Sequential gene-targeting

The most frequently used method to create loss-of-function mutations is to sequentially target both alleles of a gene in ES cells. Two different drug selection markers are needed for the two targeting events (Fig. 1-1a) (te Riele, Maandag et al. 1990). An alternative way is to flank the drug selection cassette with two *lox*P sites and remove the selection marker after targeting by Cre-mediated recombination (Fig. 1-1b). The same targeting vector can then be reused to target the second allele (Abuin and Bradley 1996).


a
Targeting
N
Targeting
P
N
b
Targeting
N
Cre
Targeting
N
Cre
+/+
+/-
+/-
-/-
-/-
c
Targeting
N
High G418
concentration
N
N
+/+
+/-
-/-


d
Targeting
Passaging
N
+/+
+/-
+/+, +/-, -/-

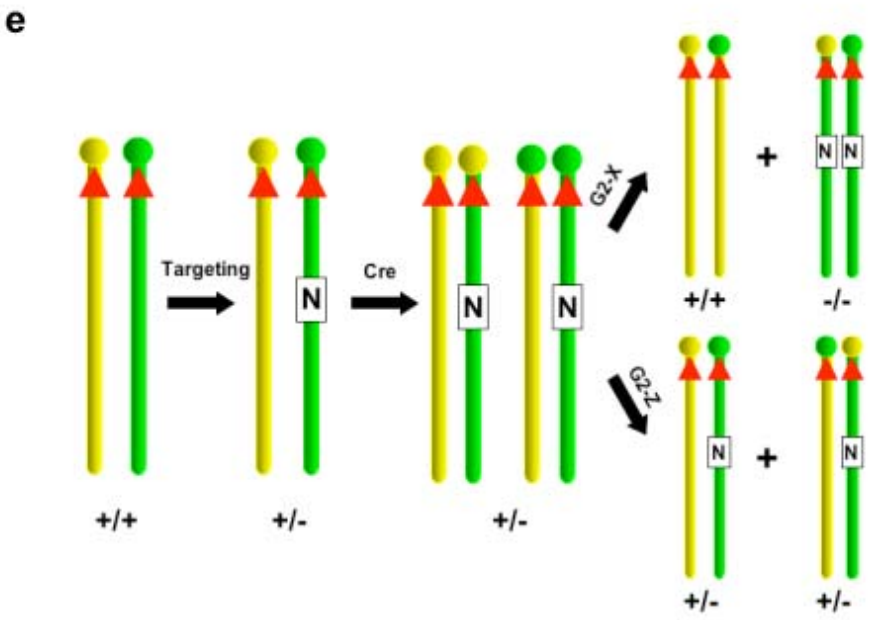

e
Targeting
Cre
N
G2-X
G2-Z
+
+/+
-/-
+/-
+/-
+/-
+/-

Fig. 1-1 Methods to generate homozygous mutations in ES cells. a. Sequential gene targeting using two different selection cassettes. Two vectors with different selection markers are used to target both alleles of a genomic locus. The two selection markers cannot be removed after targeting. **b.** Sequential gene targeting using *lox*P flanked selection cassettes. After the first allele of a genomic locus is targeted by a *lox*P flanked cassette, Cre-mediated site-specific recombination can be used to remove the cassette. The same vector can be used to target the second allele. This approach can be used to generate marker-free homozygously mutated ES cell lines. **c.** LOH induced by a high G418 concentration. After the first allele of a genomic locus is targeted by a *neo* cassette, high G418 concentration can induce the loss of the un-targeted chromosome and the subsequent duplication of the targeted one. **d.** Elevated mitotic recombination in *Blm*-deficient ES cells. The *Blm*-deficiency in mouse ES cells caused a 20-fold increase in the rate of LOH compared to wild type cells. A single allele mutation will become homozygous randomly in this genetic background. But the resulting pool of cells will be a mixture of heterozygous, homozygous and wild type clones. **e.** Induced mitotic recombination. Cre-mediated site-specific recombination can be used to induce mitotic recombination between two *lox*P sites targeted to two homologous chromosomes. Depending on the segregation pattern following the mitotic recombination, some of the daughter cells will become homozygous for the targeted locus. Yellow and green bars: homologous chromosomes; red arrow, *lox*P site; N, *neo* selection marker; P, *puro* selection marker.

The advantage for this approach is that gene targeting allows precise disruption of the gene of interest. But it also has its limitations. For targeting both alleles using different drug selection markers, the same markers cannot be reused for future steps. The selection cassette recycling approach requires an additional step of popping out the selection marker. Both methods are time-consuming and hard to scale up, so they can only be applied on a gene-by-gene basis.

1.6.3.2 High concentration G418 selection

Homozygous mutated cells also occur spontaneously from cultured mammalian somatic cells containing a heterozygous mutation, a process known as loss of heterozygosity (LOH). LOH can occur by many mechanisms including regional or whole chromosome loss, mitotic recombination and gene inactivation. But a selection strategy is needed to identify these rare events.

It has been shown that when heterozygous cells targeted with a Neomycin (*Neo*) drug resistance cassette are grown in high concentrations of G418, many of the surviving cells are homozygous for the targeted allele (Mortensen, Conner et al. 1992). This strategy provides an easy way to generate ES cells in which both alleles have been targeted. The existence of two copies of the *Neo* cassette in these cells suggests that LOH has occurred either by mitotic recombination between homologous non-sister chromatids, by chromosomal loss followed by chromosomal duplication (Fig. 1-1c) or by a local gene conversion event.

To investigate the mechanism of LOH in ES cells by high concentration G418 selection, the *Neo* cassette was targeted into six different genomic loci on four different chromosomes of a hybrid ES cell line (R1) (Lefebvre, Dionne et al. 2001). The use of a hybrid cell line allows the origin of the two homologous chromosomes to be tracked by analyzing polymorphic DNA markers. In this study, it was shown that all of the homozygous gene-targeted clones recovered by high concentration G418 selection had lost heterozygosity, not only at the targeted locus, but also at the distant linked markers. Thus LOH selected by high concentration G418 selection involves either chromosomal

loss and subsequent duplication, or mitotic recombination proximal to the locus targeted with the selection marker.

Compared to the sequential targeting method, this high G418 concentration selection approach only requires one step of gene targeting, therefore, providing a convenient way to generate homozygous mutations in ES cells. The possible mechanism underlying LOH suggests that any randomly induced mutation that lies on the same chromosome as the pre-targeted *Neo* cassette can also become homozygous under high concentrations of G418 selection. Combined with other mutagenesis methods, high G418 concentration induced LOH can be used to generate homozygous mutation in a chromosome-specific way.

However this method also has its limitations. If it is used to make targeted mutations homozygous, it will still require designing targeting vectors and probes to generate and genotype the mutations. If this method is used to make random mutations homozygous (ENU or gene-trap mutations), it will be difficult to determine the genotype of the mutated locus by just checking the genotype of the *Neo* cassette targeted locus because the range of the LOH can be different from clone to clone.

1.6.3.3 Elevated mitotic recombination in *BLM*-deficient cells

Mitotic recombination can occur spontaneously, but its efficiency is too low to be used as an efficient tool for generating homozygous mutations without strong selection. Recently, it was shown that the mitotic recombination rate can be increased in mouse ES cells that lack the function of a DNA helicase, *Blm* (Luo, Santoro et al. 2000).

Six different *Blm* knockout alleles have been published so far (Chester, Kuo et al. 1998; Luo, Santoro et al. 2000; Goss, Risinger et al. 2002; McDaniel, Chester et al. 2003). Four of these were generated by gene targeting with replacement targeting vectors(Chester, Kuo et al. 1998; Luo, Santoro et al. 2000; Goss, Risinger et al. 2002; McDaniel, Chester et al. 2003). These four alleles deleted one or more coding exons of the *Blm* gene and all of them

have been described as embryonic lethal. The other two alleles, *Blm*^{tm2Brd} and *Blm*^{tm3Brd}, are the products of an insertional gene-targeting event, which results in the duplication of exon 3 (Luo, Santoro et al. 2000). This duplication caused a frame-shift mutation. Interestingly, the *Blm*^{tm2Brd} allele is homozygous lethal but the derived *Blm*^{tm3Brd} is viable. The homozygous mice (*Blm*^{tm2Brd}/*Blm*^{tm3Brd}) exhibited genomic instability and tumor susceptibility, a phenotype mimicking the human Bloom's syndrome.

Luo et al. (2000) measured the LOH rate in *Blm*-deficient ES cells by targeting *Hprt* minigene into an autosomal genomic locus. Cells that lose the *Hprt* minigene by LOH become resistant to 6-thioguanine. By Luria-Delbruck fluctuation analysis, the rate of LOH in *Blm*-deficient ES cells was determined to be 4.2×10^{-4} (events/locus/cell/generation), compared to 2.3×10^{-5} (events/locus/cell/generation) in wild type ES cells (Luo, Santoro et al. 2000; Liu, Jenkins et al. 2002).

The *Blm* gene product is not required for cell growth or survival in culture. *Blm*-deficiency in mouse ES cells caused a 20-fold increase in the rate of LOH, which provides the basis for generating homozygous autosomal mutations from single allele mutations. By calculation, a single ES cell with a heterozygous autosomal mutation will have segregated at least one daughter cell with a homozygous mutation by the time the colony derived from this cell contains 2,000 to 5,000 cells (Guo, Wang et al. 2004). So theoretically, when an ES cell library of heterozygous mutations is expanded for more than 13 generations, the library will contain a genome-wide set of homozygous mutations (Fig. 1-1d).

Recently, two groups have used *Blm*-deficient cells to screen for recessive mutations related to the DNA mismatch repair pathway (Guo, Wang et al. 2004) and glycosylphosphatidylinasitol-anchor biosynthesis pathway (Yusa, Horie et al. 2004). The success of the two recessive genetic screens in the *Blm*-deficient ES cells has shown the utility of this system for generating genome-wide homozygous mutations that facilitate recessive genetic screens *in vitro*.

This method does not require gene targeting of the first allele for generating the homozygous alleles. So it can easily be used in combination with other large-scale mutagenesis methods, such an insertional gene-trap mutagenesis or chemical mutagenesis. However, the *Blm* system can only generate a mixture of heterozygously and homozygously mutated ES cell clones. The representation of any one particular mutation in the pool will be extremely rare (approximately 10^{-7}-10^{-8}), so the mixture can only be used when there is a strategy available to selectively isolate the clone of interest from the rest of the ES cell population. For a genetic screen without selection, mutants need to be examined one-by-one to see whether or to what extent the desired phenotype is present (Grimm 2004). So pure homozygously mutated clones are needed, *Blm*-deficient ES cells are obviously not suitable for this purpose.

1.6.4 Induced mitotic recombination in mouse

Flp/*FRT* induced mitotic recombination has provided an efficient way to generate genetic mosaics in *Drosophila*. Induced mitotic recombination makes it possible to perform F1 mosaic screens, which save the trouble of performing three generations of crosses to establish individual lines to identify potential mutants. Screening for mutations in mosaic animals also circumvents the limitation of embryonic lethality of homozygous animals, especially for the genes that might have multiple functions at different developmental stages (Theodosiou and Xu 1998).

Recently, Liu et al. (2002) have successfully developed this system for the mouse. In this study, mitotic recombination was induced in mouse ES cells via Cre-mediated recombination between targeted *lox*P sites (Fig. 1-1e). The mitotic recombination frequency varied between different genomic loci and chromosomes, ranging from 10^{-5} to 10^{-2} after transient Cre expression. However, four of five loci tested showed a relatively low frequency, ranging from 4.2 X 10^{-5} (*Snrpn*) to 5.1 X 10^{-4} (*Wnt3*) for single allelic *lox*P sites after transient expression of Cre. Even for the clones in which induced mitotic recombination did occur in G2, not all of the events were followed by X segregation. For example, only 60% of recombination events at the *D11Mit71*

locus and 23% of the events at *Snrpn* locus was a G2-X event. One explanation for the low recombination frequency and the low proportion of G2-X events compared to *Drosophila* is that homologous mouse chromosomes are not paired in the interphase.

It was also noted that the fifth locus in the study, *D7Mit178* has exceptionally high rates of induced mitotic recombination (7.0×10^{-3}). Also, for this locus, all the recombination events seemed to occur at the G2 phase and followed by X segregation (Liu, Jenkins et al. 2002). The variation of the mitotic recombination frequency and the proportion of G2-X events from one chromosome to another might be caused by the different levels of the association between homologous mouse chromosomes in interphase. It is likely that for some chromosomes, some regions may be closely associated during the S-G2 phase of the cell cycle, and this greatly promotes the recombination efficiency and the chance of G2-X segregation (Liu, Jenkins et al. 2002).

In Liu's study, it was shown that multiple allelic *lox*P sites could increase the efficiency of induced mitotic recombination. For the *D7Mit178* locus, the increase was more than seven fold (from 7.0×10^{-3} to 5.0×10^{-2}), but the proportion of G2-X segregation among the recombination events dropped from 100% to 65% (Liu, Jenkins et al. 2002).

The frequency of inducible mitotic recombination on mouse chromosome 11 is similar to the spontaneous frequency of LOH on chromosome 11 reported in *Blm*-deficient mice (Luo, Santoro et al. 2000). Considering that the Cre/*lox*P-induced mitotic recombination is achieved in a small time window by transient Cre expression, it is likely that the recombination efficiency can be significantly increased if Cre is expressed constitutively.

A possible limitation to the application of inducible mitotic recombination in mice is genome imprinting. A number of regions on several different mouse chromosomes have been identified to have imprinting effects, ranging from early embryonic lethality to various developmental defects (Cattanach and

Jones 1994). If the daughter cells generated by induced mitotic recombination carry two maternally or two paternally imprinted chromosomes, these cells will either over-express the imprinted gene(s) or not express it at all. If these imprinted chromosomes are used for genetic mosaic experiments, mitotic recombination should only be induced later at a developmental stage so that imprinted gene(s) do not cause any visible phenotypes. On the other hand, inducible mitotic recombination also provides a good way to study genomic imprinting both *in vivo* and *in vitro*. ES cells with two paternally or two maternally imprinted chromosomes can be made by mitotic recombination and injected into wild-type blastocysts or tetraploid embryos to study the contribution of these cells into different cell lineages. These ES cells can also be differentiated *in vitro* to study the effect of imprinting on the development.

It is suspected that nonspecific Cre-mediated recombination between cryptic genomic *lox*P sites could induce DNA damage and cause background problem. However, so far no dada can support this hypothesis. Transient expression of Cre recombinase should be able to minimize the effect even it does exist. Induced mitotic recombination is useful not only for generating genetic mosaic *in vivo*, but also for making homozygous mutations *in vitro*. It is compatible with a wide range of mutagenesis methods, including gene targeting and gene-trapping. Unlike the homozygous mutant clones generated in *Blm*-deficient ES cells, induced mitotic recombination can be used to generate pure homozygous clones instead of a pool of heterozygous and homozygous clones.

1.7 Gene-trap Mutagenesis

1.7.1 A brief history

Ever since the mouse fanciers began to collect mice, there have been numerous records of spontaneous mutants. When genetics became a formalized science, mouse geneticists around the world were no longer satisfied with the simple collection and documentation of spontaneous mutants (Stanford, Cohn et al. 2001). Instead, methods have been developed to generate large number of mutants in an efficient way.

X-ray mutagenesis was the first high-efficiency method that was applied to generate mutants for mouse genetic studies (Stanford, Cohn et al. 2001). The X-ray mutation rate (13~50 X 10^{-5} per locus) is about 20-100 times higher than the spontaneous mutation rate (5 X 10^{-6} per locus) in the mouse. What's more, X-rays cause chromosomal rearrangements, which leaves a molecular marker for localizing the mutated gene. But the chromosomal rearrangements often affect several genes close to the break points. Also the dosage of X-ray is limited because the high dosage required for germline mutagenesis induce massive levels of cell death to the animal.

Chemical mutagenesis with ethylnitrosourea (ENU) generates mostly point mutations, and thus affects only single genes. It is also much more efficient than the X-ray mutagenesis, with a typical mutation rate of around 150 X 10^{-5} per locus. Besides, the drug is easy to administrate. But the limitation of this approach is that it leaves no markers in the mouse genome. A complicated mating strategy is therefore needed to map the mutations.

In 1976, exogenous retroviruses were shown to be able to transmit through the mouse germline (Jaenisch 1976). The subsequent observations that the retroviral integration can disrupt endogenous genes and alter their expression have led to the widespread use of insertional mutagenesis in mouse. The integration of a retrovirus can produce a loss-of-function mutation if it integrates into the coding region of a gene. Retroviral integration can also generate gain-of-function mutations because the viral LTR contains a strong enhancer element. Wild-type retroviruses are not very efficient mutagens because the vast majority of the insertions are in the non-coding parts of the genome. Viral integrations in these locations will neither activate nor inactivate a gene, and thus these are often phenotypically "neutral" to cells.

Gene-trapping technology has successfully circumvented the limitation of insertional mutagenesis using wild-type retroviruses. Gene-trapping vectors contain a non-functional selection cassette and/or a reporter cassette.
The selection and/or reporter cassettes are designed so that they are only activated if integration occurs in the vicinity of an endogenous gene capturing

the transcriptional elements. The requirement that selection markers in the virus are only activated effectively eliminates any random integration events in non-coding regions of the genome.

Though gene targeting by homologous recombination has made it possible to precisely knockout all the genes in the mouse genome, there are still some limitations of this technology. The main drawback is that, like all the reverse genetics approaches, it is hard to predict the biological pathway from the observed phenotypes. Also in some cases, single gene knockouts might not have any obvious phenotype because of genetic redundancy. The embryonic lethality caused by ablation of some developmentally important genes also prevents their function in the adults from being characterized (Stanford, Cohn et al. 2001). So random mutagenesis is still of great interest to mouse geneticists to address many important biological questions.

1.7.2 Gene-Trap vectors

Various gene-trap vectors have been designed for different purposes. They can be divided into three main groups: enhancer-trap, promoter-trap and PolyA-trap vectors .

1.7.2.1 Enhancer-trap vectors

Enhancer-trap vectors contain a minimal promoter that is not functional. The vectors are activated when they insert next to a *cis*-acting endogenous enhancer element, which activates expression of the selection and reporter cassettes (Fig. 1-2a). Enhancer-trap vectors have not been widely used in the mouse because loss-of-function mutations generated by this type of vector are very rare. The enhancer elements are often located far away from the coding region of a gene. Analysis of the integration sites from ES cell lines that show reporter expression *in vivo* has indicated that most insertions are not in the coding regions (Gossler, Joyner et al. 1989). Thus, enhancer-trap vectors rarely disrupt the normal expression of a gene.

1.7.2.2 Promoter-trap vectors

The essential part of a promoter-trap vector is a promoter-less reporter gene that has a strong splice acceptor (SA) site immediately upstream of it. The expression of the reporter can only be driven by an endogenous promoter and the enhancer elements of a trapped gene (Fig. 1-2b). A fusion transcript is thus generated from 5' end of the endogenous gene and the reporter gene. The fusion transcript can be used to clone the 5' end of the trapped gene using the Rapid Amplification of cDNA Ends (RACE) technique. Since the splice acceptor can effectively capture the transcription from the endogenous promoter, transcription is stopped at the transcription terminator sequences in the gene-trap vector. Thus, promoter-trap vectors efficiently generate loss-of-function mutations.

The main disadvantage of promoter-trap is that insertion events which activate the selection cassettes mostly occur in introns. Alternative splicing can sometimes bypass the trapping cassette. Even very low levels of wild-type transcripts can sometimes partially rescue the phenotype and result in hypomorphic alleles (Gossler, Joyner et al. 1989; Friedrich and Soriano 1991). However, hypomorphic alleles can be very useful to characterize a gene's function, especially if the null allele causes early embryonic lethality.

The design of promoter-trap vectors restrict their use to genes which are expressed in the cell type of interest, for example undifferentiated ES cells. ES cells transcribe an abnormally high number of genes, but there are still many genes that do not express in ES cells or express at too low levels for them to activate a gene-trap cassette to a level suitable for selection. So other methods, such as PolyA-traps are needed to cover the rest of the genome.

Nevertheless, promoter-trap vectors are the most widely used trapping vectors. The International Gene-trap Consortium (IGTC; http://www-igtc.ca), a joint program of several academic groups, has successfully used a variety of plasmid and retroviral trapping vectors to achieve 32% genome coverage with 27,000 tags (Skarnes, von Melchner et al. 2004).

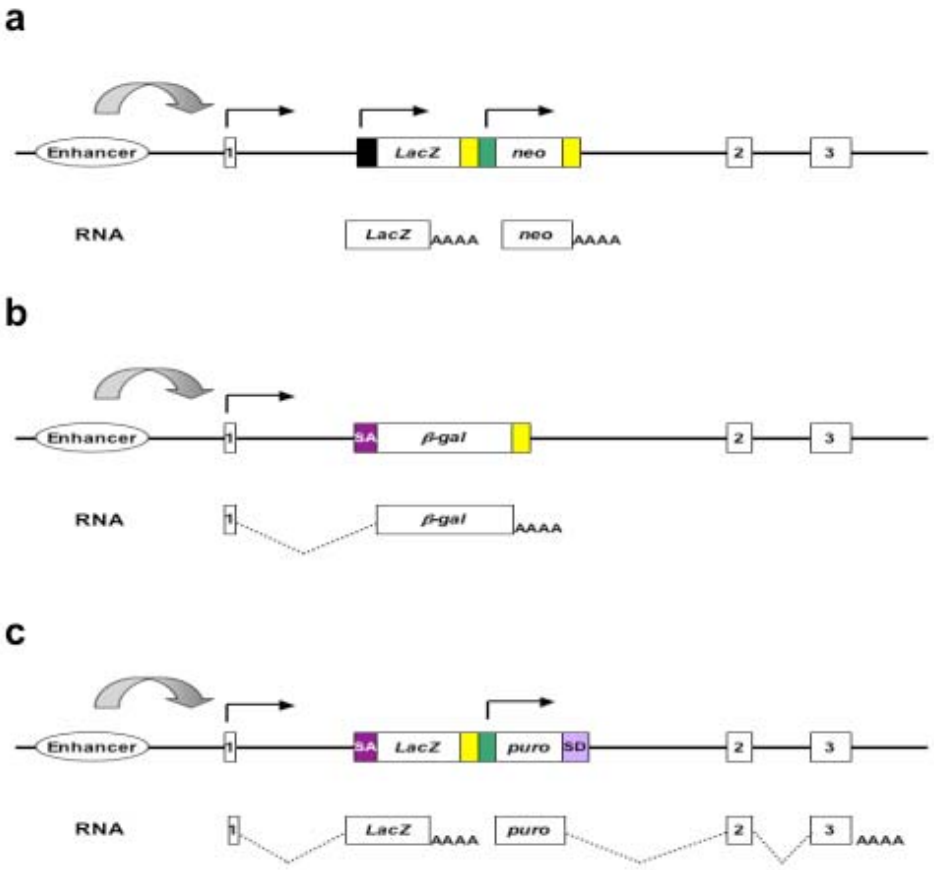


Fig. 1-2 Three main gene trap strategies. a. Enhancer trap. *LacZ* and *neo* reporter genes are driven by minimal promoters. The expression of reporter genes requires an endogenous enhancer. **b.** Promoter trap. The reporter gene doesn't have its own promoter. The expression of the reporter can only be driven by the promoter and enhancer elements of an endogenous gene. **c.** PolyA trap. The reporter gene doesn't have its own polyadenylation signal. *Puro* is transcribed from an autonomous promoter and spliced from the splice donor (SD) into an endogenous gene. Black and green boxes, different promoters; yellow box, poly A; SA, splicing acceptor; SD, Splicing donor.

1.7.2.3 PolyA-trap vectors

A polyA-trap vector contains a reporter gene lacking a polyadenylation signal, but possess a “strong” splice donor (SD) site. The reporter gene has its own promoter, but the reporter transcript is not stable unless the vector inserts into an endogenous gene upstream of a splice acceptor and a polyA signal (Fig. 1-2c). Usually, these vectors are designed so that termination codons of all the three reading frames follow the reporter gene, which prevents translation of the 3’ end of the trapped gene.

In contrast to a promoter-trap, a polyA-trap vector can mutate genes that do not normally express in undifferentiated ES cells, since the reporter is driven by an exogenous promoter, which is active in most genomic locations. The 3’ RACE technology can be used to clone the downstream exons, which is more reliable and robust than 5’ RACE.

One drawback of polyA-trap vectors is that they trap some pseudo splice acceptors and polyadenylation signals in the mouse genome. There are many fossilized gene fragments in the genome from old gene duplication events. Alternative splicing at the 3’ end of the gene is another potential problem. The mutagenicity rate of this kind of vectors is still controversial. Lexicon Genetics, a US-based biotechnology company, first used this method to create sequence-tagged mutations on a large scale (Zambrowicz, Friedrich et al. 1998). Lexicon has now achieved 60% genome coverage with 200,000 sequence tags (Zambrowicz, Abuin et al. 2003). It is interesting to note that about one-fifth of the genes trapped by IGTC are not represented in Lexicon’s tags (Skarnes, von Melchner et al. 2004). Since the number of sequence tags attained by the public domain is still relatively small, it is hard to predict how many genes trapped by Lexicon will never be represented in the promoter trapping approach pursued by IGTC. Nevertheless, the two efforts combined together have already trapped nearly two-thirds of all mouse genes (Skarnes, von Melchner et al. 2004) and provided a invaluable resource for mouse functional genomics and genetic studies.

1.7.3 Application of gene-trap mutagenesis in genetic screens

1.7.3.1 Expression screens

A genetic screen is an important method to identify a gene in any pathway or developmental event. Compared with phenotypic screens using ENU mutagenesis, gene-trapping in ES cells is a relatively inefficient method to generate germline mutations (Gossler, Joyner et al. 1989; Friedrich and Soriano 1991). But it has the advantage that the reporter, *βgeo* and human placental alkaline phosphatase (PLAP), can serve as tag for the expression of the trapped genes. The temporal and spatial expression pattern of a gene provides clues for its function. Some developmentally important genes often show highly restricted expression patterns during development.

Wurst et al. (1995) generated about 300 aggregation chimaeras using ES cell lines that contained gene-trap insertions. X-gal staining was then used to examine the expression patterns of the mutated genes in chimaeric embryos. About two-thirds of the chimaeric embryos expressed *lacZ*, which was temporally and spatially restricted for many lines. In a similar screen, Stoykova et al. (1998) has analysed 64 mouse lines generated from the gene-trap ES cell lines. About 75% of these lines showed embryonic *lacZ* expression (Stoykova, Chowdhury et al. 1998). Interestingly, for both screens, a large portion of the genes trapped in undifferentiated ES cells show *lacZ* expression in the developing nervous system.

It was noted in these studies that many gene-trap clones show widespread *lacZ* expression in *vivo*. Since many groups studying developmental questions are interested in genes that have highly restricted expression patterns in specific cell lineages, a pre-screen to enrich for genes with these characteristics would be valuable to save time and effort. For this reason, libraries of gene-trapped ES cell clones have been induced by specific growth/differentiation factors or physiological stimuli, such as nerve growth factor, retinoic acid, engrailed homeobox proteins and γ-irradiation. Those gene-trap integrations that are found to be either activated or repressed by

one of these factors are more likely to be linked to a specific signalling pathway. *LacZ* staining during embryogenesis has shown a strong enrichment of gene-trap clones that have restricted patterns *in vivo* after the induction screen procedure (Bonaldo, Chowdhury et al. 1998).

Gene-trap vectors have also been designed to trap specific classes of genes. Skarnes et al. (1995) designed a vector to trap genes that encode secreted and transmembrane proteins. In this secretory-trap vector, a transmembrane (TM) domain was placed between the splice acceptor (SA) and the *βgeo* reporter. The transmembrane domain will result in the sequestration of the *βgeo* fusion protein into the lumen of the endoplasmic reticulum (ER) of genes that encode non-secretary proteins, and thus abolish the *β-gal* activity. The *βgeo* reporter is expressed when an additional secretory signal (SS) from the trapped gene results in the *βgeo* portion of the fusion protein being positioned in the cytosol (Skarnes, Moss et al. 1995). Recently, this secretory-trap design was modified to identify and mutate receptors and ligands controlling neuronal axon guidance. Leighton et al. (2001) used a human placental alkaline phosphastase (PLAP) reporter, which is co-expressed with the *LacZ* gene-trap reporter, to label neuronal projections. By *β-gal* and PLAP staining, genes with restricted expression patterns in neuronal axons were identified. Recently, Chen et al. (2004) have used an inducible gene trapping system to screen gene trap events responding to retinoic acid (RA). 65 gene traps were identified using this method. *In vivo* analysis revealed that 85% of the retinoic acid-inducible gene traps trapped developmentally regulated genes.

1.7.3.2 Phenotype-driven screens

Phenotype-driven screens in diploid genomes require a strategy to obtain homozygous mutations. So it is difficult to perform a phenotype-driven screen in ES cells. Recently, Guo et al. (2004) has utilized the elevated mitotic recombination rate of the *Blm*-deficient ES cells to generate a genome-wide homozygous mutation library of gene-traps. This library was then used to screen for genes involved in the DNA mismatch repair pathway (Guo, Wang

et al. 2004). This strategy has provided a new way to carry out recessive genetic screens in ES cells.

1.7.3.3 Genotype-driven screens

Sequence-based screens have been made possible by the availability of large resource of ES cell clones with defined mutation. Several academic groups in the International Gene-trap Consortium (IGTC; http://www.igtc.ca) have initiated a genome-wide gene-trap program aimed at generating an international resource of embryonic stem cells with gene-trap insertions in most mouse genes. Gene-trap cell lines generated by the IGTC are freely available to the public and all the sequence tags are finely mapped on Ensemble mouse genome browser (http://www.ensembl.org/Mus_musculus) (Skarnes, von Melchner et al. 2004). ES cell lines which carry gene-trap insertions are now available for nearly 40% of mouse genes. A parallel effort is also carried out by Lexicon Genetics, but the cell lines and the mouse line derived from them are available with a fee and limitations of future work.

1.7.4 Electroporation versus retroviral infection

Trapping vectors can be introduced into the genome by either electroporation or retroviral infection. Both methods have their advantages and disadvantages. Since gene-trap screens using both methods have been reported, it is possible now to assess the two methods.

1.7.4.1 Electroporation

The simplest way to perform gene-trap mutagenesis is to electroporate the linearized gene-trap vector directly into mammalian cells. The gene-trap vectors introduced by electroporation integrate into the genome randomly, while the retroviral vectors tend to insert into the 5' portion of the gene. One advantage of the electroporation method is that it does not require the construction of a virus, which has numerous constraints discussed later. Scaling up electroporation is relatively easy because large amounts of DNA can be easily prepared. Last but not least, theoretically there is no limitation on the size of the trapping vector. Multiple reporter cassettes can be incorporated into one vector, which can be tailored for specific usages.

The biggest disadvantage of the electroporation method is that integrations are always accompanied by DNA concatermerization, though conditions can be optimized such that concatermers occur in less than 20% of the cells (Stanford, Cohn et al. 2001). Tandem insertions into the same locus happen through a recobinational process to form a concatomer followed by non-homologous end joining DNA repair (NHEJ) as the vector inserts into the genome. Multiple copies of the gene-trap vector in one locus can result in ectopic reporter expression, aberrant splicing, and can complicate the identification of the gene-trap mutations by 5' RACE. Sometimes, the gene-trap vector can be truncated during electroporation due to exonuclease digestion. The loss of different amounts of flanking sequence makes the cloning of the flanking genomic sequence by Inverse PCR problematic.

1.7.4.2 Retroviral infection

1.7.4.2.1 Retroviral life cycle

The typical retrovirus genome consists of two copies of a single-stranded RNA molecule of about 8-12 kb. The wild type murine leukaemia virus genome encodes three major proteins, Gag, Pol and Env. Gag is processed to make the core proteins. Pol has the reverse transcriptase, RNase H and integrase activities. Env encodes the viral envelope protein. A mature viral particle consists of the virus nucleoprotein core and the outer lipid-protein shell of the viral envelope (Fig. 1-3a).

Viral particles infect host cells by binding to cell surface receptors, which is mediated by the envelope proteins of the retrovirus. Infection is followed by injection of the virus nucleoprotein core into the host cell. After this, a double-stranded DNA is generated from the viral genomic RNA by the viral reverse transcriptase. Finally, the newly transcribed double strand viral DNA integrates into the host chromosome, which is catalyzed by the viral integrase. The integrated viral DNA is known as proviral DNA.

Once integrated, the virus is ready to initiate a new round of replication and infection. Full-length genomic RNA is transcribed from proviral DNA by the host cell RNA polymerase. The genomic RNA is either processed to generate mRNAs, which are translated into the viral proteins by the host protein synthesis machinery, or the full length genomic RNA remains unspliced and is packaged into new viral particles and released from the host cell by budding from the plasma membrane .

1.7.4.2.2 Recombinant retroviral, viral packaging cell lines

A retroviral vector can be used to transfer exogenous DNA into eukaryotic cells. Because the retrovirus can efficiently integrate into the host genome, exogenous genes carried by the retrovirus can be expressed. However, the wild type retrovirus is not ideal for this purpose because of the size limit of the genomic RNA that can be efficiently packaged into the virus particle.

Many recombinant retroviral vectors have been constructed. A typical recombinant retroviral vector includes the 5' long terminal repeat (5' LTR), the 3' long terminal repeat (3' LTR) and viral RNA packaging signal, known as Ψ. All the other essential components of the wild type retrovirus are deleted to make space for the exogenous DNA (Fig. 1-3b). The recombinant retrovirus vector itself is replication-deficient. To produce infectious retrovirus, the vector needs to be transfected and transcribed in a viral packaging cell line, which can express all three proteins that are required for viral reproduction, Gal/Pol and Env.

When a recombinant retrovirus construct is transiently transfected into the viral packaging cell line, the transcribed genomic RNA is recognised and assembled as an infectious particle with the viral proteins. The derived replication-deficient retrovirus particle can infect any cells that have the receptor for the virus and the vector can then integrate into the host genome. However, there is only one infection cycle because the recombinant virus lacks the Gal/Pol and Env products (Fig. 1-3c).

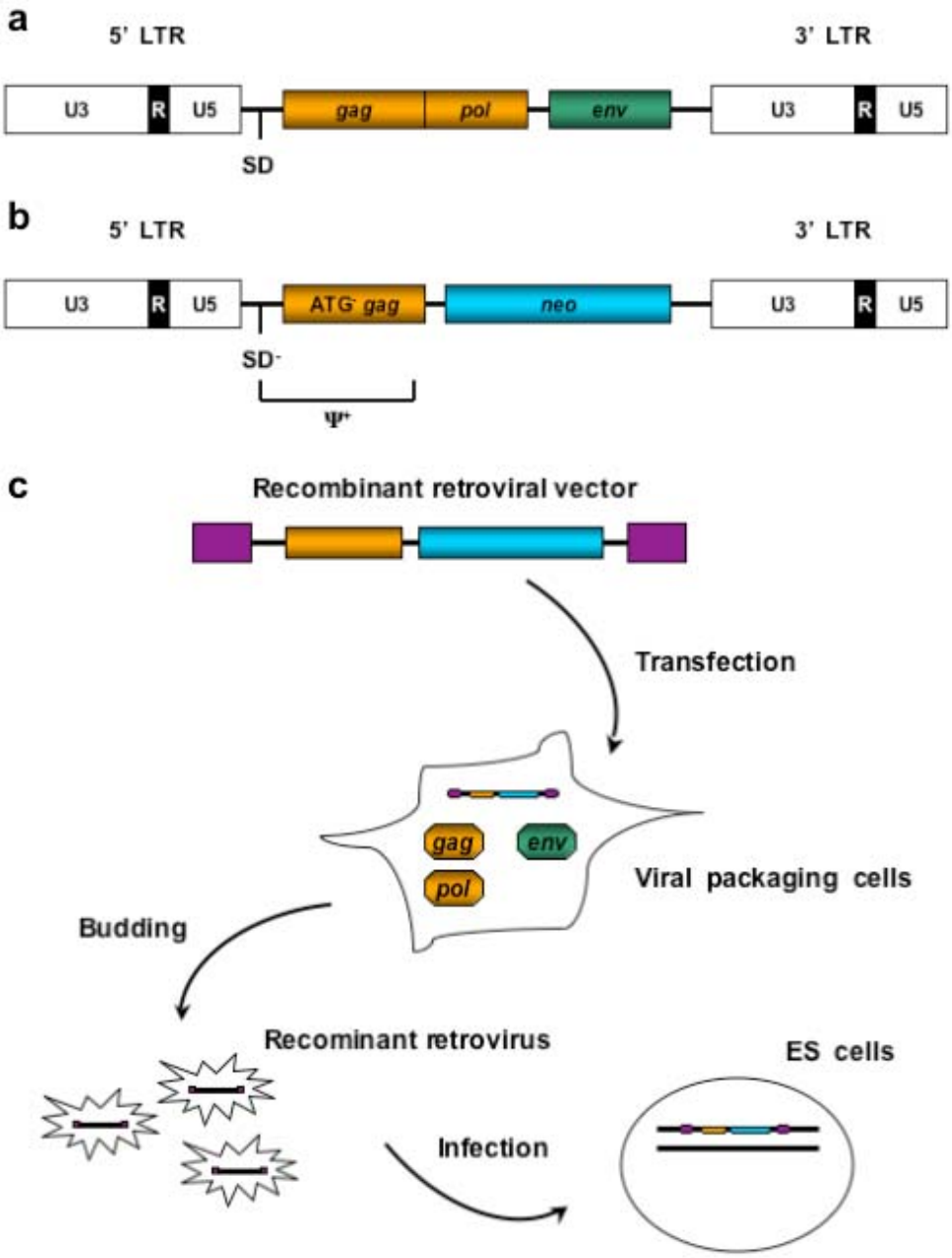

a
5' LTR
3' LTR
U3
R
U5
gag
pol
env
U3
R
U5
SD
b
5' LTR
3' LTR
U3
R
U5
ATG- gag
neo
U3
R
U5
SD-
Ψ+
c
Recombinant retroviral vector
Transfection
gag
env
pol
Viral packaging cells
Budding
Recominant retrovirus
ES cells
Infection

Fig. 1-3 Retroviral vectors and viral production. a. The structure of a wild type retrovirus. The virus genome contains genes that encode viral protein core (*gag*), reverse transcriptase (*pol*), and viral envelope protein (*env*). SD, viral splicing donor. **b.** The structure of a recombinant retroviral vector, pBabe. The viral genes *pol* and *env* are deleted. A *neo* selection marker is inserted to allow the selection of retroviral integration. The splicing donor and the *gag* gene are kept to facilitate the viral packaging. The splicing donor is mutated (SD^-) and the initiation codon of gag gene is deleted (ATG^-*gag*) to avoid the interference of internal gene expression. Ψ, viral packaging signal. **c.** Production of the recombinant retrovirus. The recombinant retrovirus is replication-deficient. To produce infectious retrovirus, the vector need to be transfected into a viral packaging cell line, which can express all the three proteins that are required for viral reproduction, Gal/Pol and Env. In the packaging cell line, the viral genomic RNA is packaged and infectious viral particles are released. The recombinant retrovirus produced can be used to infect ES cells.

1.7.4.2.3 Retroviral based gene-traps

Von Melchner et al. (1989) developed the first retroviral gene-trap vector. In this design, the gene-trap cassette is inserted in the U3 region of the 3' LTR and replaces the viral enhancer. After viral replication and integration, the provirus carries a duplicated gene-trap cassette in both of the 5' and 3' LTRs. The cassette in the 5' LTR is situated just 30 nt from the host genome and is activated by transcriptional read-through rather than splicing (von Melchner and Ruley 1989). Friedrich et al. (1991) designed another version of a retroviral gene-trap vector, called ROSA (reverse orientation splice acceptor). In the ROSA vector, the gene-trap cassette was placed between viral LTRs in the opposite orientation relative to viral transcription. This reverse orientation was essential to avoid removal of the viral packaging sequence Ψ from the full length genomic RNA by splicing from the upstream viral splice donor sequence directly to the splice acceptor in the gene-trap cassette (Friedrich and Soriano 1991). In this design, the cassette is activated only by a splicing event.

1.7.4.2.4 Advantages and disadvantages of Gene-trap via retroviral infection

Gene-trap mutagenesis using a retroviral vector has advantages compared to electroporation. In contrast to electroporation, only a single copy of retrovirus will integrate into one genomic locus. The provirus has a predictable structure which is the same in every clone, which makes the cloning of virus insertion site by PCR based methods very reliable. The trapped exons can be determined by RACE. The flanking genomic fragment of the insertion site can also be cloned by inverse PCR (Suzuki, Shen et al. 2002) or splinkerette PCR (Mikkers, Allen et al. 2002). Once the virus trapping titre is determined, it is easy to control the virus infection conditions so that most of the cells will only contain a single copy of the gene-trap vector. Retroviruses have a propensity to integrate into 5' portion of a gene. Virus insertion in the 5' untranslated region and first few introns is more likely to generate null alleles.

Retroviral vectors also have their limitations. First, the packaging size of the retroviruses is highly limiting. Even within the limit, the virus packaging efficiency drops significantly as the size of the virus increases. Second, the virus insertion can induce retroviral-mediated gene silencing. Deleting the LTR enhancer sequences can solve this problem. Third, viral splice donor sequence in the 5' LTR can cause ectopic reporter expression. This problem can be solved by putting the reporter gene in the reverse orientation. Fourth, non-random retroviral integration results in trapping "hot spots", but the same problem also exists for the plasmid-based gene-traps (Hansen, Floss et al. 2003).

1.7.4.3 Gene-trap "hot spots"

Although it was noticed a long time ago that there are preferred integration sites, or "hot spots" for gene-trap mutagenesis, the data available was not sufficient to systematically assess the problem. Recently, the German Gene-trap Consortium (GGTC) reported the generation of over 11,000 independent gene-trapped ES clones using four different gene-trap vectors, including two electroporation-based vectors and two retrovirus-based vectors. 5,142 sequence tags were obtained from gene-trap insertions, which made it possible to do a systematic analysis of gene-trap "hot spots" by both methods (Hansen, Floss et al. 2003).

It was found that there was a direct correlation between the number of gene-trap insertions on a given chromosome and the number of the genes on that chromosome, which suggests that there is no obvious bias to a single chromosome. Of all the recovered UniGene clusters, 75% of appeared only once, while the remaining 25% were hit multiple times. This data suggested that most mouse genes are randomly accessible to gene-trap mutagenesis (Hansen, Floss et al. 2003).

45% of hot spots were hit by more than one of the four vectors, suggesting that these hot spots might be caused by locus-specific factors, for example secondary chromatin structure (Hansen, Floss et al. 2003). Considering that

over half of the hot spots are vector-specific, it seems that each gene-trap vector design also has its own pool of trappable genes. Achieving saturation mutagenesis with gene-trap vectors will require the use of a combination of different gene-trap vector designs.

1.8 ES cell *in vitro* differentiation

1.8.1 Introduction

In mammals, the fertilized oocyte and the blastomeres of 2-, 4- and 8-cell stage embryos are totipotent. They can generate a complex organism of hundreds of different specialized cell types (Wobus 2001). On the other hand, the embryonic stem cells, which are derived from inner cell mass (ICM) of blastocysts, are only pluripotent. When transferred back to blastocysts, they can contribute to all the different cell types, except the placental tissues, thus they are not able to generate a complex organism by themselves.

The ability of ES cells to give rise to various cell types including germline cells has laid the basis of gene targeting technology (Bradley, Evans et al. 1984). However, ES cells not only differentiate *in vivo*, they can also form three-dimensional embryo-like aggregates which contain cells of the endodermal, ectodermal and mesodermal lineages (Wobus 2001). These three germ layers can further differentiate into a variety of specialized cell types including cardiac muscle, smooth muscle, skeletal muscle, haematopoietic, pancreatic, cartilage and neuronal cells (Czyz, Wiese et al. 2003). ES cell *in vitro* differentiation can therefore recapitulates the early mouse embryogenesis to a degree.

1.8.2 *In vitro* differentiation potential of ES cells

When ES cells are cultured on feeder layers and/or in medium supplemented with differentiation inhibitory factors such as LIF, they can remain undifferentiated indefinitely. But, once the ES cells are deprived of differentiation inhibitory factors, they will commit to a differentiation fate.

Many different protocols have been established for the *in vitro* differentiation of ES cells into different terminally differentiated cell types by the “hanging

drop" method (Wobus, Wallukat et al. 1991) (Fig. 1-4), by the "mass culture" method (Doetschman, Eistetter et al. 1985), by cultivation in methylcellulose (Wiles and Keller 1991), by stromal cell co-culture method (Nakano, Kodama et al. 1994; Kawasaki, Mizuseki et al. 2000) or by adherent monoculture method (Nishikawa, Nishikawa et al. 1998; Ying, Stavridis et al. 2003). Treatment of differentiation cultures with soluble growth factors can also help to drive differentiation into specific directions.

Compared to the "hanging drop" and "mass culture" methods, which require the generation of three-dimensional embryoid bodies (EB), the stromal cell co-culture and adherent monoculture methods are much simpler and highly efficient in inducing ES cells to differentiate into neuronal or haematopoietic lineages. The success of these methods proves that the three-dimensional structure in the embryoid body is not requisite for blood cell, endoderm and neuron cell differentiation. By adding exogenous growth factors (adherent monoculture method) or the secretion of endogenous factors (stromal cell co-culture method), ES cells can be coaxed into a specific lineage without the spatial information of the embryo (or embryoid bodies).

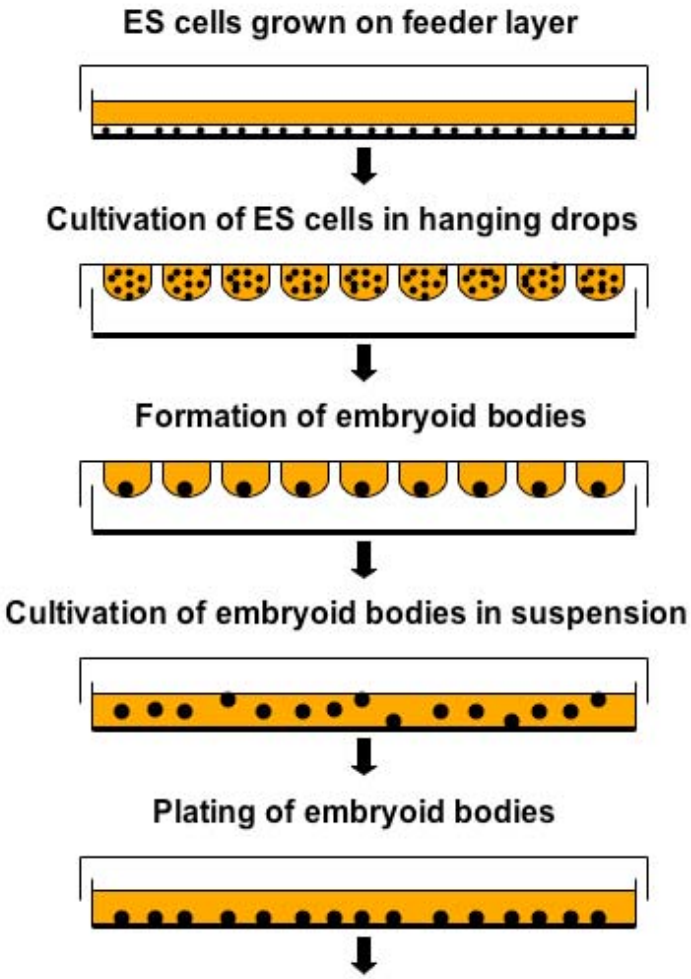
As the spatial information of the embryo is absent in the culture of ES cells, the lineage and stage of differentiation may conveniently be determined by the cell surface markers (Nishikawa, Nishikawa et al. 1998; Ying, Stavridis et al. 2003) or the GFP reporter genes tagged to an intracellular markers (Nishikawa, Nishikawa et al. 1998; Ying, Stavridis et al. 2003). Fluorescence Activated Cell Sorting (FACS) can be used to quantitatively compare the differentiation efficiency of ES cell lines in different culture conditions or with different genetic modifications.

However, these methods are used to differentiate ES cells into some specific lineages. The culture conditions, the addition of differentiation inducers and/or the stromal cell lines are different from one protocol to another. If more than one cell lineage is studied, ES cell lines need to be differentiated in many different ways which increases the complexity of the experimental design. On

the other hand, “hanging drop” method provides a simple and universal way to differentiate ES cells into multiple lineages.

Embryoid bodies (EB) can undergo specific and reproducible morphological changes. First, an outer layer of endoderm-like cells are developed over the primitive ectoderm layer. This is followed by the formation of an ectodermal rim and by the formation of the mesodermal cells. When the EBs are plated and extrinsic differentiation factors are added, differentiated specialized cells can develop in the outgrowth area of the EBs (Wobus 2001).

The morphological changes are accompanied by a dynamic change in the expression pattern of a set of lineage- and tissue-specific genes. During the first several days of EB development, the primitive ectoderm specific genes such as *Oct4* and *Fgf5* express at a high level. This is followed by up-regulation of genes characteristic for early postimplantation stages, such as *Nodal* and early endodermal genes *vHnf1*, *Hnf3*β and *Hnf4*. At the same time, genes that are characteristic of gastulation and the early mesodermal differentiation show maximum expression, such as *Brachyury*, *Goosecoid* and *Bmp4*. Tissue-specific genes that show developmental regulated expression patterns will begin to express after that. At the terminal differentiation stage, genes that are expressed only in specialized cell types are detected (Rohwedel, Guan et al. 2001).



Differentiation of endodermal, ectodermal and mesodermal cells.

RT-PCR, in situ hybridization and immunohistochemistry was performed to detect the developmentally controlled expression of mRNA and proteins.

Fig. 1-4 *In vitro* differentiation of ES cells using the "hanging drop" method. Certain number of undifferentiated ES cells are cultivated in hanging drops for 2 days to form the cell aggregates, embryoid bodies (EB). The EBs are cultivated in suspension for another 3 days before they are plated onto gelatin-coated tissue culture plates. The cellular phenotypes of differentiated cells derived from endodermal, ecotodermal and mesodermal lineages can be characterized by a variety of methods.

Leahy et al. (1999) has performed *in situ* hybridization on EBs using probes for germ layer markers (*Oct4*, *Fgf5*, *Gata4*, *Nodal* and *Brachyury*) as well as cell lineage-specific markers (*Flk-1*, *Nkx2.5*, *Eklf* and *Msx3*). Since the expression patterns of these markers has already been extensively characterized *in vivo*, the marker expression during EB formation and early embryogenesis can be correlated. By this method, different stages of EB *in vitro* differentiation can be linked to different stages of embryogenesis *in vivo* (Leahy, Xiong et al. 1999). Markers that show highly reproducible temporal and spatial distribution can thus be used in genetic screens for mutations that disrupt the normal expression pattern of these markers.

This fundamental work has made it possible to use ES cell *in vitro* differentiation as an alternative to study the functions of the genes that are important in early embryogenesis. The mutations that are caused by over-expression (gain-of-function mutations) or homozygous targeting (loss-of-function mutations) in ES cells can be analyzed in this system. The loss-of-function approach is very important for functional studies, especially for those genes that cause early embryonic lethality when both alleles are disrupted (Wobus 2001).

In vitro differentiation of ES cells can not only be used as a model system to study early embryogenesis, it also provides a promising way to generate terminally differentiated cell types for therapy. Some cell types, such as cardiomyocytes, neuronal and glial cells, and pancreatic cells, are of potential therapeutical relevance because they can be used for the treatment of cardiac diseases, neurodegenerative disorders and diabetes, respectively.

However, cystic EBs are heterogeneous, they consist of various differentiated cell types as well as undifferentiated ES cells. Numerous experiments have demonstrated that the *in vitro* differentiation of ES cells can be directed into certain lineages by controlling various parameters such as the starting ES cell numbers to form the EB aggregates, the composition of the differentiation media, the type, concentration or combinations of the growth factors,

differentiation induction factors and genetic “gain-of-function” and “loss-of-function” manipulations (Czyz, Wiese et al. 2003).

Though much progress has been made in determining the best conditions for ES cells to differentiate into certain lineages, this technology is still far from mature enough to be used for transplantation therapy. Many important developments are required, for example improving the efficiency of the differentiation protocols, the purity of cell population of the desired cell types, limiting the potential tumorigenicity of the undifferentiated ES cells, controlling the donor/recipient immune-compatibility and achieving the long-term functional engraftment of differentiated cells *in vivo* (Czyz, Wiese et al. 2003). Considering the ethical concerns surrounding the human ES cells isolated from *in vitro* fertilized human embryos and the much more advanced stages of research in the mouse system, these questions need to be answered in mouse ES cells first and then confirmed in their human counterpart.

1.8.3 Using genetic screens to study the ES cell *in vitro* differentiation

Relatively little is known about the genetic pathways that control ES cell *in vitro* differentiation and cell lineage determination. Though efficient differentiation protocols have been established for some cell types, we still do not know exactly why and how the change of concentration of certain factors can dramatically increase the percentage of certain cell types in the whole EB cell population. The lack of knowledge of the genetic pathways underlying the *in vitro* differentiation process significantly impedes further improvements in these protocols.

In vitro differentiation experiments have been performed with many mutant ES cell lines, and the phenotypes of these have been described. But in most cases, the technique was used to study the function of the genes that caused early embryonic lethality when both alleles were disrupted. Different differentiation methods (hanging drops or mass culture), different ES cell lines (R1, E14.1, and AB1) and different assay methods (RT-PCR, whole mount *in situ*, immunohistology and physiological analysis) all make the data from

different experiments not directly comparable. So a systematic approach is needed to thoroughly study the ES cell *in vitro* differentiation process.

A genetic screen is always the most powerful way to study a complex system. Screens have been successfully applied to many model organisms and mammalian cultured cell lines. Many important genetic pathways have been identified and well characterized using this approach. But very few genetic screens have been carried out on ES self-renewal and *in vitro* differentiation (Chambers, Colby et al. 2003). There has no loss-of-function screen been reported so far. One of the main reasons for this is the difficulty of producing numerous mutations in ES cells, especially homozygous mutations that are the key for recessive genetic screens.

There are many different ways to create homozygous mutations in ES cell, which have been described in chapter 1.6.3. However, selection can hardly be performed in *in vitro* differentiation studies, so single mutant ES clones need to be differentiated separately and checked one by one. *Blm*-deficient ES cells can not be used for this purpose because the resulting pool of cells will be a mixture of ES cells with heterozygous mutations with a few rare homozygous mutants. Targeting both alleles using marker recycling or different selection markers is both time-consuming and labour-intensive, limiting throughput. The high G418 concentration induction method can save the trouble of targeting the second allele, but it still requires the targeting of the first one and a genotyping strategy to distinguish the homozygous clones from the heterozygous ones. All these limitations make them unsuitable for conducting a recessive genetic screen to identify genes required for ES cell *in vitro* differentiation.

Induced mitotic recombination can be used to make homozygous mutations in a chromosome-specific way. Since all mutations distal to the mitotic recombination selection cassettes will become homozygous after Cre-induced mitotic recombination, the genotype of the mutations anywhere on the chromosome can be simply determined by the drug resistance as well as by genotyping results of the locus where the induced mitotic recombination

cassettes are targeted. But the problem remains, how does one introduce mutations on one specific chromosome which has been designed to undergo induced mitotic recombination. Gene targeting by homologous recombination is one choice, but the throughput will be limited. Insertional and ENU mutagenesis can create a lot of mutations in a random way, but most of these mutations will not be on the right chromosome, so they will not become homozygous after Cre-induced mitotic recombination. To identify clones with homozygous mutations from such a high background without selection is difficult for insertional mutagenesis and almost impossible for chemical mutagenesis. So clearly, a pre-screening strategy is needed to accumulate mutations on the desired chromosome before chromosome-specific mitotic recombination is induced by Cre expression.

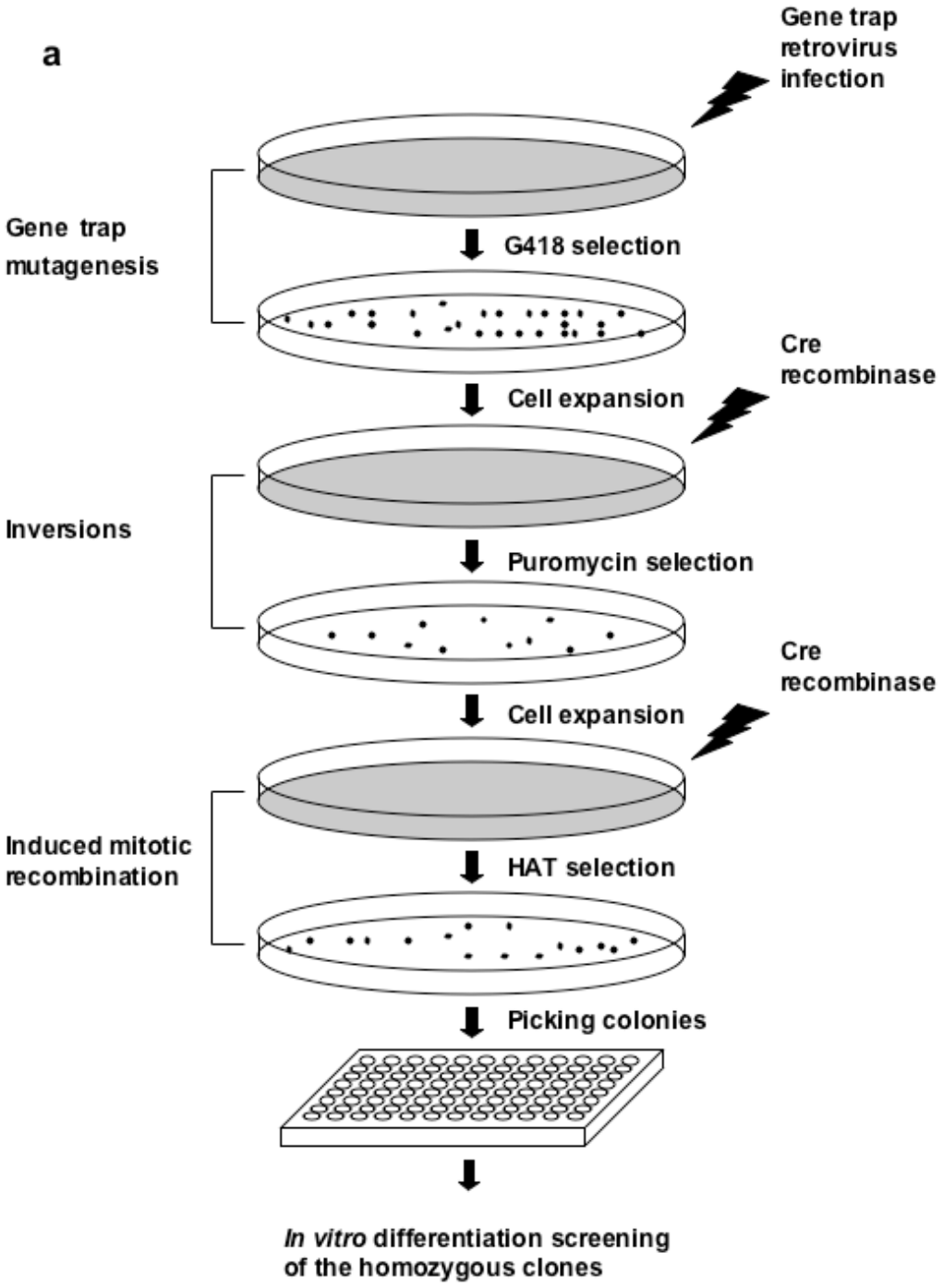
1.9 Thesis project

The primary goal of the project was to generate a large number of homozygous mutations in a genomic region of interest on chromosome 11 in mouse embryonic stem cells by exploring induced mitotic recombination and regional trapping mutagenesis methods and to investigate the application of a recessive genetic screen for genes involved in ES cell *in vitro* differentiation.

Mouse chromosome 11 was chosen for this project because of some unique characteristics. First, the distal half of the mouse chromosome 11 exhibits highly conserved linkage with human chromosome 17. Almost every gene mapped to human chromosome 17 is found on mouse chromosome 11. Second, numerous genetic tools have already been created on this chromosome (Liu, Zhang et al. 1998; Zheng, Sage et al. 1999; Su, Wang et al. 2000; Zheng, Sage et al. 2000; Liu, Jenkins et al. 2003). To date, 18 deletions and 3 inversions/balancers have been made which cover the chromosome (Yu and Bradley 2001). These resources make the downstream functional characterization of the mutations recovered from a genetic screen much easier. Third, the distal part of mouse chromosome 11 has a very high gene density which makes it an ideal target for mutagenesis studies. Forth, this chromosome is not imprinted, eliminating potential complexity associated with mono-allelic expression from maternal or paternal chromosomes.

A disadvantage of using mouse chromosome 11 for the ES cell differentiation study is that this chromosome is often found to be amplified in mouse ES cells (Nichols, Evans et al. 1990; Liu, Wu et al. 1997). ES cell clones that are trisomic for all or part of chromosome 11 exhibit accelerated cell growth and decreased efficiency of germ line transmission. Extensive engineering of this chromosome might increase the possibility of accumulating trisomic clones. These trisomic clones are likely to have abnormal differentiation potential and thus complicate the interpretation of the phenotypes. However, these clones are relatively rare in the whole culture population. If care is taken in genotyping the cell lines used to generate homozygous gene-trap mutations, it is possible to distinguish these trisomic clones by their abnormal growth rate and by Southern analysis (the ratio between the signals of targeted and un-targeted restriction fragments).

In the previous sections, the principles of induced mitotic recombination and regional trapping mutagenesis have already been discussed. The design of my project was: (i) Generate a cell line engineered to undergo induced mitotic recombination on chromosome 11 and capture a subset of gene-trap mutations that were generated on this chromosome. (ii) Generate a set of genome-wide gene-traps in this cell line and isolate those that are located on chromosome 11 using a Cre/*lox*P mediated inversion strategy. (iii) Make the inversions homozygous by Cre-induced mitotic recombination. (iv) Assess the homozygous clones for their developmental potential by an *in vitro* differentiation assay. (v) Confirm the mutation by BAC rescue or by re-generating the mutation followed by re-confirmation of the phenotype (Fig. 1-5a and b).


a
Gene trap retrovirus infection
Gene trap mutagenesis
G418 selection
Cre recombinase
Cell expansion
Inversions
Puromycin selection
Cre recombinase
Cell expansion
Induced mitotic recombination
HAT selection
Picking colonies
In vitro differentiation screening of the homozygous clones

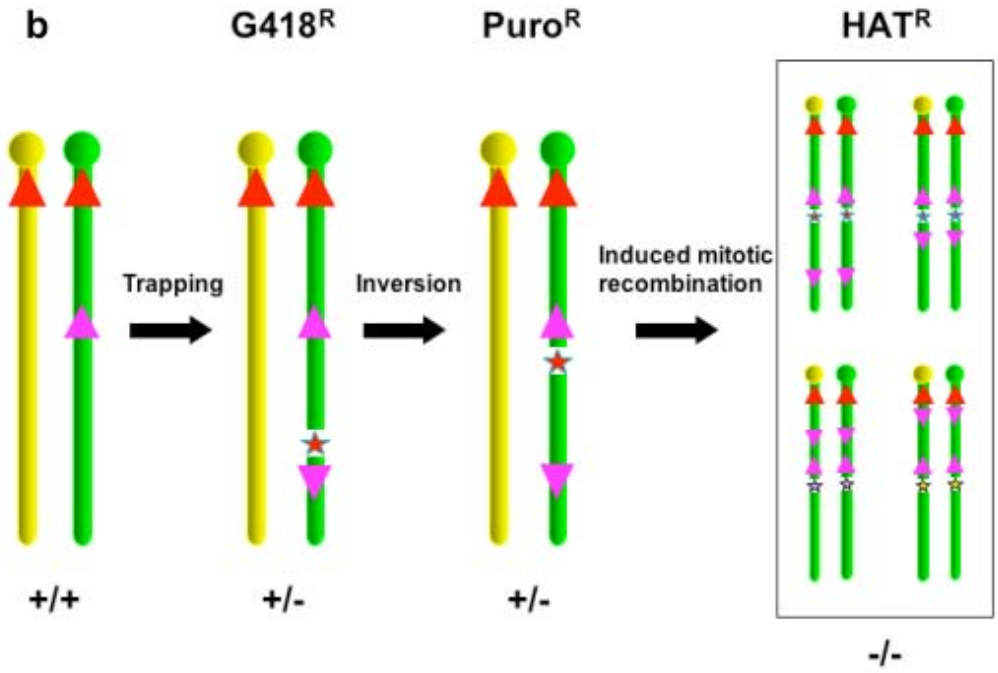


Fig. 1-5 Thesis Project. a. Schematic illustration of the screening strategy for homozygously mutated ES cells. The whole project is composed of three independent steps, gene trap mutagenesis, regional accumulation of the gene trap mutations by inversion and the induced mitotic recombination. Since each step is selected by a different drug selection marker, no genotyping is necessary for the intermediate products. **b.** Schematic illustration of the chromosomal structure of the products of each step. A cell line was engineered to undergo mitotic recombination and to capture the subset of gene trap mutations that were generated on this chromosome. A large panel of genome-wide gene traps were generated in this cell line and those that are located on chromosome 11 were specifically selected via Cre/*lox*P mediated inversions. The inversions were then made homozygous by induced mitotic recombination. Red arrows, mutant *lox* sites; pink arrows, wild type *lox*P sites; Yellow and green bars: homologous chromosomes; stars of different colours, gene-trap insertions.

2 Material and methods

2.1 Vectors

2.1.1 Vectors for induced mitotic recombination

pL330 (*Hprt* MΔ3') and pL341 (*Hprt* MΔ3') are kind gifts from Dr. Pentao Liu (Liu, Jenkins et al. 2002). These vectors contain *Neo*^{R} and *Puro*^{R} selection markers flanked by three lox site variants: *lox*5171, *lox*2272 and either *lox*66 (*Hprt* MΔ3') or *lox*71 (*Hprt* MΔ3') sites. A wild-type *lox*P site is generated by site-specific recombination between *lox*66 and *lox*71 sites. Because the wild-type *lox*P site will be used for regional trapping, *lox*66 and *lox*71 sites need to be deleted from these mitotic recombination vectors. Several vectors were constructed for this purpose.

pWW15 (*Pol II-Neo-bpA* cassette):
pL341 was cut with *Hin*dIII and *Not*I, a 2.2 kb fragment was gel purified and digested again with *Spe*I, a 1.8 kb fragment was gel purified and cloned into pBluescript (pBS) plasmid (Stratagene) digested with *Hin*dIII and *Spe*I to make pWW15.

pWW22 (*PGK-Puro-bpA* cassette):
1) pL330 was digested with *Eco*RI and *Hin*dIII sequentially, a 0.5 kb fragment was gel purified; 2) pL330 was digested with *Eco*RI and *Hin*dIII sequentially, a 1.2 kb fragment was gel purified and digested again with *Spe*I, 1.0 kb fragment was gel purified; 3) pBS was digested with *Eco*RI and *Spe*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW22.

pWW23 (*PGK*-5' *Hprt* cassette):
1) pL341 was digested with *Hin*dIII and *Bgl*II sequentially, a 0.7 kb fragment was gel purified; 2) pL341 was digested with *Hin*dIII and *Bgl*II sequentially, a 1.0 kb fragment was gel purified and digested again with *Spe*I, a 0.8 kb fragment was gel purified; 3) pBS was digested with *Hin*dIII and *Spe*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW23.

pWW24:
A variant polylinker site *Hin*dIII-*Spe*I-*Xba*I-*Nde*I-*Pst*I-*Eco*RI-*Not*I (pWW24) was generated by cutting pBS with *Hin*dIII and *Not*I and ligating to a pair of complementary oligonucleotids, Oligo-(*Hin*dIII-*Not*I)-For and Oligo-(*Hin*dIII-*Not*I)-Rev (Table 2-1).

pWW37 (multi *lox* sites):
A polylinker site with *lox5171*, *lox2272* and a *FRT* site (pWW37) was generated by cutting pWW24 with *Hin*dIII and *Xba*I and ligating to a pair of complementary oligonucleotids, *lox*P-(*Hin*dIII-*Xba*I)-For and *lox*P-(*Hin*dIII-*Xba*I)-Rev (Table 2-1).

pWW43:
1) pL341 was linearized by *Not*I first and then partially digested with *Hin*dIII, a 4.5 kb fragment was gel purified; 2) pWW15 was digested with *Hin*dIII and *Spe*I, a 2 kb fragment was gel purified; 3) pWW37 was digested with *Spe*I and *Not*I, a 0.2 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way ligation to make pWW43.

pWW48 (multi *lox* sites-3' *Hprt* cassette):
1) pWW37 was digested with *Sal*I and *Eco*RI, a 0.2 kb fragment was gel purified; 2) pWW22 was digested with *Eco*RI and *Hin*dIII, a 0.5 kb fragment was gel purified; 3) pWW22 was digested with *Hin*dIII and *Spe*I, a 1 kb fragment was gel purified; 4) pWW37 was digested with *Spe*I and *Nde*I, a 0.2 kb fragment was gel purified; 5) pL330 was digested with *Sal*I and *Nde*I, a 4.9 kb fragment was gel purified. The five fragments mentioned above were ligated together in a five-way ligation to make pWW48.

Table 2.1: Primers used for oligo ligation.

Primer Pair Name	Description	Forward Primer Name	Forward Primer Sequence	Reverse Primer	Reverse Primer Sequence	Plasmid
Oligo-(*Hin* dIII-*Not* I)	A variant polylinker site *Hin* dIII-*Spe* I-*Xba* I-*Nde* I-*Pst* I-*Eco* RI-*Not* I	Oligo-(*Hin* dIII-*Not* I)-For	5'-AGC TTA TCC ACT AGT CAC GGC CGC CAA AAG CGC TCT GAA GTT CCT ATA CTT TCT AGA GAA TAG GAAC TTC GGA ATA GGA ACT TCA AAG CGC ATA TGT CTG CAG GAA TTC GAT GAT CCA CTA GAG C-3'	Oligo-(*Hin* dIII-*Not* I)-Rev	5'-GGC CGC TCT AGT GGA TCA TCG AAT TCC TGC AGA CAT ATG CGC TTT GAA GTT CCT ATT CCG AAG TTC CTA TTC TCT AGA AAG TAT AGG AAC TTC AGA GCG CTT TTG GCG GCC GTG ACT AGT GGA TA 3'	pVWV24
lox P-(*Hin* dIII-*Xba* I)	A polylinker site with *lox5171* , *lox2272* and *FRT* sites	*lox* P-(*Hin* dIII-*Xba* I)-For	5'-AGC TTA TCC ACT AGT TAG GGA TAA CAG GGT AAT TCT AGT ATA ACT TCG TAT AAT GTG TAC TAT ACG AAG TTA TTC TAG TAT AAC TTC GTA TAA AGT ATC CTA TAC GAA GTT ATT-3'	*lox* P-(*Hin* dIII-*Xba* I)-Rev	5'-CTA GAA TAA CTT CGT ATA GGA TAC TTT ATA CGA AGT TAT ACT AGA ATA ACT TCG TAT AGT ACA CAT TAT ACG AAG TTA TAC TAG AAT TAC CCT GTT ATC CCT AAC TAG TGG ATA-3'	pVWV37
lox P-(*Xba* I-*Sal* I)	A polylinker site with *lox* P site	*lox* P-(*Xba* I-*Sal* I)-For	5'-CTA GAT AAC TTC GTA TAG CAT ACA TTA TAC GAA GTT ATG-3'	*lox* P-(*Xba* I-*Sal* I)-Rev	5'-TCG ACA TAA CTT CGT ATA ATG TAT GCT ATA CGA AGT TAT-3'	pVWV144
lox P-(*Eco* RI-*Hin* dIII)	A polylinker site with *lox* P site	*lox* P-(*Eco* RI-*Hin* dIII)-For	5'-AAT TCA TAA CTT CGT ATA GCA TAC ATT ATA CGA AGT TAT-3'	*lox* P-(*Eco* RI-*Hin* dIII)-Rev	5'-AGC TAT AAC TTC GTA TAA TGT ATG CTA TAC GAA GTT ATG-3'	pVWV205

pWW49 (5' *Hprt*-multi *lox* sites cassette):
1) pWW43 was partially digested with *Nde*I and then digested with *Sal*I, a 5 kb fragment was gel purified; 2) pWW23 was partially digested with *Spe*I and then digested with *Sal*I, a 1.5 kb fragment was gel purified. 3) pWW37 was digested with *Spe*I and *Nde*I, a 0.2 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW49.

pL325 is a gift from Dr. Pento Liu. To make this plasmid, a *D11Mit71* genomic fragment from pBZ84 (Zheng, Sage et al. 2000) was cloned into a vector containing the MC1-*tk* negative selection marker. A polylinker containing *Xho*I and *Not*I digestion sites was used to replace a 0.8 kb *Nco*I fragment. A selection cassette can be cloned into this polylinker to make a *D11Mit71* targeting vector.

pWW74 (multi *lox* sites-3' *Hprt* cassette, *D11mit71* targeting vector):
1) pL325 was digested with *Cla*I and *Not*I, a 6 kb fragment was gel purified; 2) pL325 was digested with *Cla*I and *Xho*I, a 4 kb fragment was gel purified; 3) pWW48 was digested with *Sal*I and *Not*I, a 3.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW74.

pWW75 (5' *Hprt*-multi *lox* sites cassette, *D11mit71* targeting vector):
1) pL325 was digested with *Cla*I and *Not*I, a 6 kb fragment was gel purified; 2) pL325 was digested with *Cla*I and *Xho*I, a 4 kb fragment was gel purified; 3) pWW49 was digested with *Sal*I and *Not*I, a 3.8 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW75.

2.1.2 Vectors for *E_2DH* end point targeting

pWW63 (*PGK* promotor):
pWW48 was digested with *Eco*RI and *Bgl*II, a 0.5 kb fragment was gel purified and cloned into pBS digested with *Eco*RI and *Bam*HI to make pWW63.

pWW144 (*PGK-lox*P):
1) pWW63 was digested with *Hin*dIII and *Xba*I, a 0.5kb fragment was gel purified; 2) a pair of complementary oligonucleotids, *lox*P-(*Xba*I-*Sal*I)-For and *lox*P-(*Xba*I-*Sal*I)-Rev, were annealed together (Table 2-1); 3) pBS was digested with *Hin*dIII and *Sal*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW144.

pL313 is a kind gift from Dr. Pentao Liu. It contains a *PGK-EM7-Bsd-bpA* cassette. This cassette can be selected both in *Escherichia coli* (75 μg/ml) and in eukaryotic cells (10 μg/ml) using Blasticidin S HCl.

pWW146 (*PGK-lox*P-*EM7-Bsd-bpA*):
1) pWW144 was digested with *Hin*dIII and *Sal*I, a 600 bp fragment was gel purified; 2) pL313 was partially digested with *Xho*I, and then digested with *Hin*dIII, a 3.7 kb fragment was gel purified. The two fragments were ligated together to make pWW146.

pL10 and pL11 are two vectors that contain 5' and 3' genomic insert of E_2DH locus, respectively (Liu, Zhang et al. 1998).

pWW183 (E_2DH targeting vector with *PGK-lox*P-*EM7-Bsd-bpA* cassette, without MC1-*tk*):
1) pL10 was digested with *Sac*II and *Xho*I, a 4.4 kb fragment was gel purified; 2) pL11 was digested with *Sac*II and *Not*I, a 6.8 kb fragment was gel purified; 3) pWW146 was digested with *Sal*I and *Not*I, a 1.4 kb fragment was gel purified. The three fragments were ligated together in a three-way-ligation to make pWW183.

pL253 is a kind gift from Dr. Pentao Liu. It contains a MC1-*tk* cassette, which can be used for negative selection in mammalian cells.

pWW190 (*E$_2$DH* targeting vector with *PGK*-*lox*P-*EM7*-*Bsd*-*bpA* cassette, with MC1-*tk*):
1) pWW183 was partially digested with *Bam*HI, and then digested with *Sac*II, a 9.6 kb fragment was gel purified; 2) pL253 was partially digested with *Sac*II, and then digested with *Bam*HI, a 5 kb fragment was gel purified. The two fragments were ligated together to make pWW190.

2.1.3 Trapping vectors

2.1.3.1 Promoter trapping vectors

pWW38 (SA-β*geo* cassette):
pSAβ*geo* (Friedrich and Soriano 1991) was cut with *Xho*I, a 4.3 kb fragment was gel purified and cloned into pBS digested with *Xho*I and *Sal*I to make pWW38. The desired orientation of the insert was determined by digestion with *Eco*RI.

pWW62 (*Puro*-*bpA* with multi *lox* sites):
pWW48 was digested with *Bgl*II and *Eco*RI, a 1.2 kb fragment was gel purified and cloned into pBS digested with *Bam*HI and *Eco*RI to make pWW62.

pWW202 (promoter-less *Puro*-*bpA*):
pWW62 was cut with *Hin*dIII and *Xho*I, a 1 kb fragment was gel purified and cloned into pBS digested with *Hin*dIII and *Xho*I to make pWW202.

pWW205 (promoter-less *lox*P-*Puro*-*bpA*):
1) pWW202 was digested with *Hin*dIII and *Xho*I, a 1kb fragment was gel purified; 2) a pair complementary of oligonucleotids, *lox*P-(*Eco*RI-*Hin*dIII)-For and *lox*P-(*Eco*RI-*Hin*dIII)-Rev, were annealed together (Table 2-1); 3) pBS was digested with *Eco*RI and *Xho*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW205.

pWW237 (plasmid-based 5' trapping vector):

1) pWW38 was digested with *Bgl*II and *Eco*RI a 4.4 kb fragment was gel purified; 2) pWW205 was digested with *Eco*RI and *Xho*I, a 1 kb fragment was gel purified; 3) pBS was digested with *Bam*HI and *Xho*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW237.

pWW239 (5' trapping retrovirus):
1) pWW38 was digested with *Bgl*II and *Eco*RI, a 4.4 kb fragment was gel purified; 2) pWW205 was digested with *Eco*RI and *Xho*I, a 1 kb fragment was gel purified; 3) pMSCV-Neo (Clontech) was digested with *Bam*HI and *Xho*I, a 5.1 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW239.

2.1.3.2 PolyA trapping vectors

CAG promoter and EF1α promoter subcloning vectors are kind gifts from Dr. Haydn Prosser. pVecH1S (regional trapping vector) is a kind gift from Dr. Meredith Wentland. pYTC31 (*PGK-Bsd-bpA*) is a kind gift from Dr. You-Tzung (Bob) Chen.

pWW12 (promoter-less *Bsd-bpA*):
pYTC31 was cut with *Xho*I and *Pst*I, a 0.6 kb fragment was gel purified and cloned into pBS digested with *Xho*I and *Pst*I to make pWW12.

pWW18 (3' trapping retrovirus with *PGK* promoter):
1) pVecH1S was digested with *Hin*dIII, a 1.9 kb fragment was gel purified and digested again with *Bam*HI , a 0.6 kb frament was gel purified; 2) pVecH1S was digested with *Xho*I, a 4.7 kb fragment was gel-purified and digested again with *Hin*dIII, a 4.7 kb fragment was gel purified; 3) pWW12 was digested with *Bam*HI and *Xho*I, a 0.6 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way-ligation to make pWW18.

pWW41 (*CAG* promoter):

CAG promoter subcloning vector was digested with *Hin*cII and *Eco*RI, a 1.6 kb fragment was gel purified and cloned into pBS digested with *Hin*cII and *Eco*RI to make pWW41.

pWW42 (EF1α promoter):
EF1α promoter subcloning vector was digested with *Hin*dIII and *Eco*RI , a 1.3 kb fragment was gel purified and cloned into pBS digested with *Hin*dIII and *Eco*RI to make pWW42.

pWW44 (3' trapping virus with *CAG* promoter, alternative version):
1) pWW18 was digested with *Eco*RI and *Nhe*I, a 4 kb fragment was gel purified; 2) pWW18 was digested with *Eco*RI and *Bgl*II, a 0.7 kb fragment was gel purified; 3) pWW18 was digested with *Nhe*I and *Sal*I, a 0.6 kb fragment was gel purified; 4) pWW41 was digested with *Bam*HI and *Sal*I, a 1.6 kb fragment was gel purified. The four fragments mentioned above were ligated together in a four-way ligation to make pWW44.

pWW45 (3' trapping retrovirus with EF1α promoter, alternative version):
1) pWW18 was digested with *Eco*RI and *Nhe*I, a 4 kb fragment was gel purified; 2) pWW18 was digested with *Eco*RI and *Bgl*II, a 0.7 kb fragment was gel purified; 3) pWW18 was digested with *Nhe*I and *Sal*I, a 0.6 kb fragment was gel purified; 4) pWW42 was digested with *Bam*HI and *Sal*I, a 1.3 kb fragment was gel purified. The four fragments mentioned above were ligated together in a four-way ligation to make pWW45.

pWW59 (PolII-*Neo*-*bpA* with multi *lox* sites):
pWW49 was cut with *Hin*dIII and *Not*I, a 2 kb fragment was gel purified and cloned into pBS digested with *Hin*dIII and *Not*I to make pWW59.

pWW64 (3' trapping virus with *CAG* promoter, final version):
1) pWW44 was partially digested with *Bam*HI, a 7 kb fragment was gel purified and digested again with *Nhe*I and *Not*I, a 5 kb fragment was gel purified; 2) pWW44 was digested with *Xho*I and *Not*I, a 1.5 kb fragment was

gel purified; 3), pWW12 was digested with *Bam*HI and *Xho*I, a 0.5 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way ligation to make pWW64.

pWW65 (3' trapping virus with EF1α promoter, final version):
1) pWW45 was partially digested with *Bam*HI, a 6.7 kb fragment was gel purified and digested again with *Nhe*I and *Not*I, a 4.7 kb fragment was gel purified; 2) pWW45 was digested with *Xho*I and *Not*I, a 1.5 kb fragment was gel purified; 3), pWW12 was digested with *Bam*HI and *Xho*I, a 0.5 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way ligation to make pWW65.

pWW201 (promoter-less *Neo-bpA*):
pWW59 was cut with *Eco*RI and *Xho*I, a 1.1 kb fragment was gel purified and cloned into pBS digested with *Eco*RI and *Xho*I to make pWW201.

pWW238 (plasmid-based 3' trapping vector with *CAG* promoter):
1) pWW64 was digested with *Cla*I and BamHI, a 2.7 kb fragment was gel purified; 2) pWW201 was digested with *Bam*HI and *Xho*I, a 1.1 kb fragment was gel purified; 3) pBS was digested with *Cla*I and *Xho*I, a 2.9 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way ligation to make pWW238.

pWW240 (3' trapping retrovirus with *CAG* promoter in pMSCV backbone):
1) pWW64 was digested with *Cla*I and *Bam*HI, a 2.7 kb fragment was gel purified; 2) pWW201 was digested with *Bam*HI and *Xho*I, a 1.1 kb fragment was gel purified; 3) pMSCV-Neo (Clontech) was digested with *Cla*I and *Xho*I, a 4.1 kb fragment was gel purified. The three fragments mentioned above were ligated together in a three-way ligation to make pWW240.

2.2 Cell culture

2.2.1 ES cell culture condition

ES cell culture was performed as described before (Ramirez-Solis, Davis et al. 1993). Briefly, AB2.2 (129 S7/SvEv$^{\text{Brd-}Hprtb\text{-}m2}$) wild-type ES cells and their

derivatives were always maintained on SNL76/7 feeder cell layers mitotically inactivated treated by γ-irradiation. ES cells were grown in M15 medium (Table 2-2). Cells were cultured at 37 °C with 5% CO_2. If not specified, ES cell medium was changed daily.

When ES cells reached 80-85% confluence, they were ready for passaging. The media was changed about two hours before passaging. After two hours, media was aspirated off, and the plate was washed once with PBS. 2 ml of trypsin was added to each 90-mm plate. The plate was incubated in a TC incubator at 37 °C for 15 minutes. 8 ml of fresh M15 media was added to each well. The cells were dispersed by pipetting up and down vigorously. The ES cell suspension was then evenly distributed to three to four 90-mm feeder plates. The plates were incubated in a TC incubator at 37 °C.

2.2.2 Chemicals used for selection of ES cells

Blasticidin: Blasticidin S HCl (Invitrogen), 1000X stock (5 mg/ml) was made in Phosphate Buffered Saline (PBS). After mixing, the 1000X stock solution was filter sterilized through a 0.2 mm syringe filter.

FIAU: 1-(2'-deoxy-2'-fluoro-β-D-arabinofuranosyl)-5-iodouracil, 1000X stock (200 μM) was made in PBS and 5 M NaOH was added dropwise until it is dissolved. After mixing, the 1000X stock solution was filter sterilized through a 0.2 mm syringe filter.

G418: Geneticin (Invitrogen), was bought as a sterile stock solution containing 50 mg/ml active ingredient.

Puromycin: ($C_{22}H_{29}N_7O_5$.2HCL, Sigma) 1000X stock (3 mg/ml) was made in MiliQ water. After mixing, the 1000X stock solution was filter sterilized through a 0.2 mm syringe filter.

HAT: 50X HAT supplement (Hypoxanthine-aminopterin-thymidine) (Invitrogen) was bought as a sterile stock solution containing 5 mM Hypoxathine, 20 μM Aminopterin and 0.8 mM Thymidine.

HT: 50X HT supplement (Hypoxanthine-thymidine) (Invitrogen) was bought as a sterile stock solution containing 5 mM Hypoxathine and 0.8 mM Thymidine.

Trypsin: For 5 L, add 35 g NaCl, 5 g D-glucose, 0.9 g $Na_2HPO_4.7H_2O$, 1.85 g KCl, 1.2 g KH_2PO_4, 2 g EDTA, 12.5 g Trpsin (1:250), 15 g Tris base. Adjust the pH from 8.71 to 7.6 with HCl, add phenol to get pink colour. Filter-sterilized and aliquoted into 50 ml falcon tubes, and store at -20 °C.

Table 2.2: Cell culture medium.

Medium Name	Recipe	Purpose
M15	Knockout Dulbecco's Modified Eagle's Medium (DMEM, Gibco/Invitrogen), supplemented with 15% foetal bovine serum (FBS, Gibco/Invitrogen), 2 mM L-glutamine, 50 units/ml penicillin, 40 μg/ml streptomycin and 100 μM β-Mercaptoethanol (β-ME)	Culture of undifferentiated ES cells
M10	Knockout Dulbecco's Modified Eagle's Medium (DMEM, Gibco/Invitrogen) supplemented with 10% foetal bovine serum (FBS, Gibco/Invitrogen), 2 mM L-glutamine, 50 U/ml penicillin, and 40 μg/ml streptomycin	Culture of feeder cells and Phoenix cells
Viral Production Medium	M10 medium supplemented with heat-inactivated FBS	Virus production
Differentiation Medium	For 100 ml Knockout Dulbecco's Modified Eagle's Medium (DMEM, Gibco/Invitrogen), 25 ml FBS (Gibco/Invitrogen), 1.25 ml 200 mM 100X L-glutamine stock (Gibco/Invitrogen), 1.25 ml 10mM β-ME stock (10mM) and 1.25 ml 100X nonessential amino acids (NEAA) stock (Gibco/Invitrogen) were added	ES cell *in vitro* differentiation

2.2.3 Transfection of DNA into ES cells by electroporation

DNA used for ES cell transfection was normally prepared using a Qiagen Plasmid Purification Kit (Qiagen). If DNA was used for gene targeting, it would be linearized by digestion with an appropriate enzyme under the conditions recommended by the manufacturers. If DNA was used for transient expression, the linearization step would be omitted. Before electroporation, DNA was purified by ethanol precipitation and air-dried briefly in a tissue culture (TC) hood. The air-dried DNA was then dissolved in sterile 1X TE buffer (pH 8.0) to a final concentration of about 1 μg/μl. Unless specified, 20 μg DNA was used for each electroporation.

ES cell electroporation was performed according to standard protocols (Ramirez-Solis, Davis et al. 1993). Briefly, ES cells (80% confluent) were fed 2-3 hours before harvesting. Immediately before electroporation, ES cells were trypsinized and resuspend in M15 media. The cells were collected by centrifuging and washed once in PBS. The cells were resuspended in PBS to a final concentration of $1X10^7$ cells/ml. $1X10^7$ ES cells were transferred into a 0.4 cm gap curvette (Biorad) together with 20 μg DNA. The electroporation was carried out using a Biorad “Gene Pulser” at 230 V, 500 μF. After electroporation, ES cells were plated onto a 90-mm feeder plate and unless stated otherwise, were cultured for 10 days to allow the formation of single ES cell colonies. Drug selection was usually initiated 24 hours post-electroporation.

2.2.4 Picking ES cell colonies

50 μl of trypsin was added to each well of a 96-well round bottom plate by using a multi-channel pipette. After washing a 90-mm tissue culture plate for picking with PBS, about 8 ml PBS was added to cover the plate. The colonies were picked from the 90-mm plate by using a P20 Pipetman set at 10 μl and transferred into the wells with trypsin. After completing a 96-well plate, the plate was incubated in a TC incubator at 37 °C for 10 to 15 minutes. After that, 150 μl of fresh M15 media was added to each well. The colonies were

broken up by pipetting up and down vigorously. The ES cell suspension was then transferred to a 96-well feeder plate. The plate was incubated in a TC incubator at 37 °C.

2.2.5 Passaging ES cells

When ES cells in most wells on a 96-well plate reached 80-85% confluence (determined both by the change of the medium colour and by checking the plate under a microscope), the plate was judged ready for passaging. The media was changed about two hours before passaging. After two hours, media was aspirated off, and the plate was washed once with PBS. 50 µl of trypsin was added to each well of a 96-well plate by using a multi-channel pipette. The plate was incubated in a TC incubator at 37 °C for 15 minutes. 150 µl of fresh M15 media was added to each well. The cells were separated by pipetting up and down vigorously. The ES cell suspension was then evenly distributed to three to four 96-well feeder/gelatinized plates. The plates were incubated in a TC incubator at 37 °C.

2.2.6 Freezing ES cells

When ES cells reached 80-85% confluence, they were ready for freezing. The media was changed about two hours before passaging. After two hours, media was aspirated off, and the plate was washed once with PBS. 50 µl of trypsin was added to each well of a 96-well plate by using a multi-channel pipette. The plate was incubated in a TC incubator at 37 °C for 15 minutes. 50 µl of 2X Freezing Media (60% DMEM, 20% FCS, 20% DMSO) was added to each of the wells and the cells were broken up by pipetting up-and-down. 100 µl of filter-sterilized (0.22 um) Mineral Oil was added to each well. The plate was put into a polystyrene box with lid and frozen at –80 °C.

2.2.7 Thawing ES cells

To thaw frozen ES cell clones, the 96-well plate was taken out of the -80 °C freezer and placed immediately into the 37 °C incubator. After all of the wells thawed completely, the clones were transferred to appropriately labelled wells in 24-well feeder plates pre-equilibrated with 2 ml of M15 media per well. For

maximum recovery of sample, another 200 µl of M15 was added to rinse each well and the cell suspension was transferred to the appropriate wells in the 24-well feeder plates. The plates were incubated in a TC incubator at 37 °C.

2.2.8 Cre-mediated recombination to pop out the selection cassettes

20 µg of the Cre expression plasmid pCAAG-Cre (Araki, Araki et al. 1995) was electroporated into $1X10^7$ ES cells. After electroporation, the cells were serially diluted in M15 and about 1,000 ES cells were plated onto a 90-mm feeder plate and cultured for 10 days to allow the formation of single ES cell colonies. 96 ES cell clones were picked into a 96-well feeder plate. To identify clones with Cre-mediated recombination events, the 96 well plates were replicated. Sib-selection was performed to identify ES clones with correct drug resistance pattern. The right clones were expanded and confirmed by Southern analysis.

2.2.9 Generation of targeted ES cell lines

WW14 (AB2.2 targeted with pWW74):

20 µg of pWW74 was linearized with *Sca*I and electroporated into AB2.2 cells (#239, passage 17). The transfectants were selected with puromycin and FIAU simultan*eo*usly for 8 days. 96 puromycin resistant clones were picked and expanded on a 96-well feeder plate. Genomic DNA was extracted and digested with *Bam*HI for Southern analysis using a *D11Mit71* 3' probe to identify gene-targeting events. The expected sizes of the detected restriction fragments were 14.1 kb for wild-type allele and 10.3 kb for the targeted allele. The correctly targeted clones were expanded and named WW14.

WW16 (AB2.2 targeted with pWW75):

20 µg of pWW75 was linearized with *Sca*I and electroporated into AB2.2 cells (#239, passage 17). The transfectants were selected with G418 and FIAU simultaneously for 8 days. 96 G418 resistant clones from each cell line were picked and expanded on a 96-well feeder plate. Genomic DNA was extracted and digested with *Bam*HI for Southern analysis using a *D11Mit71* 3' probe to identify gene-targeting events. The expected sizes of the detected restriction

fragments were 14.1 kb for the wild-type allele and 17.9 kb for the targeted allele. The correctly targeted clones were expanded and named WW16.

WW24 (WW14 targeted with pWW75):
20 μg of pWW75 was linearized with *Sca*I and electroporated into WW14. The transfectants were selected with G418 and FIAU simultaneously for 8 days. 96 G418 resistant clones were picked and expanded on a 96-well feeder plate. Genomic DNA was extracted and digested with *Bam*HI for Southern analysis using a *D11Mit71* 3' probe to detect gene-targeting events. The double-targeted clones would have a 10.3 kb restriction fragment (3' *Hprt* targeting) and a 17.9 kb restriction fragment (5' *Hprt* targeting). The correctly targeted clones were expanded and named WW24. The function of the 5' *Hprt* and 3' *Hprt* cassettes were tested by transient Cre expression and subsequent HAT selection.

WW25 (WW16 targeted with pWW74):
20 μg of pWW74 was linearized with *Sca*I and electroporated into WW16. The transfectants were selected with puromycin and FIAU simultaneously for 8 days. 96 puromycin resistant clones were picked and expanded on a 96-well feeder plate. Genomic DNA was extracted and digested with *Bam*HI for Southern analysis using a *D11Mit71* 3' probe to detect gene-targeting events. The double-targeted clones would have a 10.3 kb restriction fragment (3' *Hprt* targeting) and a 17.9 kb restriction fragment (5' *Hprt* targeting). The correctly targeted clones were expanded and named WW25. The function of the 5' *Hprt* and 3' *Hprt* cassettes were tested by transient Cre expression and subsequent HAT selection.

Table 2.3: ES Cell lines constructed for the project.

ES Cell Line	Targeted Locus	Starting Cell line	Targeting Vector / Cre expression plasimid	Linearization Enzyme	Drug Selection	Genomic DNA digestion	Probe and sizes of fragments	Drug Resistance and Sensitivity
WW14	*D11Mit71*	AB2.2	pWW74	*Sca* I	puromycin+FIAU	*Bam* HI (*D11Mit71* 5' probe), *Xba* I (*D11Mit71* 3' probe)	*D11Mit71* 5' probe: 7.1 kb (W.T.), 4.9 kb (3' *Hprt*). *D11Mit71* 3' probe: 14.1 kb (W.T.), 10.3 kb (3' *Hprt*).	Neo^S, $Puro^R$, Bsd^S, HAT^S
WW16	*D11Mit71*	AB2.2	pWW75	*Sca* I	G418+FIAU	*Bam* HI (*D11Mit71* 5' probe), *Xba* I (*D11Mit71* 3' probe)	*D11Mit71* 5' probe: 7.1 kb (W.T.), 6.4 kb (5' *Hprt*). *D11Mit71* 3' probe: 14.1 kb (W.T.), 17.9 kb (5' *Hprt*).	Neo^R, $Puro^S$, Bsd^S, HAT^S
WW24	*D11Mit71*	WW14-B2	pWW75	*Sca* I	G418+FIAU	*Bam* HI (*D11Mit71* 5' probe), *Xba* I (*D11Mit71* 3' probe)	*D11Mit71* 5' probe: 6.4 kb (5' *Hprt*), 4.9 kb (3' *Hprt*). *D11Mit71* 3' probe: 17.9 kb (5' *Hprt*), 10.3 kb (3' *Hprt*).	Neo^R, $Puro^R$, Bsd^S, HAT^S
WW25	*D11Mit71*	WW16-B2	pWW74	*Sca* I	puromycin+FIAU	*Bam* HI (*D11Mit71* 5' probe), *Xba* I (*D11Mit71* 3' probe)	*D11Mit71* 5' probe: 6.4 kb (5' *Hprt*), 4.9 kb (3' *Hprt*). *D11Mit71* 3' probe: 17.9 kb (5' *Hprt*), 10.3 kb (3' *Hprt*).	Neo^R, $Puro^R$, Bsd^S, HAT^S
WW45	*D11Mit71*	WW24-A1	pCCAG-Cre	N.A.	No selection	*Xba* I (*D11Mit71* 5' probe)	*D11Mit71* 5' probe: 6.4 kb (5' *Hprt*), 4.9 kb (3' *Hprt*).	Neo^S, $Puro^S$, Bsd^S, HAT^S
WW46	*D11Mit71*	WW25-C1	pCCAG-Cre	N.A.	No selection	*Xba* I (*D11Mit71* 5' probe)	*D11Mit71* 5' probe: 6.4 kb (5' *Hprt*), 4.9 kb (3' *Hprt*).	Neo^S, $Puro^S$, Bsd^S, HAT^S
WW69	*E_2DH*	WW45-B2	pWW190	*Sac* II	blasticidin+FIAU	*Eco* RI (*E_2DH* 5' probe), *Nde* I (*E_2DH* 3' probe)	*E_2DH* 5' probe: 14.9 kb (W.T.), 9.2 kb (targeted). *E_2DH* 3' probe: 13.1 kb (W.T.), 9.6 kb (targeted).	Neo^S, $Puro^S$, Bsd^R, HAT^S
WW93	*E_2DH*	WW69-D6	pCCAG-Cre	N.A.	HAT/HT	*Eco* RI (*E_2DH* 5' probe), *Nde* I (*E_2DH* 3' probe)	*E_2DH* 5' probe: 9.2 kb (targeted). *E_2DH* 3' probe: 9.6 kb (targeted).	Neo^S, $Puro^S$, Bsd^R, HAT^R
WW99	Random retroviral insertion	WW69-D6	pWW239 derived gene trap virus	N.A.	G418	*Kpn* I (*LacZ* probe)	*LacZ* probe: 6.0/6.9 kb (proviral insertion).	Neo^R, $Puro^S$, Bsd^R, HAT^S
WW103-RT	Cre induced inversion	WW99	pCCAG-Cre	N.A.	puromycin	*Kpn* I (*LacZ* probe)	*LacZ* probe: 19.0 kb (inversion).	Neo^R, $Puro^R$, Bsd^R, HAT^S
WW103	Cre induced mitotic recombination	WW103-RT	pCCAG-Cre	N.A.	HAT/HT	*Nde* I (*E_2DH* 3' probe), *Eco* RI & *Spe* I (*LacZ* probe)	*E_2DH* 3' probe: 9.6 kb (targeted). LacZ probe: restriction fragments of various lengths (proviral/host junction fragment).	Neo^R, $Puro^R$, Bsd^R, HAT^R
WW100	Random vector insertion	WW69-D6	pWW237	*Sca* I	G418	N.A.	N.A.	Neo^R, $Puro^S$, Bsd^R, HAT^S
WW104	Cre induced inversion	WW100	pCCAG-Cre	N.A.	puromycin	N.A.	N.A.	Neo^R, $Puro^R$, Bsd^R, HAT^S
WW106	Cre induced mitotic recombination	WW104	pCCAG-Cre	N.A.	HAT/HT	*Nde* I (*E_2DH* 3' probe), *Eco* RI & *Spe* I (*LacZ* probe)	*E_2DH* 3' probe: 9.6 kb (targeted). LacZ probe: restriction fragments of various lengths (vector/host junction fragment)	Neo^R, $Puro^R$, Bsd^R, HAT^R

WW45 (WW24 selection markers pop-out):
20 μg of supercoiled pCAAG-Cre was electroporated into WW24. About 1,000 ES cells were plated onto a 90-mm feeder plate and cultured for 10 days to allow the formation of single ES cell colonies. 96 clones were picked and expanded on a 96-well feeder plate. The 96-well plate was replicated and sib-selection was performed to identify ES clones in which both *Neo* and *Puro* cassettes were popped out, but no recombination had happened between the two half *Hprt* cassettes. The correct recombinants should be G418 sensitive, puromycin sensitive and HAT sensitive. The clones showing this pattern of sensitivity were expanded and confirmed by Southern analysis using a *D11Mit71* 3' probe. The double-targeted clones would have a 10.3 kb *Bam*HI restriction fragment (3' *Hprt* targeting) and a 15.9 kb *Bam*HI restriction fragment (5' *Hprt* targeting and *Neo* pop-out). The function of the 5' *Hprt* and 3' *Hprt* cassettes were tested by transient Cre expression and subsequent HAT selection.

WW46 (WW25 selection markers pop-out):
20 μg of supercoiled pCAAG-Cre was electroporated into WW25. About 1,000 ES cells were plated onto a 90-mm feeder plate and cultured for 10 days to allow the formation of single ES cell colonies. 96 clones were picked and expanded on a 96-well feeder plate. The 96-well plate was replicated and sib-selection was performed to identify ES clones in which both *Neo* and *Puro* cassettes were popped out, but no recombination had happened between the two half *Hprt* cassettes. The correct recombinants should be G418 sensitive, puromycin sensitive and HAT sensitive. The clones showing this pattern of sensitivity were expanded and confirmed by Southern analysis using a *D11Mit71* 3' probe. The double-targeted clones would have a 10.3 kb *Bam*HI restriction fragment (3' *Hprt* targeting) and a 15.9 kb *Bam*HI restriction fragment (5' *Hprt* targeting and *Neo* pop-out). The function of the 5' *Hprt* and 3' *Hprt* cassettes were tested by transient Cre expression and subsequent HAT selection.

WW69 (WW45 targeted with pWW190):
20 μg of pWW190 was linearized with *Sca*I and electroporated into WW45. The transfectants were selected with blasticidin and FIAU simultaneously for 8 days. 96 blasticidin resistant clones were picked and expanded on a 96-well feeder plate. Genomic DNA was extracted and Southern analysis was performed using several different probes to determine the genotype of the clones: 1) *Bam*HI digestion, hybridized with a *D11Mit71* 3' probe, the correct clones would have a 10.3 kb restriction fragment (3' *Hprt* targeting) and a 15.9 kb restriction fragment (5' *Hprt* targeting and *Neo* pop-out). 2) *Xba*I digestion, hybridized with a *D11Mit71* 5' probe (pWW116), the correctly targeted clones would have a 6.4 kb restriction fragment (5' *Hprt* targeting and *Neo* pop-out) and a 5.0 kb restriction fragment (3' *Hprt* targeting and *Puro* pop out); 3) *Eco*RI digestion, hybridized with an E_2DH 5' probe (pL16, (Liu, Zhang et al. 1998)), the targeted restriction fragment is 9.2 kb and the wild-type restriction fragment is 14.9 kb; 4) *Nde*I digestion, hybridized with an E_2DH 3' probe (pL17, (Liu, Zhang et al. 1998)), the targeted restriction fragment is 9.6 kb and the wild-type restriction fragment is 13.1 kb. Two correctly targeted clones were identified by Southern analysis, WW69-C8 and WW69-D6.

To use induced mitotic recombination to make the homozygous mutations, the E_2DH end point targeting cassette needs to be on the same chromosome as the 3' *Hprt* cassette. To determine the location of the *PGK-lox*P-*Bsd-bpA* cassette, the WW69-C8 and D6 clones were expanded and supercoiled pCAAG-Cre was electroporated into both. The recombinants were selected with HAT for 6 days and HT for another 4 days. 36 HAT resistant clones were picked from each electroporation and cultured on a 96-well feeder plate. Genomic DNA was extracted and Southern analysis was performed using the *D11Mit71* 5' and 3' probe as well as the E_2DH 5' and 3' probe. The cell line WW69-D6 was determined to have both the right genotype and *PGK-lox*P-*Bsd-bpA* cassette location. Single cell subclones were isolated to avoid possible contamination of other cells. The subclones were confirmed by Southern and sib-selection. The correct recombinants should be G418 sensitive, puromycin sensitive, HAT sensitive and blasticidin resistant.

WW93 (WW69-D6 induced mitotic recombination):
20 μg of supercoiled pCAAG-Cre was electroporated into WW69-D6. The recombinants were selected with HAT for 6 days and HT for another 4 days. 96 clones were picked and expanded on a 96-well feeder plate. The plate was replicated and sib-selection was performed to determine the drug resistance of the clones. The correct recombinants should be G418 sensitive, puromycin sensitive, HAT resistant and blasticidin resistant. The clones with this combination of drug resistance and sensitivity were expanded and confirmed by Southern analysis using the *E_2DH* 5' and 3' probe: 1) *Eco*RI digestion, hybridized with an *E_2DH* 5' probe. The correct recombinants would only have the 9.2 kb targeted restriction fragment but not the 14.9 kb wild-type restriction fragment; 2) *Nde*I digestion, hybridized with *E_2DH* 3' probe, The correct recombinants would only have the 9.6 kb targeted restriction fragment but not the 13.1 kb wild-type restriction fragment. One of the correct clones, WW93-A12 was expanded and single cell subcloned to avoid possible contamination by other cells. This cell line was used as control for the ES cell *in vitro* differentiation.

2.2.10 Retroviral approaches

2.2.10.1 Retrovirus production

The Phoenix ecotropic retroviral packaging cell line (Grignani, Kinsella et al. 1998), a derivative of human embryonic kidney 293T line expressing retroviral *gal*, *pol* and *env* proteins, was obtained from the American Tissue Culture Collection (Manassas, Virginia, USA). Cells were cultured according to the protocols on Dr. Garry Nolan's lab webpage (http://www.stanford.edu/group/nolan). Briefly, the Phoenix cells were cultured in M10 medium (Table 2-2) at 37°C with 5% CO_2. The medium was changed every 2-3 days. Cells were split 1:5 when they reached 70-80% confluence.

24 hours prior to transfection, Phoenix cells were plated at a density of 2X 10^6 cells per 90-mm plate in M10. 2-3 hours before transfection, cells were fed with 14 ml fresh M10 medium (at this time the cells were about 60% confluent).

CalPhos™ Mammalian Transfection Kit (BD Bioscience) was used for transient transfection of the Phenoix retroviral packaging cell line. Briefly, DNA prepared with the Qiagen Plasmid Purification Kit (Qiagen) was precipitated with ethanol, air-dried and then dissolved in appropriate volume of TE. For each transfection of cells on each 90-mm plate, 25 μg DNA was mixed with 86.8 μl 2 M Calcium Phosphate Solution. Sterile water was added to make a final volume of 700 μl. The calcium solution containing DNA was added dropwise to 700 μl 2X HEPES-buffered Saline (HBS) solution, while being mixed quickly by bubbling vigorously with a 1 ml sterile pipette and an autopipettor.

The DNA mixture was incubated at room temperature for 20 minutes, vortexed gently and then added dropwise to the culture plate medium. 24 hours after transfection, the calcium phosphate-containing medium was removed, plates were washed twice with PBS and 10 ml of fresh Viral Production Medium (Table 2-2) was added to each plate. Viral supernatant was harvested 36, 48, 60 and 72 hours after transfection and stored immediately in a –80°C freezer.

2.2.10.2 Viral Infection

ES cells were plated at a density of 3X 10^6 cells per 90-mm feeder plate about 24 hours before infection. The viral supernatant collected from all the time points was mixed together and filtered through a 0.45 μm filter. Heat-inactivated FBS was added to the viral supernatant to make the final concentration of FBS up to 15%. Polybrene (Hexadimethrine Bromide, Sigma) was added to the viral supernatant to a final concentration of 4 μg/ml. 12 ml viral supernatant was added to each plate of ES cells. The viral supernatant was replaced every 12 hours with fresh supernatant. After 48 hours of infection, the viral supernatant was removed and fresh M15 medium was added. The drug selection was applied 24 hours after infection was stopped.

2.2.10.3 Titration of the retrovirus

ES cells were plated at a density of 3X 10^6 cells per 90-mm feeder plate. 24 hours later, 1 ml or 10 ml of viral supernatant was applied to each plate. For the virus carrying a *Neo* cassette, G418 selection (180 μg/ml) was initiated 24 hours after viral infection and continued for 8 days. The drug-resistant ES colonies were stained with 2% methylene blue in 70% ethanol and counted. The titre of the retrovirus is defined as the number of drug resistant ES cell colonies per milliliter of viral supernatant used to infect the cells.

2.2.11 Gene trap mutagenesis using the retroviral vector

2.2.11.1 Gene trapping

WW99 (WW69-D6 infected with pWW239-derived retrovirus):
Gene-trap retrovirus was produced by transient transfection of Phoenix viral packaging cells. A total of 2000 ml of viral supernatant was harvested and filtered through 0.45 μm filters. WW69-D6 ES cells were plated on a total of twenty 90-mm feeder plates at a density of 3X 10^6 cells per plate (WW99-1 to WW90-20). 24 hours later, each plate of cells was infected with 12 ml of viral supernatant. Viral supernatant was replaced by fresh supernatant every 12 hours. After 48 hours, the viral supernatant was removed and fresh M15 medium was added to each plate. G418 selection (180 μg/ml) was initiated 24 hours after the viral infection terminated. Drug selection was continued for 10 days until the G418 resistant colonies were clearly visible. One plate (WW99-20) was stained with 2% methylene blue in 70% ethanol to determine the number of gene-trap clones obtained. The G418 resistant ES cell colonies from each of the remaining 19 retrovirus infected plates were separately trypsinized, resuspended in M15 medium and plated as a pool onto 19 feeder plates (WW99-1 to WW99-19). These cells were selected with G418 until they reached about 80% confluence. 1X10^7 cells were used for the Cre-mediated inversions. The rest of the cells were frozen down for the stock.

2.2.11.2 Cre-mediate inversion

WW103-RT (WW99 regional trapping):

20 μg of supercoiled pCAAG-Cre was electroporated into $1X10^7$ cells from the WW99-1 to WW99-19 pools. Puromycin selection (3 μg/ml) was initiated 24 hours after the electroporation. The drug selection was continued for 6 days until the colonies were visible under microscope. Selection was then released, and the colonies were grown in M15 medium for another 4 days. The puromycin resistant ES cell colonies from each of the 19 plates were trypsinized, resuspended in M15 medium and maintained as 19 separate pools on 19 feeder plates (WW103-RT-1 to WW103-RT-19). These cells were selected with puromycin until they reached about 80% confluence. $1X10^7$ cells were used for the Cre-induced mitotic recombination. The rest of the cells were frozen down for the stock.

2.2.11.3 Cre-induced mitotic recombination

WW103 (WW103-RT induced mitotic recombination):

20 μg of supercoiled pCAAG-Cre was electroporated into $1X10^7$ cells of WW103-RT-1 to WW103-RT-19 pools. HAT selection was initiated 24 hours after the electroporation. The drug selection continued for 6 days until the colonies were visible under microscope and the colonies were grown in M15 medium with HT supplement for another 4 days. 48 HAT resistant ES cell colonies from each of the 19 plates were picked and expanded on 96-well feeder plate.

All of the 96-well plates were replicated and sib-selection was performed to determine the drug resistance of the clones. Cells on a 96-well feeder plate were split 1:5 onto 5X gelatinized 96-well tissue culture plates. These five plates were selected with M15, M15+G418, M15+puromycin, M15+HAT, and M15+blasticidin, respectively. Once most drug resistant clones on the plates grew to about 100% confluence, these plates were stained with 2% methylene blue in 70% ethanol, and drug resistance of each clone was scored. The correct recombinants should be G418 resistant, puromycin resistant, HAT resistant and blasticidin sensitive.

Genomic DNA was extracted and Southern analysis was performed using the *E_2DH* 3' probe and a *lacZ* probe (a 800 kb *Bam*HI-*Cla*I fragment from pWW239): 1) *Nde*I digestion, hybridized with the *E_2DH* 3' probe, the correct recombinants would only have the 9.6 kb targeted restriction fragment but no 13.1 kb wild-type restriction fragment; 2) *Eco*RI digestion, hybridized with the *lacZ* probe; 3) *Spe*I digestion, hybridized with the *lacZ* probe. All clones that are homozygous for the *E_2DH* locus presumably also carry homozygous mutations at the trapped locus. Individual trapping events were identified by their unique proviral/host junction generated by two different restriction enzyme digestions (*Eco*RI and *Spe*I)

All the homozygous clones from the 19 plates were grouped according to the sizes of their proviral junction fragments. For the groups that have more than one clone, at least 2 independent clones were expanded. For the groups that only have one clone, the clone was expanded. Genomic DNA and RNA were extracted from all the expanded clones. Southern analysis was carried out using different probes and enzyme digestions to confirm the clones and determine their genotypes: 1) *Eco*RI digestion, hybridized with the *E_2DH* 5' probe, the correct recombinants would only have the 9.2 kb targeted restriction fragment but not the 14.9 kb wild-type restriction fragment; 2) *Nde*I digestion, hybridized with the *E_2DH* 3' probe, the correct recombinants would only have the 9.6 kb targeted restriction fragment but not the 13.1 kb wild-type restriction fragment; 3) *Eco*RI digestion, hybridized with the *lacZ* probe; 4) *Spe*I digestion, hybridized with the *lacZ* probe; 5) *Kpn*I digestion, hybridized with the *lacZ* probe. The *Kpn*I digestion and hybridization using the *lacZ* probe was used to determine whether the clones carry homozygous inversions.

2.2.12 Gene trap mutagenesis using plasmid based vector

WW100 (WW69-D6 cells electroporated with pWW237):
20 μg *Sca*I linearized pWW237 DNA was electroporated into $1X10^7$ WW69-D6 ES cells. Ten electroporations were carried out and the cells were plated on ten 90-mm feeder plates. G418 selection (180 μg/ml) was initiated 24

hours after electroporation. The drug selection continued for 10 days when the G418 resistant colonies were clearly visible. The G418 resistant ES cell colonies from each of the 10 plates were trypsinized, resuspended in M15 medium and maintained as separate pools on 10X 90-mm feeder plates (WW100-1 to WW100-10). These cells were selected with G418 until they reached about 80% confluence. $1X10^7$ cells were used for the Cre-mediated inversions. The rest of the cells were frozen down for the stock.

WW104 (WW100 regional trapping):
20 μg of supercoiled pCAAG-Cre was electroporated into $1X10^7$ cells of WW100-1 to WW100-10 pools. Puromycin selection (3 μg/ml) was initiated 24 hours after the electroporation. The drug selection continued for 6 days until the colonies were visible under microscope and the colonies were transferred into M15 medium for another 4 days. The puromycin resistant ES cell colonies from each of the 10 plates were trypsinized, resuspended in M15 medium and maintained as 10 separate pools on 10X 90-mm feeder plates (WW104-1 to WW104-10). These cells were selected with puromycin until they reached about 80% confluence. $1X10^7$ cells were used for the Cre-induced mitotic recombination. The rest of the cells were frozen down for the stock.

WW106 (WW104 induced mitotic recombination):
20 μg of supercoiled pCAAG-Cre was electroporated into $1X10^7$ cells of the pools WW104-1 to WW104-10. HAT selection was initiated 24 hours after the electroporation. The drug selection continued for 6 days until the colonies were visible under microscope and the colonies were transferred into M15 medium with HT supplement for another 4 days. 48 HAT resistant ES cell colonies from each of the 19 plates were picked and expanded on 96-well feeder plate.

Sib-selection and Southern analysis were carried out in essentially the same way as the gene-trap mutagenesis using the retrovirus.

2.2.13 ES cell *in vitro* differentiation

Embryoid bodies were established and cultured as described before (Wobus, Guan et al. 2002). In brief, ES cells were grown on 90-mm or 6-well feeder plates until they reached 70-80% confluence. The cells were fed 2-3 hours before trypsinization. The plates were washed in PBS and trypsinized for 15 minutes. The cells were resuspended in M15 and counted using a Coulter Counter (Beckman). The cells were diluted in Differentiation Medium (Table 2-2) to a final concentration of 600 cells per 20 μl. 20 μl drops of ES cell suspension was placed on the bottom of 100-mm bacteriological Petri dishes. The bacteriological dishes were inverted (upside down) and the hanging drops of ES cell aggregates cultured at 37°C with 5% CO_2.

After two days (Day 2), 15 ml Differentiation Medium was put into each bacteriological dishes, and the aggregates were rinsed off the bottom into the media. The aggregates were cultured in suspension at 37°C with 5% CO_2. After another three days (day 5), the EBs from each dish were transferred into a 15-ml falcon tube. The EBs sedimented by gravity and the medium was discarded and replaced with Differentiation Medium supplemented with 10^{-8} M RA. The EBs were resuspended by inverting for several times and transferred to gelatinized 90-mm tissue culture plates. One plate of EBs were washed in PBS and used to extract RNA at day 5. The culture medium was changed every other day during the differentiation process. RNA samples were taken at various time points.

2.3 DNA methods

2.3.1 Probes

LacZ probe: A probe for gene-trap viruses containing the SA*βgeo* gene-trap cassette, consisting of a 1.4 kb *Cla*I fragment from pSA*βgeo,* a plasmid containing the *SAβgeo* cassette in pBS (from Dr. Philippe Soriano).

Neo probe: A probe for gene-trap viruses containing the SAβgeo gene-trap cassette and consisting of a 700 bp *Pst*I /*Xba*I fragment from the *PGK-Neo* cassette.

E$_2$DH 5' probe: A 2.1 kb *Nhe*I-*Not*I genomic fragment was cloned into *Xba*I-*Not*I digested pBS vector to make pL16 (Liu, Zhang et al. 1998). pL16 was cut with *Bgl*II and *Not*I, and an 1.7 kb fragment was gel purified to be used as *E$_2$DH* 5' probe.

E$_2$DH 3' probe: An 1.7 kb *Spe*I-*Sac*I genomic fragment was cloned into *Xba*I-*Sac*I digested pBS vector to make pL17 (Liu, Zhang et al. 1998). pL17 was cut with *Eco*RI and *Sac*I, and an 1.7 kb fragment was gel purified to be used as *E$_2$DH* 3' probe.

D11Mit71 3' probe: pBZ84 was isolated from a 3' *Hprt* lirary using a pair of *D11Mit71* specific primers (Zheng, Sage et al. 2000). pBZ84 was cut with *Asc*I and *Xho*I, and a 3.6 kb fragment was gel purified to be used as *D11Mit71* 3' probe.

Genomic probes:
Genomic DNA probes were PCR amplified from AB2.2 mouse genomic DNA and used for Southern-blot analysis. PCR products were routinely cloned into TOPO TA Cloning Vector (Invitrogen). The probes were made by digestion of the plasmid DNA using appropriate enzymes and gel purified. The concentration of the probe DNA was determined by spectrophotometer (Beckman) and/or gel electrophoresis.

D11Mit71 5' probe:
D11Mit71-5' probe-F
5'-CCC TAA CCA GGA TAG ATA CTG CTT GCT TTG TG-3'
D11Mit71-5' probe-R
5'-GCT TGG GGG TCA CTA CAA CTT GAA GAA CTG-3'

Pecam trapping probe:
Pecam-trapping-F
5'-CTG GCA CCT TTC TCC AGT GAA CCG TCC-3'
Pecam-trapping-R

5'-CCT CTG GCA TCA AGG AGG TCT TGG TCT G-3'

Acly 5' trapping probe:
Acly-5' Probe-F
GCTGCGTCAAGGAGTGGAGACCTATGG
Acly-5' Probe-R
GGCTGGGTACTGAA*CAG*TGTCCT*CAG*G

Acly 3' trapping probe:
Acly-3' Probe-F
GGCCTGACCTGGGGCTGATGGG
Acly-3' Probe-R
GGTACCTGTTAGACTGGGCGCTC*CAG*

2.3.2 Southern blotting and hybridization

2.3.2.1 Southern blotting

2-5 μg genomic DNA was digested with an appropriate restriction enzyme overnight. The digested fragments were separated by electrophoresis on 0.8% agarose gel. After electrophoresis, the gel was first soaked in Depurination Buffer (0.25 M HCl) for 10 minutes with gentle agitation, and then transferred into Denaturation Buffer (0.5 M NaOH, 1.5 M NaCl) for 1 hour with gentle agitation. A capillary blot was set up according to standard methods. Denaturation Buffer was used as the transfer buffer. Following overnight transfer, the blot was neutralized in Membrane Rinse Buffer (0.2 M Tris-Cl (pH7.4), 2X SSC) for 5 minutes, and baked at 80 °C for 1 hour.

2.3.2.2 Probe preparation

Probe DNA was labelled using Rediprime™ II Random Prime Labeling System (Amersham) according to the manufacturer's instructions. Briefly, 20 ng DNA was diluted in a final volume of 45 μl 1X TE buffer. The DNA sample was denatured by heating to 100 °C for 5 minutes and then placed on ice for another 5 minutes. The denatured DNA was added to a reaction tube. 5 μl Redivue [^{32}P] dCTP was added and the labelling solution was mixed

thoroughly by pippetting up and down. The tube was incubated at 37 °C for 10-30 minutes and the purified with a pre-filled G-50 column. The purified probe was denatured at 100 °C for 5 minutes, and chilled on ice for another 5 minutes before use.

2.3.2.3 Hybridization

Blots were pre-hybridized at 65 °C for at least one hour in Hybridization Buffer (1.5X SSCP, 1X Denhardts solution, 0.5% SDS, 10% Dextran Sulfate) supplemented with denatured herring sperm DNA. After pre-hybridization, the denatured probe was added and the blot was hybridized at 65°C overnight. The next day, the blot was first rinsed briefly in low stringency wash buffer (1X SSC, 0.1% SDS) at room temperature and then washed in high stringency wash buffer (0.5X SSC and 0.1% SDS) at 65 °C for 15 minutes. The blot was then exposed to X-ray film (Fuji).

2.3.3 Splinkerette PCR

2.3.3.1 Splinkerette adaptors preparation

Splinkerette PCR was carried out as described previously (Mikkers, Allen et al. 2002). 150 pmol of HMSpAa, 150 pmol of HMSpBb and 5 μl NEB Buffer 2 (New England Biolabs) were used to make a 100 μl oligonucleotide mixture. The mixture was denatured by heating to 95 °C for 3 minutes, and then annealed by slowly cooling to room temperature.

Splinkerette Oligos:
HMSpAa: 5'-CGA AGA GTA ACC GTT GCT AGG AGA GAC CGT GGC TGA ATG AGA CTG GTG TCG ACA CTA GTG G-3'
HMSpBb-*Sau*3AI
5'-gatc CCA CTA GTG TCG ACA CCA GTC TCT AAT TTT TTT TTT CAA AAA AA-3'
HMSpBb –*Xba*I
5'-ctag CCA CTA GTG TCG ACA CCA GTC TCT AAT TTT TTT TTT CAA AAA AA-3'
HMSpBb -*Eco*RI

5’-aatt CCA CTA GTG TCG ACA CCA GTC TCT AAT TTT TTT TTT CAA AAA AA-3’

2.3.3.2 Genomic DNA digestion and ligation with Splinkerette adaptors

2 μg of genomic DNA was digested with *Sau*3AI in a 30 μl volume at 37 °C for 3 hours. The *Sau*3AI enzyme was heat-inactivated by incubating at 65 °C for 20 minutes. For a 20 μl ligation mixture, 3 μl of the annealed Splinkerette adaptors, 5 μl digested genomic DNA, 2 μl 10X Ligation Buffer and 5 units T4 DNA Ligase (New England Biolabs) were added. The ligation mixture was incubated at 16 °C overnight. The T4 DNA ligase was heat-inactivated by incubating at 65 °C for 15 minutes. The ligation product was then digested with *Cla*I. For a 20 μl *Cla* I digestion mixture, 10 units *Cla*I, 4 μl 10X NEB Buffer 4 (New England Biolabs) were added. The digestion mixture was incubated at 37 °C for 2 hours. The *Cla*I enzyme was heat-inactivated by incubating at 65 °C for 20 minutes.

The digestion product was purified and desalted using Sephacryl^{TM}S-300 (Amersham). Briefly, Sephacryl^{TM}S-300 Media was mixed at a 1:1 ratio with MilliQ water. 200 μl of this mixture was added to each well of a 0.2 μm PVDF filtration plate (Corning) and spun for 2 minutes at 600 *g*. This step was repeated once. 200 μl of ddH_2O was added to each well of the filtration plate and spun for 2 minutes at 600 *g*. This step was repeated once. The digestion products were then loaded onto the Sephacryl^{TM}S-300-filled filtration plate. The purified products were collected by spinning for 2 minutes at 600 *g*.

To obtain provial/host flanking genomic fragments from as many clones as possible, genomic DNA was also digested with restriction enzyme *Eco*RI*, Xba*I*, Spe*I*, Nhe*I. The Splinkerette adapters were generated by annealing the HMSpAa with different HMSpBb oligos designed for different restriction enzymes. Because *Xba*I*, Spe*I*, Nhe*I digestion will generate the same 5’ overhang (3’-GATC-5’), these three enzymes were used to cut the genomic DNA at the same time.

2.3.3.3 First round PCR

The 5' LTR proviral flanking genomic fragments were amplified with the LTR specific primer, AB949new, and the Splinkerette primer, HMSp1. A 50 μl PCR system contains 20 μl purified ligation proucts, 1 μl AB949new (10 μM), 1 μl HMSp1 (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen), ddH_2O 20 μl. The hot-start PCR conditions were 94 °C 1.5 minutes; 2 cycles of 94 °C 1 minute, 68 °C 30 seconds, 72 °C 1 minutes; 30 cycles of 94 °C 30 seconds, 65 °C 30 seconds, 72 °C 2 minutes; 72 °C 10 minutes.

2.3.3.4 Second round PCR

The first round of PCR product was 1:100 diluted in ddH_2O and 5 μl of the diluted product was used as the template for the second round of nested PCR. A 50 μl PCR system contains 5 μl of the diluted 1st round PCR product, 1 μl HM001 (10 μM), 1 μl HMSp2 (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen), ddH_2O 35 μl. Hot-start PCR conditions were: 94 °C 1.5 minutes; 30 cycles of 94 °C 30 seconds, 60 °C 30 seconds, 72 °C 1.5 minutes; 72 °C 10 minutes. The nested PCR products were separated on a 1% agarose gel. The specific PCR fragments were gel purified using QIAquick Gel Extraction Kit (Qiagen) according to the manufacture's instructions.

2.3.3.5 Sequencing the splinkerette PCR products

Sequencing reactions were performed using ABI PRISM^{TM} Big DyeTerminator Cycle Sequencing Ready Reaction Kits (PE Applied Biosystems) according to the manufacturer's instructions. A 10 μl sequencing mix contains 5 μl gel purified PCR product, 1 μl of HM002 or HMSp3 primer (5 μM) and 4 μl Big Bye. The sequencing conditions were 94 °C for 1.5 minutes; 40 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 60 °C 4 minutes.

After the sequencing reaction, 10 μl of MilliQ water was added to each well of the 96-well plate. 50 μl of Precipitation Mix (100 ml 96% ethanol, 2 ml

Na_2OAC (3 M, pH 5.2), 4 ml EDTA (0.1 mM, pH 8.0)) was then added to each well. The precipitated sequencing products were collected by centrifugation at 4000 rpm at 4 °C for 25 minutes. The supernatant was discarded and the precipitates were washed with 100 µl of chilled 70% ethanol followed by centrifuging at 4000 rpm at 4 °C for 10 minutes. The ethanol was discarded and the samples were dried at 65°C for 2 minutes. The sequencing reactions were run on an ABI PRISM™ 3730 DNA sequencer (Perkin Elmer).

Splinkerette PCR primers:
AB949new: 5'-GCT AGC TTG CCA AAC CTA CAG GTG G-3'
HM001: 5'- GCC AAA CCT ACA GGT GGG GTC TTT-3'
HMSp1: 5'-CGA AGA GTA ACC GTT GCT AGG AGA GAC C-3'
HMSp2: 5'-GTG GCT GAA TGA GAC TGG TGT CGA C-3'

Splinkerette sequencing primers:
HM002: 5'-ACA GGT GGG GTC TTT CA-3'
HMSp3: 5'-GGT GTC GAC ACT AGT GG-3'

2.4 RNA methods

2.4.1 5' RACE

2.4.1.1 Total RNA extraction

Total RNA was extracted from ES cells grown on gelatinized 6-well tissue culture plate using RNAqueous™ Kits (Ambion) according to the manufacturer's protocol. 5 µg of total RNA was treated with 1 µl amplification grade DNase I (1 unit/µl, Invitrogen) in a 10 µl volume for 15 minutes at room temperature to eliminate the residual genomic DNA. After the DNase I treatment, 1 µl of EDTA (25 mM) was added to each reaction, and the reaction mixture was incubated at 65 °C for 15 minutes to heat-inactivate the DNase I.

2.4.1.2 First strand cDNA synthesis

3 µl *lacZ*-GSP1primer (10 µM, dissolved in DEPC-treated water) and 9 µl DEPC-treated water was added to the reaction to make up the volume to 25

μl. The RNA template was denatured by incubation at 65°C for 10 minutes and then placed on ice for 1 minute. 1 μl dNTPs (10 mM, Invitrogen), 10 μl 5X first-strand buffer, 5 μl DTT (0.1 M), 1 μl Superscript™ II (5 units/μl), 8 μl DEPC-treated water were added to denatured RNA template. The mixture was incubated at 50 °C for 1 hour. The retro-transcriptase was heat-inactivated by incubation at 70 °C for 15 min. After that, 1 μl of Ribonuclease H (2 U/μl, Invitrogen) was added. The mixture was incubated at 37 °C for 30 minutes to destroy the RNA template. The synthesized first strand cDNA was purified using QIAquick PCR purification kit (Qiagen). If first strand cDNA was synthesized on 96-well PCR plates, the samples were purified using Sephacryl™S-300 (Amersham) as described before.

lacZ-GSP1: 5'-GGG CCT CTT CGC TAT TAC GC-3'

2.4.1.3 TdT tailing

8 μl 5X TdT buffer, 2 μl dCTP (4 mM) and 1 μl TdT enzyme (Invitrogen) were added to 30 μl purified first strand cDNA. The samples were incubated for 10 minutes at 37 °C. After the reaction, the TdT enzyme was heat-inactivated by incubating the samples for 10 min at 65 °C.

2.4.1.4 First round PCR

The 5' RCAE products were amplified with the *lacZ* specific primer, *lacZ*-GSP2, and the 5' RACE Abridged Anchor Primer (AAP, Invitrogen). A 50 μl PCR system contains 10 μl purified dC-tailed cDNA, 1 μl *lacZ*-GSP2 (10 μM), 1 μl AAP (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen) and 30 μl ddH_2O. The hot-start PCR conditions were 94 °C 1.5 minutes; 35 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 72 °C 2 minutes; 72 °C 10 minutes.

2.4.1.5 Second round PCR

First-round PCR products were 1:100 diluted using ddH_2O. 5 μl of the diluted PCR product was used as the template for the second round of nested PCR. A 50 μl PCR system contains 5 μl diluted 1st round PCR product, 1 μl *lacZ*-

GSP3 (10 μM), 1 μl Abridged Universal Amplification Primer (AUAP, Invitrogen), 5 μl 10X PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 1 μl dNTP (25 mM), 0.5 μl Platinum Taq (5 units/μl, Invitrogen), 35 μl ddH_20. The hot-start PCR conditions were 94 °C 1.5 minutes; 35 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 72 °C 2 minutes; 72 °C 10 minutes. 10 μl of the nested PCR products were loaded on a 1.0 % agarose gel.

2.4.1.6 Sequencing the 5' RACE product

If the nested-PCR was performed on a small scale, the nested-PCR product was purified using QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions. If the nested-PCR was performed in a 96-well plate format, 10 μl of the nested-PCR product was treated with 1U each of Exonuclease I (Exo I, NEB) and Shrimp Alkaline Phosphatase (SAP, Amersham) for one hour at 37 °C to get rid of the unused primers and dNTPs. After the reaction, the mixture was incubated at 95 °C for 15 minutes to heat-inactivate the enzymes, and 5 μl was used for sequencing.

Sequencing reaction was performed using ABI PRISM™ Big DyeTerminator Cycle Sequencing Ready Reaction Kit (PE Applied Biosystems) according to the manufacturer's instructions. A 10 μl sequencing mix contains 5 μl purified PCR product, 1 μl of *SA*-seq primer (5 μM) and 4 μl Big Bye. The sequencing conditions were 94 °C for 1.5 minutes; 40 cycles of 94 °C 30 seconds, 55 °C 30 seconds, 60 °C 4 minutes.

After sequencing reaction, 10 μl of MilliQ water was added to each well of the 96-well plate. 50 μl of Precipitation Mix (100 ml 96% ethanol, 2 ml Na_2OAC (3 M, pH 5.2), 4 ml EDTA (0.1 mM, pH 8.0)) was then added to each well. The sequencing products were precipitated by centrifugation at 4000 rpm at 4 °C for 25 minutes. The supernatant was discarded and the precipitates were washed with 100 μl of chilled 70% ethanol followed by centrifuging at 4000 rpm at 4 °C for 10 minutes. The ethanol was discarded and the samples were dried at 65°C for 2 minutes. The sequencing reactions were run on an ABI PRISM™ 377 DNA sequencer (Perkin Elmer).

5’ RACE PCR primers:
lacZ-GSP2: 5’-ATG TGC TGC AAG GCG ATT AAG-3’
SA-GSP3: 5’-GTT GTA AAA CGA CGG GAT CCG CCA T-3’

5’ RACE sequencing primers:
SA-seq: 5’-TGTCAC AGA TCA TCA AGC TTA TC-3’

2.4.2 RT-PCR

2.4.2.1 First strand cDNA synthesis

Total RNA was prepared using an RNeasy® Mini Kit (Qiagen). The total DNA from each sample was quantified using Spectrophotometer (Beckman). 5 μg total RNA of each sample was used for each reaction, DEPC-treated water was added to each sample to bring up the final volume to 24 μl. 1 μl of Oligo-dT primer (10 μM) was added to each reaction. The RNA template was denatured by incubation at 65°C for 10 minutes and then placed on ice for 1 minute. 1 μl dNTPs (10 mM, Invitrogen), 10 μl 5X first-strand buffer, 5 μl DTT (0.1 M), 1 μl Superscript^{TM} II (5 units/μl, Invitrogen) and 8 μl DEPC-treated water were added to denatured RNA template. The retro-transcriptase was heat-inactivated by incubation at 70 °C for 15 min. After that, 1 μl of Ribonuclease H (2 U/μl, Invitrogen) was added. The mixture was incubated at 37 °C for 30 minutes to destroy the RNA template. The resultant cDNA was diluted at a ratio of 1:5 with ddH_2O and 5 μl was used for each PCR reaction.

2.4.2.2 RT-PCR

The first strand cDNA was amplified with the gene-specific primers designed. A 50 μl PCR system contains 5 μl diluted cDNA, 1 μl Forward Primer (10 μM), 1 μl Reverse Primer (10 μM), 5 μl 10x PCR buffer, 1.5 μl $MgCl_2$ (50 mM), 0.5 μl dNTPs (25 mM), 0.5 μl PlatinumTaq (5 units/μl, Invitrogen) and 35.5 μl ddH_2O. The hot-start PCR conditions were 94 °C 1.5 minutes; 25-35 cycles (depends on the primers) of 94 °C 30 seconds, 55-65 °C (depends on the primers) 30 seconds, 72 °C 1 minute; 72 °C 10 minutes.

Oligo-dT primer: GGC CAC GCG TCG ACT AGT AC $(T)_{17}$

Other germ layer and cell lineage specific marker: Table 2.4

Table 2.4: RT-PCR primers used for *in vitro* differentiation assay.

Gene Name	Forward Primer	Reverse Primer	Annealing Temperature	Length of PCR product	Reference
Alk-3	TCACCGAAAGCCCAGCTACG	TCACCGAAAGCCCAGCTACG	55°C	700 bp	Wiles MV, Johansson BM. *Exp Cell Res.* 1999 Feb 25;247(1):241-8.
Brachyury	ATGCCAAAGAAAGAAACGAC	AGAGGCTGTAGAACATGATT	55°C	836 bp	
Fyn	CAACCGGGGAAACTGGTTAC	GCTCATGTACTCCGTGACGA	55°C	645 bp	
Goosecoid	GCACCATCTTCACCGATGAG	AGGAGGATCGCTTCTGTCGT	55°C	179 bp	
Nodal	TCACCGTCCCCTCTGGCGTA	ACTCCTCCCCCACAGGGTTA	60°C	773 bp	
Noggin	TGGCGGCCGCCTTCCCAAGT	AGCCCGGGGGATCCATCAAG	60°C	365 bp	
Pax-6	CAGTCACAGCGGAGTGAATC	CGCTTCAGCTGAAGTCGCAT	55°C	658 bp	
Scleraxis	GTGGACCCTCCTCCTTCTAATTCG	GACCGCACCAACAGCGTGAA	63°C	375 bp	Kramer J, Hegert C, Guan K, Wobus AM, Muller PK, Rohwedel *J. Mech Dev.* 2000 Apr;92(2):193-205.
Pax-1	TTCTCGGTGTTTGAAGGTCATTGCCG	GATGGAAGACTGGGCGGGTGTGAA	60°C	318 bp	
Sox-9	TCTTTCTTGTGCTGCACGCGC	TGGCAGACCAGTACCCGCATCT	57°C	135 bp	
Aggrecan	TCCTCTCCGGTGGCAAAGAAGTTG	CCAAGTTCCAGGGTCACTGTTACCG	60°C	270 bp	
Collagen II	AGGGGTACCAGGTTCTCCATC	CTGCTCATCGCCGCGGTCCTA	60°C	432 bp (splice variant A) and 225 bp (splice variant B)	
b-Tubulin	GGAACATAGCCGTAAACTGC	TCACTGTGCCTGAACTTACC	54°C	317 bp	
HPRT	GCCTGTATCCAACACTTCG	AGCGTCGTGATTAGCGATG	63°C	507 bp	
PECAM	GTCATGGCCATGGTCGAGTA	CTCCTCGGCATCTTGCTGAA	55°C	260 bp	Vittet D, Prandini MH, Berthier R, Schweitzer A, Martin-Sisteron H, Uzan G, Dejana E. *Blood.* 1996 Nov 1;88(9):3424-31.
Flk-1	TCTGTGGTTCTGCGTGGAGA	GTATCATTTCCAACCACCCT	55°C	269 bp	
Tie-1	CCTTCCTACCTGCTA	CCACTACACCTTTCTTTACA	55°C	441 bp	
Tie-2	CTCACTGCCCTCCTGACTGG	CGATGTACTTGGATATAGGC	55°C	228 bp	
VE-Cadherin	GGATGCAGAGGCTCACAGAG	CTGGCGGTTCACGTTGGACT	55°C	226 bp	
HPRT	GCTGGTGAAAAGGACCTCT	CACAGGACTAGAACACCTGC	55°C	248 bp	
a-Cardiac myosin heavy chain	CTGCTGGAGAGGTTATTCCTCG	GGAAGAGTGAGCGGCGCATCAAGG	64°C	301 bp	Fassler R, Rohwedel J, Maltsev V, Bloch W, Lentini S, Guan K, Gullberg D, Hescheler J, Addicks K, Wobus. *AM.J Cell Sci.* 1996 Dec;109 (Pt 13):2989-99.
b-Cardiac myosin heavy chain	TGCAAAGGCTCCAGGTCTGAGGGC	GCCAACACCAACCTGTCCAAGTTC	64°C	205 bp	
Myosin light chain isoform 2V	TGTGGGTCACCTGAGGCTGTGGTTCAG	GAAGGCTGACTATGTCCGGGAGATGC	64°C	189 bp	
Atrial natriuretic factor	TGATAGATGAAGGCAGGAAGCCGC	AGGATTGGAGCCCAGAGTGGACTAGG	64°C	203 bp	
Cardiac-specific a1-subunit of the L-type calcium channel	GTTCCTGAAGGAGGTGTGCTGGACG	AAAGGCAGTTCCCATGCCGG	62°C	183 bp	
Skeletal muscle-specific a1-subunit of the L-type calcium channel	GATCACCAGCCAATAGAAGACC	GGCGAGGTCATGGACGTGGACG	62°C	200 bp	
b-Tubulin	GGAACATAGCCGTAAACTGC	TCACTGTGCCTGAACTTACC	64°C?	317 bp	
NFL	CCAGGAAGAGCAGACAGAGGT	GTTGGGAATAGGGCTCAATCT	59°C	302 bp	Rohwedel J, Kleppisch T, Pich U, Guan K, Jin S, Zuschratter W, Hopf C, Hoch W, Hescheler J, Witzemann V, Wobus AM. *Exp Cell Res.* 1998 Mar 15;239(2):214-25.
NFH	AGGACCGTCATCAGGCAGACATTGC	CTTCTGTCACTCCTTCCGTCACCCG	59°C	368 bp	
Synaptophysin	TACCGAGAGAACAACAAAGGGC	GCCTGTCTCCTTGAACACGAAC	60°C	287 bp	
Tau	CCGCACTCCCCCTAAGTCACCATC	TGCCGTGGAGATGTGTCCCCAGAC	60°C	440 bp(splice variant A) and 578 bp (splice variant B)	
S-laminin	TGGCTGTCTACCTGGCATCTGG	GCGACCACCATCTTGAGAACCC	58°C	187 bp	
AChR e-unit	ATTTCTGGCTTGGTGCTGCTCGC	GAGTCGTTGGCGTCCTCAAAGATACG	59°C	246 bp	

Table 2.4 (cont): RT-PCR primers used for *in vitro* differentiation assay.

Gene Name	Forward Primer	Reverse Primer	Annealing Temperature	Length of PCR product	Reference
68 kDa neurofilament protein (=NFL)	CCAGGAAGAGCAGACAGAGGT	GTTGGGAATAGGGCTCAATCT	59°C	302 bp	Rohwedel J, Guan K, Zuschratter W, Jin S, Ahnert-Hilger G, Furst D, Fassler R, Wobus AM. *Dev Biol.* 1998 Sep 15;201(2):167-84.
200 kDa neurofilament protein (=NFH)	AGGACCGTCATCAGGCAGACATTGC	CTTCTGTCACTCCTTCCGTCACCCG	59°C	368 bp	
Synaptophysin	TACCGAGAGAACAACAAAGGGC	GCCTGTCTCCTTGAACACGAAC	60°C	287 bp	
Tau	CCGCACTCCCCCTAAGTCACCATC	TGCCGTGGAGATGTGTCCCCAGAC	60°C	440 bp(splice variant A) and 578 bp (splice variant B)	
Brachyury	GAGAGAGAGCGAGCCTCCAAAC	GCTGTGACTGCCTACCAGAATG	59°C	230 bp	
Pax-6	GCTTCATCCGAGTCTTCTCCCTTAG	CCATCTTTGCTTGGGAAATCCG	59°C	312 bp	
Mash-1	CTCGTCCTCTCCGGAACTGATG	CGACAGGACGCCCGCCTGAAAG	62°C	301 bp	
BMP-4	ATTCTCTGGGATGCTGCTGAGG	CCGAGCCAACACTGTGAGGAGT	59°C	114 bp	
Wnt-1	GATTGCGAAGATGAACGCTGTTTC	TCCTCCACGAACCTGTTGACGG	54°C	266 bp	
S-laminin	TGGCTGTCTACCTGGCATCTGG	GCGACCACCATCTTGAGAACCC	58°C	187 bp	
b-Tubulin	GGAACATAGCCGTAAACTGC	TCACTGTGCCTGAACTTACC	60°C	317 bp	
HPRT	GCCTGTATCCAACACTTCG	AGCGTCGTGATTAGCGATG	59°C	507 bp	
B7-1	ATGGCTTGCAATTGTCAGTT	ATCAGGAGGGTCTTCTGGGGGT	55°C	728 bp	Ling V, Munroe RC, Murphy EA, Gray GS. *Exp Cell Res.* 1998 May 25;241(1):55-65.
B7-2	ATCGCCAACTTCAGTGAACC	TCTCACTGCCTTCACTCTGC	55°C	525 bp	
CD28	ACTCAGGCTGCTGTTCTTGG	TCGTTGTCTAGGTAAGGCGG	55°C	375 bp	
CTLA-4	CACAACACTGATGAGGTCCG	TGAGTTCCACCTTGCAGAGG	55°C	210 bp	
b-Actin	GTCGTCGACAACGGCTCCGGCATGTG	CATTGTAGAAGGTGTGGTGCCAGAT	55°C	252 bp	
K18	TTGTCACCACCAAGTCTGCC	TTTGTGCCAGCTCTGACTCC	60°C	213 bp	Bagutti C, Wobus AM, Fassler R, Watt FM. *Dev Biol.* 1996 Oct 10;179(1):184-96.
K14	GTGTCCACTGGCGATGTGAACGTGG	GCTGCCGCAGTAGCGACTCTACTGT	60°C	330 bp	
K10	CGCAAGGATGCTGAAGAGTGGTTC	TGGTACTCGGCGTTCTGGCACTCGG	60°C	278 bp	
Involucrin	GGTGTACAGAAGCTTCCAAGATGTCC	GGCATTGTGTAGGATGTGGAGTTGG	60°C	150 bp	
actin	GTTTGAGACCTTCAACACCCC	GTGGCCATCTCCTGCTCGAAGTC	60°C	320 bp	
Cx40	CCACGGAGAAGAATGTCTTCA	TGCTGCTGGCCTTACTAAGG	N.D.	447 bp	Oyamada Y, Komatsu K, Kimura H, Mori M, Oyamada M. *Exp Cell Res.* 1996 Dec 15;229(2):318-26.
Cx43	TGGGGGAAAGGCGTGAG	CTGCTGGCTCTGCTGGAAGGT	N.D.	1.3 kb	
Cx45	ATCATCCTGGTTGCAACTCC	CTCTTCATGGTCCTCTTCCG	N.D.	168 bp	
MHC-a	CTGCTGGAGAGGTTATTCCTCG	GGAAGAGTGAGCGGCGCATCAAGG	N.D.	302 bp	
MHC-b	TGCAAAGGCTCCAGGTCTGAGGGC	GCCAACACCAACCTGTCCAAGTTC	N.D.	205 bp	
MLC-2V	GCCAAGAAGCGGATAGAAGG	CTGTGGTTCAGGGCTCAGTC	N.D.	499 bp	
TTR	CCTCTGATGGTCAAAGTCCTGGATGCTG	CCTGGTCCTCCTGGGCTGAGTCTCTC	60°C	379 bp	Wei Wang, unpublished data
AFP	GGACATTTGTGTATAAGGAATGAAGCAAGCCC	GCAGTTACAGTTAAGCCAAAAGGCTCACACC	60°C	463 bp	
AFM	GGCAGCCCCTCAGCTCCCCAT	GGACTGAAACAGGACTCAGGCTCCTGC	60°C	465 bp	
HNF1	GGCCTCCACTCAGGCACAGAGCG	CGGAACTCTGATACAACACCAGGCTGC	60°C	488 bp	
vHNF1	CGGGAGGAGGACTGCCTCCCG	CAGGGCCCTCTCCTGGGCTCCC	60°C	574 bp	
HNF4	CGGCTGCGGTCACAGGTGCAAG	AGGTGCTCTTCTGAGGGTATGAGCCAGC	60°C	554 bp	
HNF3b	GGCCAGACCACGCGAGTCCTACG	TGAGCCGCTCATGCCCGCCAT	60°C	377 bp	
Oct3/4	GTGTAAGCTGCGGCCCCTGCTGG	GCTTCCATAGCCTGGGGTGCCAAAGT	60°C	388 bp	
EKLF	CTGGGACCTGGGACTGTGGCCAC	GGCCCATCTTTTGGGATACGGTCC	60°C	449 bp	
FGF5	CTGCAGAGTGGGCATCGGTTTCC	GCTCGGACTGCTTGAACCTGGGTAGG	60°C	396 bp	
GATA4	GGACACTACCTGTGCAATGCCTGTGG	GACAGGAGATGCATAGCCTTGTGGGG	60°C	532 bp	
Msx3	CCACACAGAGCACGGACCACTCCCCT	CCTGAGCCTAACCAAGAGGTTAGGGCTT	60°C	598 bp	
Nkx2.5	GCCTGGCGTCTGGGGACCTGTCTG	TGCGCGTGGTCTCTCGGCGC	60°C	327 bp	
PECAM	CCAGTGCAGAGCGGATAATTGCCATTCC	TAAGGTGGCGATGACCACTCCAATGAC	60°C	428 bp	

3 Library construction

3.1 Introduction

The aim of my project is to generate a panel of homozygous mutations for a recessive genetic screen. As I described in the first chapter, there are several different mutagenesis methods to create mutations in ES cells, and each of them has their own advantages and disadvantages. The bottleneck for all the gene-targeting based methods is that they require the labour-intensive and time-consuming generation of the targeting vectors as well as genotyping the targeting events. For all the random mutagenesis methods, although they can mutate the first allele of a gene efficiently, it is hard to disrupt the second allele and make the mutation homozygous. So a new method was needed to make the generation of homozygous mutations high-throughput, so that more focus can be put on the design of sophisticated functional genetic screens.

The original design of my project was: 1) create a cell line for induced mitotic recombination on chromosome 11; 2) target some important developmental genes in this cell line by homologous recombination; 3) induce mitotic recombination by transient expression of Cre to generate homozygous mutant ES cell clones; 4) use ES cell *in vitro* differentiation to study the function of these genes. Though this strategy still requires the genotyping of each locus that has been disrupted, it effectively saves one step of targeting and genotyping. Since I chose to use *E. coli* recombination to make gene-targeting constructs, which allows longer homology arms and has less chance of introducing point mutations, it is possible to achieve high targeting efficiency (Copeland, Jenkins et al. 2001; Liu, Jenkins et al. 2003). By calculation, if the targeting efficiency of a construct is 10%, then half of these (5%) will happen on the same chromosome as the 3' *Hprt* cassette, and 60% of the HAT resistant clones (3%) will carry homozygous mutations (Liu, Jenkins et al. 2002). So in principle, it is possible to omit the genotyping of the first targeting event, pool the transformants and induce mitotic recombination directly. The homozygous clones can be efficiently recovered from the HAT resistant clones.

During the process of making the starting cell line for induced mitotic recombination, Meredith Wentland, a former PhD student in Dr. Allan Bradley's lab, successfully isolated gene-trap insertions in a chromosome-specific way using a method called "regional trapping". Given this observation, we decided to combine the two methods (induced mitotic recombination and regional trapping) to generate random homozygous gene-trap mutations in a chromosome-specific way and thus completely save the trouble of constructing different targeting vectors and designing genotyping strategies for targeting of each different locus.

Since both studies have been carried out on mouse chromosome 11 (Liu, Jenkins et al. 2002), all the targeting vectors are readily available and the efficiencies of induced mitotic recombination and regional trapping on chromosome 11 are already known, we decided to perform this experiment still on chromosome 11. The design of my project was changed accordingly to: 1) create a cell line for induced mitotic recombination on chromosome 11; 2) perform random mutagenesis in this cell line by gene-trapping; 3) induce inversion by transient expression of Cre to select for trapping events on chromosome 11; 4) induce mitotic recombination by transient expression of Cre to generate homozygous mutant ES cell clones; 5) screen these ES cell clones by an *in vitro* differentiation assay to study the function of trapped genes.

Some modifications of Liu and Wentland's original experiment design needed to be made to suit my purpose. For example, both experiments use Cre/*lox*P system, and it is necessary to use different *lox* sites for the two steps so that they will not interfere with each other. There are also considerations about the trapping strategy: 1) use 5' trapping or 3' trapping; 2) use electroporation-based trapping or retrovirus-based trapping.

3.1.1 5' trapping versus 3' trapping

In Wentland's original work, she used the 3' trapping (polyA trap) strategy. The bioinformatic analysis of the trapped sequences generated in her experiment has shown that half of the 3' RACE sequences did not match any

known genes, and therefore termed “novel”. Some of these RACE sequences were shown to be transcribed in normal tissues, suggesting that these proviral integrations were located in transcribed genes. More detailed analysis has shown that the other sequences are not functional genes, consisting of processed pseudogenes and partially duplicated genomic sequences. Moreover, there are very few papers describing 3’ trapping strategies and the alleles generated by them. So it is difficult to assess the mutagenicity of 3’ trapping vectors. The transcripts initiated in the gene-trap vector can sometimes capture cryptic splice acceptors and PolyA signals downstream of the integration sites of the trapping vector. Also, if alternative terminal exons are trapped, only the transcripts that utilize the trapped exons will be disrupted. It is reasonable to predict that some of the trapped clones might not disrupt any functional genes and thus will not cause any phenotype. But on the other hand, 3’ trapping does not require the trapped genes to express in the undifferentiated ES cells which increases the possibility of the trapping of genes that express later in the developmental process.

On the other hand, most of the gene-trap mutagenesis experiments in mouse ES cells published so far use 5’ trapping (promoter trap) strategy. Considerable amount of data has been accumulated so that comparisons of the relative efficiency of different 5’ trapping constructs can be made based on statistical analysis of the trapped loci (Hansen, Floss et al. 2003). Also the mutagenicity of 5’ trapping strategy has been well documented and a significant portion of the genes trapped by promoter trap vectors were found to express in early embryogenesis (Gajovic, Chowdhury et al. 1998). In many cases, disruption of these genes will cause severe phenotypes *in vivo*. However, the nature of the promoter gene-trap design restricts 5’ trapping to those genes expressed in undifferentiated ES cells. Also, some genes that express at low levels are unlikely to be trapped.

So at the beginning stage of my project, vectors for both 5’ trapping and 3’ trapping have been constructed and functionally tested in ES cells. For reasons that will be discussed later, the 5’ trapping was chosen to develop the final strategy.

3.1.2 Electroporation versus retroviral vector

Trapping vectors can be introduced into the genome by either electroporation or retroviral infection. As I discussed before, both methods have their advantages and limitations.

Electroporation of a linearized gene-trap vector directly into mammalian cells is both simple and highly reproducible. There is almost no limitation on the size and structure of the vector, which allows flexibility for sophisticated vector design. But the structure of the mutant alleles created by this method is unpredictable. The integrations are always accompanied by DNA concatermerization. Multiple copies of the gene-trap vector in one locus can make the identification of the mutations difficult.

Gene trap mutagenesis using a retroviral vector results in a single copy of the retrovirus integrating into one genomic locus, which makes the cloning of the virus insertion site by PCR based methods reliable. Retroviruses have a tendency to integrate into 5' portion of a gene and thus are more likely to generate null alleles. However the packaging size of a retrovirus is limiting, and the packaging efficiency will drop significantly with an increase in the size of the exogenous DNA insert. Also, the virus preparation and transfection process are complicated and time-consuming compared to the electroporation method. Finally, it is not easy to predict which sequences can be efficiently packaged.

So at the beginning stage of my project, vectors for both electroporation-based trapping and retrovirus-based trapping have been constructed and functionally tested in ES cells. For the reasons that will be discussed later, the retrovirus-based trapping was selected.

3.2 Results

3.1.1 Construction of the inducible mitotic recombination cell line

3.2.1.1 Modification of the inducible mitotic recombination cassettes

The two original inducible mitotic recombination cassettes, pL330 (Multi *lox*-*Hprt* ∆5') and pL341 (Multi *lox*-*Hprt* ∆3'), contain *Neo* and *Puro* selection markers flanked by three *lox* site variants: *lox*5171, *lox*2272 and either *lox*66 (Multi *lox*-*Hprt* ∆3') or *lox*71 (Multi *lox*-*Hprt* ∆5') sites (Fig. 3-1) (Liu, Jenkins et al. 2002). A wild-type *lox*P site will be generated by site-specific recombination between *lox*66 and *lox*71 sites (Fig. 3-2a), which might recombine with the wild-type *lox*P sites used for regional trapping. Therefore *lox*66 or *lox*71 need to be deleted from the cassettes (Fig. 3-2b). Two pairs of oligonucleotides were used to generate the construct, pWW37 (Fig. 3-2c), which has *lox*5171, *lox*2272 and FRT sites flanked by several convenient enzyme cutting sites for the future cloning steps. The modified *lox* sites were ligated with *Neo*, *Puro*, 5' *Hprt* and 3' *Hprt* cassettes to make the modified inducible mitotic recombination cassettes, pWW48 (multi *lox* sites flanked *Puro*-3' *Hprt*) and pWW49 (5' *Hprt*-multi *lox* sites flanked *Neo*).

To fully sequence the two modified inducible mitotic recombination cassettes, pWW48 and pWW49 plasmids were digested with various restriction enzymes and subcloned into pBS. The sequencing results confirmed the intactness of the *lox* sites flanking the selection cassettes.

The *Neo*, *Puro*, 5' *Hprt* and 3' *Hprt* cassettes in pWW48 and pWW49 were functionally tested in ES cells. The linearized pWW48 and pWW49 plasmids were co-electroporated into AB2.2 ES cells lines with or without supercoiled pCAAG-Cre plasmid. The transformants were selected in M15 supplemented with HAT, G418 or puromycin, respectively. HAT resistant colonies were only generated when the two cassettes were co-electroporated with the Cre expression plasmid. Puromycin and G418 resistant colonies were generated by pWW48 and pWW49 as expected. The function test has proved the intactness of all the cassettes.

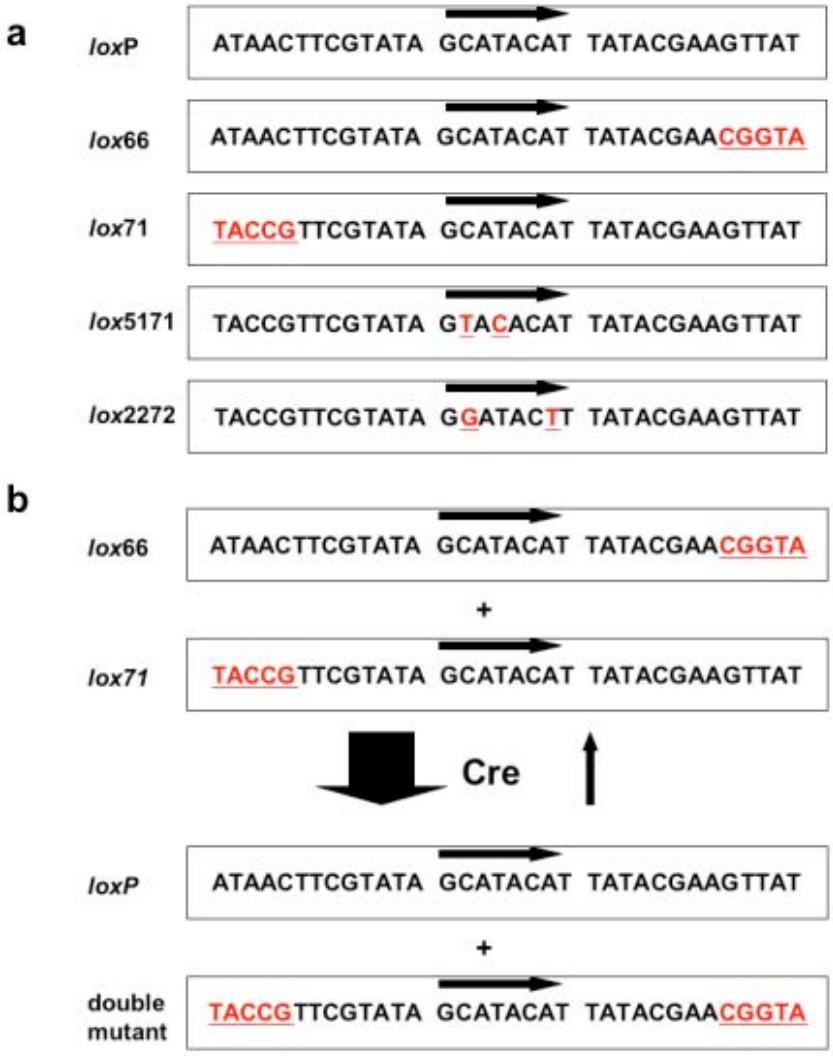


Fig. 3-1 *lox* sites used in the experiment. a. Nucleotide sequences of *lox* sites used in the experiment. The mutated nucleotides in the *lox* variants are underlined and marked as red. **b.** Site-spacific recombination between *lox*66 and *lox*71 sites. The Cre-mediated recombination between *lox*66 and *lox*71 will create a wild type *lox*P site, as well as a double mutant *lox* site.

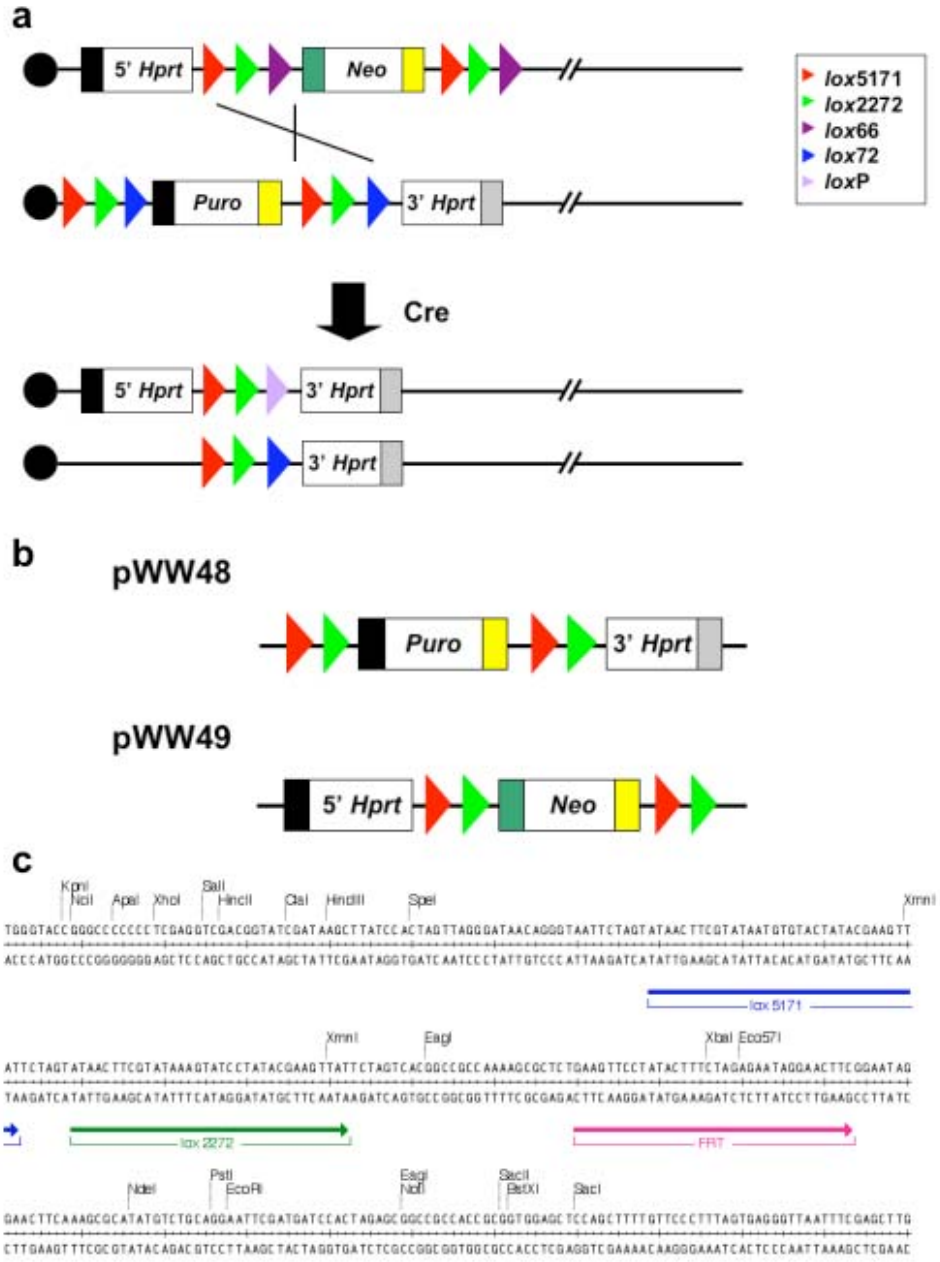

a
5' Hprt
Neo
Puro
3' Hprt
lox5171
lox2272
lox66
lox72
loxP
Cre
5' Hprt
3' Hprt
3' Hprt
b
pWW48
Puro
3' Hprt
pWW49
5' Hprt
Neo
c
KpnI
NcoI
ApaI
XhoI
SalI
HincII
ClaI
HindIII
SpeI
XmnI
TGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTATCCACTAGTTAGGGATAACAGGGTAATTCTAGTATAACTTCGTATAATGTGTACTATACGAAGTT
ACCCATGGCCCGGGGGGGAGCTCCAGCTGCCATAGCTATTCGAATAGGTGATCAATCCCTATTGTCCCATTAAGATCATATTGAAGCATATTACACATGATATGCTTCAA
lox5171
XmnI
EagI
XbaI
Eco57I
ATTCTAGTATAACTTCGTATAAAGTATCCTATACGAAGTTATTCTAGTCACGGCCGCCAAAAGCGCTCTGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCGGAATAG
TAAGATCATATTGAAGCATATTTCATAGGATATGCTTCAATAAGATCAGTGCCGGCGGTTTTCGCGAGACTTCAAGGATATGAAAGATCTCTTATCCTTGAAGCCTTATC
lox2272
FRT
NdeI
PstI
EcoRI
EagI
NotI
SacII
BstXI
SacI
GAACTTCAAAGCGCATATGTCTGCAGGAATTCGATGATCCACTAGAGCGGCCGCCACCGCGGTGGAGCTCCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTG
CTTGAAGTTTCGCGTATACAGACGTCCTTAAGCTACTAGGTGATCTCGCCGGCGGTGGCGCCACCTCGAGGTCGAAAACAAGGGAAATCACTCCCAATTAAAGCTCGAAC

Fig. 3-2 Induced mitotic recombination cassettes. a. Induced mitotic recombination cassettes described in Liu et al. (2002). The multi *lox*-5' *Hprt* cassette contains a *PGK* promoter (black box), 5' half of the *Hprt* minigene, tandem *lox* site variants, a *Pol*II promoter(green box) that drives *Neo* expression, and a bovine growth hormone polyA site (yellow box). The multi *lox*-3' *Hprt* cassette contains a *PGK* promoter (black box) that drives *Puro* expression, a bovine growth hormone polyA site (yellow box), tandem *lox* site variants, 3' half of the *Hprt* minigene and an SV40 polyA site (gray box). Arrows of different colour denote variant *lox* sites. Note that Cre-induced mitotic recombination reconstitutes a wild type *Hprt* minigene and leaves a wild type *lox*P site (purple arrow) on the same chromosome as the 3' *Hprt* cassette. **b.** Modified induced mitotic recombination cassettes with multi *lox* sites. The structure of the modified cassettes is essentially the same as the original cassettes, except that the *lox*66 and *lox*71 sites were deleted. These mutant *lox* sites will not recombine with each other or with wild type *lox*P site. **c.** Sequence of the tandem *lox* sites. Two pairs of oligos were used to construct the tandem *lox* sites. The *lox* sites were used to flank the *Puro* and *Neo* selection markers so that the selection markers can be popped-out afterwards. The tandem *lox* sites can also increase the efficiency of the induced mitotic recombination.

3.2.1.2 Targeting of the inducible mitotic recombination cassettes

After all the selection markers and the *lox* sites in pWW48 and pWW49 were confirmed to be functional *in vivo*, the 5' *Hprt* and 3' *Hprt* cassettes were cloned into pL325 to make the final targeting vectors, pWW74 (3' *Hprt*) and pWW75 (5' *Hprt*). The two vectors were linearized and co-electroporated into AB2.2 ES cells with or without pCAAG-Cre plasmid to confirm the functionality of the selection cassettes. HAT resistant colonies were only recovered from the co-transformation with the Cre expression plasmid, which proved that the two cassettes were functional (data not shown).

Linearized pWW74 was electroporated into AB2.2 ES cells and the transformants were selected with puromycin and FIAU. The correctly targeted clones were identified by Southern analysis using a *D11Mit71* 3' probe. The expected sizes of the detected restriction fragments were 14.1 kb for the wild-type allele and 10.3 kb for the targeted allele. The correctly targeted clones were expanded and confirmed by Southern analysis using a *D11Mit71* 5' probe. The expected sizes of detected restriction fragments were 7.1 kb for the wild-type allele and 4.9 kb for the targeted allele when digested with *Xba*I (Fig. 3-3). The targeted clones were named as WW14.

Linearized pWW75 was electroporated into AB2.2 ES cells and the transformants were selected with G418 and FIAU. The correctly targeted clones were identified by Southern analysis using a *D11Mit71* 3' probe. The expected sizes of detected restriction fragments were 14.1 kb for the wild-type allele and 17.9 kb for the targeted allele. The correctly targeted clones were expanded and confirmed by Southern analysis using a *D11Mit71* 5' probe. The expected sizes of detected restriction fragments were 7.1 kb for the wild-type allele and 6.4 kb for the targeted allele when digested with *Xba*I (Fig. 3-4). The targeted clones were named as WW16.

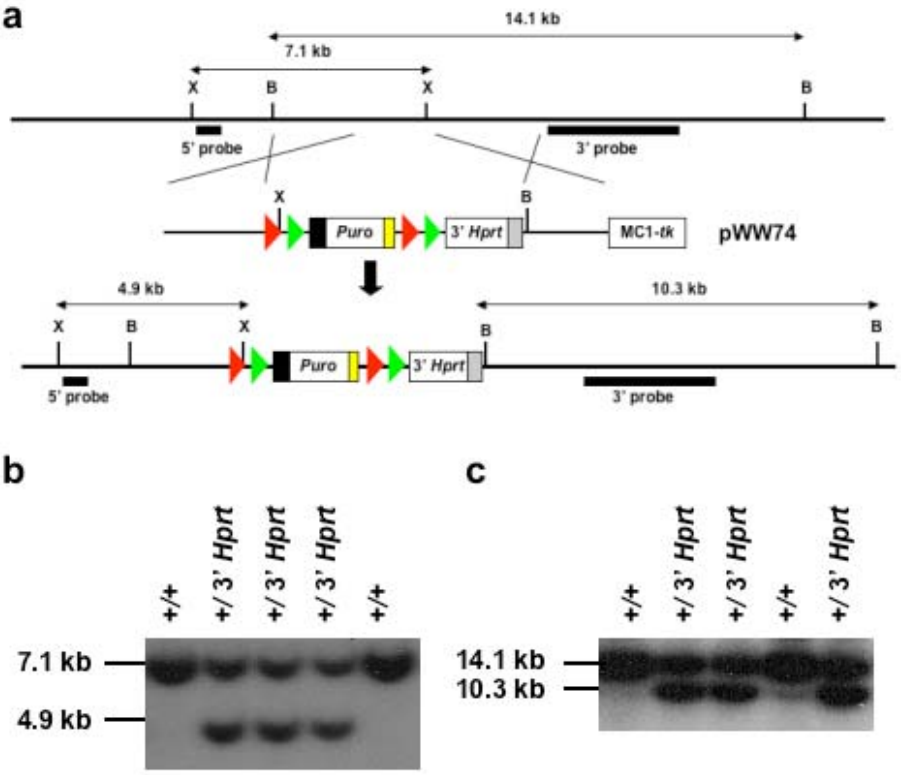


Fig. 3-3 Targeting 3' *Hprt* cassette to *D11Mit71* locus. a. Schematic illustration of targeting 3' *Hprt* cassette to *D11Mit71* locus. A 2.1 kb genomic fragment at *D11Mit71* locus was replaced by the 3' *Hprt* cassette (pWW48) to make the 3' *Hprt* targeting vector (pWW74). The MCI-*tk* cassette was used as the negative selection marker. Arrows of different colours denote multi *lox* sites used to flank the *Puro* selection cassettes. B, *Bam*HI; X, *Xba*I. **b.** Southern analysis of the WW14 cell line using a *D11Mit71* 5' probe. Genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, the detected restriction fragments were 7.1 kb for the wild type allele and 4.9 kb for the targeted allele. **c.** Southern analysis of the WW14 cell line using a *D11Mit71* 3' probe. Genomic DNA was digested with *Bam*HI and hybridized with a *D11Mit71* 3' probe, the detected restriction fragments were 14.1 kb for the wild type allele and 10.3 kb for the targeted allele.

To confirm the pluripotency of the cell lines, several subclones from WW14 and WW16 cell lines were injected into C57^{TyrBrdC1} blastocysts and germline transmissions were obtained for both of them. Germline transmission of the targeted alleles was confirmed by Southern analysis using the restriction enzymes and probes described above. Mice carrying the targeted alleles were crossed to a Cre expression mouse line (Su, Mills et al. 2002) and a FLP-expressing mouse line (Farley, Soriano et al. 2000; Su, Mills et al. 2002). The multi *lox* sites flanked *Neo* and *Puro* cassettes can be popped out in mice which carry both the inducible mitotic recombination cassettes and the Cre or Flp transgene. The two mouse lines can be crossed together to generate genetic mosaics *in vivo* by induced mitotic recombination (data not shown).

To make a cell line that carries both the 5' *Hprt* and the 3' *Hprt* cassettes targeted to allelic positions at the *D11Mit71* locus on chromosome 11, linearized pWW75 was electroporated into WW14. The transfectants were selected with G418 and FIAU and the correctly targeted clones were identified by Southern analysis using a *D11Mit71* 3' probe. The double-targeted clones were expected to have 10.3 kb (3' *Hprt* targeting) and 17.9 kb (5' *Hprt* targeting) *Bam*HI restriction fragments. The correctly targeted clones were expanded and confirmed by Southern analysis using a *D11Mit71* 5' probe. The double-targeted clones were expected to have 4.9 kb band (3' *Hprt* targeting) and 6.4 kb band (5' *Hprt* targeting) *Xba*I restriction fragments. The double-targeted ES cell clones were named as WW24 (Fig. 3-5a & b). Similarly, linearized pWW74 was electroporated into WW16 to make the double-targeted cell line WW25. The Southern screening strategy and the sizes of the detected restriction fragments are the same as described for WW24.

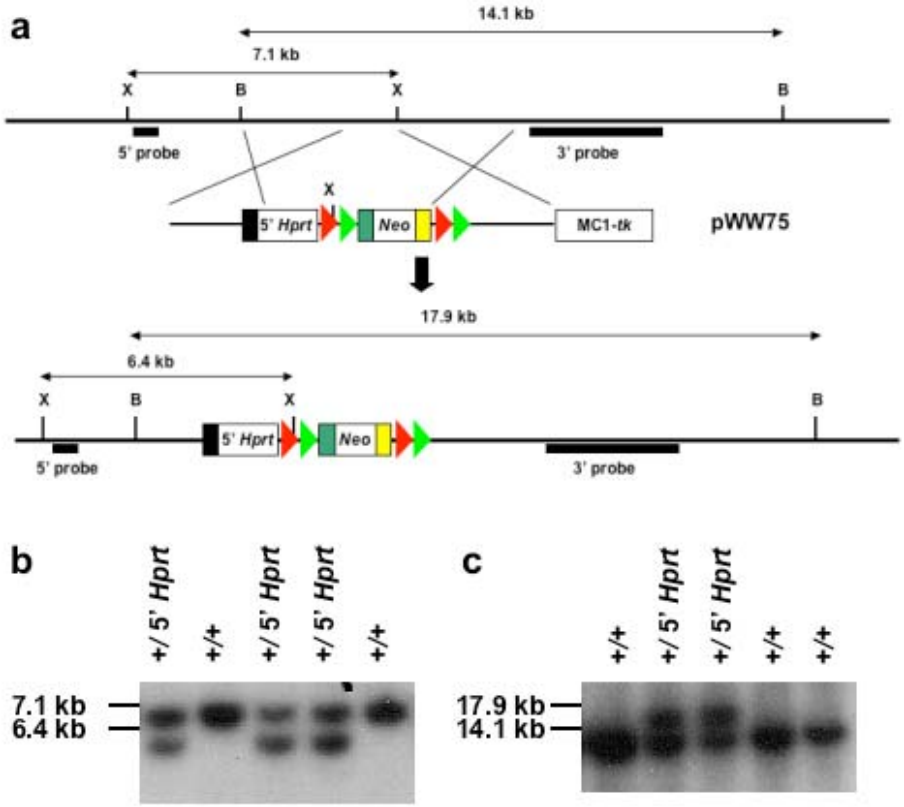


Fig. 3-4 Targeting 5' *Hprt* cassette to *D11Mit71* locus. a. Schematic illustration of targeting 5' *Hprt* cassette to *D11Mit71* locus. A 2.1 kb genomic fragment at *D11Mit71* locus was replaced by the 5' *Hprt* cassette (pWW49) to make the 5' *Hprt* targeting vector (pWW75). The MCI-*tk* cassette was used as the negative selection marker. Arrows of different colours denote multi *lox* sites used to flank the *Neo* selection cassettes. B, *Bam*HI; X, *Xba*I. **b.** Southern analysis of the WW16 cell line using a *D11Mit71* 5' probe. Genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, the detected restriction fragments were 7.1 kb for the wild type allele and 6.4 kb for the targeted allele. **c.** Southern analysis of the WW16 cell line using a *D11Mit71* 3' probe. Genomic DNA was also digested with *Bam*HI and hybridized with a *D11Mit71* 3' probe, the detected restriction fragments were 14.1 kb for the wild type allele and 17.9 kb for the targeted allele.

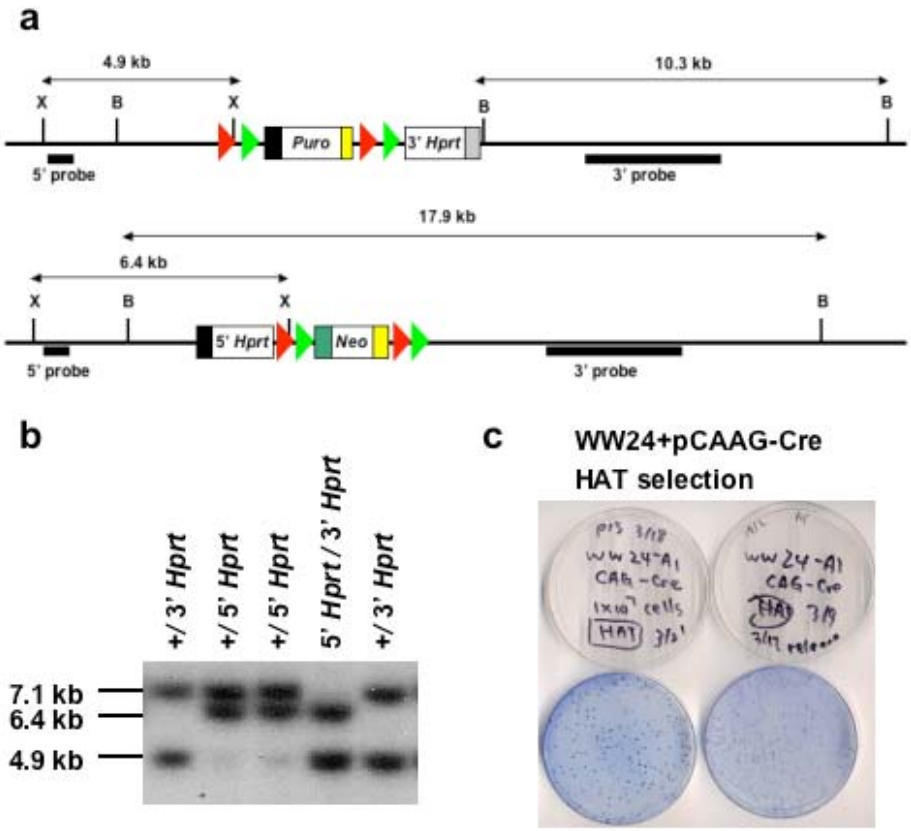


Fig. 3-5 Generating the *D11Mit71*$^{5' Hprt/3' Hprt}$ ES cell line. a. Schematic illustration of double-targeted *D11Mit71* alleles. Arrows of different colours denote multi *lox* sites used to flank the *Neo* and *Puro* selection cassettes. B, *Bam*HI; X, *Xba*I. **b.** Southern analysis of the WW24 cell line. Genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, the detected restriction fragments for double targeted clones were 6.4 kb for the 5' *Hprt* targeted allele and 4.9 kb for the 3' *Hprt* targeted allele. For some clones, the 5' *Hprt* cassette targeting resulted into the replacement of the 3' *Hprt* targeted allele, instead of the wild type allele. **c.** Functional test of the inducible mitotic recombination cassettes. pCAAG-Cre plasmid was electroporated into WW24 ES cell and the recombinants were selected by M15 supplemented with HAT. HAT resistant colonies were recovered as expected. The experiment was repeated twice.

To test the function of the 5' *Hprt*, 3' *Hprt* and the *lox* sites, pCAAG-Cre plasmid was electroporated into WW24 and WW25 cells. HAT resistant colonies were recovered from both cell lines when the Cre was transiently expressed. This experiment confirmed the functional intactness of the 5' *Hprt*, 3' *Hprt* and the *lox* sites (Fig. 3-5c).

To pop-out the *lox* site flanked *Neo* and *Puro* cassettes from WW24, pCAAG-Cre was electroporated into WW24 cells (Fig. 3-6). The cells were plated at low density and allowed to form single colonies without drug selection. The clones that lost both *Neo* and *Puro* selection markers were identified by sib-selection using M15, M15+G418, M15+puromycin and M15+HAT, respectively. The correct recombinants should be G418 sensitive, puromycin sensitive and HAT sensitive (Fig. 3-6). The function of the 5' *Hprt* and 3' *Hprt* cassettes were tested by transient Cre expression and subsequent HAT selection (Fig. 3-7a). The appropriate clones were expanded and confirmed by Southern analysis using a *D11Mit71* 5' probe. The popped-out double-targeted clones were expected to have 4.9 kb band (3' *Hprt* targeting) and 6.4 kb band (5' *Hprt* targeting) *Xba*I restriction fragments (Fig. 3-7b).

3.2.1.3 Targeting of the end point cassette for regional trapping

The original regional trapping design of Wentland et al. (unpublished data) utilized the two non-functional half *Hprt* cassettes to select for inversion events (Fig. 3-8a). Since we have already used these cassettes for inducible mitotic recombination, we elected to use the split promoter and selection marker strategy to achieve efficient recovery of inversion events. In brief, a *PGK-lox*P-*Bsd* cassette was first targeted to the E_2DH locus to fix the inversion end point. A promoter-less, non-functional *lox*P-*Puro* cassette was then introduced by retroviral integration. By Cre-mediated site-specific recombination, the two cassettes recombined to produce a functional *PGK-lox*P-*Puro* cassette which will be resistant to puromycin selection (Fig. 3-8b).

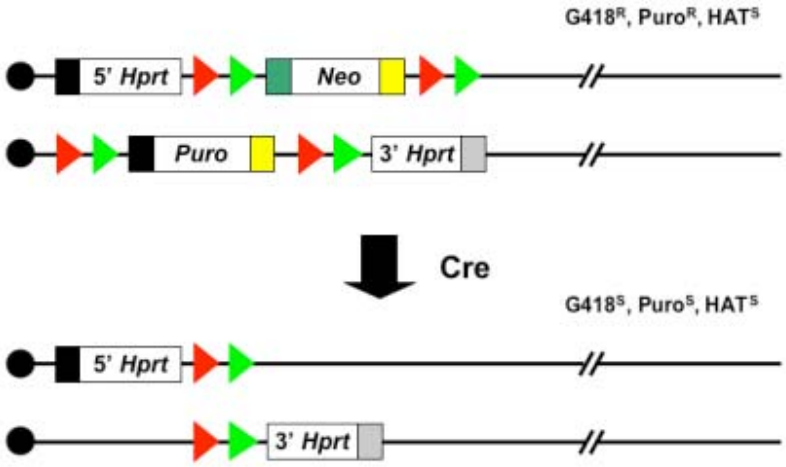


Fig. 3-6 Selection marker pop-out of the *D11Mit71*$^{5'\ Hprt/3'\ Hprt}$ ES cell line. Schematic illustration of selection marker pop-out of the double-targeted *D11Mit71* alleles. Arrows of different colour denote multi *lox* sites used to flank the *Neo* and *Puro* selection cassettes.

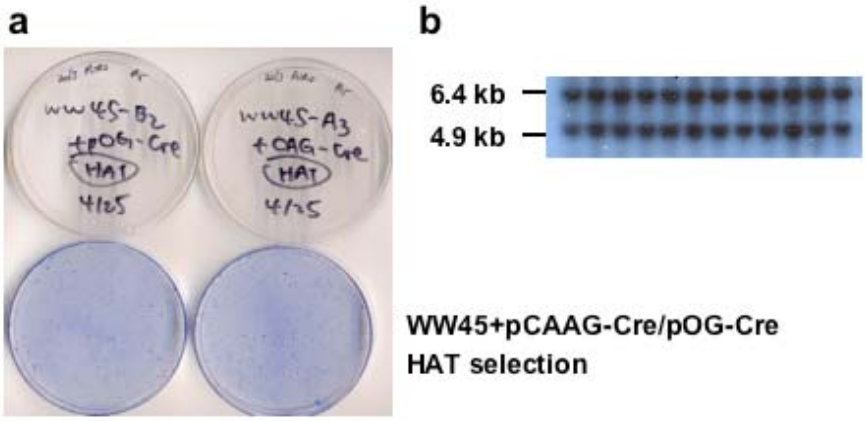


Fig. 3-7. **Test of the WW45 cell line a.** Functional test of the induced mitotic recombination cassettes. pCAAG-Cre or pOG-Cre plasmid was electroporated into WW45 ES cells and the recombinants were selected by M15 supplemented with HAT. HAT resistant colonies were recovered as expected. The experiment was repeated twice. **b.** Southern analysis of the WW45 cell line. Genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, all the Puro^s G418^s clones carry the 6.4 kb restriction fragment for the 5' *Hprt* targeted allele and 4.9 kb restriction fragment for the 3' *Hprt* targeted allele.

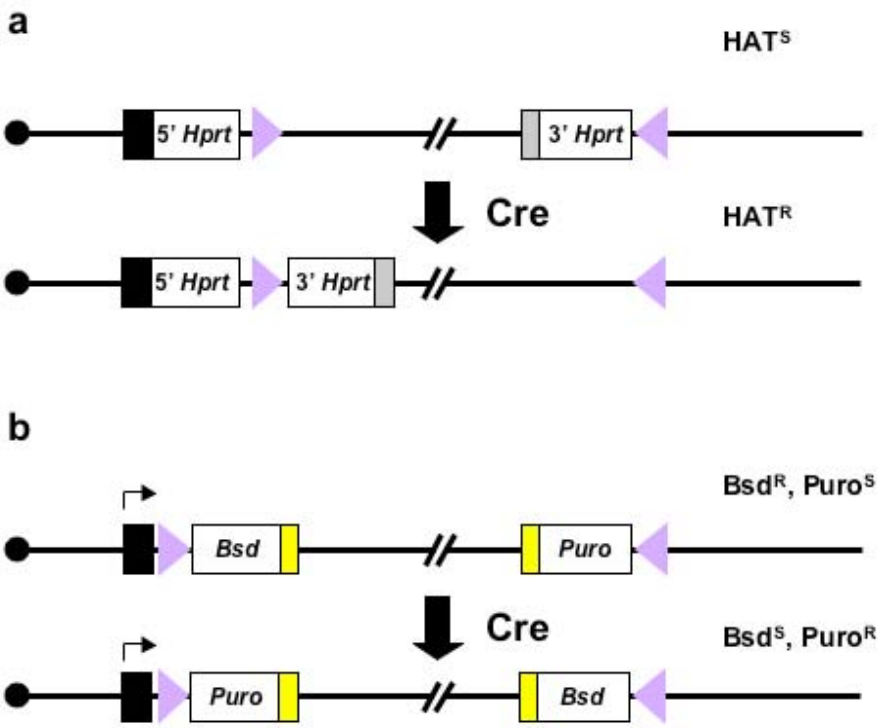
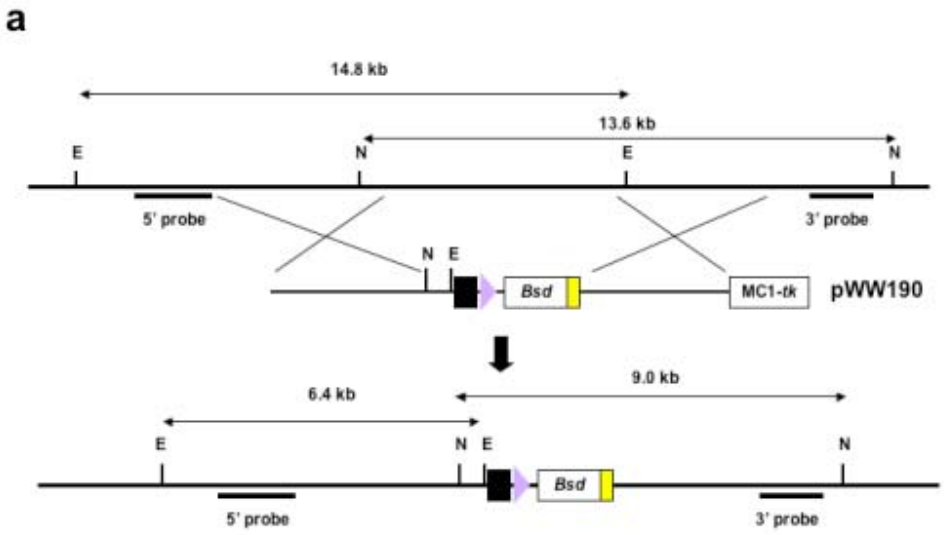


Fig. 3-8 Regional trapping strategy. a. Schematic illustration of the original regional trapping selection strategy by Wentland *et al.* (unpublished data). The Cre-mediated recombination can reconstitute a functional *Hprt* gene. **b.** Schematic illustration of the regional trapping selection strategy used in my project. The Cre-mediated recombination will reconstitute a functional *Puro* cassette, and disrupt the original *Bsd* cassette. Black box, *PGK* promoter; yellow box, bovine growth hormone polyA site; gray box, SV40 polyA site; purple arrow, wild type *lox*P site.

A *PGK-lox*P-*EM7-Bsd-bpA* (pWW146) cassette was constructed by ligating a pair of oligonucleotides containing a wild-type *lox*P site into pL313 (*PGK-EM7-Bsd-bpA*). The function of this cassette was tested by electroporating the linearized plasmid DNA into AB2.2 ES cells and selected in M15 medium supplemented with blasticidin. Blasticidin resistant colonies were recovered from the selection, which proved that the insertion of the *lox*P site did not interfere with the function of the *Bsd* selection cassette. The intactness of the inserted *lox*P site between the *PGK* promoter and the *Bsd* ORF was confirmed by sequencing (data not shown).

After the cassette was tested *in vivo*, it was ligated with two E_2DH homology arms and a MC1-HSV*tk* negative selection marker to make the final targeting vector, pWW190. Linearized pWW190 was electroporated into WW45 cells and the transfectants were selected with blasticidin and FIAU. Correctly targeted clones were identified by Southern analysis (Fig. 3-9a). Genomic DNA was digested with *Eco*RI and hybridized with an E_2DH 5' probe. The targeted restriction fragment was 9.2 kb and the wild-type restriction fragment was 14.9 kb (Fig. 3-9b). Genomic DNA was also digested with *Nde*I and hybridized with an E_2DH 3' probe. The targeted restriction fragment was 9.6 kb and the wild-type restriction fragment was 13.1 kb (Fig. 3-9c). To confirm that the targeted clones also carried the two induced mitotic recombination cassettes, genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, all the blasticidin resistant clones had both a 6.4 kb restriction fragment (5' *Hprt* targeting and *Neo* pop-out) and a 5.0 kb restriction fragment (3' *Hprt* targeting and *Puro* pop out) (Fig. 3-9d). Two correctly targeted clones, WW69-C8 and WW69-D6, were identified by Southern analysis. Sib-selection was carried out to determine the functional intactness of all the cassettes using M15, M15+G418, M15+puromycin, M15+HAT and M15+blasticidin, respectively. The two clones were G418 sensitive, puromycin sensitive, HAT sensitive and blasticidin resistant (Fig. 3-9e).


a
14.8 kb
13.6 kb
E
N
E
N
5' probe
3' probe
N E
Bsd
MC1-tk
pWW190
6.4 kb
9.0 kb
E
N E
N
Bsd
5' probe
3' probe

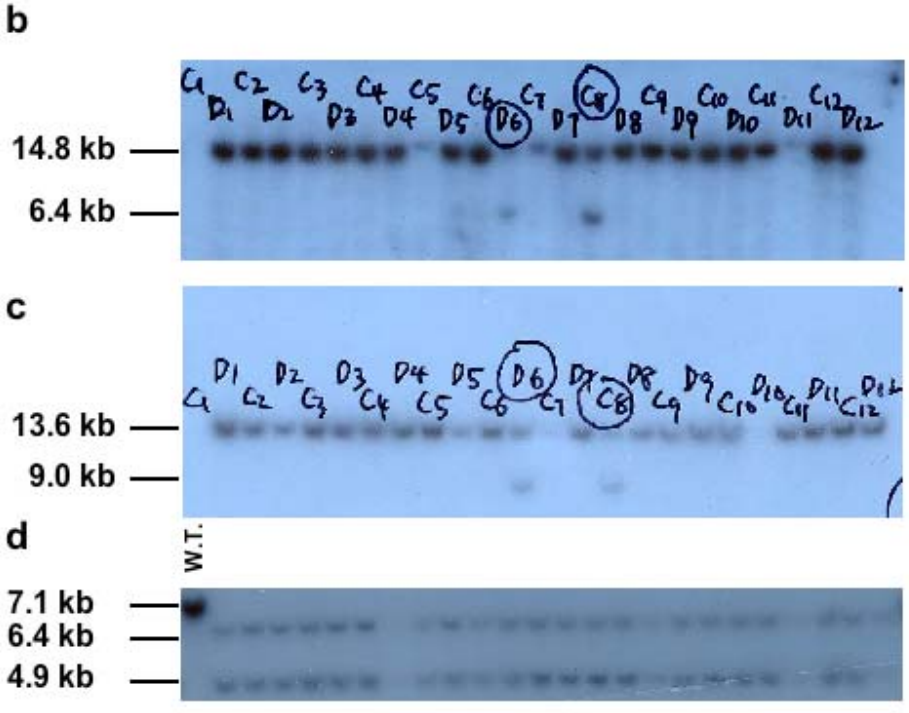

b
14.8 kb
6.4 kb
c
13.6 kb
9.0 kb
d
W.T.
7.1 kb
6.4 kb
4.9 kb

Fig. 3-9 Generating end-point targeting ES cell line. a. Schematic illustration of targeting the *PGK-lox*P-*Bsd-bpA* cassette to the E_2DH locus. Arrow denotes the wild-type *lox*P site used to separate the *PGK* promoter and the *Bsd* ORF. E, *Eco*RI; N, *Nde*I. **b.** Southern analysis of the WW69 cell line using a E_2DH 5' probe. Genomic DNA was digested with *Eco*RI and hybridized with an E_2DH 5' probe, the targeted allele was 6.4 kb and the wild type allele was 14.8 kb. **c.** Southern analysis of the WW69 cell line using a E_2DH 3' probe. Genomic DNA was digested with *Nde*I and hybridized with a E_2DH 3' probe, the targeted allele was 9.0 kb and the wild type allele was 13.6 kb. **d.** Southern analysis of the WW69 cell line using a *D11Mit71* 5' probe. Genomic DNA was digested with *Xba*I and hybridized with a *D11Mit71* 5' probe, the detected restriction fragments were 6.4 kb for the 5' *Hprt* targeted allele, 4.9 kb for the 3' *Hprt* targeted allele and 7.1 kb for the wild type allele. DNA from a wild type cell line was loaded on the left as a control.

To use induced mitotic recombination to make homozygous mutations, the disruption of the first allele should happen on the same chromosome as the 3' *Hprt* cassette. A direct way to determine the location of the end point targeting cassette is to induce mitotic recombination by transient Cre expression, and check the genotypes of the recombinants recovered. If most of the HAT resistant colonies are homozygous for the targeted *E_2DH* allele, the end point targeting cassette is targeted to the right chromosome (Fig. 3-10a). On the other hand, if no clone is homozygous for targeted *E_2DH* allele, the end point targeting cassette is targeted to the wrong chromosome (Fig. 3-10b).

To determine the location of the *PGK-lox*P-*Bsd-bpA* cassette, a Cre expression plasmid was electroporated into WW69-C8 and D6 cell lines and transfectants were selected in M15 supplemented with HAT. 36 HAT resistant colonies were picked from each electroporation and Southern analysis was performed using the *D11Mit71* 5' probe, *D11Mit71* 3' probe, *E_2DH* 5' probe and *E_2DH* 3' probe, respectively (Fig. 3-11). Sib-selection was carried out to determine the drug resistance of each recombinant using M15, M15+HAT and M15+blasticidin, respectively. Clones that carried two targeted *E_2DH* alleles were recovered from recombinants of WW69-D6. Of the 25 clones that carried at least one copy of *PGK-lox*P-*Bsd-bpA* cassette (resistant to blasticidin selection), 18 of them were homozygous for targeted *E_2DH* locus (X segregation) and 7 of them were heterozygous (Z segregation). The percentage of G2-X segregation is 72%, which is comparable to the data described before (Liu, Jenkins et al. 2002). 2 of the clones that were homozygous for the targeted *E_2DH* allele also carried two copies of 5' *Hprt* cassettes, suggesting that the induced mitotic recombination was followed by second around of mitotic recombination between the centromere and the *D11Mit71* locus or a chromosome loss/duplication event (Fig. 3-12).

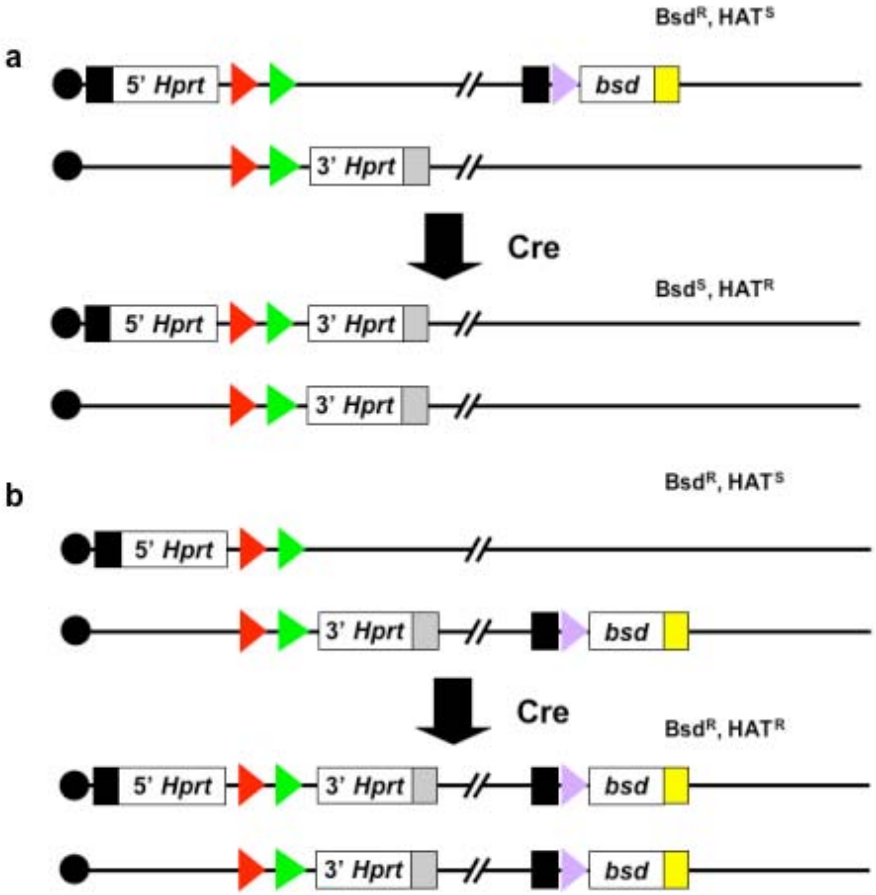


Fig. 3-10 The location of the end point targeting cassette. a. Schematic illustration of the outcome of the induced mitotic recombination when the end-point targeting cassette is on the same chromosome as the 5' *Hprt* cassette. Note that only G2-X segregation is shown here, no clones homozygous for the targeted E_2DH allele will be recovered in this situation. **b.** Schematic illustration of the outcome of the induced mitotic recombination when the end-point targeting cassette is on the same chromosome as the 3' *Hprt* cassette. Note that only G2-X segregation is shown here, clones homozygous for the targeted E_2DH allele will be recovered in this situation. Black box, *PGK* promoter; yellow box, bovine growth hormone polyA site; gray box, SV40 polyA site; arrows of different colors denote wild type and mutant *lox*P sites.

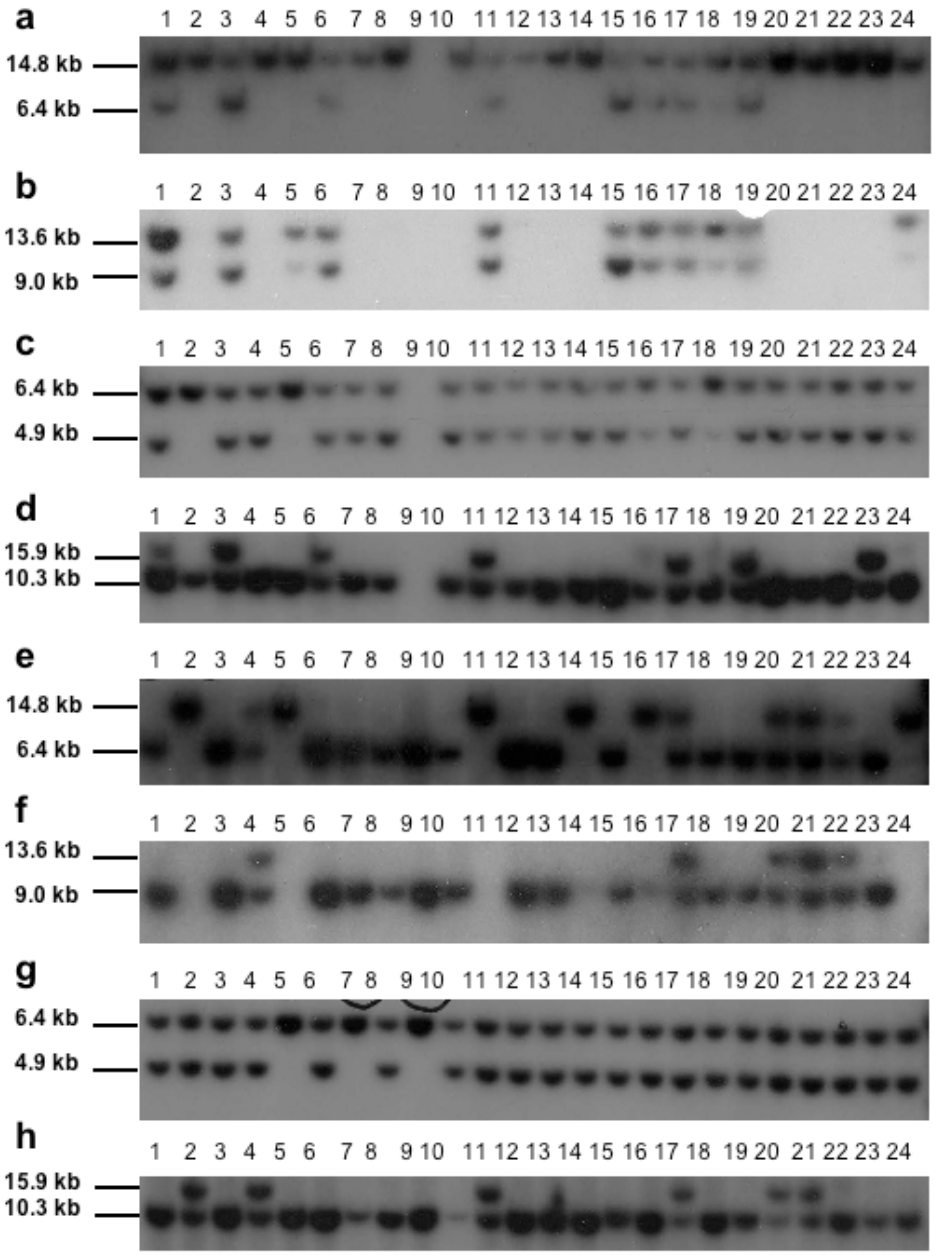

a
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
14.8 kb
6.4 kb
b
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
13.6 kb
9.0 kb
c
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
6.4 kb
4.9 kb
d
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
15.9 kb
10.3 kb
e
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
14.8 kb
6.4 kb
f
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
13.6 kb
9.0 kb
g
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
6.4 kb
4.9 kb
h
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
15.9 kb
10.3 kb

Fig. 3-11 Determination of the location of the end point targeting cassette by induced mitotic recombination. a. Genomic DNA from WW69-C8 recombinants was digested by *Eco*RI and hybridized with a E_2DH 5' probe, the targeted restriction fragment was 9.2 kb and the wild type restriction fragment was 14.9 kb. **b.** Genomic DNA from WW69-C8 recombinants was digested by *Eco*RI and hybridized with a E_2DH 3' probe, the targeted restriction fragment was 9.6 kb and the wild type restriction fragment was 13.1 kb. **c.** Genomic DNA from WW69-C8 recombinants was digested by *Xba*I, hybridized with a *D11Mit71* 5' probe, the 5' *Hprt* cassette targeted fragment was 6.4 kb and 3' *Hprt* cassette targeted fragment was 4.9 kb fragment . **d.** Genomic DNA from WW69-C8 recombinants was digested by *Bam*HI, hybridized with a *D11Mit71* 3' probe, the 5' *Hprt* cassette targeted fragment was 10.3 kb and 3' *Hprt* cassette targeted fragment was 15.9 kb. **e.** Genomic DNA from WW69-D6 recombinants was digested by *Eco*RI and hybridized with a E_2DH 5' probe, the targeted restriction fragment was 9.2 kb and the wild type restriction fragment was 14.9 kb. **f.** Genomic DNA from WW69-D6 recombinants was digested by *Eco*RI and hybridized with a E_2DH 3' probe, the targeted restriction fragment was 9.6 kb and the wild type restriction fragment was 13.1 kb. **g.** Genomic DNA from WW69-D6 recombinants was digested by *Xba*I, hybridized with a *D11Mit71* 5' probe, the 5' *Hprt* cassette targeted fragment was 6.4 kb and 3' *Hprt* cassette targeted fragment was 4.9 kb fragment . **h.** Genomic DNA from WW69-D6 recombinants was digested by *Bam*HI, hybridized with a *D11Mit71* 3' probe, the 5' *Hprt* cassette targeted fragment was 10.3 kb and 3' *Hprt* cassette targeted fragment was 15.9 kb. Note that for **b** and **f**, the Southern analysis was carried out on blasticidine resistant clones. So all the clones that are sensitive to the drug were killed during the selection.

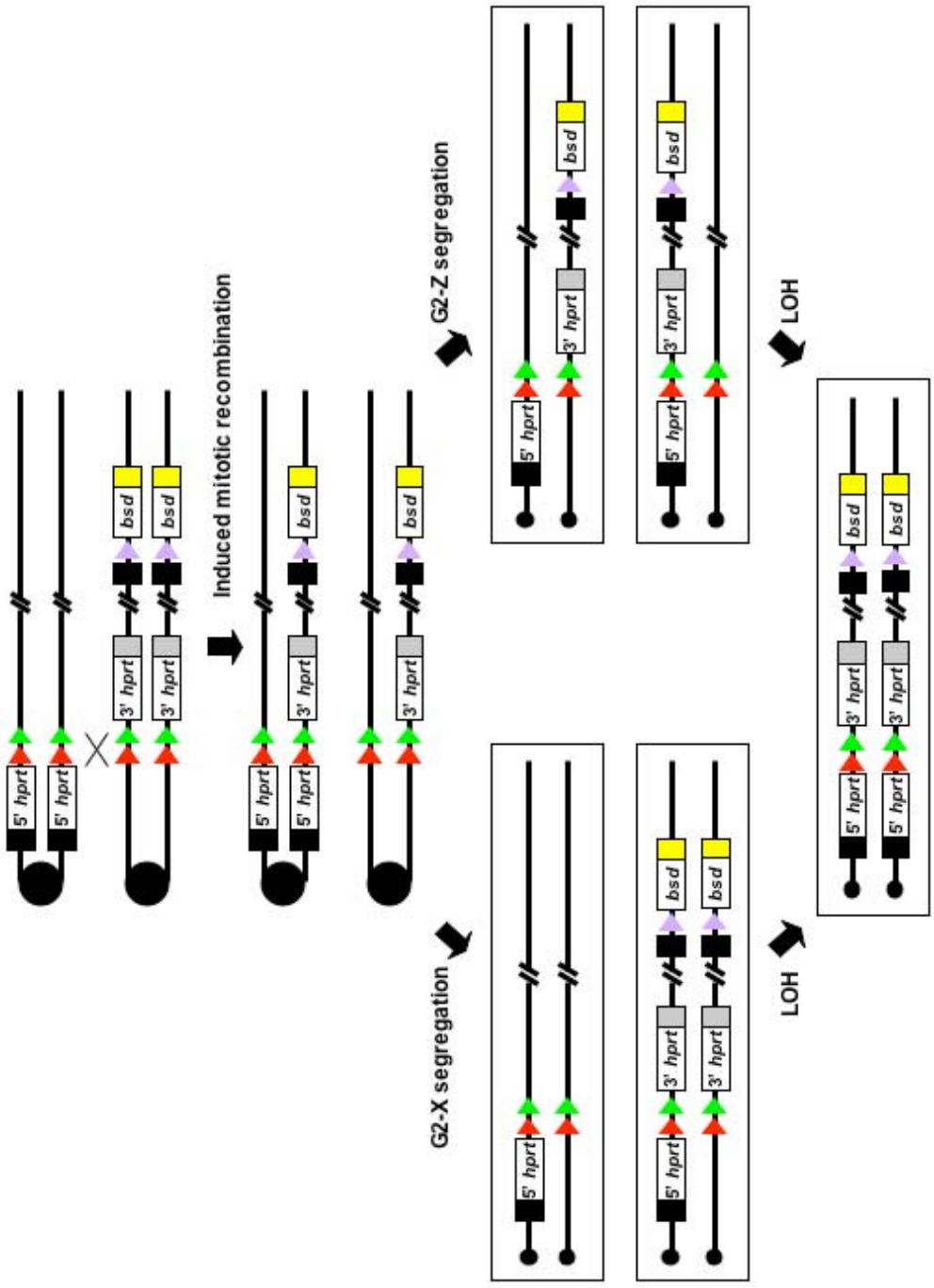

5' hprt
3' hprt
bsd
Induced mitotic recombination
G2-Z segregation
G2-X segregation
LOH
LOH

Fig. 3-12 Segregation pattern of recombinant chromatides after G2 recombination. Note that G2-X segregation will make the *PKG-lox*P-*bsd-bpA* cassette homozygous. But some HAT resistant clones carry two 5' *hprt* cassettes, two 3' *hprt* cassettes and two *PKG-lox*P-*bsd-bpA* cassettes. These cones might derive from the LOH event or a second mitotic recombination following the induced mitotic recombination. Black box, *PGK* promoter; yellow box, bovine growth hormone polyA site; gray box, SV40 polyA site; arrows of different colors denote wild type and mutant *lox*P sites.

11 clones were sensitive to blasticidin and carried two wild-type *E_2DH* alleles, suggesting that they derive either from untargeted WW45 cells carried over with *PGK-lox*P-*Bsd-bpA* targeted WW69 cells or a spontaneous mitotic recombination occurred somewhere between the *D11Mit71* and *E_2DH* loci before the Cre-induced mitotic recombination. Since the spontaneous mitotic recombination rate is very low (10^{-7} to 10^{-8}), these clones are most likely derived from untargeted WW45 cell contamination. To eliminate this background, WW69-D6 cells were plated at low density to form single colonies. The subclones were confirmed by Southern analysis and used for future experiments.

For WW69-C8, no HAT resistant colonies were homozygous for the targeted *E_2DH* allele, therefore the *PGK-lox*P-*Bsd-bpA* cassette was targeted to the wrong chromosome. Linearized pWW190 was also electroporated into another *D11Mit71*$^{5'\ Hprt/3'\ Hprt}$ cell line, WW46. Several targeted clones were identified using the same Southern screening strategy. But the subsequent Cre-induced mitotic recombination proved that for all these clones, the *PGK-lox*P-*Bsd-bpA* cassette was also targeted to the wrong chromosome (data not shown). Therefore, all the future experiments were carried out in WW69-D6 cell line.

3.2.2 Construction of the regional trapping vectors

The original gene-trap retrovirus of Wentland et al. (unpublished data) contained a 3' trapping cassette and a *lox*P-3' *Hprt* cassette for the selection of inversions. Since we elected to use the split promoter and selection marker strategy to select for the regional inversion events, some modifications were required to the original virus. A 5' trapping retrovirus was also constructed to compare the trapping efficiency and the subsequent inversion efficiency.

3.2.2.1 5' trapping vectors

The 5' trapping virus (pWW239) has a SAβ*geo* cassette (Friedrich and Soriano 1991) and a promoter-less *lox*P-*Puro-bpA* cassette (Fig. 3-13a). To test the function of the two cassettes, linearized pWW239 and pWW183 (*PGK-lox*P-*EM7-Bsd-bpA*) plasmids were co-electroporated into wild-type

AB2.2 ES cells with or without a Cre expression plasmid, and the transfectants were selected with M15+G418, M15+blasticidin, or M15+puromycin, respectively.

As a control, linearized pWW239 plasmid was also electroporated into AB2.2 and selected with M15+G418, or M15+puromycin, respectively. G418 resistant colonies were recovered, which confirmed that the 5' trapping cassette worked in ES cells. No puromycin resistant colonies were recovered when pWW239 was electroporated into AB2.2 ES cells alone. This result confirmed that promoter-less *lox*P-*Puro*-*bpA* cassette of the virus does not function. Importantly, puromycin resistant colonies were only recovered when linearized pWW239 and pWW183 (*PGK*-*lox*P-*EM7*-*Bsd*-*bpA*) plasmids were co-electroporated into AB2.2 ES cells with the Cre expression plasmid. This result confirms that when the non-functional *Puro* cassette gains the *PGK* promoter by Cre-mediated recombination, it becomes functional. This test has confirmed that both halves of the virus were functional *in vivo* (Fig. 3-14). This virus construct was used for future experiments.

The size of the exogenous DNA fragment that can be inserted into the virus backbone is limited, and the virus packaging efficiency drops significantly when the size of the virus increases. The two cassettes in the pMSCV (Clontech) virus backbone are about 5.3 kb in size, so it is reasonable to predict that the trapping titre will be low. Compared with the virus-based trapping vectors, there is little limit on the size of the electroporation-based trapping vectors. So electroporation-based 5' trapping vectors (pWW237) were also constructed. The two cassettes in pWW237 are essentially the same as in pWW239. However, instead of cloning these two cassettes into a virus backbone, they were cloned into pBS vector. pWW237 was functionally tested in the same way as pWW239. The results confirmed that the trapping and inversion cassettes were functional *in vivo*.

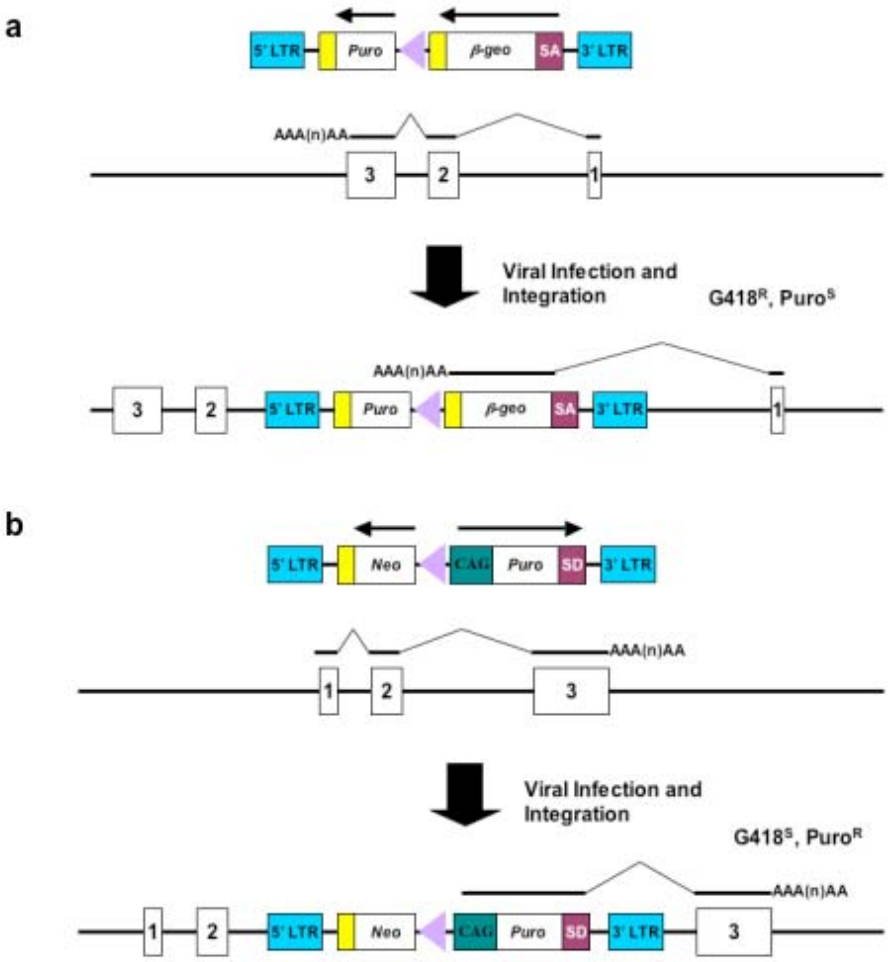


Fig. 3-13 The structure of gene trapping virus. a. Schematic illustration of the structure of the 5' trapping virus and the way it integrates into the genome. **b.** Schematic illustration of the structure of the 3' trapping virus and the way it integrates into the genome. LTR, long terminal repeat; SA, splicing acceptor; SD, splicing donor; yellow box, bovine growth hormone polyA site; purple arrow, wild type *lox*P sites.

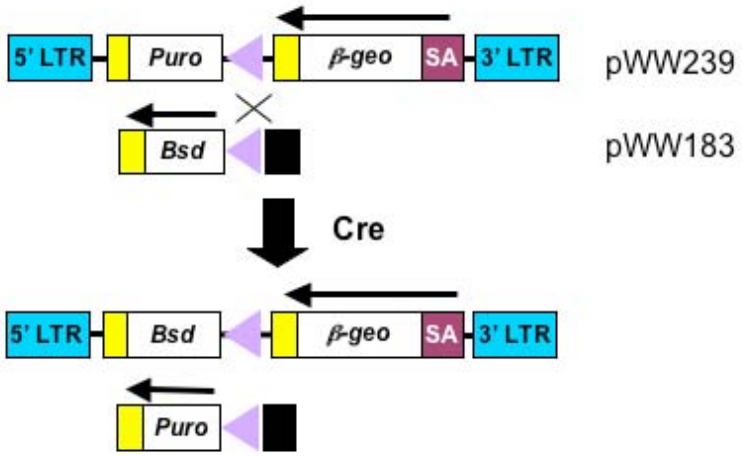


Fig. 3-14 Functional test of the 5' trapping virus. Schematic illustration of the outcome of Cre-mediated recombination between *PGK-lox*P-*Bsd-bpA* cassette and the promoter-less *lox*P-*Puro-bpA* cassette. LTR, long terminal repeat; SA, splicing acceptor; yellow box, bovine growth hormone polyA site; black box, *PGK* promoter; purple arrow, wild type *lox*P site.

3.2.2.2 3' trapping vectors

The original gene-trapping virus of Wentland et al. (unpublished data) contained a *PGK*-*Puro*-SD cassette. To improve the 3' trapping efficiency, different promoters were used to make a series of trapping vectors, and the trapping efficiency of different constructs was compared by *in vivo* functional test. pWW64 (*CAG* promoter), pWW65 (*EF1*α promoter) and pWW73 (*PGK* promoter) were linearized and then electroporated into AB2.2 ES cells. The transfectants were selected in M15 supplemented with puromycin. More puromycin resistant colonies were recovered from pWW64 than the other two plasmids, which suggested that the *CAG* promoter was much stronger than the other two promoters *in vivo* (Fig. 3-15). So the *CAG*-*Puro*-SD trapping cassette was used for future experiments.

The *CAG*-*Puro*-SD cassette and a promoter-less *lox*P-*Neo*-*bpA* cassette were cloned into pMSCV (Clontech) to make the final 3' trapping virus (pWW240), (Fig-3-13b). To test the function of the two cassettes, linearized pWW240 and pWW183 (*PGK*-*lox*P-*EM7*-*Bsd*-*bpA*) plasmids were co-electroporated into AB2.2 ES cells with or without a Cre expression plasmid, and the transfectants were selected with M15+G418, M15+blasticidin or M15+puromycin, respectively. As a control, linearized pWW240 plasmid was also electroporated into AB2.2 cells and these cells were selected with M15+G418 or M15+puromycin, respectively.

Puromycin resistant colonies were recovered when linearized pWW240 was electroporated into AB2.2 ES cell alone. This result confirmed the function of the 3' trapping cassette of the virus. No G418 resistant colonies were recovered when pWW240 was electroporated into AB2.2 ES cell alone. This result confirmed that promoter-less *lox*P-*Neo*-*bpA* cassette of the virus does not function. G418 resistant colonies were recovered when linearized pWW240 and pWW183 (*PGK*-*lox*P-*EM7*-*Bsd*-*bpA*) plasmid was co-electroporated into AB2.2 ES cells with a Cre expression plasmid. This result confirmed that when the non-functional *Neo* cassette gains the *PGK* promoter by Cre-mediated recombination, it becomes functional. This test has

confirmed that both halves of the virus were functional *in vivo* (Fig 3-16). This 3' trapping virus construct was used for future experiment.

For the same reason as 5' trapping, electroporation-based 3' trapping vector (pWW238) was also constructed. The two cassettes in pWW238 are the same as pWW240, but instead of cloning these two cassettes into a virus backbone, they were cloned into pBS vector. pWW238 was also tested in the same way as the pWW240. The results confirmed that both the trapping and inversion cassettes worked *in vivo*.

3.2.3 Retrovirus transfection

The gene-trap retroviral vector was used to transiently transfect the Phoenix (REF) viral packaging cells using the Calcium Phosphate method. In brief, Phoenix cells were transfected with supercoiled pWW239 DNA. Viral supernatant was harvested at different time points, combined together and filtered through 0.45 μm filters to remove the Phoenix cells in viral supernatant. 1 ml of viral supernatant was used to infect wild-type AB2.2 ES cells. The infected ES cells were selected in M15 supplemented with G418. The G418 resistant colonies were stained with Methylene Blue (Fig. 3-17a). The trapping titre for the 5' trapping virus was around 10 trapping events/ml virus supernatant.

Phoenix cells were transiently transfected with pWW240 DNA to package the 3' trapping retrovirus. Viral supernatant was harvested and 1 ml of viral supernatant was used to infect wild-type AB2.2 ES cells. The infected ES cells were selected in M15 supplemented with puromycin, and puromycin resistant colonies were stained with Methylene Blue (Fig. 3-17b). The trapping titer for the 5' trapping virus was around 5 trapping events/ml virus supernatant. Because the trapping titer was slightly lower than the 5' trapping virus and a portion of the insertions might represent trapping of cryptic splice acceptors and polyA signals, we decided to use the 5' trapping strategy.

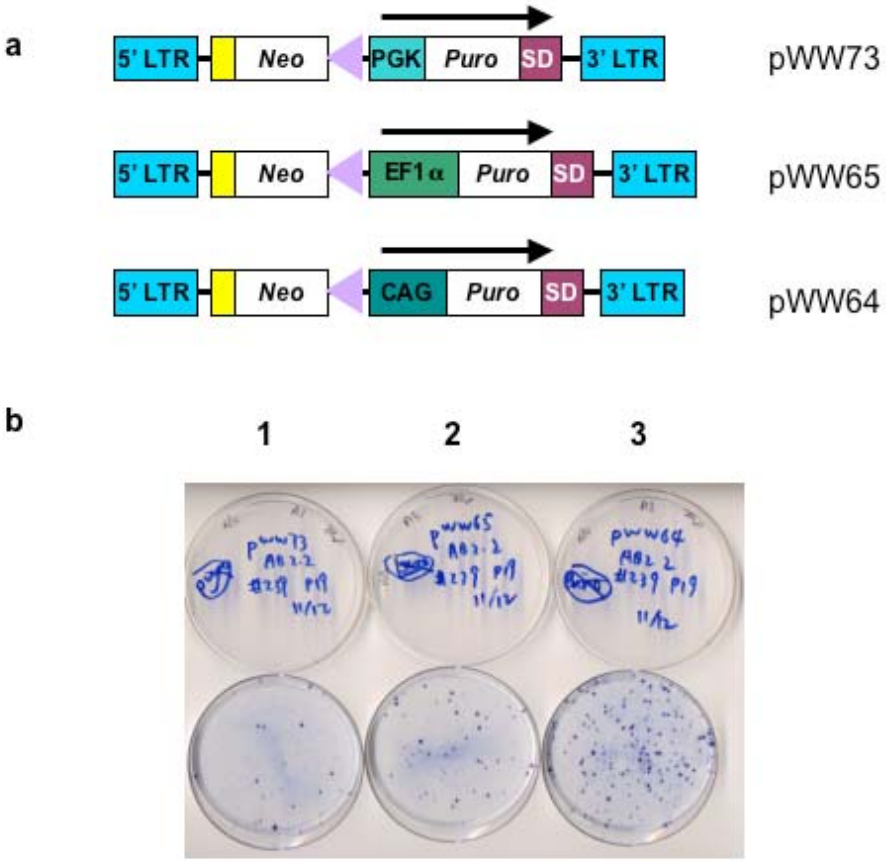


Fig. 3-15 Promoter test of the 3' trapping virus. a. Schematic illustration of the 3' trapping viruses sing different promoters. LTR, long terminal repeat; SD, splicing donor; yellow box, bovine growth hormone polyA site; purple arrow, wild type *lox*P site. **b.** Functional test of the 3' trapping virus construct.1: pWW73 (PGK promoter); 2: pWW65 (EF1α promoter); 3: pWW64 (CAG promoter). 20 μg of linearized pWW73, pWW64 and pWW65 was electroporated into AB2.2 ES cells, and the transfectants were selected with M15+puromycin.

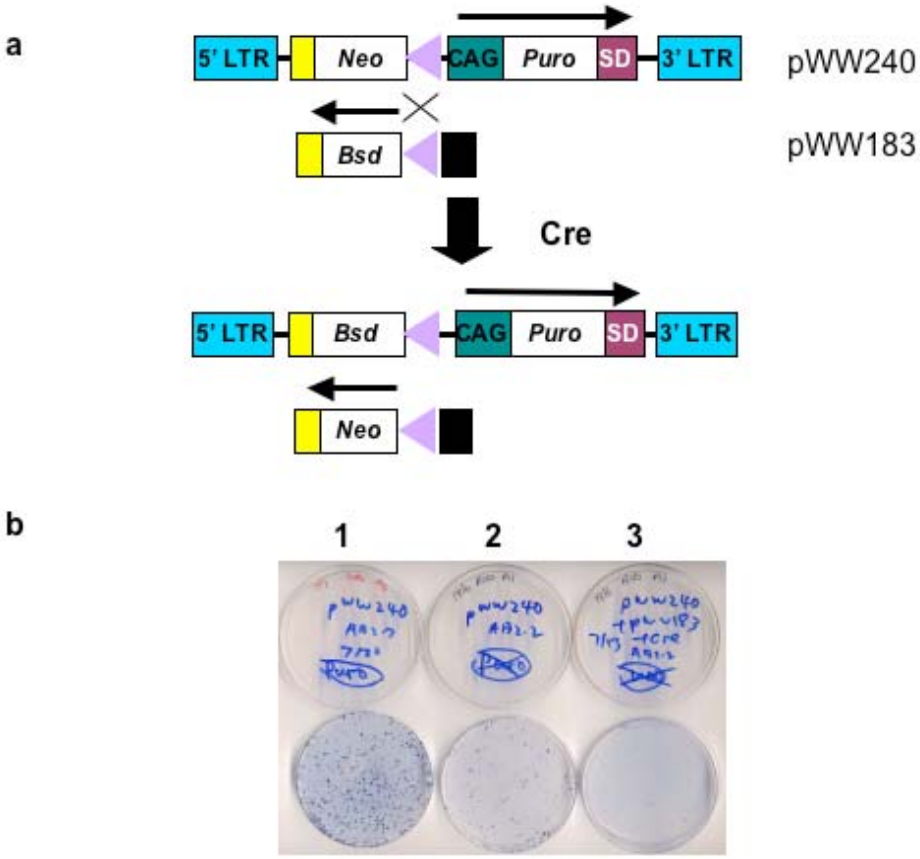


Fig. 3-16 Functional test of the 3' trapping virus. a. Schematic illustration of the outcome of the Cre-mediated recombination between the *PGK-lox*P-*Bsd-bpA* cassette and the promoter-less *lox*P-*Neo-bpA* cassette. LTR, long terminal repeat; SD, splicing donor; yellow box, bovine growth hormone polyA site; purple arrow, wild type *lox*P site. **b.** Functional test of the 5' trapping virus construct, pWW240. 1and 2: Linearized pWW240 was electroporated into AB2.2 ES cells and selected with M15+puromycin. Puromycin resistant colonies were recovered as expected, which confirm the function of the trapping cassette. 3: Linearized pWW240 and pWW183 plasmid DNA were co-electroporated into AB2.2 ES cells with the Cre expression plasmid which were selected with M15+G418. G418 resistant colonies were recovered as expected, which confirm the function of the inversion cassette.

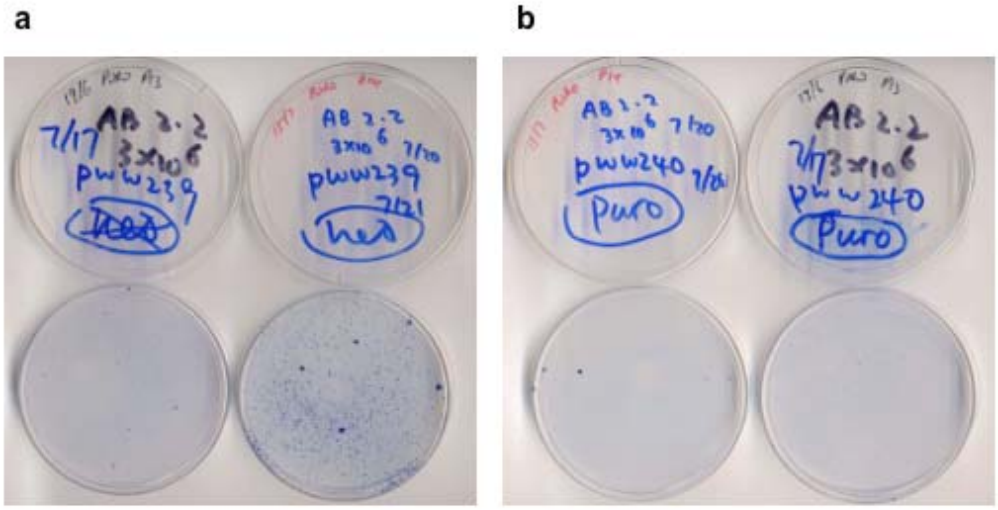


Fig. 3-17 The trapping titer of the 5' and 3' trapping retrovirus. a. 5' trapping retrovirus. Phoenix cells transfected with pWW239 DNA were used to make the 5' trapping retrovirus. Viral supernatant was harvested at different time points, combined together and filtered through 0.45 μm filters to get rid of the Phoenix cells. 1 ml of viral supernatant was used to infect the wild type AB2.2 ES cells. The infected ES cells were selected in M15 supplemented with G418. The G418 resistant colonies were stained with Methylene Blue. The virus infection experiment was repeated twice. 7 and 5 colonies were found on the two plates. Note that in the second experiment, G418 selection was released after 7 days. Many background colonies were found. So in the large-scale experiment, all the G418 colonies were selected for 10 days to reduce the background. **b.** 3' trapping retrovirus. Phoenix cells transfected with pWW240 DNA were used to make 3' trapping retrovirus. Viral supernatant was harvested at different time point,s combined together and filtered through 0.45 μm filters. 1 ml of viral supernatant was used to infect the wild type AB2.2 ES cells. The infected ES cells were selected in M15 supplemented with puromycin. The puromycin resistant colonies were stained with Methylene Blue. The virus infection experiment was repeated twice. 5 and 4 colonies were found on the two plates.

3.2.4 Pilot experiment to test the regional trapping strategy

3.2.4.1 Pilot experiment to test the intactness of proviral insertion

As a pilot experiment, the retrovirus supernatant was also used to infect WW69-C6 cells. Some G418 resistant colonies were randomly picked and expanded. Genomic DNA was extracted from these clones for Southern analysis (Fig. 3-18a). Since there is a single *Eco*RI site in SA-β*geo* cassette, individual virus integration events can be discriminated by the size of their unique proviral/host junction fragments, which are determined by the location of the endogenous *Eco*RI site nearest to the 3' LTR (Fig. 3-18b). There are two *Kpn*I sites in the 5' LTR and 3' LTRs of the retrovirus, so *Kpn*I digestion of the genomic DNA and subsequent hybridization using a *LacZ* probe should detect a 6.9 kb *Kpn*I restriction fragment from the intact provirus (Fig. 3-18c).

The Southern hybridization did identify a 6.9 kb *Kpn*I fragment for some trapped clones, but for the others, the hybridization detected an unexpected 6.0 kb fragment. A possible explanation for this is that an alternative splicing event happened in the transcription process when the retrovirus was replicated in the Phoenix cells. As the result, a part of the retrovirus would be skipped as an intron. If the alternative splicing event had occurred in the trapping or inversion cassettes, it would affect the trapping or the subsequent inversion. But if the splicing had occurred in a non-essential region in the virus backbone, it would not have any effect on the following steps.

Since the 6.0 kb *Kpn*I fragment was detected in the G418 resistant clones, the alternative splicing did not inactivate the trapping cassette. To determine whether the inversion cassette had been inactivated, the G418 clones were expanded and a Cre expression plasmid was electrorated into the trapped clones. The recombinants were selected in M15 supplemented with puromycin. Puromycin resistant colonies were recovered from the clones with the 6.0 kb *Kpn*I fragment, as well as the clones with the 6.9 kb fragment. So the alternative splicing event must have occurred somewhere on the virus backbone though this did not interfere with virus packaging and integration.

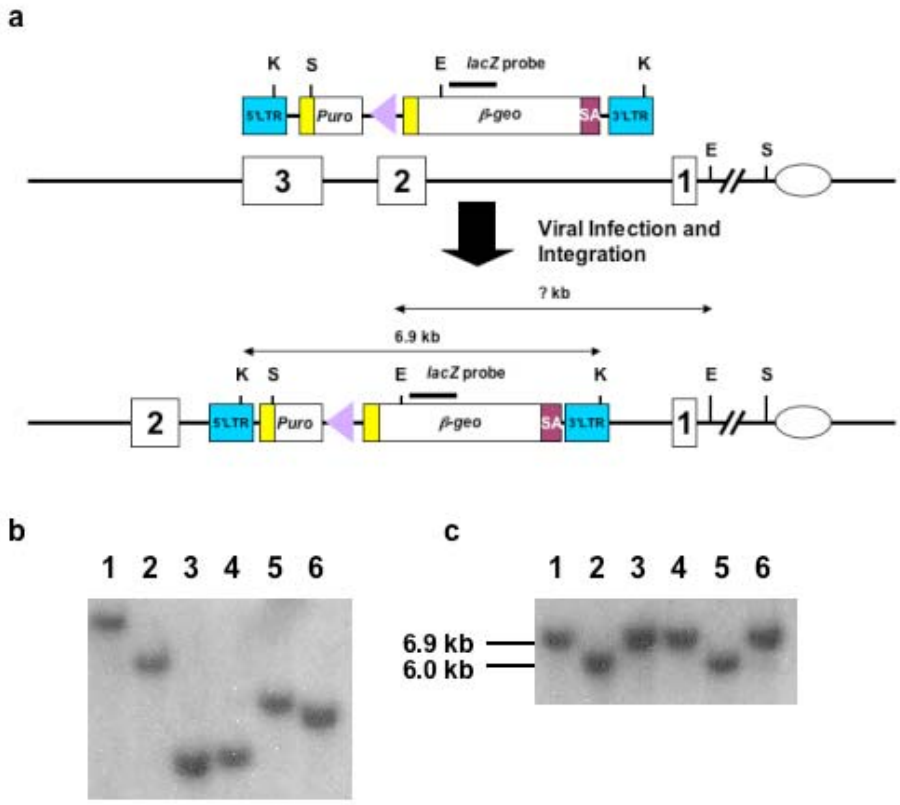
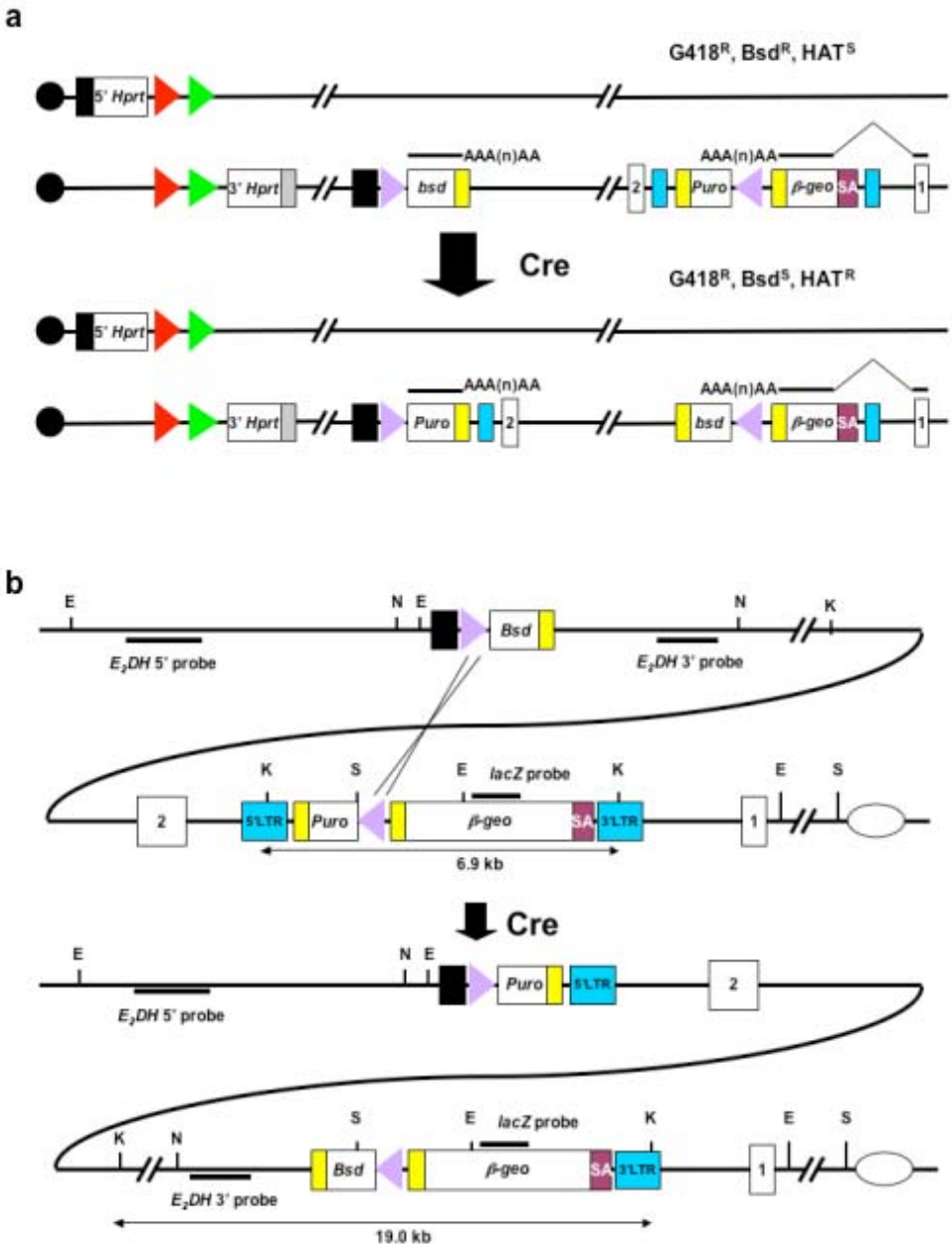


Fig. 3-18 Identification of individual virus integration events. a. Schematic illustration of retrovirus integration. LTR, long terminal repeat; black arrow, wild type *lox*P site; K, *Kpn*I, E, *Eco*RI. **b.** Genomic DNA from G418 resistant gene-trap clones were cut with *Eco*RI and hybridized to a *lacZ* probe. Individual insertion events will have unique proviral junction fragments, and these can be distinguished by different sizes of the restriction fragments. **c.** Genomic DNA from G418 resistant gene-trap clones cut with *Kpn*I and hybridized to a *lacZ* probe. A 6.9 kb intact proviral fragment was found for some of the clones. But for the others, an unexpected 6.0 fragment was found, which might result from alternative splicing during retrovirus replication process.

3.2.4.2 Pilot experiment to test the Cre-mediated inversion

Mouse chromosome 11 contains 1797 known or predicted genes, which consist 6.4% of all the 28069 mouse genes (NCBI m33 mouse assembly, freeze May 27, 2004, strain C57BL/6J), so there is a 6.4% chance that a gene-trap will occur on chromosome 11. Half of gene traps on chromosome 11 (3.2% of the total) are expected to occur on the end point cassette targeted homolog of chromosome 11. Half of the gene traps on the targeted homolog (1.6% of the total) will be in the correct orientation for an inversion. One-third of the gene traps (0.5% of the total) will be in the vicinity of the end point (*E_2DH*) for an inversion event to happen efficiently. Thus a strong selection strategy is needed to select for these rare events. In my project, a split promoter and selection marker was used to achieve efficient recovery of the inversion events (Fig. 3-19a).

To test whether the selection strategy works, a pilot experiment was carried out. The 5' gene-trap retrovirus was used to infect WW69-A12 ES cells, and 100 G418 resistant gene-trap clones were picked and pooled together. A Cre expression plasmid was electroporated into the pool to induce the inversion. The cells were selected in M15 supplemented with puromycin and 16 puromycin resistant colonies were randomly picked and expanded. Genomic DNA was extracted from these clones for Southern analysis (Fig. 3-19b). Genomic DNA was digested with *Kpn*I and hybridized with a *lacZ* probe to identify the inversion events. When a Cre-mediated inversion occurs, the 5' region of the *E_2DH* end point targeting cassettes will move adjacent to the 3' region of the retrovirus. In such puromycin resistant clones, the 6.9 kb or 6.0 kb proviral insertion fragment detected by the *lacZ* probe will be replaced by a 19.0 kb inversion fragment (Fig. 3-19c). Genomic DNA was also digested with *Eco*RI and hybridized with a *lacZ* probe to identify independent gene-trap events (Fig. 3-19d). In this pilot experiment, the puromycin resistant clones had a limited repertoire of proviral junction fragments; it is likely that clones with the same-sized *Eco*RI fragments are derived from the same gene-trap clone.


a
G418R, BsdR, HATS
5' Hprt
3' Hprt
bsd
AAA(n)AA
2
Puro
β-geo
SA
1
Cre
G418R, BsdS, HATR
b
E
N
Bsd
K
E2DH 5' probe
E2DH 3' probe
S
lacZ probe
5'LTR
3'LTR
6.9 kb
19.0 kb

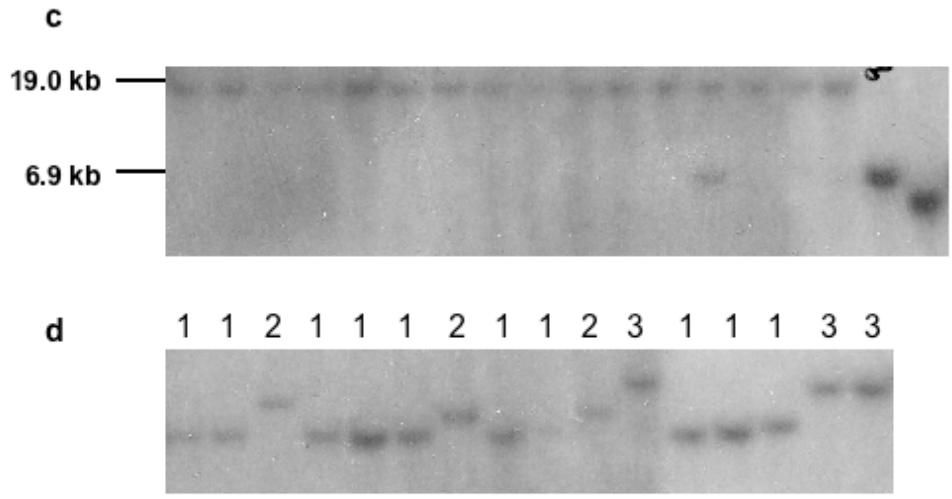


Fig. 3-19 Identification of individual inversion events. a. Schematic illustration of regional trapping induced by Cre transient expression. Black box: *PGK* promoter, yellow box: a bovine growth hormone polyA site; grey box: SV40 polyA site; blue box, long terminal repeat; SA, splicing acceptor; arrows of different color, variant *lox* sites. **b.** Schematic illustration of Southern screening strategy to identify individual inversion events. LTR, long terminal repeat; black arrow, wild type *lox*P site; K, *Kpn*I; E, *Eco*RI; N, *Nde*I; S, *Spe*I. **c.** Genomic DNA from Puromycin resistant clones was cut with *Kpn*I and hybridized to a *lacZ* probe. A 19.0 kb inversion fragment was detected. Note that one of the clones has both the 6.9 kb proviral insertion fragment and the 19.0 kb inversion fragment. The last two lanes are gene trap clones used as control. **d.** Genomic DNA from puromycin resistant gene-trap clones was cut with *Eco*RI and hybridized to a *lacZ* probe. Individual events will have unique proviral junction fragments, which can be resolved by different sized restriction fragments. Note that several clones share the same sized proviral junction fragments (group 1, 2 and 3), these are probably derived from the same trapping clone.

Clones that have rare proviral junction fragments might harbour inefficient recombination events, such as a balanced translocation, a balanced deletion/duplication, or an inversion over a long genetic distance. The balanced translocation and deletion/duplication do not involve loss of any genetic material, so these cells are expected to be viable at this step. Such clones can not become homozygous by induced mitotic recombination because this will result in the loss of a significant part of the chromosome 11 (Fig. 3-20).

3.2.5 Large-scale regional trapping experiment using gene-trap retrovirus

To perform large-scale experiments, 2000 ml of viral supernatant was harvested from 20X 90-mm plates of pWW239-transfected Phoenix cells. The viral supernatant was used to infect WW69-D6 ES cells plated on a total of twenty 90-mm feeder plates. Fresh viral supernatant was used to replace the old one every 12 hours for 3 days to increase the chance of viral infection. The gene-trap clones were then selected in M15 medium supplemented with G418. One plate was stained with Methylene Blue to count the number of the G418 resistant colonies. About 500 G418 resistant colonies were found on this plate. By extrapolation, there are around 10,000 independent gene-trap clones in the gene-trap library. The G418 resistant ES cell colonies from the remaining 19 retrovirus-infected plates were maintained as 19 subpools (WW99-1 to 19).

A Cre expression plasmid was electroporated into the subpools of gene-trap clones, WW99-1 to 19. The recombinants were selected in M15 supplemented with puromycin. Most of the plates had more than 100 puromycin resistant colonies. Some plates (WW103-RT-3, 10, 17 and 19) had less than 100 colonies, while WW103-RT-16 only had about 10 puromycin resistant colonies. The variation in puromycin resistant colony number from plate to plate might represent the variation in the proportion and position of gene-traps on the right chromosome. The puromycin resistant ES cell colonies were pooled together to make WW103-RT-1 to 19.

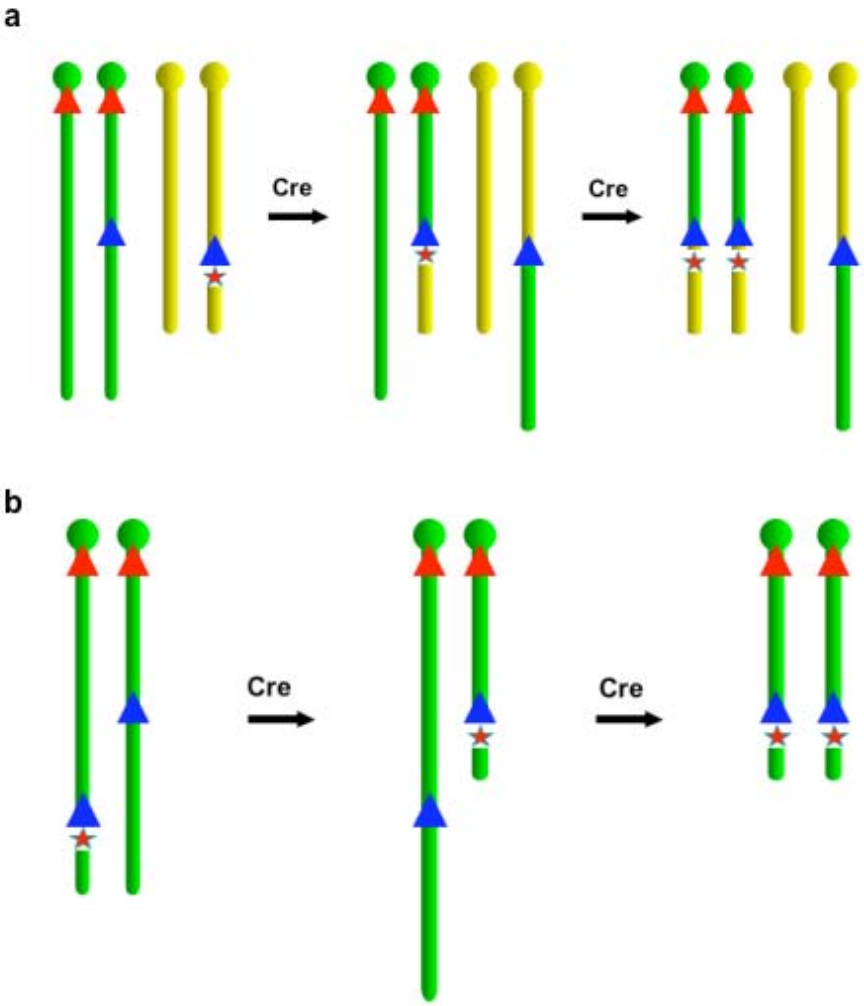


Fig. 3-20 Balanced translocation and deletion/duplication events can not be made homozygous. a. Balanced translocation events cannot become homozygous because this will cause loss of one copy of the distal part of chromosome 11. **b.** Balanced deletion/duplication events cannot become homozygous because this will cause loss of both copies of the distal part of chromosome 11. Red arrow, mutant *lox* sites used for induced mitotic recombination; blue arrow, wild type loxP sites used for regional trapping; red star, gene-trap mutation; green and yellow bar, different chromosomes.

3.2.6 Regional trapping experiment using electroporation-based plasmid

Linearized pWW237 plasmid DNA was electroporated into WW69-C6 ES cells. The recombinants were plated on 10X 90-mm feeder plates and selected in M15 supplemented with G418. The G418 resistant colonies were maintained as 10 subpools (WW100-1 to 10).

Cre expression plasmid was electroporated into the subpools of plasmid-based gene-trap clones, WW100-1 to 10. The recombinants were selected in M15 supplemented with puromycin. The puromycin resistant ES cell colonies were pooled together to make WW104-1 to 10.

3.2.7 Induced mitotic recombination

Both the induced mitotic recombination and the regional trapping are mediated by the Cre/*lox*P system. The Inducible mitotic recombination cassettes contain the mutant *lox* sites, *lox*5171 and *lox*2272, and the regional trapping cassettes contain the wild-type *lox*P site. These variant *lox* sites were used because they can efficiently recombine with themselves but not with each other. This reduces the chance that the two events (inversion and mitotic recombination) will interfere with each other.

However, there is possibility that the mitotic recombination and the inversion events can happen simultaneously at the first round of Cre expression, which will produce HAT and puromycin double resistant clones. In such a circumstance, when the puromycin resistant colonies are pooled, every single cell from the double resistant colonies will be able to survive the HAT selection. If this happens, these cells will dominate the pool and it will be almost impossible to identify other HAT resistant clones from the same pool after mitotic recombination induced by the second round of Cre-mediated recombination.

The induced mitotic recombination rate at the *D11Mit71* locus, which I used to create the induced mitotic recombination cell line, is $3.5 \pm 1.8 \times 10^{-4}$/cell

electroporated (Liu, Jenkins et al. 2002). And the efficiency of Cre recombination over a physical distance of 34 Mb (*Wnt3-p53*) is 2.2 ± 0.6 X 10^{-3}/cell electroporated (Zheng, Sage et al. 2000). So it is reasonable to predict that the chance of the two events happening simultaneously is very low.

In the large-scale experiments, Cre expression plasmid was electroporated into the pools of inversion clones, WW103-RT-1 to 19. The recombinants were selected in M15 supplemented with HAT. Most of the plates had around 1000 HAT resistant colonies, which is comparable with the colony number obtained from the parent cell line WW69-C6 after Cre transient expression. But for the pools WW103-RT-1 and 15, the cell density was too high to form single colonies. So no colonies were picked for these two plates. It is possible that for these pools, prior to the second round of Cre-mediated recombination, some clones were already HAT resistant. For the remaining plates, 48 colonies were randomly picked and cultured on 96-well feeder plates.

Genomic DNA was extracted from the 96-well plates and Southern analysis was carried out to determine the genotype of every HAT resistant clone. Because the inversions become homozygous when mitotic recombination is induced by Cre expression, both end points of the inversions are homozygous for the targeted alleles (Fig. 3-21). It is impossible to identify the genotype of the trapped loci without knowing the proviral integration sites. It is not feasible to carry out 1,000 inverse PCR or splinkerette PCR to identify the trapped locus for every HAT resistant clone and design allele specific probes to verify the genotypes of each one of them. So the genotype of the *E_2DH* locus was used to determine the genotype of the other ends of the inversions.

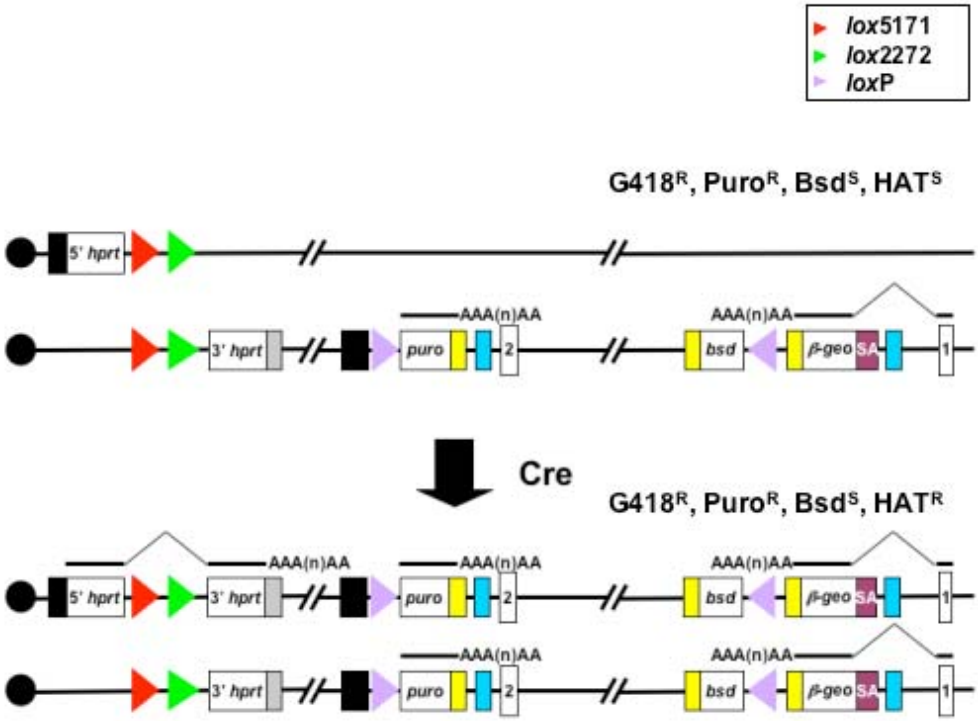


Fig. 3-21 Induced mitotic recombination to make the inversions homozygous. Schematic illustration of mitotic recombination to make the inversions homozygous. Black box, *PGK* promoter; yellow box, bovine growth hormone polyA site; grey box: SV40 polyA site; blue box, long terminal repeat; SA, splicing acceptor; arrows of different colours, variant *lox* sites. Note that the targeted E_2DH locus also becomes homozygous after induced mitotic recombination. So the genotype of the E_2DH locus can be used to represent the genotypes of the other ends of the inversions.

Individual recombination events were identified by their unique proviral/host junction fragments generated by *Eco*RI and *Spe*I digestion (*Eco*RI and *Spe*I are unique sites in the virus). If two clones from a same pool exhibit the same-sized proviral junctions fragments by two different restriction enzyme digestions (*Eco*RI and *Spe*I), they were considered as daughter clones from the same gene-trap and recombination event, and were therefore grouped together.

For most pools, homozygous clones were identified by Southern analysis using an E_2DH 3' probe (Fig. 3-22a and b). But from pool WW103-2, all of the HAT resistant clones are heterozygous and they all have the same-sized *Eco*RI proviral/host junction fragments (Fig. 3-22c and d). One possible explanation for this is that in this pool, a G2-Z event occurred at the same time as the inversion event, which resulted into a HAT resistant, puromycin resistant heterozygous clone. This double resistant heterozygous clone would expand with the pool and these cells would be much more numerous than the other recombinants after the second Cre-mediated recombination event. Homozygous clones were recovered from the other 16 pools (WW103-3 to WW103-14, WW103-16 to WW103-19).

The homozygous clones from the 16 plates were classified according to the sizes of their proviral junction fragments. For the groups that were represented by more than one clone, at least 2 independent clones were expanded. For the groups that had only one clone, the clone was expanded. Genomic DNA and RNA were extracted from all of the expanded clones. Southern analysis was carried out using different probes and enzyme digestions to confirm their genotypes (Fig. 3-23).

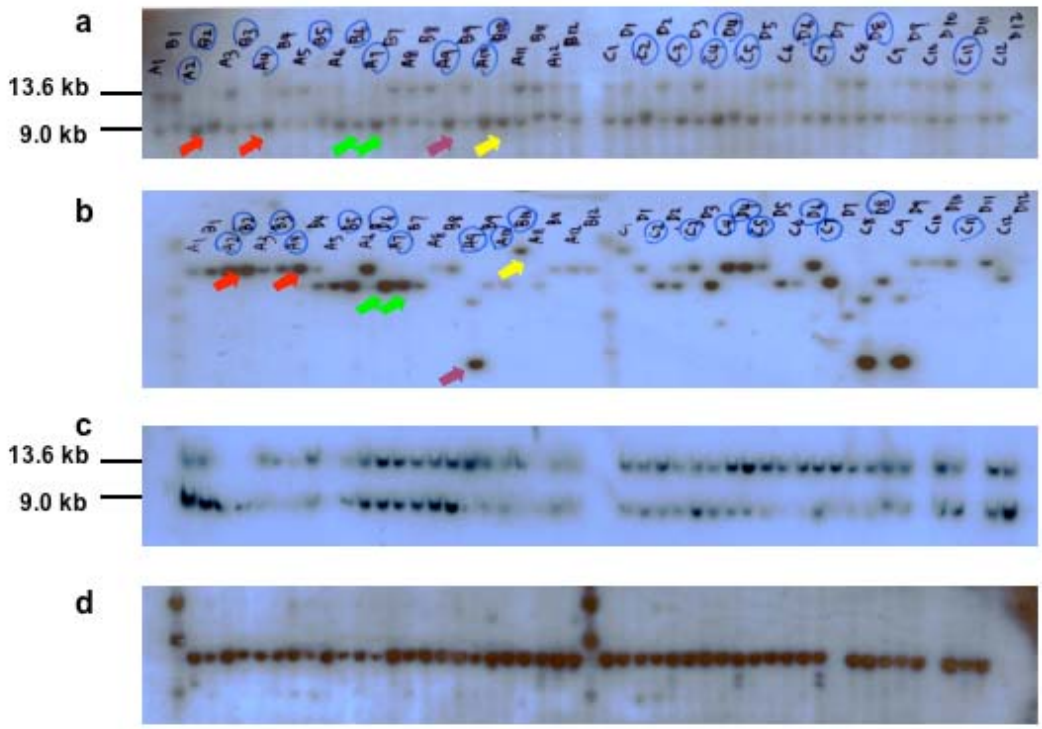


Fig. 3-22 Identification of the induced mitotic recombination clones. **a.** Mini-Southern analysis of clones from WW103-7. Genomic DNA was digested with *Nde*I and hybridized with the E_2DH 3' probe. Homozygous clones only have the targeted restriction fragment of 9.6 kb, and lack the wild type restriction fragment (13.1 kb). **b.** Mini-Southern analysis of clones from WW103-7. Genomic DNA was digested with *Eco*RI and hybridized with the *lacZ* probe. Individual recombination events were grouped according to their unique proviral junction fragments. Arrows of the same colour identify clones with the same sized *Eco*RI junction fragments. **c.** Mini-Southern analysis of clones from WW103-2. Genomic DNA was digested with *Nde*I and hybridized with the E_2DH 3' probe. No homozygous clones were identified. **d.** Mini-Southern analysis of clones from WW103-2. Genomic DNA was digested with *Eco*RI digestion and hybridized with the *lacZ* probe. All the clones showed the same sized *Eco*RI junction fragments.

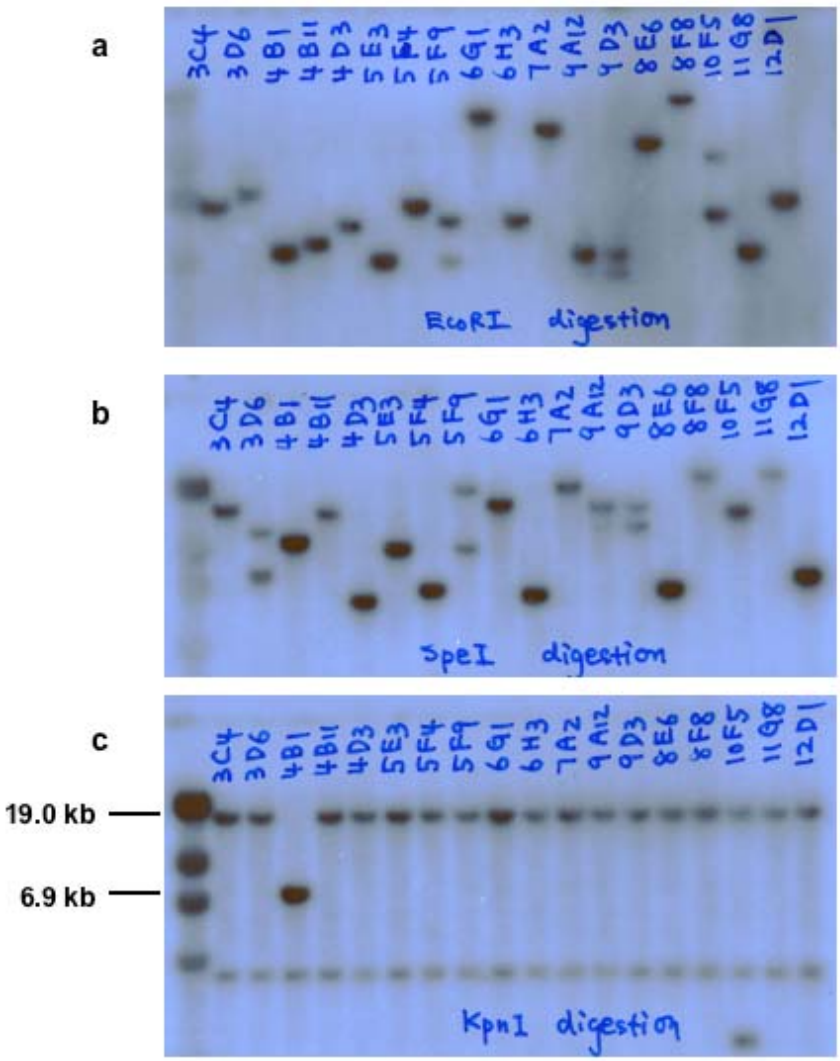


Fig. 3-23 Confimation of the homozygous clones. a. Genomic DNA extracted from WW103 homozygous clones was digested with *Eco*RI and hybridized with a *lacZ* probe. **b.** Genomic DNA from WW103 homozygous clones was digested with *Spe*I and hybridized with a *lacZ* probe. **c.** Genomic DNA from WW103 homozygous clones was digested with *Kpn*I and hybridized with a *lacZ* probe. Note that WW103-4B1 has only 6.9 kb proviral insertion fragment , and all the other clones only have the 19.0 kb inversion fragment .

Most of the clones only have a 19 kb *Kpn*I inversion fragment detected by the lacZ probe, some clones have only the 6.9 kb proviral insertion fragment, while the rest will have both. Sometime, these three different genotypes were found in clones of the same group. The explanation for this is that expression of Cre to induce mitotic recombination will in some case revert the inversion, especially the small ones. Depending on the sequence of these two events, the resulting clones can carry two gene-trap alleles or one inversion allele and one gene-trap allele (Fig. 3-24). These clones with reverted inversions can not be distinguished from other inversion clones by drug selection because they still have one functional *Puro* selection marker. A total of 146 clones, which could be classified into 66 independent groups, were expanded and DNA and RNA samples were taken to identify the virus integration sites and the trapped exons.

Cre expression plasmid was also electroporated into pools WW104-1 to 10. The recombinants were selected in M15 supplemented with HAT. 48 HAT resistant colonies were picked and Southern analysis was carried out to genotype the clones using the same strategy described before for the WW103 pools. The *E_2DH* 3' probe successfully identified homozygous inversions. However, the *lacZ* probe failed to identify unique junction fragments. Irrespective of which restriction enzyme was used, multiple bands were detected for almost all the clones. It was therefore difficult to group the clones according to their digestion pattern. In principle, 5' RACE followed by sequence analysis could be used to identify the trapping exons of all the homozygous clones. But this approach is very labour-intensive, so no further characterization was carried out for the clones generated by plasmid-based gene-trap strategy.

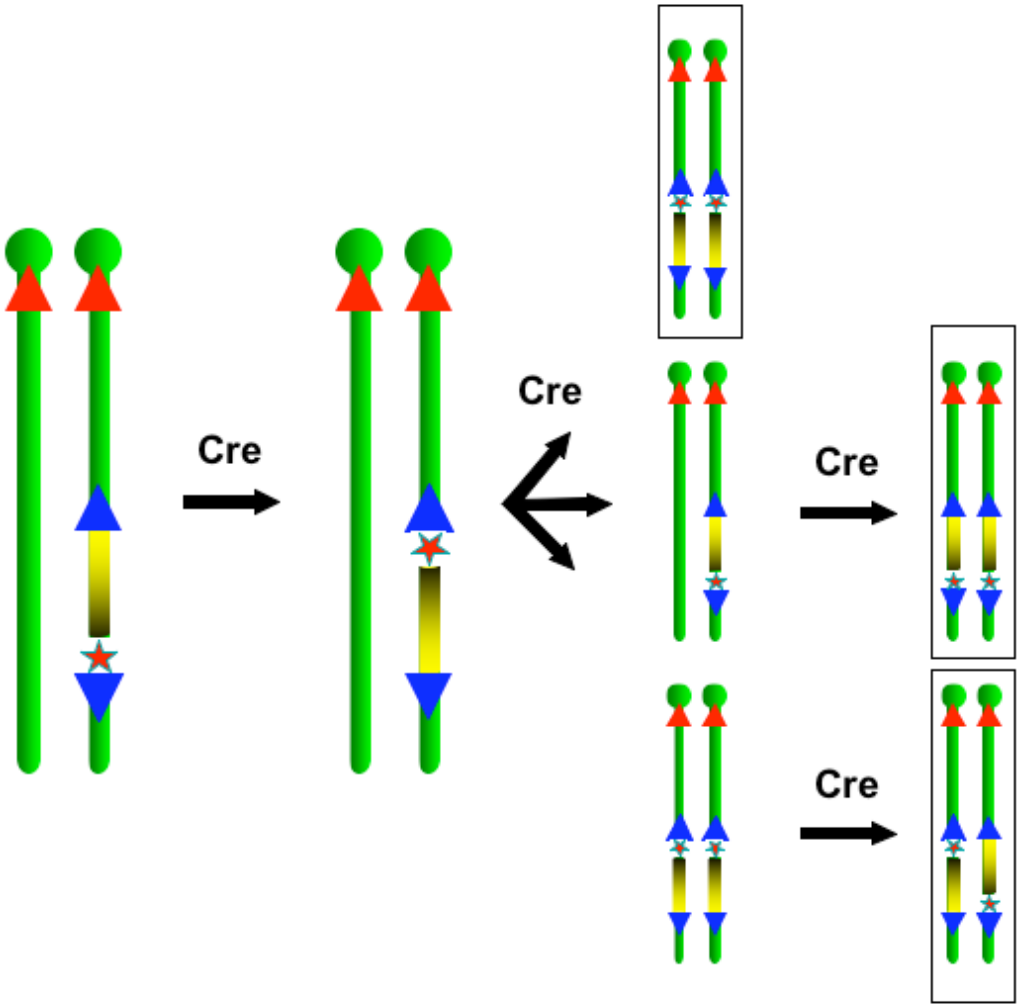


Fig. 3-24 Revertible inversion. The regional inversion can be inverted back with transient expression of Cre. The efficiency of the reverse inversion will be extremely high if the original inversion happens over a short physical distance. Depending on the sequence of induced mitotic recombination and the reverse inversion, the resulting homologous clones can carry either two trapping alleles or one trapping allele and one inversion allele. Red arrow, mutation *lox* sites used for induced mitotic recombination; blue arrow, wild loxP sites used for regional trapping; red star, gene-trap mutations; green bar, chromosome 11; yellow bar, the inversion region.

3.3 Discussion

As discussed in the previous chapters, homozygous mutant mouse ES cells are a very important resource for functional studies *in vitro*. The existing methods to generate homozygous clones require designing and constructing targeting and genotyping strategies for each different gene, which is difficult to scale up. This has greatly limited the effort to utilise homozygous mutant ES cells for genetic screens.

We have designed a strategy to circumvent this bottleneck. Heterozygous mutations were generated by regional trapping, and these mutations were converted to homozygosity by induced mitotic recombination. Strong selection strategies, a split *Hprt* minigene and a split *PGK*/*Puro*, were chosen to recover these rare events. Variant *lox* sites were used to avoid interference between the two separate selection systems.

For regional trapping, the first *lox*P site was targeted to the E_2DH locus, and the second *lox*P site was introduced into the genome by a retroviral vector. As the result of random integration of the retrovirus, the direction of the two *lox*P sites can be either the same or opposite. The *lox*P sites can also be located on the same chromosome (*cis*) or on different ones (*trans*). The outcome of the recombination event is directly determined by the location and the direction of the second *lox*P site (Fig. 3-25).

Of the four possible recombination products generated from a cell in G1 phase (G2 events will be discussed in the next chapter), inversion, balanced deletion/duplication and translocation do not result in loss of genetic material, and are therefore viable, unless the chromosomal breakpoints disrupt gene(s) that are essential for ES cell self-renewal. Dicentric/acentric chromosomes are not viable. For the deletion, the viability of the resulting recombinant depends on the size of the deletion and the genes in the deletion region. In the nested deletion experiment (Su, Wang et al. 2000), most of the deletions recovered were mapped within 1 cM distal or proximal to the anchor point. It is possible that deletions larger than that will cause haplosufficiency. Other efforts to generate large deletions also encountered the same problem (Liu,

Zhang et al. 1998; Zheng, Sage et al. 2000). Although the recombination can still occur, the resulting recombinants sometimes duplicate the wild-type chromosome to compensate for the loss caused by the large deletion.

However, no homozygous deletion clones have been identified in my experiment. If the two *lox*P sites are in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is proximal to *Hsd17b1,* the resulting deletion cells can not survive subsequent puromycin selection because the *Puro* cassette is deleted (Fig. 3-26a). If the two *lox*P sites are in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is distal to *Hsd17b1*, the resulting deletion cells can survive subsequent puromycin selection. However, even if the heterozygous (before mitotic recombination) and homozygous (after mitotic recombination) deletions do not cause haplosufficiency or homozygous lethality in ES cells, these cells will not be identified by Southern because the β-*geo* cassette is deleted (Fig. 3-26b) and will therefore be discarded from the screen.

In a pilot experiment carried out to test the experimental design. 100 gene-trap clones were pooled and inversions were induced by Cre transient expression. From this experiment, puromycin resistant clones were successfully recovered and Southern analysis confirmed the recombination events. It is interesting to notice that most of the clones in this pilot had the same-sized proviral junction fragments. These clones are likely derived from a recombination event of high efficiency, most likely a small deletion or inversion. Those clones that have rare proviral junction fragments might represent inefficient recombination events, such as balanced translocations, balanced deletion/duplications, or large inversions.

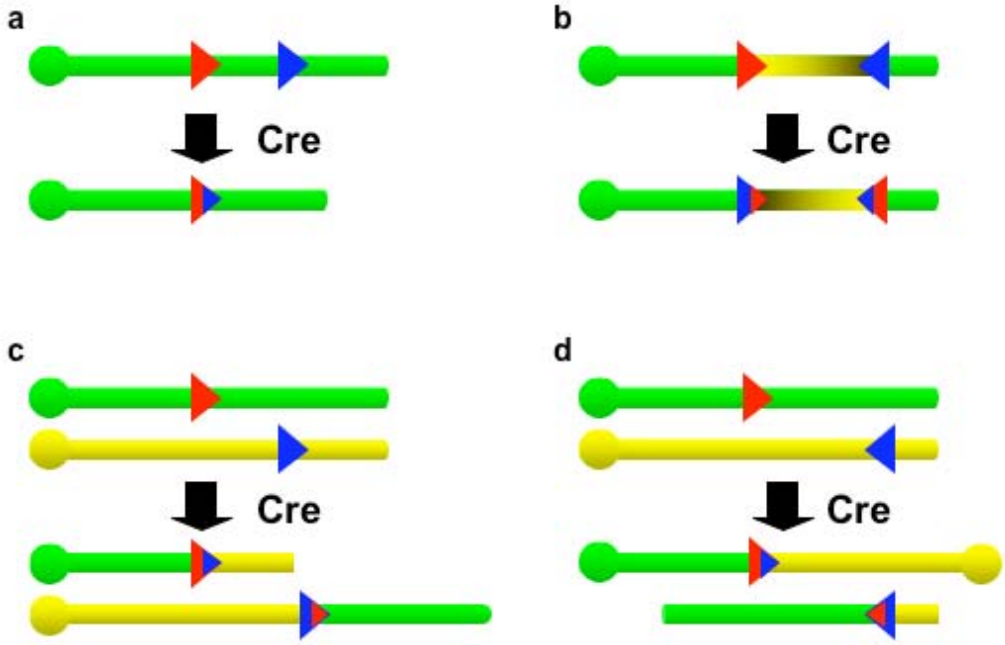


Fig. 3-25 Possible results of the Cre-mediated recombination (G1). a. *lox*P sites in direct orientation on the same chromosome (*cis*). A deletion will be generated by Cre-mediated recombination. The deletion might cause haplosufficiency depending on the size of the deletion. **b.** *lox*P sites in inverted orientation on the same chromosome (*cis*). An inversion will be generated by Cre-mediated recombination. **c.** *lox*P sites in direct orientation on different chromosomes (*trans*). A balanced deletion/duplication will be generated by Cre-mediated recombination. **d.** *lox*P sites in inverted orientation on different chromosomes (*trans*). A deletion will be create by Cre-mediated recombination. A dicentric and an acentric chromosome will be generated by Cre-mediation recombination, and the resulting cells are not viable. Red arrow,end point *lox*P site; blue arrow, *lox*P site introduced by retrovirus.

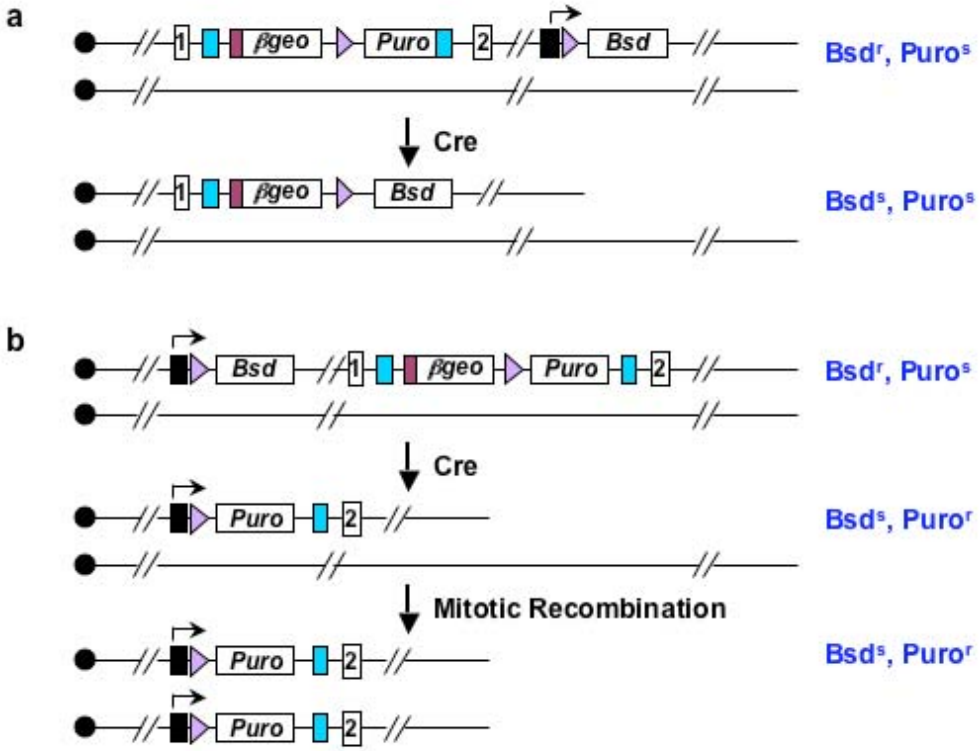


Fig. 3-26 Deletion. ***a.*** *lox*P sites in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is proximal to *Hsd17b1*. A deletion will be generated by Cre-mediated recombination. The resulting cells can not survive subsequent puromycin selection because the *Puro* cassette is deleted. ***b.*** *lox*P sites in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is distal to *Hsd17b1*. A deletion will be generated by Cre-mediated recombination. The resulting cells can survive subsequent puromycin selection. However, even if the heterozygous (before mitotic recombination) and homozygous (after mitotic recombination) deletions do not cause haplosufficiency or homozygous lethality in ES cells, they will not be identified by Southern because the *β-geo* cassette is deleted. Black box, *PGK* promoter; lavender triangle, wild-type *lox*P site; blue box, virus long terminal repeat; plum box, splice acceptor.

A balanced translocation or deletion/duplication will not result in loss of genetic material, and therefore should be recovered after the first Cre-mediated recombination and subsequent puromycin selection. However the Cre-induced mitotic recombination and HAT selection will select against those clones with translocations, because one copy of the distal part of chromosome 11 will be lost after mitotic recombination (Fig. 3-19). For a balanced deletion/duplication clone, depending on the position of the deletion chromosome, mitotic recombination will either result in a homozygous deletion (deletion occurs on the same chromosome as the 3' *Hprt* cassette) or a homozygous duplication clone (deletion occurs on the same chromosome as the 5' *Hprt* cassette). While the homozygous deletion clones are not viable, the homozygous duplication clones should survive because they do not loss any genetic material. This possibility will be discussed in more detail in the next chapter.

When the induced mitotic recombination clones from the same pool were analyzed by Southern, most of the homozygous clones exhibited a limited number of proviral junction fragments. It is likely that those clones carry very small inversions. The efficiency of generating small inversions is so high that most of the puromycin resistant colonies have arisen from one parental clone. Assuming such clones are viable after mitotic recombination, they will also dominate the population of HAT resistant colonies. The dominance of these small inversions has the potential to limit the coverage of the screen.

There are two ways to avoid this bias. First, more HAT resistant clones can be picked from each induced mitotic recombination pool for Southern analysis. Second, a smaller pool of trapped clones can be used for regional trapping (Wentland, unpublished data). By calculation, less than 1% the trapped clones will be on the right chromosome and in the right direction for an inversion. So in a pool of 100 trapped events, 0-1 inversion is expected after the Cre electroporation. If a pool yields many puromycin resistant colonies, these are likely to be the same inversion represented by many subclones. If just a few puromycin resistant colonies are recovered, they might reflect large inversion, deletion/duplication or translocation events. After induced mitotic

recombination, clones derived from the large inversion will dominate the pool of HAT resistant clones because they have a selection advantage over the translocations and deletion/duplications. The second method proved to work very well in Wentland's experiment. Inversions as big as 100 Mb were recovered (Wentland, unpublished data), but this method greatly increases the number of electroporation needed. For a total of 10,000 trapped clones, 100 electroporations are needed for regional trapping and another 100 electroporations for induced mitotic recombination. The extra effort can not guarantee the generation of 10 times more unique homozygous clones. So we decided to use a much larger pool of 500 trapped clones for the experiment. It is likely that most large inversions will be eliminated in the selection process.

Another issue of concern is that regional trapping and induced mitotic recombination can happen at the same time. Though variant *lox* sites were used to avoid interference between the two events, they can still take place simultaneously and generate puromycin and HAT double resistant clones after the first Cre-mediated recombination event. If this happens, the descendents of the double resistant colonies will dominate the newly formed HAT resistant colonies after the second Cre-mediated recombination. The use of Flp/*FRT* system is a possible alternative for one of the events. If Cre/*lox*P is used for regional trapping and Flp/*FRT* is used for induced mitotic recombination, then the two events can be separated. To test this system, an *FRT* site was inserted into multi-*lox* site linker (Fig. 3-2c) and a FPL expression plasmid, p*CAG*G-FLPe (Genbridge), was electroporated into the D11Mit71$^{5'\ Hprt\ /\ 3'\ Hprt}$ cell line, WW45. and recombinants were selected in M15 supplemented with HAT. Though HAT resistant colonies were recovered, the efficiency is about 2 to 3 orders of magnitude lower than Cre/*lox*P system (data not shown). So in the large-scale experiments, Cre was used for both regional trapping and induced mitotic recombination.

Homozygous inversions were identified from the HAT resistant clones from 16 out of 19 pools. For the other three pools, two (WW103-1 and 15) had too many HAT resistant colonies and one (WW103-2) was composed of the

heterozygous clones with the same proviral junction fragment. Most likely, in these pools, mitotic recombination was induced at about the same time as the inversion. Because the induced mitotic recombination will generate a functional *Hprt* mini-gene, it is possible to select against the double resistant colonies using 6-thioguanine (6-TG) after the first Cre mediated recombination because in cells with functional *Hprt* gene, 6-TG can be used to produce 2'-deoxy-6-thioguanosine-triphosphate, the active guanine nucleotide analogue in DNA synthesis, and thus kill the HAT^R cells.

Both 5' trapping and 3' trapping constructs were designed and tested. The titre of the 3' trapping virus is slightly lower than the 5' trapping virus. Considering that 3' trapping is more likely to trap cryptic splice acceptors and pseudo polyadenylation signals scattering throughout the genome, the mutagenicity of 3' trapping is not as high as 5' trapping. So we decided to choose 5' trapping strategy for the large-scale experiments. However, 3' trapping can trap genes that do not normally express in undifferentiated ES cells, which is an advantage for *in vitro* differentiation studies. For example, a gene required for mesoderm formation but not expressed in undifferentiated ES cells can only be mutated by 3' trapping strategy.

Both retroviral- and plasmid-based trapping constructs were designed and used in the large-scale experiments. The classification of trapped clones into groups according to their insertion/host junction fragments was straightforward for the clones generated from the virus, however the clones generated by electroporation of linearized trapping vectors always displayed multiple fragments which made them difficult to be classified by their Southern pattern. The multiple fragments detected can be caused by either concatermerization at a single insertion site or multiple insertions throughout the genome. Though conditions can be optimized to minimize the possibility of concatermerization, it will still occur in about 20% cells (Stanford, Cohn et al. 2001). Concatermerization can also result in ectopic reporter expression leading to expression of the reporter without trapping an endogenous gene. Also, the gene-trap vectors can be randomly truncated when they integrate into the genome. The differing lengths of the truncation make the cloning of

the flanking genomic sequence by Inverse PCR problematic. Since trapping by the retroviral vector generated enough clones for downstream analysis, the clones generated by electroporation have not been further characterized.

Southern analysis of the proviral junction fragments identified 66 different homozygous clones from a total of 16 plates of gene-trap clones. Each original trapping plate contains about 500 independent trapping events. So the proportion of homozygous gene-trap clones recovered from 8000 gene-trap events 0.75%. This efficiency is almost the same as we predicted before the experiment, thus the two selection strategies used to induce regional trapping and mitotic recombination are efficient enough to isolate these rare events.

In the chapter, I described the variables I tested and the strategy I chose to isolate homozygously mutated ES cell clones by regional trapping and induced mitotic recombination. Genotyping is not required for the trapping and inversion events. A common genotyping strategy was used at the final stage to genotype all the clones that underwent induced mitotic recombination. This has greatly simplified the genotyping procedure for identifying a large number of homozygous mutant clones. The different lengths of proviral/host flanking fragments were used to group the homozygous clones from the same pool. This procedure reduces the redundancy of the clones which is caused by the use of pooling to handle large numbers of clones. In turn, this reduces the number of clones that needed to be identified by 5’ RACE and splinkerette PCR. The results obtained proved that our strategy can efficiently isolated homozygous mutant clones without any previous knowledge of the loci that have been disrupted. This is an obvious advantage compared to traditional methods to generate homozygous mutant clones. Conditions can still be optimized to improve the yield. Therefore, I have shown that this strategy can be used to isolate homozygous mutant ES cell clones in an efficient way.

4 Analysis of gene traps on chromosome 11

4.1 Introduction

In the previous chapter, I showed that I could efficiently generate homozygous gene-trap mutations on mouse chromosome 11. By Southern analysis, independent gene-trap events from different pools were identified according to the sizes of the proviral/host junction fragments generated by different restriction enzymes. A total of 66 different homozygous gene-trap events have been isolated. 146 ES cell clones representing these 66 gene traps were expanded, and DNA and RNA samples were used to identify the virus integration sites and the trapped exons.

Using the retroviral vector 5' gene-trap strategy, it is possible to identify the gene-trap locus by Splinkerette PCR, Inverse PCR and 5' RACE. All these methods have been tested in large-scale insertion mutagenesis experiments (Mikkers, Allen et al. 2002; Suzuki, Shen et al. 2002; Hansen, Floss et al. 2003). By comparison of the results of a pilot experiment, I decided to use a combination of Splinkerette PCR and 5' RACE for the large-scale experiment.

4.1.1 Splinkerette PCR versus inverse PCR

Inverse PCR is the original method to clone proviral/host junction fragments. First, a restriction endonuclease that cuts only once within the provirus is used to digest the genomic DNA. The completely digested DNA is then self-ligated to form circles at a low DNA concentration and the flanking fragments are amplified using proviral DNA specific primers. Nested PCR is performed to improve the sensitivity of the reaction and specificity of the inverse PCR products. The final products are cloned into plasmids to facilitate sequencing (Fig. 4-1). The inverse PCR conditions can be optimized to amplify fragments as large as 12 kb (Li, Shen et al. 1999). Obtaining large flanking fragments was a great advantage before the mouse genomic sequence was finished. The bigger the fragment, the better chances are that a sequencing result can match a known gene or EST sequence, and determine the proviral insertion site.

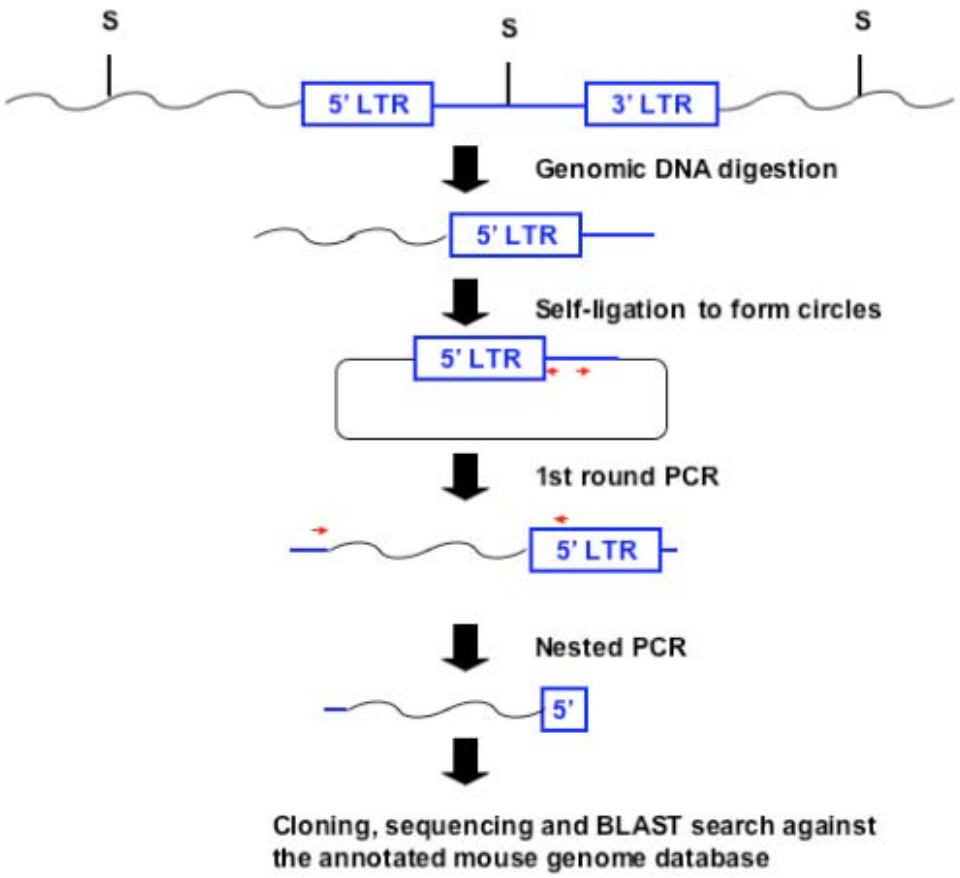


Fig. 4-1 Inverse PCR. First, a restriction enzyme is used to cut the genomic DNA. The completely digested genomic DNA is then self-ligated at low concentration to form circles. The flanking fragments are amplified using proviral DNA specific primers. Nested PCR is performed to improve the sensitivity and specificity of the Inverse PCR products. The final products are cloned into a plasmid to facilitate sequencing.

However, the inverse PCR strategy also has its disadvantages. First, to amplify a fragment big enough to determine the insertion site, a relatively rare cutting restriction enzyme is chosen for digestion of the genomic DNA. Long-range PCR amplification not only limits the recovery rate of proviral flanking sequences, it also significantly increases the cost. Second, the inverse PCR fragments need to be cloned for sequencing. This step limits the speed of the isolation of proviral-flanking sequences.

Splinkerette PCR was recently introduced as an alternative method to clone the proviral flanking fragments. A splinkerette is a pair of oligonucleotides that are partially complementary. One of the oligonucleotides contains a hairpin loop that prevents nonspecific PCR amplification by inhibiting new DNA strand synthesis from the adaptor. The other oligonucleotide contains the bind sites for the two primers used for the two rounds of PCR amplification (Fig. 4-2a). First, a restriction endonuclease that cuts only once within the 5' LTR of the provirus is used to digest the genomic DNA. The completely digested genomic DNA is then ligated to the splinkerette adaptor. The flanking proviral fragments are amplified using a pair of primers homologous to the splinkerette and the 5' LTR, respectively. Nested PCR is performed to improve the sensitivity and specificity of the PCR products (Fig. 4-2b). The final products are purified from the gel and sequenced directly (Mikkers, Allen et al. 2002).

Compared to inverse PCR, splinkerette PCR has some advantages. First, genomic DNA can be digested with a frequent cutter to get smaller fragments for amplification. Because the mouse genome is virtually complete, the size of the amplified flanking fragment is no longer a bottleneck for locus mapping. The development of the SSAHA (Sequence Search and Alignment by Hashing Algorithm) search engine makes it possible to find an exact or "almost exact" match between two sequences, even when the size of matched sequence is very small (Ning, Cox et al. 2001).

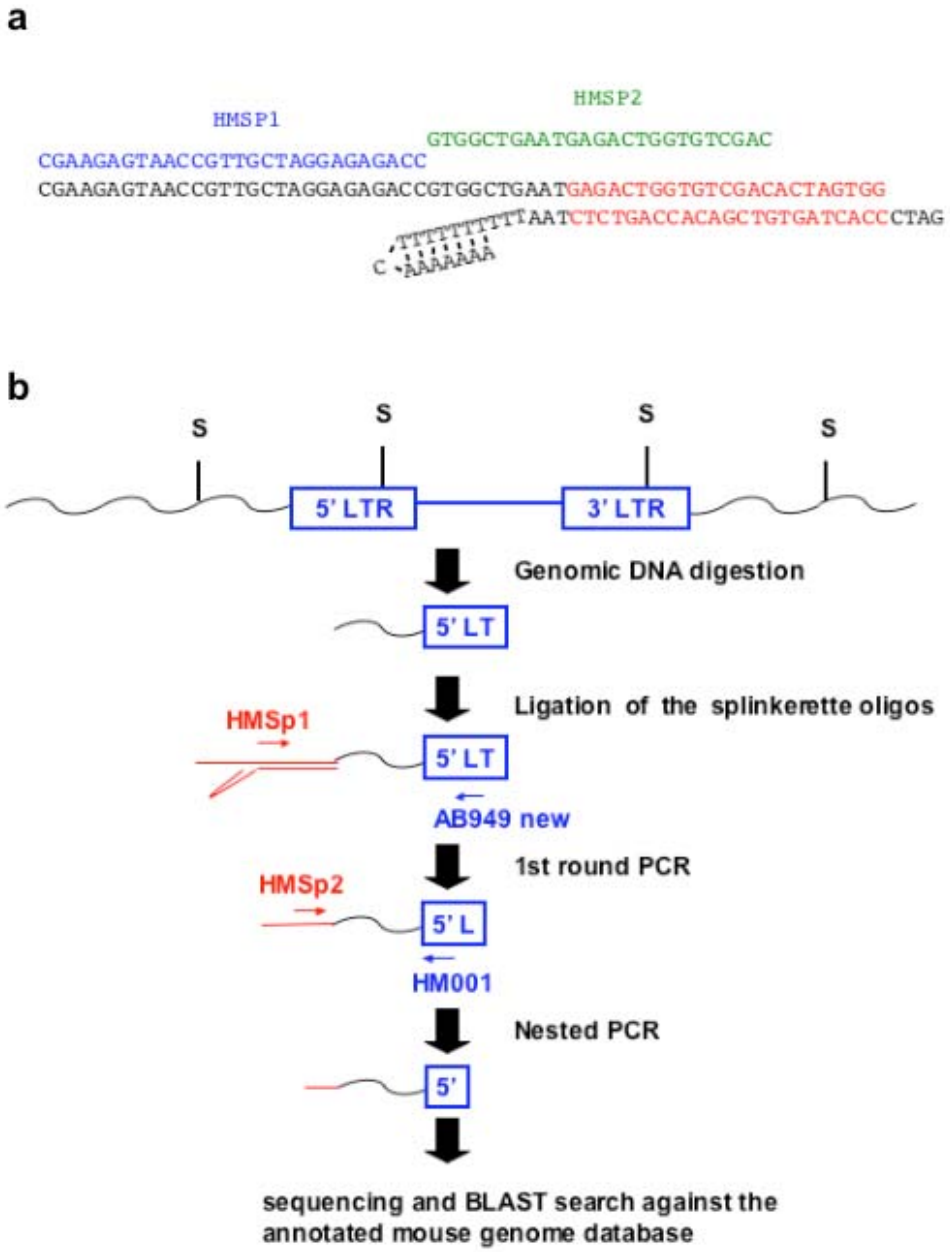

a
HMSP2
HMSP1
GTGGCTGAATGAGACTGGTGTCGAC
CGAAGAGTAACCGTTGCTAGGAGAGACC
CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGACACTAGTGG
TTTTTTTTTTAATCTCTGACCACAGCTGTGATCACCCTAG
C AAAAAAA
b
S
S
S
S
5' LTR
3' LTR
Genomic DNA digestion
5' LT
Ligation of the splinkerette oligos
HMSp1
5' LT
AB949 new
1st round PCR
HMSp2
5' L
HM001
Nested PCR
5'
sequencing and BLAST search against the
annotated mouse genome database

Fig. 4-2 Splinkerette PCR. a. The Structure of the splinkerette. A splinkerette is a pair of partially complementary oligos. One of the oligonucleotides contains a hairpin loop that prevents nonspecific PCR amplification by inhibiting new DNA strand synthesis from the adaptor. The other oligonucleotide contains the bind sites for the two primers used for the two rounds of PCR amplification. **b.** Splinkerette PCR. First, a restriction enzyme is used to cut the genomic DNA. The completely digested DNA is then ligated to the splinkerette adaptor. The flanking proviral fragments are PCR amplified. Nested PCR is performed to improve the sensitivity and specificity of the PCR products. The final products are purified from the gel and sequenced directly.

In principle, the PCR products generated by splinkerette PCR can be sequenced directly, but several background amplification products always coexist with the specific proviral insertion product because of endogenous viral sequences. So the PCR products need to be purified from a gel. This step has become the major bottleneck for splinkerette PCR.

4.1.2 5’ RACE

5’ Rapid Amplification of cDNA Ends (5’ RACE) is a method to amplify the 5’ region from an mRNA template between a defined internal site and the 5' end. To specifically amplify a rare template in a complex mixture usually requires two sequence-specific primers flanking the region of interest. This is not compatible with the need to amplify an unknown region with only one known end. 5' RACE methodologies offer a convenient way to solve this problem.

5' RACE, or “anchored” PCR, can be used to isolate and characterize 5' ends of low-copy mRNA templates. Although the 5’ RACE protocols vary from user to user, the general strategy is the same. First, a gene-specific primer is used for first strand cDNA synthesis. This step not only decreases the non-specific amplification, but also increases the possibility of obtaining the 5' end of a long mRNA template. Terminal deoxynucleotidyl transferase (TdT) is then used to add a homopolymeric tail to the 3' end of the cDNA. The 5’ end of the mRNA is then amplified using a pair of primers homologous to the homopolymeric tail and the internal anchor region, respectively. Nested PCR is performed to improve the product yield and specificity of the PCR product (Fig. 4-3). The 5’ RACE procedure can be utilized to amplify and characterize unknown coding sequences in gene-trap mutagenesis. It is an especially important technology for 5’ trapping strategy based on electroporation, because the flanking genomic fragments are difficult to isolate.

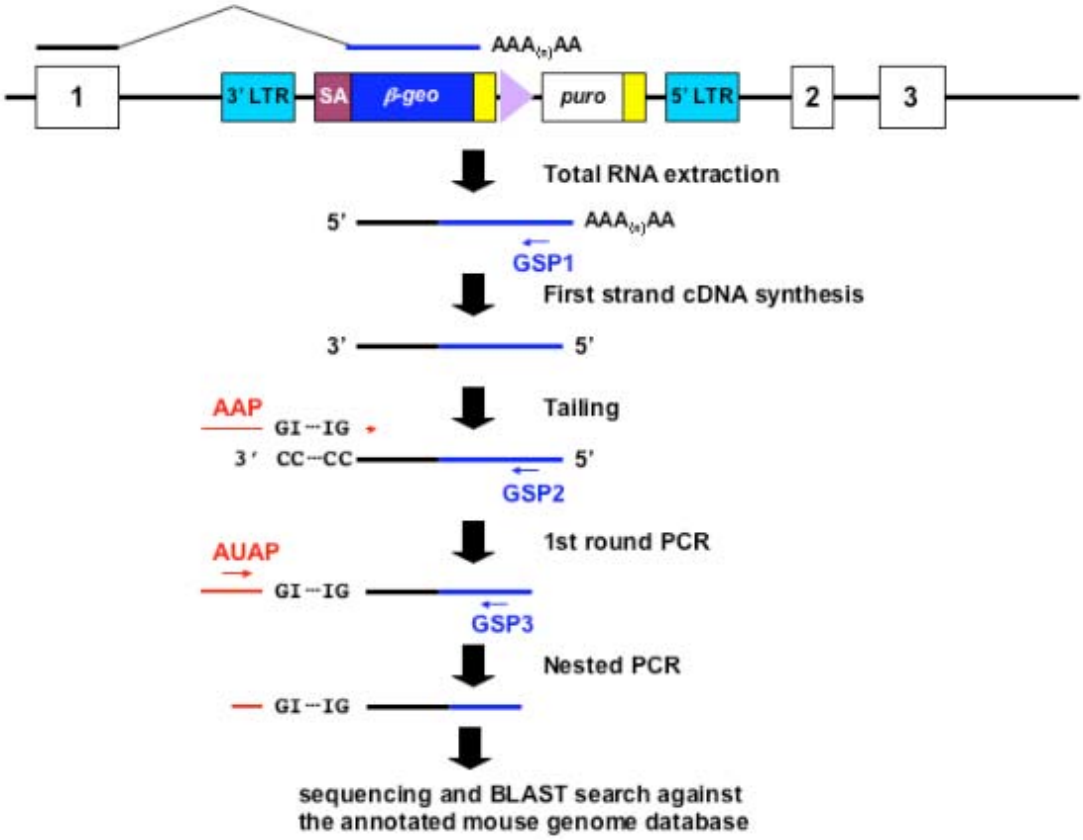


Fig. 4-3 5' RACE. First, a gene-specific primer is used for first strand cDNA synthesis. Terminal deoxynucleotidyl transferase (TdT) is then used to add homopolymeric tails to the 3' ends of the cDNA. The 5' end region of the mRNA is amplified by PCR. Nested PCR is performed to improve the sensitivity and specificity of the PCR products. SA, splicing acceptor; yellow box; bpA; purple arrow, *lox*P site, GSP, gene-specific primer; AAP, abridged anchor primer; AUAP, abridged universal amplification primer.

4.1.3 Distribution of the trapped genes on Chromosome 11

From the drug selection and Southern results, I determined that I had generated homozygous inversions on chromosome 11. Based on the knowledge of Cre efficiency over long genetic distances (Zheng, Sage et al. 2000), it was expected that a large number of the isolated homozygous gene traps would be within 10 Mb proximal and distal to the *E_2DH* locus (20 Mb in total). The results of the original regional trapping experiment (Wentland et al. unpublished data) has shown that 86% of the gene-traps were concentrated on the distal part of chromosome 11, and fell within a 43 Mb region surrounding the *E_2DH* locus. So the range of the regional trapping technology is much higher than expected.

Although I have chosen a different trapping and drug selection strategy for recovering the inversion events, I expected that the regional trapping efficiency should be comparable to the original regional trapping experiment. However, because I used a much larger pool to select out regional trapping events, I expected that I would not be able to isolate big inversions (the largest inversions isolated in the original regional trapping experiment is 82 Mb in size).

4.1.4 Orientation of transcription of the trapped genes

To generate an inversion, the two *lox*P sites must be in opposite orientation on the same chromosome. Therefore, the trapped genes should be transcribed from the antisense strand of the chromosome 11 (from telomere to centromere). The results of the original regional trapping experiment (Wentland et al. unpublished data) has shown that 17 out of the 21 trapped genes on chromosome 11 were transcribed from the antisense strand. However, the other four appeared to be transcribed from the sense strand (from centromere to telomere), but the drug resistance of these clones is the same as the other inversion clones. By fluorescent *in situ* hybridization (FISH), Wentland et al. (unpublished data) has found that the clones that have the expected drug resistance (HAT^R, $Puro^R$ and $G418^R$) but wrong transcription direction of the trapped gene, either have a duplication

chromosome and a wild-type chromosome, or balanced deletion/duplication chromosomes.

These results further confirmed the observations of the previous studies that large heterozygous chromosomal deletions will cause ES cell lethality (Su, Wang et al. 2000; Zheng, Sage et al. 2000). Cells with large heterozygous deletions can only survive if a second genetic change occurs to compensate for the loss of the genetic material caused by the deletion. Unbalanced deletions are rescued by a partial trisomy of two wild-type and one deletion chromosome. The majority of the *trans* recombination products will result in balanced deletion/duplication chromosomes.

4.2 Results

Splinkerette PCR was carried out for each of the expanded clones. To increase the possibility of generating PCR products with suitable size for sequencing, the genomic DNA extracted from the expanded ES cell clones was digested using *Sau*3AI, *Eco*RI or a combination of *Spe*I, *Xba*I and *Nhe*I, respectively. The derived splinkerette PCR products were separated on a 1% agarose gel. The specific PCR fragments were gel purified and sequenced using a pair of primers specific to the splinkerette and the 5' LTR of the retrovirus. The sequences were searched against the annotated mouse genome databases, Ensembl (http://www.ensembl.org/Mus_musculus) and NCBI (http://www.ncbi.nlm.nih.gov/BLAST).

5' RACE was carried out for at least one subclone from each group. RACE products were treated with Exonuclease I and Shrimp Alkaline Phosphatase to destroy the unused primers and dNTPs. The products were sequenced using a primer specific to the Splice Acceptor (SA) region of the trapping cassette. The sequences were also searched against the annotated mouse genome databases, Ensembl (http://www.ensembl.org/Mus_musculus) and NCBI (http://www.ncbi.nlm.nih.gov/BLAST).

5' RACE and/or Splinkerette PCR products were obtained from 49 of the 66 groups of independent recombination events (Table 4-1). The sequences from

44 of the groups matched sequences on chromosome 11. The other 5 groups matched sequences on chromosomes other than chromosome 11. For the remaining 17 groups, either no sequence information was obtained, or the sequence information from 5' RACE and Splinkerette PCR was inconsistent. So the exact identities of these clones are designated as "unknown".

Table 4.1: Gene trap loci.

Group	ES Cell Clone ID	Splinkerette product	RACE product	Invesion / Trapping	Gene Name	Position on the chromosome	Orientation of the transcirption
1	WW103-3A5	ENSMUSG00000018428		Inversion	Akap1 (A kinase (PRKA) anchor protein 1)	88501918 - 88535621 bp (88.5 Mb)	-
	WW103-3B3	ENSMUSG00000018428		Inversion	Akap1 (A kinase (PRKA) anchor protein 1)	88501918 - 88535621 bp (88.5 Mb)	-
	WW103-3B5			Inversion			
	WW103-3D7			Inversion			
2	WW103-3A8			Inversion			
	WW103-3D6	ENSMUSG00000035086		Inversion	Becn1 (coiled-coil, myosin-like BCL2-interacting protein)	100957075 - 100973390 bp (101.0 Mb)	-
3	WW103-3C4	ENSMUSG00000020737	ENSMUSG00000020737	Inversion	Hn1 (hematological and neurological expressed sequence 1)	115168484 - 115185516 bp (115.2 Mb)	-
	WW013-3D5	ENSMUSG00000020737		Inversion	Hn1 (hematological and neurological expressed sequence 1)	115168484 - 115185516 bp (115.2 Mb)	-
4	WW103-4A2			Inversion			
	WW103-4A10		ENSMUSG00000034520	Inversion	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4B3		ENSMUSG00000034520	Inversion/Trapping	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4B7			Inversion			
	WW103-4B11		ENSMUSG00000034520	Inversion	Gja7 (gap junction membrane channel protein alpha 7)	102470710 - 102490284 bp (102.5 Mb)	-
	WW103-4C5			Inversion/Trapping			
5	WW103-4A6	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
6	WW103-4A12						
7	WW103-4B1	ENSMUSG00000010342		Trapping	Tex14 (testis expressed gene 14)	87006730 - 87157487 bp (87.0 Mb)	+
8	WW103-4D11			Inversion			
	WW103-4B11			Inversion			
9	WW103-4D3	ENSMUSG00000020935		Inversion	6720485C15Rik (RIKEN cDNA 6720485C15 gene)	102665181 - 102688272 bp (102.7 Mb)	-
10	WW103-5E2			Trapping			
	WW103-5E3	ENSMUSG00000020923		Inversion	Ubtf (upstream binding transcription factor, RNA polymerase I)	101977115 - 101990113 bp (102.0 Mb)	-
	WW103-5F1	ENSMUSG00000020923		Inversion/Trapping	Ubtf (upstream binding transcription factor, RNA polymerase I)	101977115 - 101990113 bp (102.0 Mb)	-
11	WW103-5E4	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
	WW103-5H2			Inversion			
12	WW103-5E5	unknown locuc on chromosome 11		Inversion	unknown	74223628 - 74223809 (74.2 Mb)	-
13	WW103-5F4	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
14	WW103-5F9	ENSMUSG00000056649	ENSMUSG00000056649	Inversion	2810410L24Rik (RIKEN cDNA 2810410L24 gene)	119856472 - 119857805 bp (119.9 Mb)	-
15	WW103-5G11			3 Bands?			
16	WW103-5H4			Inversion			
17	WW103-6E2	ENSMUSG00000014195		Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-6E11	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-6F7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
18	WW103-6E5	Chr.X		strange band?			
	WW103-6F8	Chr.X		Trapping			
	WW103-6G4	Chr.X		Trapping			
19	WW103-6F3	ENSMUSG00000020715		Inversion/Trapping	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
	WW103-6H3	ENSMUSG00000020715		Inversion	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
	WW103-6H9	ENSMUSG00000020715		Inversion	Ern1 (endoplasmic reticulum (ER) to nucleus signalling 1)	106068745 - 106158921 bp (106.1 Mb)	-
20	WW103-6G1	ENSMUSG00000013415		Inversion	Igf2bp1 (insulin-like growth factor 2, binding protein 1)	95633855 - 95677035 bp (95.6 Mb)	-
21	WW103-6G2			Inversion/Trapping			
22	WW103-6H5			Inversion			
23	WW103-6G9	ENSMUSG00000001552		Inversion	Jup (junction plakoglobin)	100041744 - 100068862 bp (100.0 Mb)	-
24	WW103-7A2	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-7A4	ENSMUSG00000014195		Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-7C4	ENSMUSG00000014195	ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
25	WW103-7A7			Inversion			
	WW103-7B5			Inversion/Trapping			
	WW103-7C7			Inversion			
26	WW103-7A9	Chr.19		3 bands			
27	WW103-7B10			Inversion			
28	WW103-8E3	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8E6	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8F5	ENSMUSG00000020717		Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
	WW103-8G9	ENSMUSG00000020717	ENSMUSG00000020717	Inversion	Pecam1 (platelet/endothelial cell adhesion molecule 1)	106325342 - 106421753 bp (106.3 Mb)	-
29	WW103-8E10			Inversion/Trapping			
	WW103-8H5			Inversion/Trapping			
	WW103-8H9			Inversion			
30	WW103-8E11			Inversion			
31	WW103-8F2	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-8F6	ENSMUSG00000014195	ENSMUSG00000014195	Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-8G11	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
32	WW103-8F8	ENSMUSG00000038268	ENSMUSG00000038268	Inversion	OVCA2 (Ovarian Cancer-associated Gene 2)	74788498 - 74803029 bp (74.8 Mb)	-
33	WW103-8H8			Inversion			
34	WW103-9A1			Inversion/Trapping			
	WW103-9A2						
	WW103-9A9	ENSMUSG00000020530		Inversion	Gqnbp2 (gametogenetin binding protein 2)	84434394 - 84472403 bp (84.4 Mb)	-
	WW103-9B4			Inversion			
	WW103-9D4	ENSMUSG00000020530	ENSMUSG00000020530	Inversion	Gqnbp2 (gametogenetin binding protein 2)	84434394 - 84472403 bp (84.4 Mb)	-
35	WW103-9A5			Trapping			
	WW103-9A10	ENSMUSG00000019173		Inversion/Trapping	Rab5c (RAB5C, member RAS oncogene family)	100386134 - 100409313 bp (100.4 Mb)	-
	WW103-9A12			Inversion			
	WW103-9C6	ENSMUSG00000019173		Inversion	Rab5c (RAB5C, member RAS oncogene family)	100386134 - 100409313 bp (100.4 Mb)	-
	WW103-9D6			Inversion			
36	WW103-9B9			Inversion/Trapping			
	WW103-9C4			Trapping			
37	WW103-9D3			Inversion			
38	WW103-10E1			Inversion			
	WW103-10E2			Inversion			
	WW103-10F5		ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-

Table 4.1 (cont): Gene trap loci.

Group	ES Cell Clone ID	Splinkerette product	RACE product	Invesion / Trapping	Gene Name	Position on the chromosome	Orientation of the transcirption
39	WW103-10E4	Chr.9		Trapping			
40	WW103-10H3			Inversion			
	WW103-10H10			Inversion			
41	WW103-11E2	Chr.6 ?		Inversion			
42	WW103-11E3	ENSMUSG00000051378	ENSMUSG00000051378	Inversion	3000004C01Rik (RIKEN cDNA 3000004C01 gene)	102576655 - 102596217 bp (102.6 Mb)	-
	WW103-11F2	ENSMUSG00000051378	ENSMUSG00000051378	Inversion	3000004C01Rik (RIKEN cDNA 3000004C01 gene)	102576655 - 102596217 bp (102.6 Mb)	-
43	WW103-11G3			Inversion			
44	WW103-11G8	ENSMUSG00000040430		Inversion	Pitpnc1 (phosphatidylinositol transfer protein, cytoplasmic 1)	106882559 - 107008762 bp (106.9 Mb)	-
45	WW103-11H1			Trapping?			
46	WW103-12A1	ENSMUSG00000018537		Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
	WW103-12A2	ENSMUSG00000018537		Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
	WW103-12A3	ENSMUSG00000018537	ENSMUSG00000018537	Inversion	Pcgf2 (polycomb group ring finger 2)	97360948 - 97371469 bp (97.4 Mb)	-
47	WW103-12A7	ENSMUSG00000034247		Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
	WW103-12D1	ENSMUSG00000034247	ENSMUSG00000034247	Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
	WW103-12D2	ENSMUSG00000034247	ENSMUSG00000034247	Inversion	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
48	WW103-12C8	Chr.3		Trapping			
	WW103-12D7	Chr.3		Trapping			
49	WW103-13A8		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13B1			Inversion			
	WW103-13B3			Inversion/Trapping			
	WW103-13B8		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13C4		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13C10		ENSMUSG00000017119	Inversion	Brca1/Nbr1, bidirectional promoter region	101227295 - 101253068 bp (101.2 Mb)	-
	WW103-13D3			Inversion			
50	WW103-13A11	ENSMUSG00000034247	ENSMUSG00000034247	Inversion/Trapping	Plekhm1 (pleckstrin homology domain containing, family M (with RUN domain) member 1)	103036219 - 103083789 bp (103.0 Mb)	-
51	WW103-13C12	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074-88759792 bp (88.7 Mb)	+
	WW103-13D10	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074-88759792 bp (88.7 Mb)	+
	WW103-13D12	Mus musculus hypothetical LOC217071		Trapping	Gm525 (gene model 525, NCBI)	88745074 - 88759792 bp (88.7 Mb)	+
52	WW103-14E7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-14E8	ENSMUSG00000014195	ENSMUSG00000014195	Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-14F1	ENSMUSG00000014195	ENSMUSG00000014195	Inversion/Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
53	WW103-14F3	ENSMUSG00000001552		Inversion	Jup (junction plakoglobin)	100041744 - 100068862 bp (100.0 Mb)	-
54	WW103-14F11	ENSMUSG00000056649	ENSMUSG00000056649	Inversion	2810410L24Rik (RIKEN cDNA 2810410L24 gene)	119856472 - 119857805 bp (119.9 Mb)	-
55	WW103-16A3			Inversion			
	WW103-16B1	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16B2	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16D2	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
	WW103-16D4	ENSMUSG00000018727		Inversion	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
56	WW103-16B3	ENSMUSG00000010342	ENSMUSG00000010342	Trapping	Tex14 (testis expressed gene 14)	87006730 - 87157487 bp (87.0 Mb)	+
57	WW103-18E2	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G2	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G10	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
	WW103-18G12	ENSMUSG00000037992		Trapping	Rara (retinoic acid receptor, alpha)	98608821 - 98646063 bp (98.6 Mb)	+
58	WW103-18E8		ENSMUSG00000018362	Inversion	Kpna2 (karyopherin (importin) alpha 2)	106659761 - 106670589 bp (106.7 Mb)	-
59	WW103-18F1			Inversion			
	WW103-18F9		ENSMUSG00000020917	Inversion/Trapping	Acly (ATP citrate lyase)	100147484 - 100199024 bp (100.1 Mb)	-
	WW103-18F11		ENSMUSG00000020917	Inversion	Acly (ATP citrate lyase)	100147484 - 100199024 bp (100.1 Mb)	-
60	WW103-18G5	ENSMUSG00000014195		Trapping	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
	WW103-18G7	ENSMUSG00000014195		Inversion	Dnajc7 (DnaJ (Hsp40) homolog, subfamily C, member 7)	100253974 - 100291291 bp (100.3 Mb)	-
61	WW103-18G8	ENSMUSG00000041629		Inversion	D11Wsu99e (DNA segment, Chr 11, Wayne State University 99, expressed)	113332450 - 113355190 bp (113.3 Mb)	-
62	WW103-19A1	ENSMUSG00000006931		Inversion	1110036O03Rik (RIKEN cDNA 1110036O03 gene)	100079579 - 100085819 bp (100.1 Mb)	-
63	WW103-19A2			Inversion			
	WW103-19A9			Inversion			
64	WW103-19A6	ENSMUSG00000020917		Inversion/Trapping	D11Ertd636e (DNA segment, Chr 11, ERATO Doi 636, expressed)	113369301 - 113381146 bp (113.4 Mb)	-
65	WW103-19A8	ENSMUSG00000018160		Inversion	Pparbp (peroxisome proliferator activated receptor binding protein)	97823286 - 97864383 bp (97.8 Mb)	-
	WW103-19B1	ENSMUSG00000018160		Inversion	Pparbp (peroxisome proliferator activated receptor binding protein)	97823286 - 97864383 bp (97.8 Mb)	-
66	WW103-19D3	ENSMUSG00000040430		Inversion	Pitpnc1 (phosphatidylinositol transfer protein, cytoplasmic 1)	106882559 - 107008762 bp (106.9 Mb)	-

4.2.1 Gene trapping hot spots

The 44 unique events selected on chromosome 11 trapped 30 different loci. Several loci on chromosome 11 were trapped more than once. One locus was trapped 6 times (*Dnajc7*), a second locus was trapped 3 times (*Igf2bp1*), and 7 loci were trapped twice (*Pecam*, *2810410L24Rik*, *Tex14*, *Jup*, *Plekhm1*, *D11Erd636e* and *Pitpnc1*). These loci probably represent gene-trap or viral insertion hot spots. The Splinkerette PCR results showed that when different gene-trap insertions occurred at the same locus, they either occurred at different positions in the same intron or in different introns. This result shows that the bias is locus-specific, instead of sequence-specific (Hansen, Floss et al. 2003). One example is given below.

Igf2bp1 (insulin-like growth factor 2, binding protein 1) is also known as *CRD-BP* (c-myc mRNA coding region instability determinant binding protein) (Tessier, Doyle et al. 2004). This protein is a multifunctional RNA-binding protein, which can bind to *c-myc*, insulin-like growth factor II, β*-actin* and H19 mRNAs. By binding to different RNA substrates, this protein can affect their localization, translation, or stability. The protein level of *Igf2bp1* is high during foetal development and almost undetectable in normal adult tissues. But the expression of *Igf2bp1* is reactivated in some adult human tumours including breast, colon, and lung tumours (Tessier, Doyle et al. 2004), though the significance of this is not clear.

Two independent gene-trap events were found in this gene locus. A *Spe*I/*Xba*I/*Nhe*I splinkerette PCR product was obtained from one subclone of the group, WW103-6G1. Sequence of the PCR product matched the second intron of the gene (Fig. 4-4a). *Spe*I/*Xba*I/*Nhe*I and *Sau*3AI splinkerette PCR products were obtained from one subclone of the group, WW103-5F4. Sequences of both products matched the seventh intron of the gene (Fig. 4-4b).

a


IgDp1 Exon 3
EcoRI
PgII
EcoRI
PgII
EcoI
SacI
XbaI
Spliceosetta Fragment
SphI
Spliceosetta Fragment
Spliceosetta Fragment
PgII
Spliceosetta Fragment
Spliceosetta Fragment

b

Fig. 4-4 Two independent gene traps at a same locus, *Igf2bp1*. a. *Spe*I/*Xba*I/*Nhe*I splinkerette PCR product was obtained from one subclone of the first group, WW103-6G1. Sequence of the PCR product mapped to the second intron of *Igf2bp1*. **b.** *Spe*I/*Xba*I/*Nhe*I and *Sau*3A1 splinkerette PCR products were obtained from one subclone of the second group, WW103-5F4. Sequence of both products mapped to the seventh intron of the gene. Splinkerette results were underlined in green, the green arrow represents the direction from the 5' LTR of the virus to the endogenous restriction site. The exons of *Igf2bp1* are underlined in blue, the blue arrow represents the transcription direction of *Igf2bp1*.

4.2.2 Distribution of trapped genes on chromosome 11

Using the Ensembl and NCBI database, the sequences generated from the splinkerette PCR and 5' RACE products were mapped exclusively in a 45.7 Mb distal region of the mouse chromosome 11, surrounding the *E_2DH* locus. About two thirds of the trapped genes (23/33) were clustered within 5 Mb proximal and 5 Mb distal to the *E_2DH* locus (Fig. 4-5). This distribution shows that the efficiency of the Cre inside this 11.3 Mb region is much higher than outside. The biggest inversion distal to the *E_2DH* locus was 19.2 Mb in size. The trapped locus, *2810410L24Rik* (119.9 Mb), is very close to the telomere of the chromosome (121.5 Mb). The biggest inversion proximal to the *E_2DH* locus was 26.5 Mb. So it is reasonable to expect that if a locus in the middle of the chromosome 11 was chosen, the whole region from which homozygous gene-trap clones could be recovered will be even bigger.

4.2.3 Orientation of the transcription of the trapped genes

The orientations of the transcription of the trapped genes were determined according to the Ensembl database (Fig. 4-6). An inversion can only be generated if the trapped genes are transcribed from the antisense strand (from telomere to centromere) of the chromosome 11. In this orientation, the *lox*P site introduced by the retrovirus will be in the opposite orientation to the anchor *lox*P site. Of all the 30 loci mapped to chromosome 11, 27 of them are transcribed from the antisense strand as expected. However, 3 gene traps are transcribed from the sense strand (from centromere to telomere) (Fig. 4-6a).

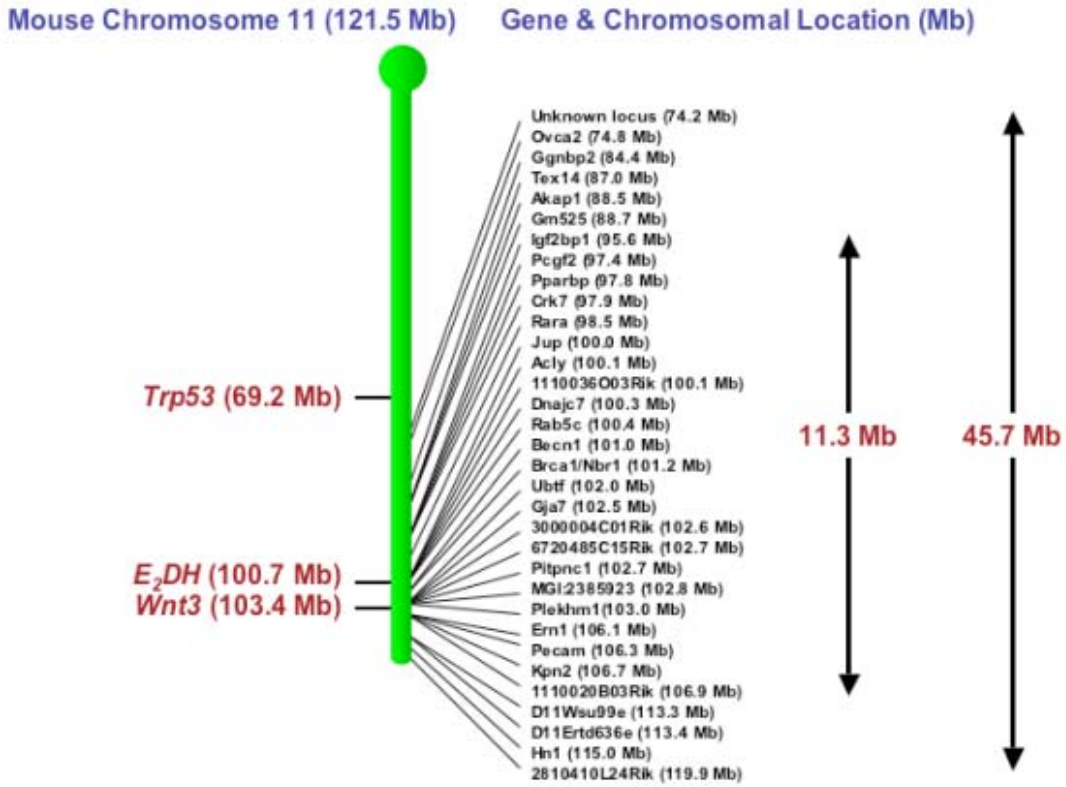


Fig. 4-5 Distribution f gene traps on chromosome 11. Splinkerette PCR and 5' RACE results were mapped on chromosome 11 according to the Ensembl database (http://www.ensembl.org/Mus_musculus/). Gene traps were isolated exclusively in a 45.7 Mb distal region of the mouse chromosome 11, surrounding the *E_2DH* locus. About two third of the trapped genes (23/33) were clustered within 5 Mb proximal and 5 Mb distal to the *E_2DH* locus.

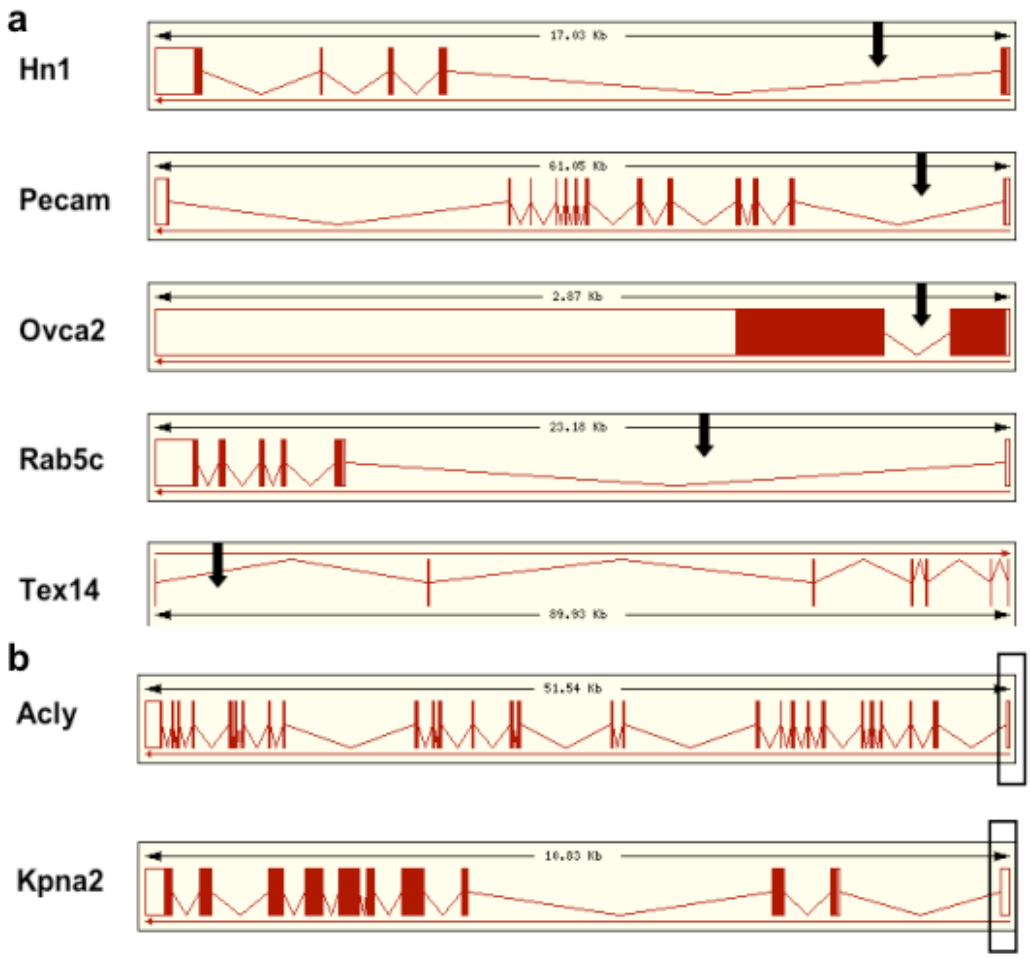


Fig. 4-6 Transcription orientation and the proviral insertion sites of the trapped trapped loci. a. The vector insertion sites of the gene traps at 5 different loci, *Hn1*, *Pecam*, *Ovca2, Rab5c* and *Tex14*, determined by the Splinkerette PCR results. Note that the Tex14is transcribed from the sense strand (from centromere to telomere), and the other four genes are transcribed from the antisense strand (from telomere to centromere). The black arrows point to the proviral insertion site. **b.** The trapped exons of gene trap at 2 different loci, *Acly* and *Kpna2*, determined by the 5' RACE results. The black box highlight the trapped exons. The chromosomal structures of the trapped loci are downloaded from Ensembl (http://www.ensembl.org/Mus_musculus/). The red arrows underneath the exon structures represent the transcription orientation of the genes.

One possibility is a *trans* recombination event between *lox*P sites in direct orientation on the two homologs of chromosome 11 in G1. This will result in a pair of balanced deletion/duplication chromosomes. If the trapped locus is distal to *E_2DH*, the chromosome with the anchor *lox*P site will become a deletion chromosome (Fig. 4-7a). And it is unlikely that such a chromosome can become homozygous after induced mitotic recombination because of the loss of genetic material. If the trapped locus is proximal to *E_2DH*, the chromosome with the anchor *lox*P site will become a duplication chromosome (Fig. 4-7b). This should be able to become homozygous after the induced mitotic recombination. But the double duplication ES cells will lose the *β-geo* cassette after the induced mitotic recombination event and no proviral/host junction fragment should be detected using *lacZ* as a probe. So the balanced deletion/duplication can not be the cause for these clones that are transcribed from the sense strand (from centromere to telomere).

Another possibility is a *trans* recombination event between *lox*P sites in direct orientation on the sister-chromatids of chromosome 11 in G2, which will result into a duplication chromosome and a wild-type chromosome. If the trapped locus is distal to *E_2DH*, the duplication chromosome will carry two *β-geo* cassettes (Fig. 4-8a), one associated with a complete proviral insertion, while the other is a half proviral insertion split by Cre-mediated recombination. But this recombinant can not survive the puromycin selection because the duplication chromosome does not have a functional *Puro* cassette on it. If the trapped locus is proximal to *E_2DH*, the duplication chromosome will carry only one *β-geo* cassette (Fig. 4-8b), which belongs to a complete proviral insertion. The recombinant should be able to survive the puromycin selection because the duplication chromosome has a functional *Puro* cassette on it. The duplication chromosome can also become homozygous after induced mitotic recombination because it does not lose genetic material.

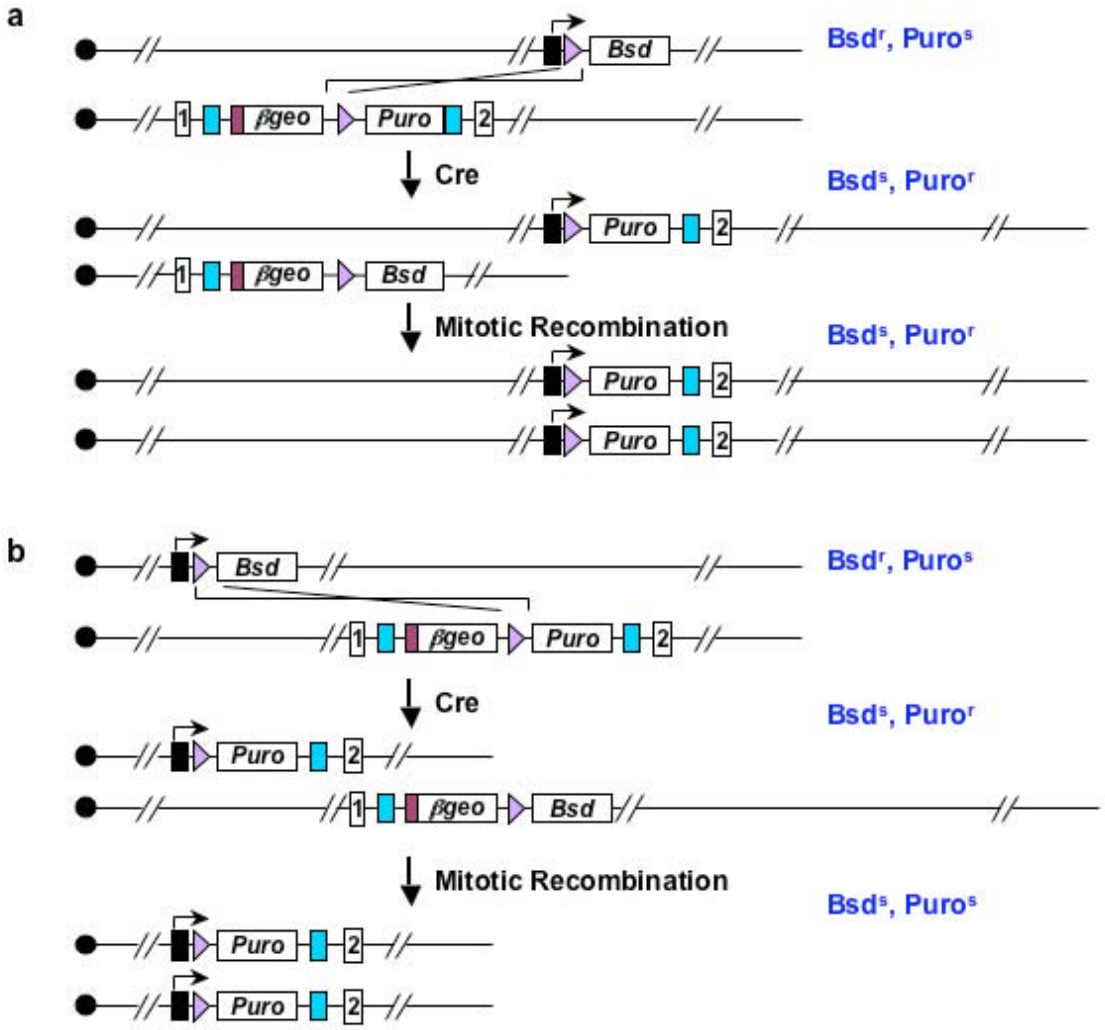

a
Bsd
Bsd^r, Puro^s
1
βgeo
Puro
2
Cre
Bsd^s, Puro^r
Puro
2
1
βgeo
Bsd
Mitotic Recombination
Bsd^s, Puro^r
Puro
2
Puro
2
b
Bsd
Bsd^r, Puro^s
1
βgeo
Puro
2
Cre
Bsd^s, Puro^r
Puro
2
1
βgeo
Bsd
Mitotic Recombination
Bsd^s, Puro^s
Puro
2
Puro
2

Fig. 4-7 G1 *trans* recombination event. a. *lox*P sites in direct orientation on different homologs of chromosome 11 (*trans*) and the trapped locus is proximal to *Hsd17b1*. A pair of balanced deletion/duplication chromosomes will be generated by Cre-mediated recombination (genetic material is not lost at this stage). The resulting cells can survive subsequent puromycin selection. After mitotic recombination, the duplication chromosome can become homozygous. However, the duplication chromosomes do not have the *βgeo* cassette, so these clones can not be picked up by Southern using a *lacZ* probe. **b.** *lox*P sites in direct orientation on different homologs of chromosome 11 (*trans*) and the trapped locus is distal to *Hsd17b1*. A pair of balanced deletion/duplication chromosomes will be generated by Cre-mediated recombination (genetic material is not lost at this stage). The resulting cells can survive subsequent puromycin selection. After mitotic recombination, the deletion chromosome will become homozygous. However, this recombination event was not observed, presumably because the homozygous deletions will be selected against because of homozygous lethality in ES cells. Black box, *PGK* promoter; lavender triangle, wild-type *lox*P site; blue box, virus long terminal repeat; plum box, splice acceptor.

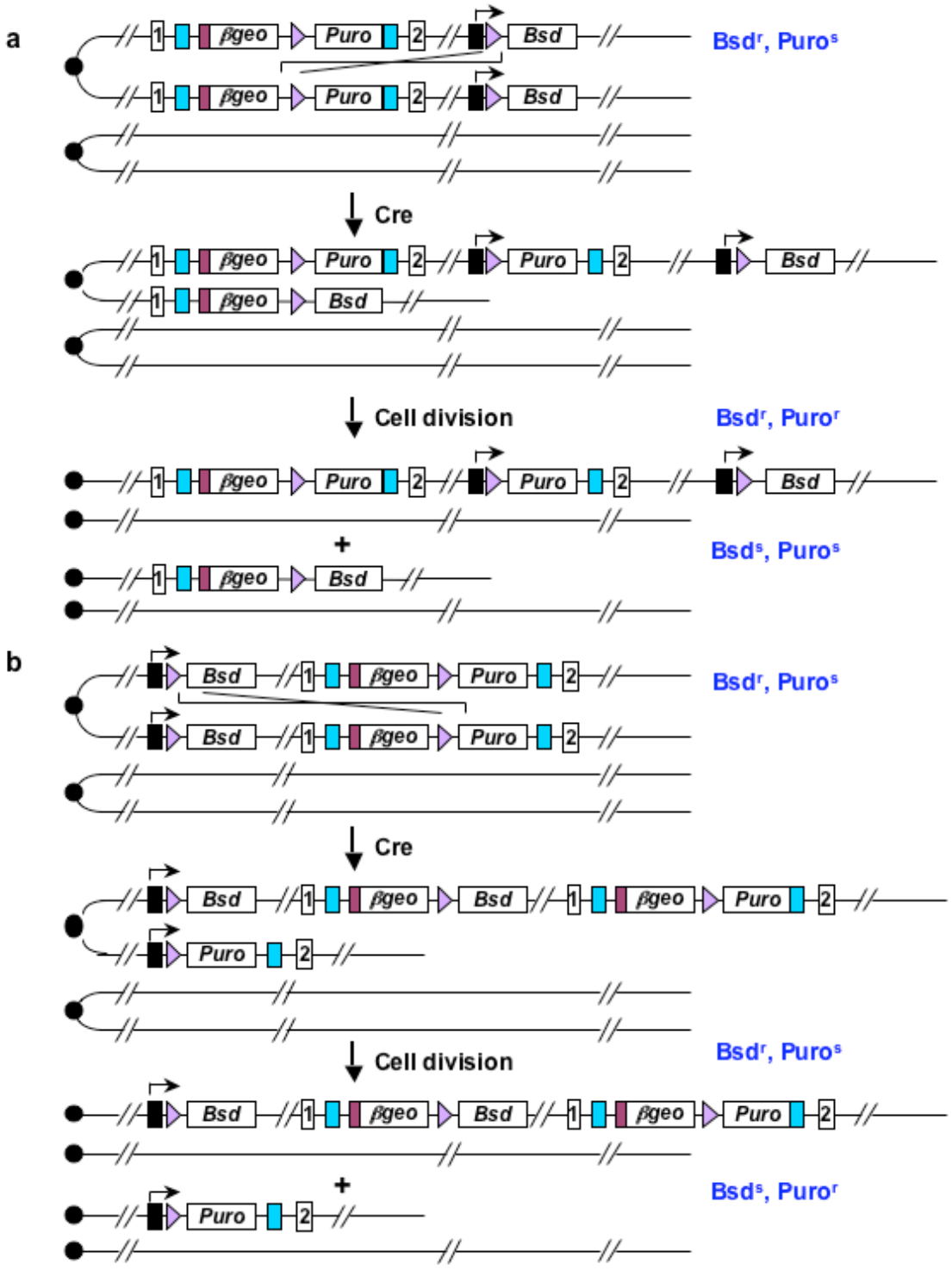

a
βgeo
Puro
Bsd
1
2
Bsd^r, Puro^s
Cre
Cell division
Bsd^r, Puro^r
Bsd^s, Puro^s
b
Bsd^r, Puro^s
Cre
Cell division
Bsd^r, Puro^s
Bsd^s, Puro^r

Fig. 4-8 G2 *cis* recombination event. a. *lox*P sites are in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is proximal to *Hsd17b1*. A pair of balanced deletion/duplication sister chromatids will be generated by Cre-mediated recombination (G2-*trans*). After cell division, the cells with the duplication chromosome can survive puromycin selection. This duplication chromosome can become homozygous because it does not lose any genetic material. **b.** *lox*P sites in direct orientation on the same homolog of chromosome 11 (*cis*) and the trapped locus is distal to *Hsd17b1*. A pair of balanced deletion/duplication sister chromatids will be generated by Cre-mediated recombination (G2-*trans*). After cell division, the cells with the deletion chromosome can survive the puromycin selection. However, this recombination event was not observed. The heterozygous (before mitotic recombination) and homozygous (after mitotic recombination) deletions will be selected against because of haplosufficiency and/or homozygous lethality in ES cells. Black box, *PGK* promoter; lavender triangle, wild-type *lox*P site; blue box, virus long terminal repeat; plum box, splice acceptor.

These homozygous duplication clones can be distinguished from other homozygous inversion clones by Southern as well as the drug selection. A *Puro* specific probe can detect the 6.0/6.9 kb *Kpn*I fragment representing proviral insertion and 3.4 kb *Kpn*I fragment representing the reconstitute *PGK-Puro-bpA* cassette. A simpler way to confirm the identity of the clones is a sib-selection using M15+puromycin and M15+blasticidine. The clones with two duplication chromosomes should be resistant both to puromycin and blasticidin, while the clones with two inversion chromosomes should be resistant to puromycin but sensitive to blasticidin.

Three trapped loci (*Tex14*, *LOC217071* and *Rara*) transcribed from the sense strand (from centromere to telomere) are all located proximal to the E_2DH locus, and *Tex14* was trapped twice. All the HAT resistant subclones from the three independent events are homozygous for the modified E_2DH locus (*Nde*I digestion, E_2DH 3' probe), and all the clones only carry the 6.9 kb proviral insertion fragment (*Kpn*I digestion, *lacZ* probe). And all these clones are resistant to both puromycin and blasticidin. These results are consistent with these clones carrying two duplication chromosomes.

4.2.4 Proviral insertion sites in trapped loci

For a large portion of the trapped loci, I have cloned the proviral/host junction fragments by Splinkerette PCR. This is informative on the chromosomal structure of the recombinants after regional trapping and induced mitotic recombination. In most of the trapped loci, the proviral insertions occur in the first or second intron, the inversion thus generates a breakpoint between the first one or two exons and the rest of the gene (Fig. 4-6a). That means, in most cases, the transcriptional regulation elements will be several megabases away from the coding region, and in many cases, the open reading frames themselves is also disrupted.

For some clones, the proviral insertion sites were not determined by Splinkerette PCR, for example when the sequence was repetitive or the Splinkerette PCR failed. In these cases, the RACE results mapped the insertion to an exon. In one case, the 5' RACE products mapped to an

unknown locus close to the *Gja7* gene (gap junction membrane channel protein alpha 7 or connexin 45). By searching the NCBI database, the 5' RACE product matched the 5' untranslated region of an alternatively spliced form of connexin 45 (AY390396). So it is likely that I have trapped an alternative spliced form that expresses in the undifferentiated ES cells. In another case, the 5' RACE product mapped to an unknown locus close to the *Brca1* gene (breast cancer 1, early onset). By searching the NCBI database, the 5' RACE product matched *Brca1*/*Nbr1* bidirectional promoter region (AF080589) and some *Brca1* EST sequence. So it is likely that I have trapped an ES cell specific alternative spliced form of *Brca1*.

In at least one case, it appears that more than one gene was disrupted by the retrovirus integration and the subsequent inversion. In this case, the trapped locus, *2810410L24Rik* (119.9 Mb), is very close to the telomere of the chromosome 11. Ensembl predicts this is a single exon gene that is transcribed from the sense strand, but the 5' RACE sequence matches two separate regions from the opposite direction, which suggests that there is another transcript from the opposite strand. By searching the NCBI database, I have identified another cDNA, *D030042H08Rik*, transcribed in a similar way as the 5' RACE product which overlaps with *2810410L24Rik*. The splinkerette results map the retroviral insertion site between the second and third exons of *D030042H08Rik* (Fig. 4-9). Another independent gene-trap was also mapped to a locus very close to *2810410L24Rik*. A hypothetical gene *LOC432619* is also transcribed from the antisense strand. *D030042H08Rik* and *LOC432619* both belong the same mouse UniGene, Mm.269766. Interestingly, the UniGene is named as "RIKEN cDNA 2810410L24 gene (2810410L24Rik)", though the transcription direction of *2810410L24Rik* is opposite to the other transcripts, *D030042H08Rik* and *LOC432619*. In fact, there is another UniGene, Mm.125044 with the same name "2810410L24Rik RIKEN cDNA 2810410L24 gene (2810410L24Rik)". And this one is composed of all the cDNAs and ESTs transcribed from the sense strand. Since *D030042H08Rik* and *LOC432619* have not been mapped in Ensembl, I still name the trapped locus as *2810410L24Rik*. But the proviral integration and the inversion will disrupt transcripts from both directions.


Exon 3
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
Exon 2
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
Exon 2
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
5F'R/3' RACE Product
Exon 1
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA
Exon 1
RIKEN 2610410L24 cDNA
RIKEN 2610410L24 cDNA

Fig. 4-9 The gene trap insertion at *2810410L24Rik* has disrupted two genes with opposite directions of transcription. *2810410L24Rik* is a single exon gene that is transcribed from the sense strand. The 5' RACE matches two separate regions from the opposite direction, which suggests another transcript from the anti-sense strand. Another gene, *D030042H08Rik*, is transcribed in a similar way as the 5' RACE product and its transcript is overlapping with *2810410L24Rik*. Splinkerette PCR results have mapped the retroviral insertion site between the second and third exons of *D030042H08Rik*. Note that the transcription of *D030042H08Rik* is not the same as the 5' RCAE results, so they might represent different splice forms of the same gene. Another independent gene-trap was also mapped to a locus very close to *2810410L24Rik*, the predicted gene *LOC432619*. This predicted gene is also transcribed from the antisense strand.

4.3 Discussion

In this study, I have mapped 49 of the 66 independent gene traps by sequence analysis of 5' RACE and/or Splinkerette PCR products. Most of the gene traps identified are located on chromosome 11 and are transcribed from the antisense strand (from telomere to centromere). But some possible translocations and other possible chromosomal rearrangements were observed. These events constituted background in the context of the goal of this study, namely to generate homozygously mutated ES cell clones for recessive genetic screens *in vitro*. No matter how these clones survived the stringent selection procedures, through a series of rare recombination events, they have lost or gained a large part of the chromosome 11. The phenotype of these clones can not be attributed to a single gene, and thus they are not suitable for the genetic screens. Importantly, they can be easily identified by the sequence of Splinekerette PCR and 5' RACE products because of the location and transcriptional orientation of the trapped genes.

Some gene-trap hot spots were found in all of the trapped loci. 9 of the 30 mapped gene-trap loci on chromosome 11 were hit more than once. One of them, *Dnajc7*, was hit 6 times. Interestingly, this locus (100.3 Mb) is very close to the anchor locus, *E_2DH* (100.7 Mb). The recombination efficiency over such small distance will be extremely high (Zheng, Sage et al. 2000). One would predict that if a gene-trap hot spot with the correct transcriptional orientation happens to be close to the anchor point, the small inversions will dominate the pool of inversion clones, and consequently the induced mitotic recombination clones. This will significantly decrease the possibility of identifying other recombination events, especially those relatively rare events (large inversions) from the same pool. It is almost impossible to screen against these clones by Southern analysis even if the hot spots are already known, because the proviral insertion bias is locus specific, not sequence specific, which means that even if the same locus is hit multiple times, the proviral insertion will occur in different introns or different positions in the same intron.

A similar situation was also noticed in other genome-wide gene trapping programs (Zambrowicz, Friedrich et al. 1998; Hansen, Floss et al. 2003; Skarnes, von Melchner et al. 2004). The efficiency of trapping new genes is not linear, and it will drop with the increase of gene-trap tags. Within the first 100,000 tags, the rate of capturing new genes declines to about one new gene every 35 tags (Skarnes, von Melchner et al. 2004). If our strategy is applied to generate genome-wide homozygous gene-trap clones, there will also be a balance point beyond which new genes can not be mutated economically. However, if various plasmid and retroviral vectors are used, it will help to overcome the bias of gene trap insertions of a single vector and increase the efficiency of gene trapping (Hansen, Floss et al. 2003).

One way to control this bias is to limit the size of the pool of original gene traps. As discussed in the previous chapter, between zero and one inversion events are expected after Cre-mediated recombination in a pool of 100 trapping events. If the starting cell number and the electroporation conditions are the same, it is possible to predict the type of recombination (inversion or translocation) and the size of the inversion (small or large) simply by counting the number of puromycin resistant colonies on each plate.

On the other hand, 21 of the 30 mapped gene-trap loci on chromosome 11 were hit only once. So the regional trapping experiment is far from reaching saturation. If the same experiment is repeated, many new homozygous gene traps on chromosome 11 will be recovered. Also, 27 of the 30 mapped gene-trap loci are transcribed from the antisense strand (from telomere to centromere), which proves that our strategy is highly efficient for generating homozygous gene-trap mutations transcribing from one strand. Simply by changing the orientation of the *lox*P site of the anchor point targeting vector, the genes in the same region but are transcribed from the other strand will be trapped.

I have compared the gene traps identified in my experiment with the ones identified in the regional trapping experiment by Meredith Wentland (Wentland et al. unpublished data). It is interesting to notice that the distribution pattern

of the gene traps is extremely similar between the two experiments. In both experiments, most gene traps occurred within a region of 40 Mb in distal region of the mouse chromosome 11. A large portion of the gene traps clustered within 5 Mb proximal and 5 Mb distal of the *E_2DH* locus (18/21 for Wentland's experiment and 20/30 for our experiment). This distribution pattern of the gene traps is expected based on the relationship between the efficiency of Cre in recombining *lox*P sites over different distances (Liu, Zhang et al. 1998; Zheng, Sage et al. 2000).

Though the distribution pattern is very similar between the two different experiments, the gene traps isolated are totally different. None of the trapped loci identified in Wentland's experiment were hit in my experiment or vice versa. Of the 21 gene traps isolated in Wentland's experiment, only 2 matched known mouse genes, 4 matched predicted transcripts, 4 matched ESTs and 11 matched unknown loci. In my experiment, most of the gene traps matched known mouse genes or transcripts, only one of them appeared to be an unknown locus on chromosome 11. This difference probably reflects the different "trappable" sets of genes for 5' and 3' trapping. 5' trapping is dependent on the expression of the trapped locus in undifferentiated ES cells, whilst 3' trapping is not. Considering that the purpose of my experiment is to mutate genes for *in vitro* screen, 5' trapping is more likely to disrupt a functional gene in ES cells and thus more likely to result in a phenotype *in vitro.*

In both experiments, only genes transcribed from one strand can be selected. However, some gene traps transcribed from the opposite strand have also survived the selection (4/21 for Wentland's experiment and 3/30 for my experiment). Wentland et al. (unpublished data) have carried out Fluorescence In situ Hybridization (FISH) to identify the alternative recombination events. She found that these clones were either cells with balanced deletion/duplication chromosomes derived from a G1 *trans* recombination event, or cells with one wild-type chromosome and one duplication chromosome derived from a G2 *trans* recombination event. By Southern analysis and drug sib-selection, the three ES cell clones in which

the trapped loci are transcribed on the sense strand are all clones that carry two duplication chromosomes. These alternative recombination events caused some background, but their efficiency is relatively low and thus did not present a serious problem for the genetic screen. Clones with these events can be easily identified from the sequence of their Splinkerette PCR and 5' RACE products or drug selection.

Cells with translocation chromosomes and duplication chromosomes are by-products of my products. They are not useful for the *in vitro* genetic screen because in these clones, a large genomic region is either deleted or duplicated, and the phenotype is very hard to be associated with a certain gene. But on the other hand, these rearranged chromosomes might be a useful resource for other experiments. ES cells with large deletions may be selected against if the deletion affects cell viability or growth (Zheng, Sage et al. 2000). For example, a 22 cM deletion distal to the *E_2DH* locus (*E_2DH* - *D11Mit69*) was found to be haploinsufficient in ES cells. In the rare cells that survived selection, the remaining wild-type chromosome was duplicated. In some of the translocation events identified in my experiment, the modified *E_2DH* locus became homozygous. That means that the resulting ES cell clones are partially trisomic for the genomic region translocated from other chromosomes, but they have lost one copy of the genomic region on chromosome 11 distal to the *E_2DH* locus, which is approximately 20 Mb in size. The Southern analysis using the *E_2DH* 3' probe has shown that these clones do not have wild-type chromosome 11. It is possible that the gain of a genomic region from another chromosome somehow can compensate the haploinsufficiency caused by the loss of the distal region of chromosome 11. Another possibility is that the chromosome which the distal region on chromosome 11 was translocated to was duplicated. I have not carried out FISH experiment to determine the exact genomic structure of the clones. But these clones can be useful to study the functional relationships between different genomic regions.

There are several reports that gene-trap insertions do not completely disrupt the normal transcription of the endogenous gene and the mutagenicity of

gene-trapping is still a controversial issue. Mitchell et al. (2001) have generated sixty mouse lines with secretory gene-trap vectors. Twenty-five of them showed visible embryonic or adult phenotypes. For 11 of the 25 gene traps that showed observable phenotypes, alleles generated by gene-targeting have also been reported. Ten of these strains showed exactly the same phenotype as the gene-targeted mutations. The remaining strain had a less severe phenotype than the gene-targeted allele but still caused embryonic lethality (Mitchell, Pinson et al. 2001). Stanford et al. (2001) has reviewed one hundred additional gene-trap insertions that have been described in the literature. Sixty percent of these insertions show “obvious” phenotypes, and 40% are recessive lethal mutations. The frequency of recessive lethal mutations and “obvious” phenotypes generated by gene-trapping is comparable to the results obtained from gene-targeting (Stanford, Cohn et al. 2001). Nevertheless, trapping alone is not always sufficient to completely block transcription. Leaky expression of wild-type transcripts can partially rescue some phenotypes and thus complicate analysis.

Our strategy not only inserts a trapping cassette in the gene, but also introduces a breakpoint at the proviral insertion site. So this technique should be more mutagenic than the other mutagenesis methods. At the same time, long-range chromosomal rearrangements can disrupt more than the trapped locus. In my experiment, in one gene-trap clone, the inversion disrupted two partially overlapped genes that are transcribed in opposite orientations. In another case, the inversion created a breakpoint at a bidirectional promoter. So the expression of genes around the breakpoint need to be checked to avoid misinterpretation of any observed phenotypes.

The identification of the gene-trap loci has proved that this strategy is useful to generate homozygous mutations in a genomic region of interest. This strategy can easily be applied to other mouse chromosomes to generate homozygous mutant ES cells in other regions of the genome. This resource will facilitate large-scale *in vitro* genetic screens.

5 Genetic screen on homozygous gene traps

5.1 Introduction

During *in vitro* differentiation, ES cells can form cystic embryo-like aggregates, embryoid bodies (EB), that contain cells of endodermal, ectodermal and mesodermal lineages, which can further differentiate into more specialized cell types. The morphological changes of embryoid bodies are accompanied, at the molecular level, by the changes in the expression of a set of lineage-specific and tissue-specific markers. By comparing the dynamic changes in the expression of these markers *in vivo* and *in vitro*, different stages of EB differentiation *in vitro* can be linked to different stages of embryogenesis *in vivo* (Leahy, Xiong et al. 1999). These properties allow us to use ES cell *in vitro* differentiation as an *in vitro* model to study early embryogenesis and this facilitates genetic approaches.

5.1.1 *In vitro* differentiation protocols

There are three main protocols for ES cell *in vitro* differentiation: the hanging drop method (Wobus, Wallukat et al. 1991); the mass culture method (Doetschman, Eistetter et al. 1985); and the methylcellulose method (Wiles and Keller 1991). All three of these have been widely used for making embryoid bodies (EB) for different purposes.

The advantage of the hanging drop method is that the starting number of ES cells in an embryoid body is defined, so the size and the differentiation pattern of the EBs generated by this method is more consistent than with the other two methods. This characteristic is particularly important for developmental studies, which require the comparisons between EBs under different culture conditions and/or with different mutations (Wobus, Guan et al. 2002). However, this method is also more complicated than the other two methods.

On the other hand, the mass culture method is useful for differentiating a large number of ES cells. By plating undifferentiated ES cells onto bacteriological Petri dishes, the cells automatically form cell aggregates, and the aggregates can differentiate into a variety of different cell types. However, the size and the differentiation pattern can vary significantly between plates or between

experiments, even when the same ES cell line is used. The methylcellulose method is used specifically for the differentiation of haematopoietic lineages, and is not suitable for other purposes.

In this project, it was necessary to compare the *in vitro* differentiation potential of a number of homozygous mutant ES cell clones. Therefore the hanging drop method was the most appropriate *in vitro* differentiation protocol to use.

5.1.2 Parameters influencing *in vitro* differentiation of ES cells

The developmental potency of ES cells in culture is dependent on a number of intrinsic and extrinsic parameters. These include the number of ES cells used to make the EBs; the composition of the differentiation medium; cellular growth factors and differentiation inducers added to the culture medium; the ES cell lines; as well as the genetic changes in the ES cell genome.

Compared to *in vivo* differentiation in the mouse, the parameters for *in vitro* differentiation are more controllable. Whichever differentiation protocol is chosen, extrinsic parameters can be effectively controlled by using defined medium and culture conditions. Variations caused by intrinsic parameters can be eliminated by choosing an appropriate control ES cell line. Thus loss-of-function or gain-of-function studies using *in vitro* differentiation can be an ideal alternative to study the phenotypes of mutations on embryogenesis and early development (Wobus, Guan et al. 2002).

5.1.3 Recessive genetic screens using ES cell *in vitro* differentiation

Genetic analysis of recessive mutations in ES cells is informative on possible functions *in vivo*, especially for mutations that result in embryonic lethality. A recessive genetic screen using ES cell *in vitro* differentiation can be used to identify important genes in the differentiation process.

The bottleneck of recessive genetic screens in ES cells is the difficulty of obtaining enough homozygous mutant ES cells. If a genetic screen is performed to identify genes involved in ES cell *in vitro* differentiation, pure homozygous mutant ES cell clones need to be differentiated individually to

check for their differentiation potential. Existing methods to generate homozygous mutations in ES cells are not ideal for this purpose. In the previous chapters, I have demonstrated that a strategy combining regional trapping and inducible mitotic recombination can be used to generate homozygous mutations in a genomic region of interest. By Splinkerette PCR and 5' RACE, proviral/host flanking genomic sequences and/or cDNA sequence were isolated to identify the proviral insertion sites and inversion breakpoints of these mutant clones.

A total of 30 different gene-trap loci on chromosome 11 that are homozygously mutated were isolated. These homozygous gene-trap clones can be used to perform a small-scale genetic screen to identify the mutations that will disrupt the normal ES *in vitro* differentiation process. Each gene-trap clone has been differentiated individually, and a set of important lineage-specific and tissue-specific markers have been checked to determine the differentiation potential of each of the homozygous mutant ES cell lines. Mutant cell lines that show an abnormal differentiation pattern have been confirmed using independent methods.

5.2 Results

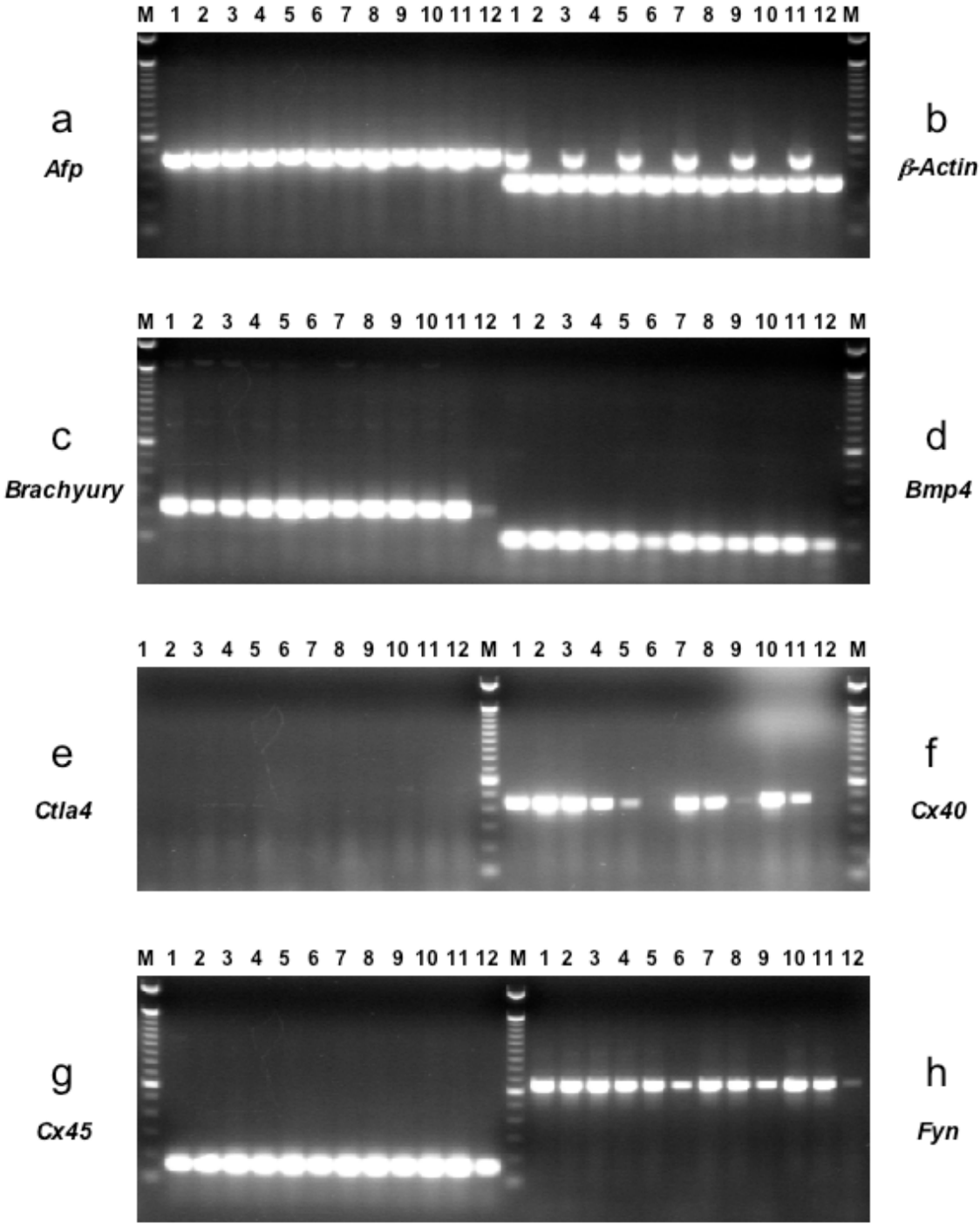
5.2.1 Primary screen

For each of the 33 mapped gene-trap loci, at least one subclone was chosen for the primary *in vitro* differentiation screen. Embryoid bodies were made and cultured as described before (Wobus, Guan et al. 2002). In brief, undifferentiated ES cells were maintained on feeder layers until they were used for *in vitro* differentiation. To setup the assay, ES cells were trypsinized and diluted to a final concentration of approximately 600 cells in 20 μl Differentiation Medium (see material and methods). 20 μl drops of the ES cell suspension were laid onto the bottom of 100-mm bacteriological Petri dishes. The Petri dishes were inverted and the ES cell aggregates were cultured in the resulting hanging drops for two days. After this, the Petri dishes were turned the right way up and Differentiation Medium was added into each dish to rinse the aggregates. The aggregates were cultured in suspension for

another three days. The sample was harvested at Day 5. At the same time, the EBs on the remaining dishes were plated out onto gelatinized 90-mm tissue culture plates. The plated EBs were subsequently cultured in Differentiation Medium supplemented with 10^{-8} M retinoic acid (RA) and the medium was changed every two days. Subsequent samples were taken at Day 8 and Day 11.

When all the samples were taken, RNA was extracted from each sample and quantified. 5 μg total RNA was used for first strand cDNA synthesis. The resulting cDNA was used as a template for RT-PCR. In the primary of screen, 16 pairs of primers were used (*Afp*, *β-Actin*, *Brachyury*, *Bmp4*, *Ctla4*, *Cx40*, *Cx45*, *Fyn*, *Gata4*, *Goosecoid*, *Hnf4*, *Nodal*, *Oct3/4*, *Pecam*, *Tie2* and *vHNF1*). All the homozygous mutant cell lines that showed abnormal expression (significant up-regulation or down-regulation compared to the WW93A12 control line) for one or more markers in the primary screen were selected for the second round screen (Fig. 5-1 and Fig. 5-2).

All the mitotic recombination clones (WW103) used in the screen carry two homologs of chromosome 11 from the same parent, either bi-paternal or bi-maternal. It is possible that because of the imprinting, the *in vitro* differentiation pattern of ES cells carrying bi-paternal or bi-maternal homologs of chromosome 11 will be different from that of the wild-type ES cells that have one paternal and one maternal homologs of chromosome 11. Also, the *in vitro* differentiation potential of ES cells homozygous for the targeted E_2DH allele has not been assessed. So an ideal control cell line for this experiment will carry two homologs of chromosome 11 from the same parent as the WW103 clones, and this control cell line is also homozygous for the targeted E_2DH allele.


M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn

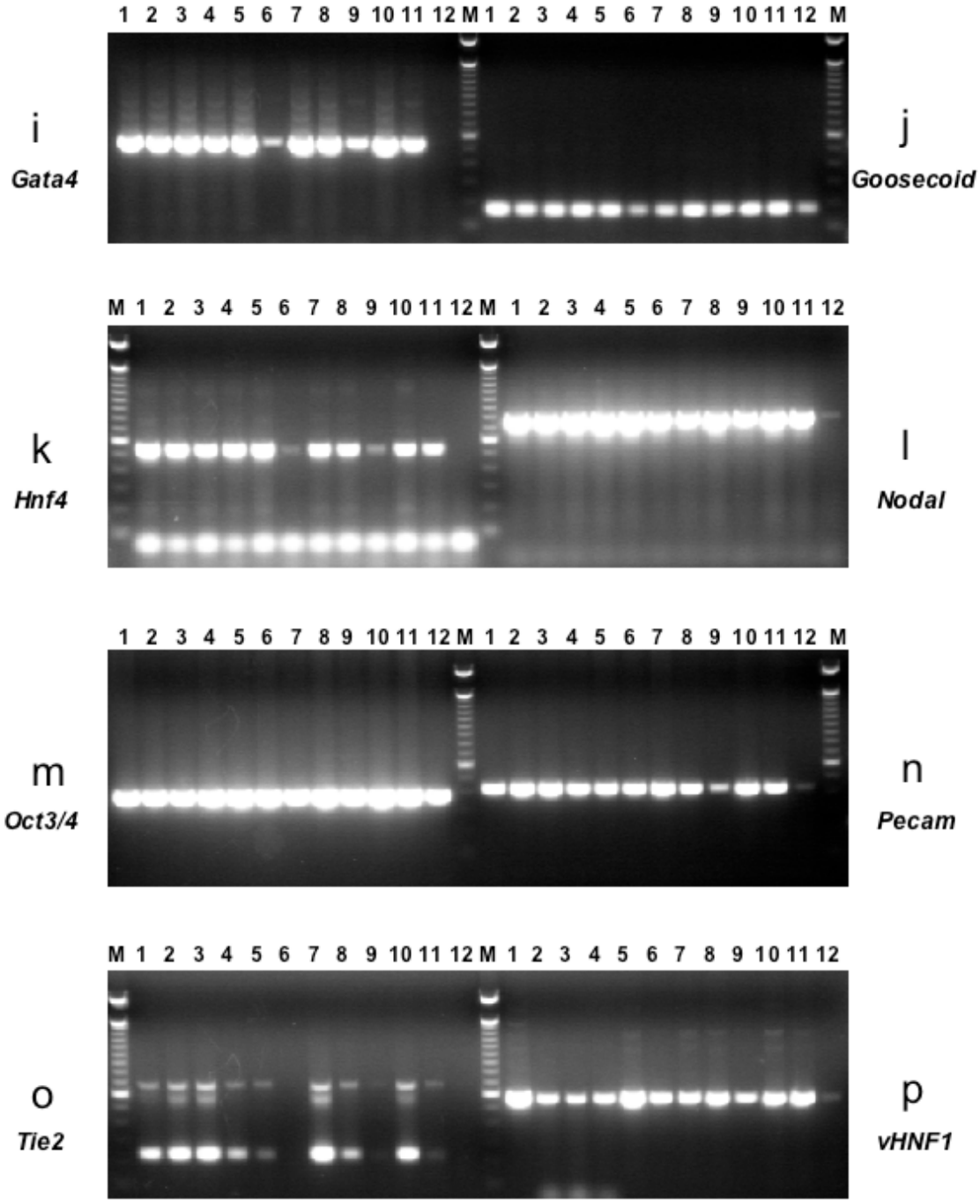
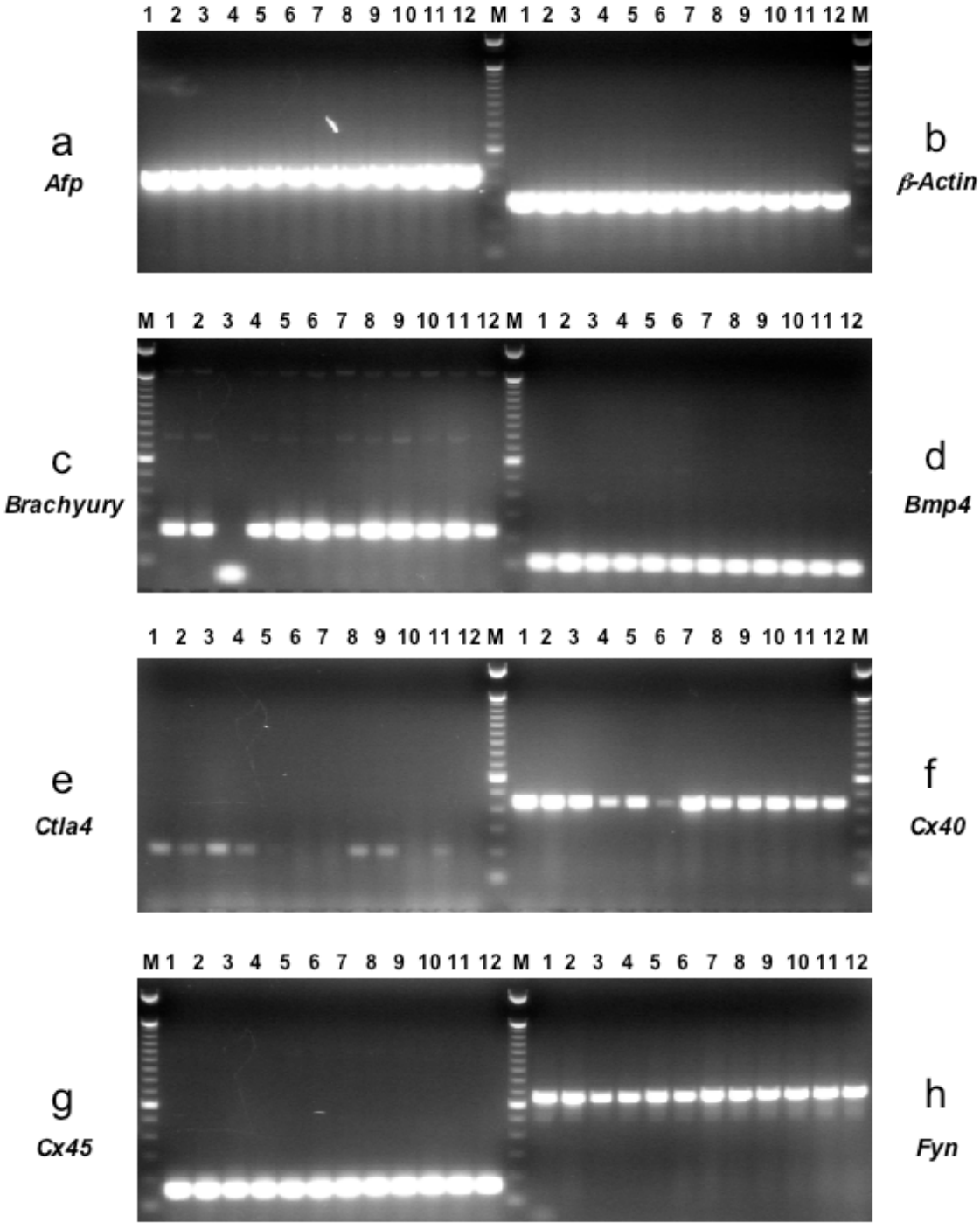
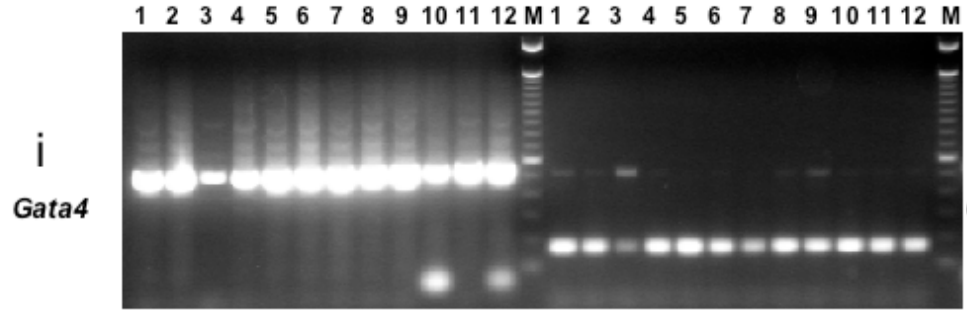

1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
i
Gata4
j
Goosecoid
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
k
Hnf4
l
Nodal
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
m
Oct3/4
n
Pecam
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
o
Tie2
p
vHNF1

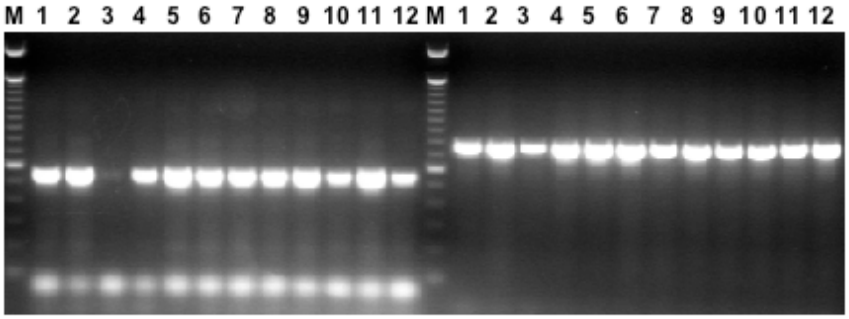
Fig. 5-1 RT-PCR results of Day 5 embryoid bodies. a. *Afp*. **b.** *β-Actin*. Note that line 1, 3, 5, 7, 9 and 11 were mistakenly mixed with PCR products from **a**. *β-Actin* was used as a loading control. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)


1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn



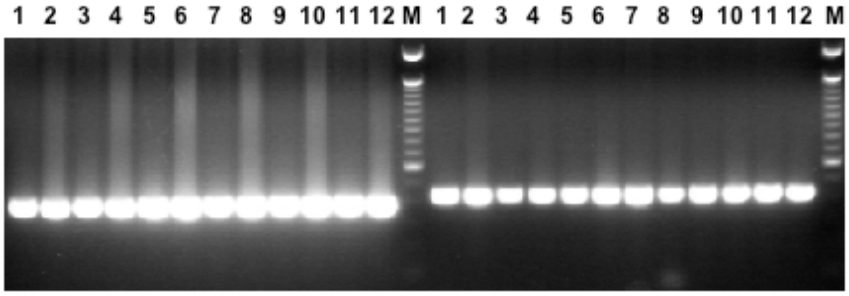
i
Gata4

j
Goosecoid



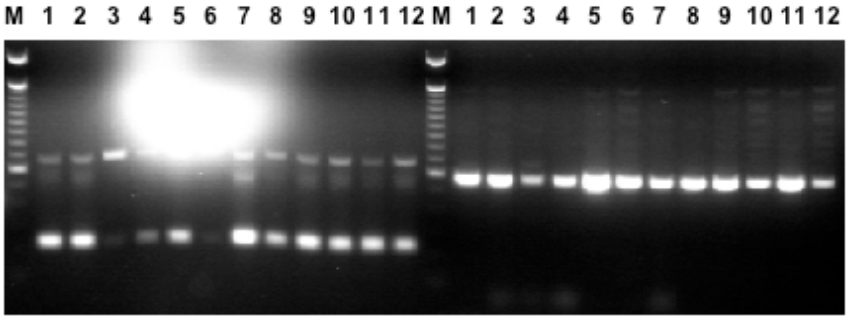
k
Hnf4

l
Nodal



m
Oct3/4

n
Pecam



o
Tie2

p
vHNF1

Fig. 5-2 RT-PCR result of Day 8 embryoid bodies. a. *Afp*. **b.** *β-Actin*. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)

To generate this control ES cell line, a Cre expression plasmid was electroporated into the WW69-D6 cell line and mitotic recombination clones were selected in M15 supplemented with HAT. The clones with the desired phenotype were identified both by sib-selection and by Southern analysis using an *E_2DH* 3' probe. The G2-X recombinants are resistant to HAT and blasticidin, but sensitive to G418 and puromycin. Southern analysis of *Nde*I digested genomic DNA will generate a 9.6 kb targeted fragment instead of the 13.1 kb wild-type fragment. One clone with the desired genotype, WW93-A12 and its subclones were used as controls in the ES cell *in vitro* differentiation screen.

5.2.2 Secondary screen

Mutant cell lines that showed an abnormal expression pattern for the markers checked in the primary screen were subcloned and single colonies were picked to avoid cross-contamination by ES cells that did not have the correct genotype. The control cell line, WW93-A12 was also subcloned. Southern analysis was performed on all the subclones to confirm their identities (Fig. 5-3).

The *in vitro* differentiation protocol for the second round screen is essentially the same as that of the first round. But more time points were taken and more molecular markers were checked using RT-PCR. The clones that still showed abnormal expression for the markers checked were characterized individually.

5.2.3 WW103-8E6 (*Pecam*)

One of the mitotic recombination clones, WW103-8E6, have overtly impaired *in vitro* differentiation potential. When the EBs were plated onto the gelatinized tissue culture plates, the EBs could not form cystic three-dimensional structures. When RT-PCR was performed using a series of molecular markers, the expression of some markers in the day 8 EBs was significantly down-regulated compared to the wild-type control, WW93-A12 (Fig. 5-4).

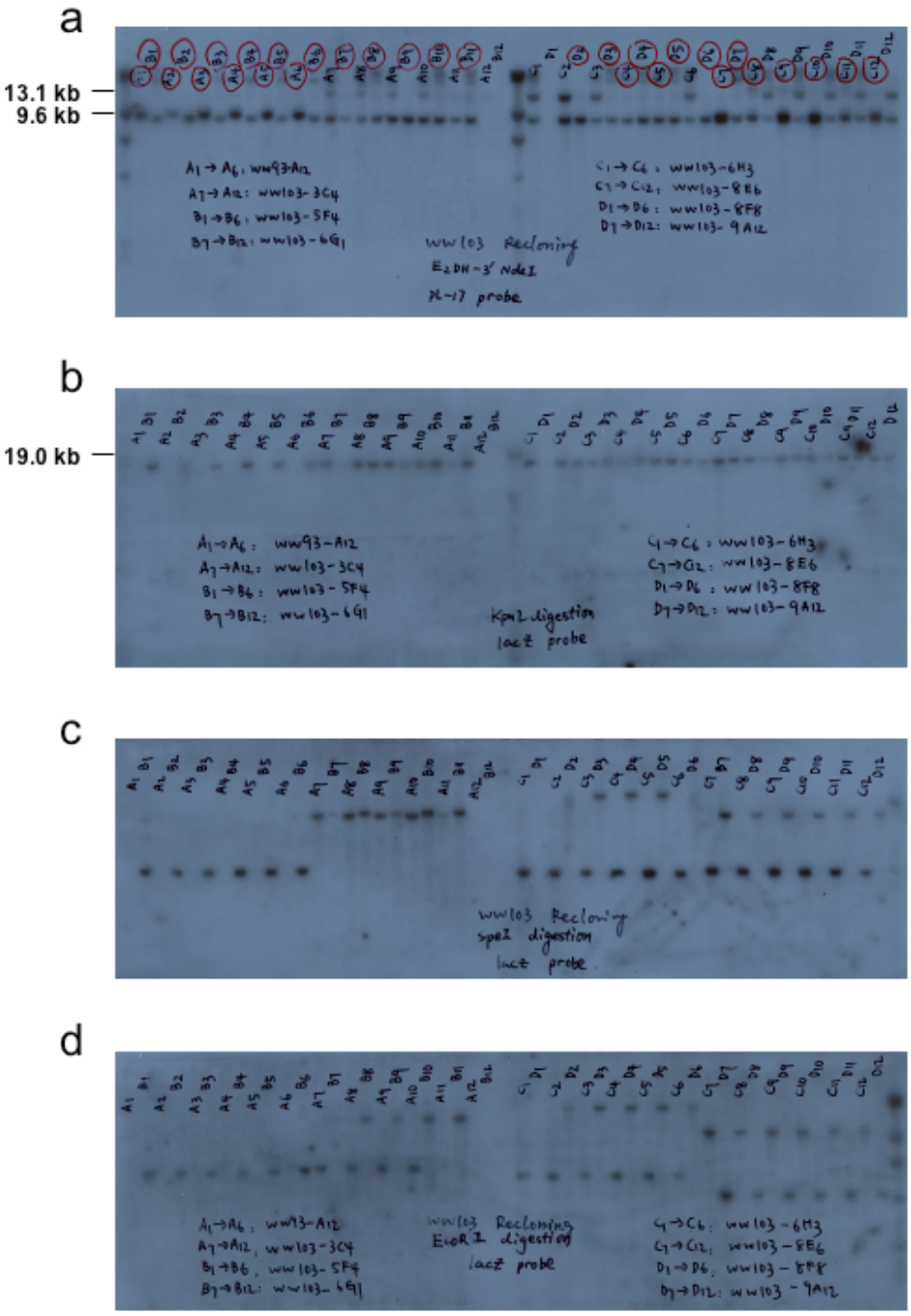
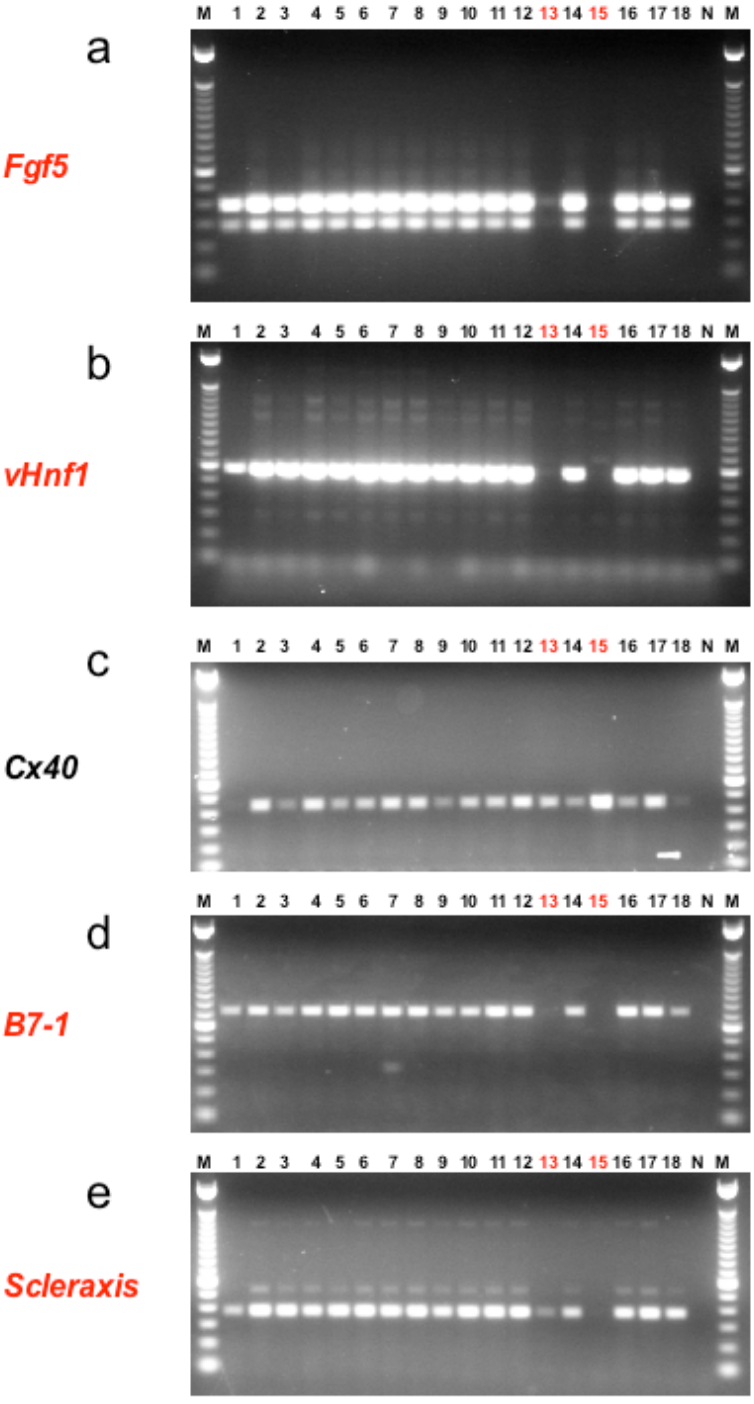
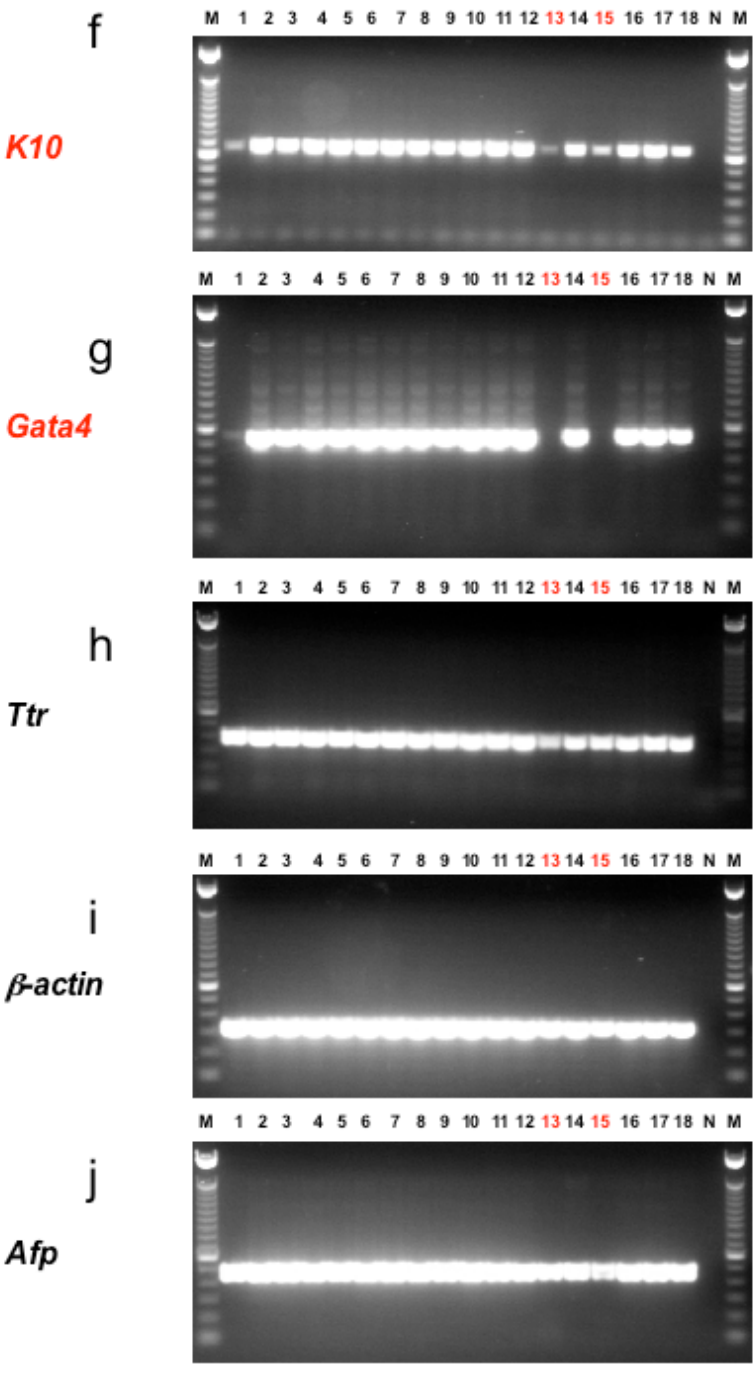

a
13.1 kb
9.6 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
E2DH-3' NdeI
PL-17 probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
b
19.0 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
KpnI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
c
ww103 Recloning
SpeI digestion
lacZ probe
d
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
EcoRI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12

Fig. 5-3 Southern analysis of subclones of homozygous mutant ES cell lines. a. Genomic DNA from the subclones was cut with *Nde*I and hybridized to an E_2DH 3' probe. The targeted restriction fragment is 9.6 kb and the wild type fragment is 13.1 kb. Note that some subclones of WW103-6H3 (c1, c2, c3 and c6) are heterozygous. **b.** Genomic DNA from the subclones was cut with *Kpn*I and hybridized to a *lacZ* probe. A 19.0 kb inversion restriction fragment was detected for all the subclones except those of the control cell line, WW93-A12. **c.** Genomic DNA from the subclones was cut with *Spe*I and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12. **d.** Genomic DNA from the subclones was cut with *Eco*RI and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12.

The *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product from this clone mapped the proviral insertion site to the first intron of *Pecam* (Platelet endothelial cell Adhesion Molecule Precursor, CD31) (Fig. 5-5a). The 5' RACE product matched an alternative spliced exon (Exon 1b) (Fig. 5-5b). In the Ensembl browser, there are at least three different spliced forms at the 5' end of this gene. *Pecam* transcripts can start from Exon 1a, Exon 1b or a site just 5' to Exon2 (Fig. 5-5c). The open reading frame (ORF) of *PECAM* starts from Exon 2. So the breakpoint in intron 1 created by the inversion would disrupt the transcripts starting from Exon1a and 1b, but it may not affect the transcripts starting from Exon 2. RT-PCR primers were used to determine the expression of different alternative spliced forms of *Pecam* in undifferentiated WW93-A12 and WW103-8E6 ES cells. This analysis revealed that none of the transcripts in undifferentiated ES cells started from Exon 1a (data not shown). In undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. However, in undifferentiated WW103-8E6 ES cells, *Pecam* transcripts starting from Exon 2 and Exon 1b were both detected, implying that the inversion did not completely block the transcription across the breakpoint (Fig. 5-5d).

The *in vitro* differentiation of another *Pecam* gene-trap clone, WW103-4A6, showed that the differentiation of this clone was not impaired by the proviral insertion and the breakage caused by inversion. The *Sau*3A1 Splinkerette PCR product fro this clone has mapped the proviral insertion site to the third intron of *Pecam* (Fig. 5-6a and b). RT-PCR analysis of WW103-4A6 during the process of differentiation showed the expression of all the molecular markers during differentiation which was the same as the control cell line, WW93-A12 (data not shown). RT-PCR using a pair of primers specifically designed to amplify Exons 6, 7 and 8 of *Pecam* showed that the Pecam expression was completely blocked in WW103-4A6. On the other hand, WW103-8E6 and WW103-8G9, subclones in the same group as 8E6, showed normal *Pecam* expression (Fig. 5-6c).


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
a
Fgf5
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
b
vHnf1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
c
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
d
B7-1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
e
Scleraxis


f
K10
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
g
Gata4
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
h
Ttr
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
i
β-actin
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
j
Afp
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M

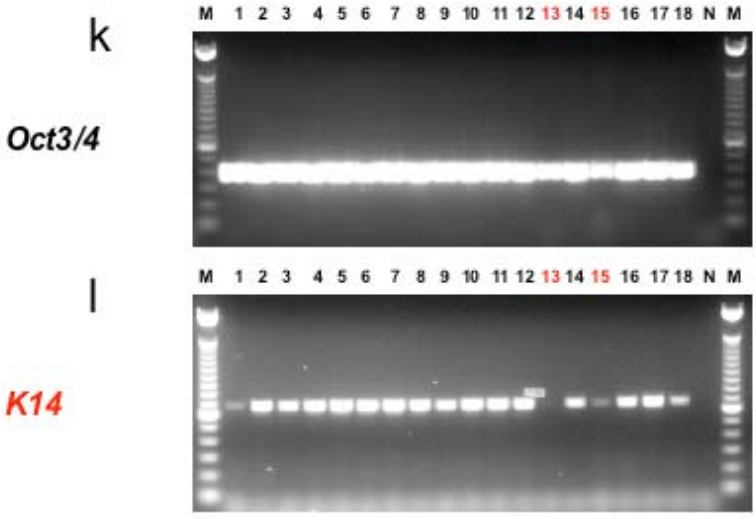
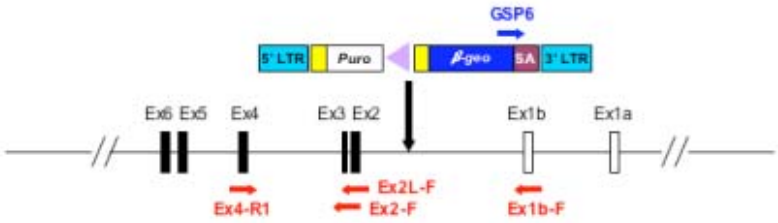


Fig. 5-4 RT-PCR results of Day 8 embryoid bodies. a. *Fgf5*. **b.** *vHnf1*. **c.** *Cx40*. **d.** *B7-1*. **e.** *Scleraxis*. **f.** *K10*. **g.** *Gata4*. **h.** *Ttr*. **i.** *β-actin*. **j.** *Afp*. **k.** *Oct3/4*. **l.** *K14*. Note that lane 13 is WW103-8E6 (marked in red). Lane 1 is the AB2.2 wild type ES cell line. Lane 2 is the WW93-A12 control line, all the other lanes are irrelevant mutant cell lines. At day 8, EBs derived from WW103-8E6 didn't express *Fgf5, vHnf, B7-1, Gata4 and K14*. The expression of *Scleraxis* and *K10* was also down-regulated compared to WW93-A12 control. Markers that are down-regulated are marked in red. Interestingly, the expression of various markers differs between EBs from AB2.2 (lane 1) and those from WW93A12 (lane 2).The expression pattern of the mutant cell lines is more similar to WW93-A12. In the whole screen, WW93A12 and its subclones were used as control.

a

b

c


GSP6
5' LTR
Puro
β-geo
SA
3' LTR
Ex6 Ex5
Ex4
Ex3 Ex2
Ex1b
Ex1a
Ex4-R1
Ex2L-F
Ex2-F
Ex1b-F

d

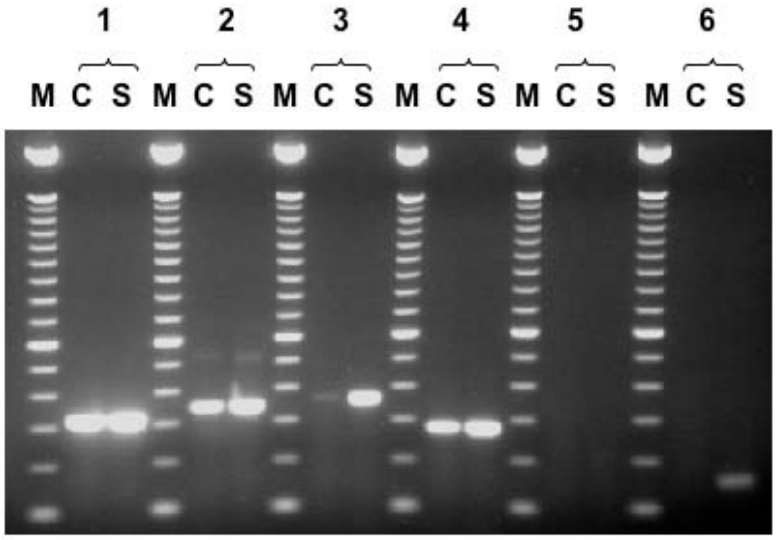
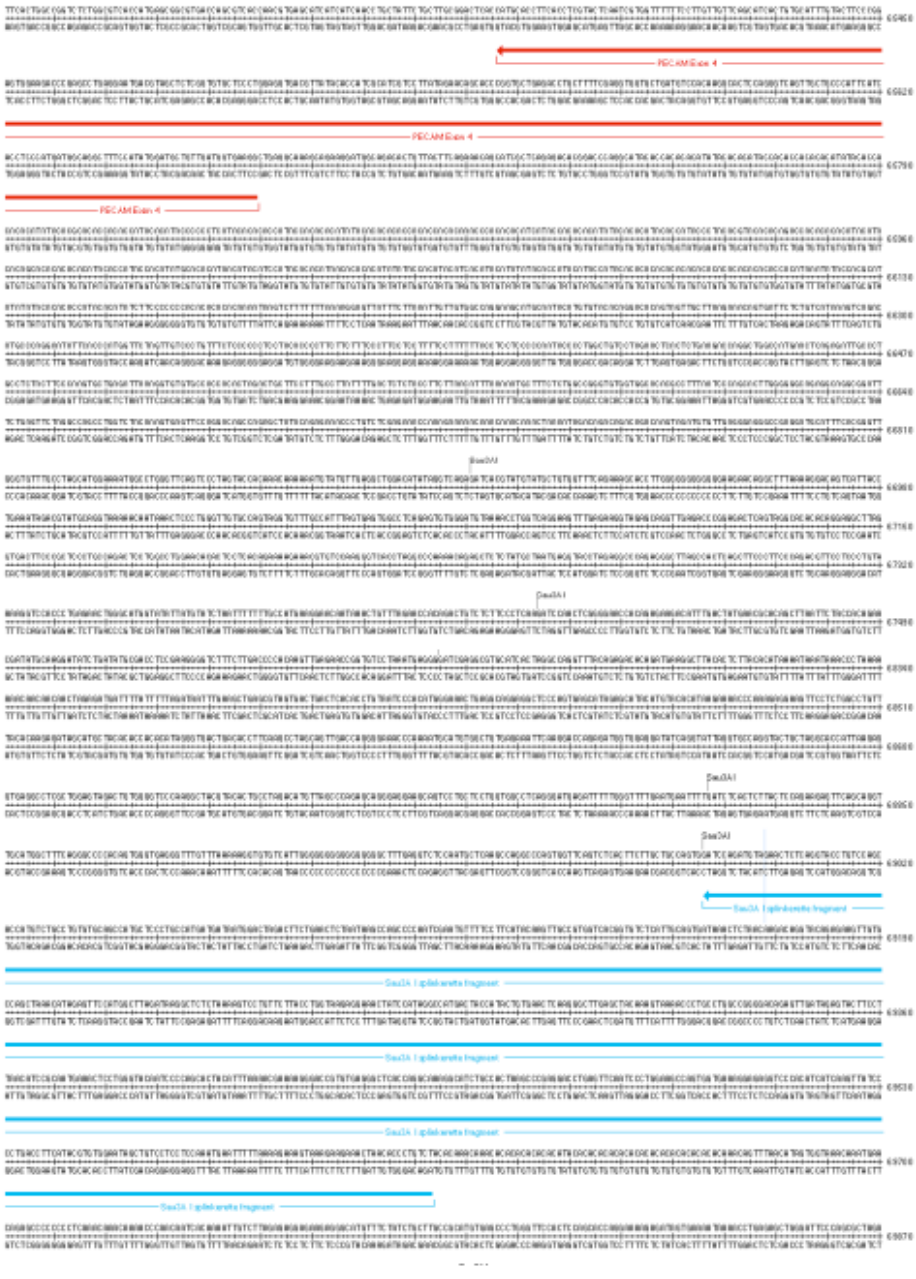

1
2
3
4
5
6
M C S M C S M C S M C S M C S M C S

Fig. 5-5 WW103-8E6. a. Splinkerette PCR result. *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product mapped the proviral insertion site to the first intron of the *Pecam* gene. The second and the third exons of *Pecam a*re marked in red, and the splinkerette PCR fragment is marked in blue. **b.** 5'RACE result. The 5' RACE product was mapped to an alternative exon of *Pecam*.The splice acceptor of the SA-βgeo cassette is marked in black, the *Pecam* alternative exon 1b is marked in red, the polyC tail is marked in green. **c.** Schematic illustration of the proviral insertion in WW103-8E6. *Pecam* gene specific primers were designed to decide the structure of the clone. PECAM-Ex1b-F, PECAM-Ex2L-F, PECAM-Ex2-F and PECAM-Ex4-R1 primers are shown as red arrows, SA-βgeo cassette specific primer GSP6 is shown as a blue arrow, the coding exons are marked in black. **d.** RT-PCR result. 1: positive control, *β-actin* RT primers; 2: PECAM-Ex1b-F/PECAM-Ex4-R1; 3: PECAM-Ex2L-F/PECAM-Ex4-R1; 4: PECAM-Ex2-F/PECAM-Ex4-R1; 5: negative control, no primers were added. 6: PECAM-Ex1b-F/GSP6. M, 100 bp ladder (Invitrogen); C: RNA extracted from undifferentiated WW93-A12 ES cells; S: RNA extracted from undifferentiated WW103-8E6 ES cells. Note that in undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. In undifferentiated WW103-8E6 ES cells, a large portion of Pecam transcripts start from Exon 2. But significant amount of transcripts started from Exon 1b were still detected.

a


PECAM Exon 4
PECAM Exon 4
PECAM Exon 4
Sau3AI
Sau3AI
Sau3AI
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment

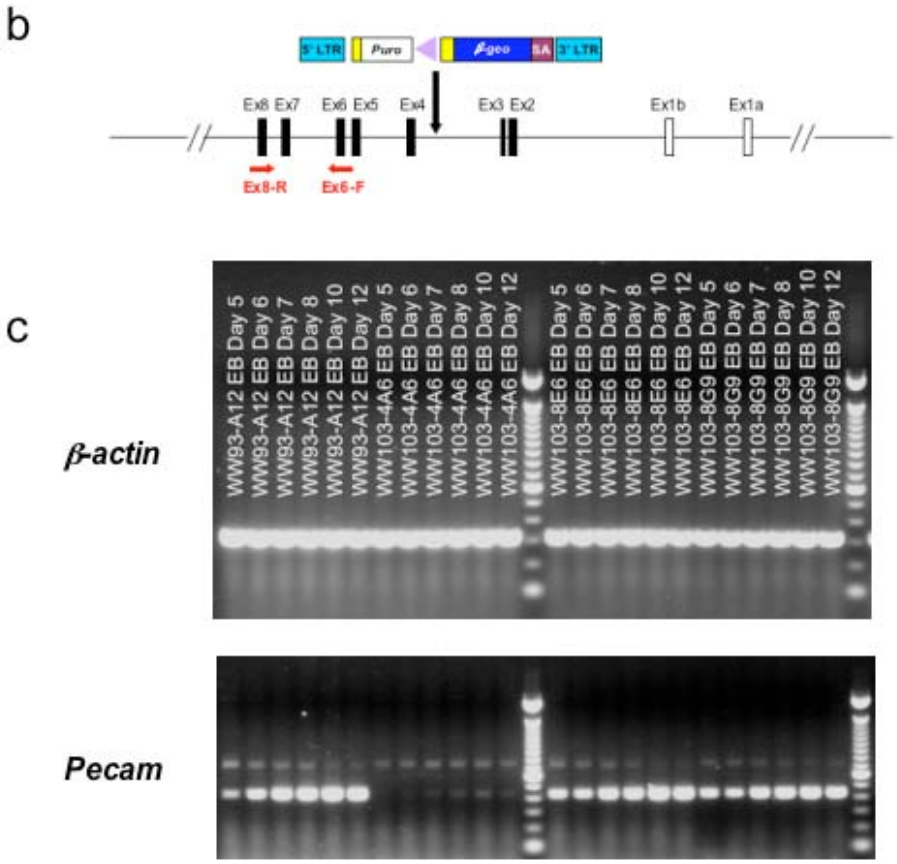


Fig. 5-6 WW103-4A6. a. Splinkerette PCR result. *Sau*3AI Splinkerette PCR product mapped the proviral insertion site to the third intron of the *Pecam* gene. The forth exon of *Pecam* is marked in red, and the splinkerette PCR fragment is marked in blue. **b.** Schematic illustration of the proviral insertion in WW103-4A6. *Pecam* gene specific primers were designed to confirm the structure of the allele. PECAM-Ex6-F and PECAM-Ex8-R primers are shown as red arrows, the coding exons are marked in black. **c.** RT-PCR result. The expression of *Pecam* during *in vitro* differentiation of WW93-A12, WW103-4A6, WW103-8E6 and WW103-8G9 was examined by RT-PCR using PECAM-Ex6-F and PECAM-Ex8-R primers. Note that the *Pecam* expression was completely blocked in WW103-4A6, but not in WW103-8E6 and WW103-8G9.

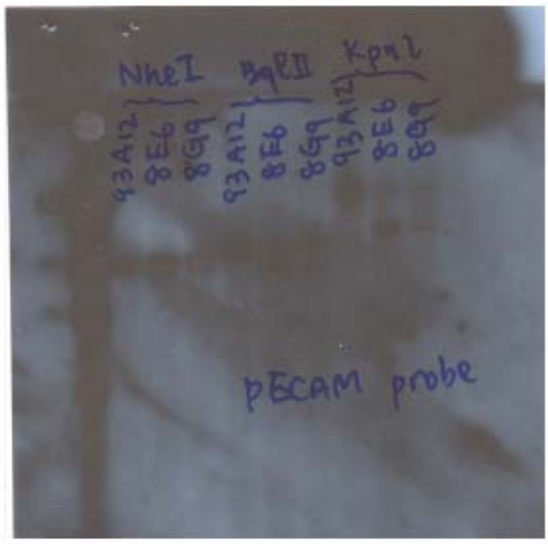
In an attempt to resolve how this situation could have occurred, Southern analysis was performed using a *Pecam* specific probe (Fig. 5-7). This revealed that both WW103-8E6 and WW103-8G9 were heterozygous for *Pecam* locus. But interestingly, the ratio between the targeted restriction fragment and the wild-type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The unexpected Southern result suggested that both clones might be trisomic. If so, it is most likely that the trisomy appeared after the end point cassette targeting and before the retrovirus infection. In this case, the original trisomy would contain two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild-type chromosome. After regional trapping, the puromycin resistant trisomy will have one 3' *Hprt* chromosome with an inversion, one 3' *Hprt* chromosome with targeted end point cassette and one 5' *Hprt* wild-type chromosome. Induced mitotic recombination can generate two different products: clones with two inversion chromosomes and one chromosome with the end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the end point cassette (WW103-8G9). In both cases, the clones will carry three targeted E_2DH alleles (end point targeting), thus Southern analysis using E_2DH probe can not distinguish these trisomies from the homozygous inversion clones.

Therefore the impaired differentiation potential of WW103-8E6 does not have any direct connection with the *Pecam* trapping and the subsequent inversion. This may be the result of the up-regulation of the chromosome 11 genes caused by the extra chromosome.

5.2.4 WW103-14F11 (*2810410L24Rik*)

As described in the previous chapter, the WW103-14F11 subclone has a proviral insertion at the *2810410L24Rik* locus (119.9 Mb) (Fig. 5-8a), which is close to the telomere of the chromosome 11. But instead of trapping the *2810410L24Rik* gene, which is transcribed from the sense strand (from centromere to telomere), the retrovirus trapped another transcript transcribed from the anti-sense strand (from telomere to centromere), *D030042H08Rik*.

a


NheI
BglII
KpnI
93A12
8E6
8G9
93A12
8E6
8G9
93A12
8E6
8G9
pECAM probe

b

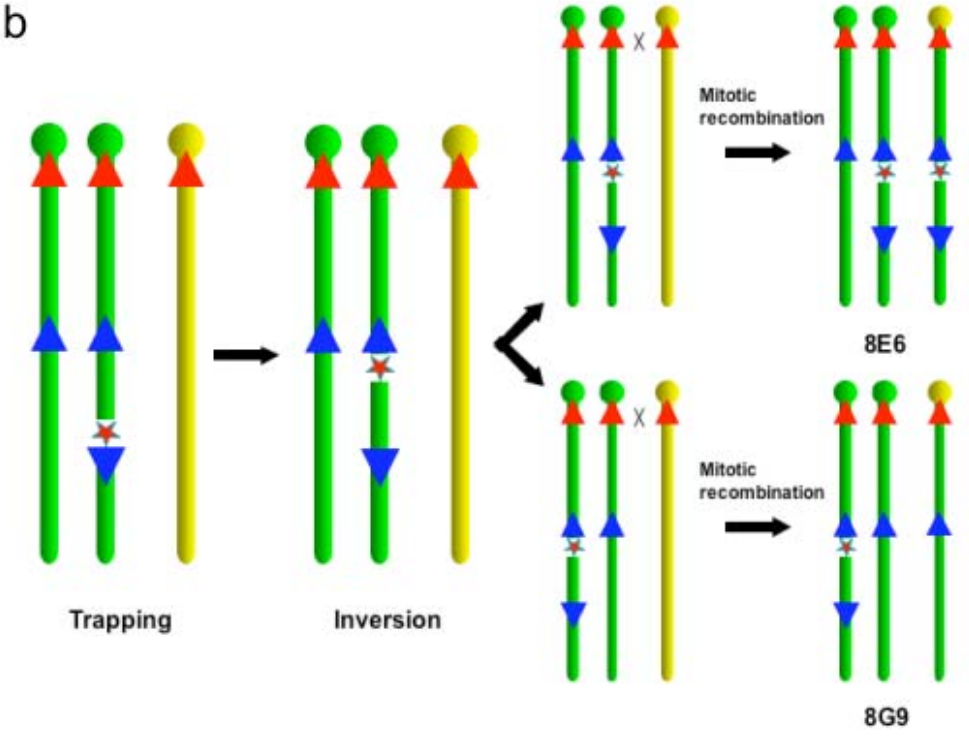
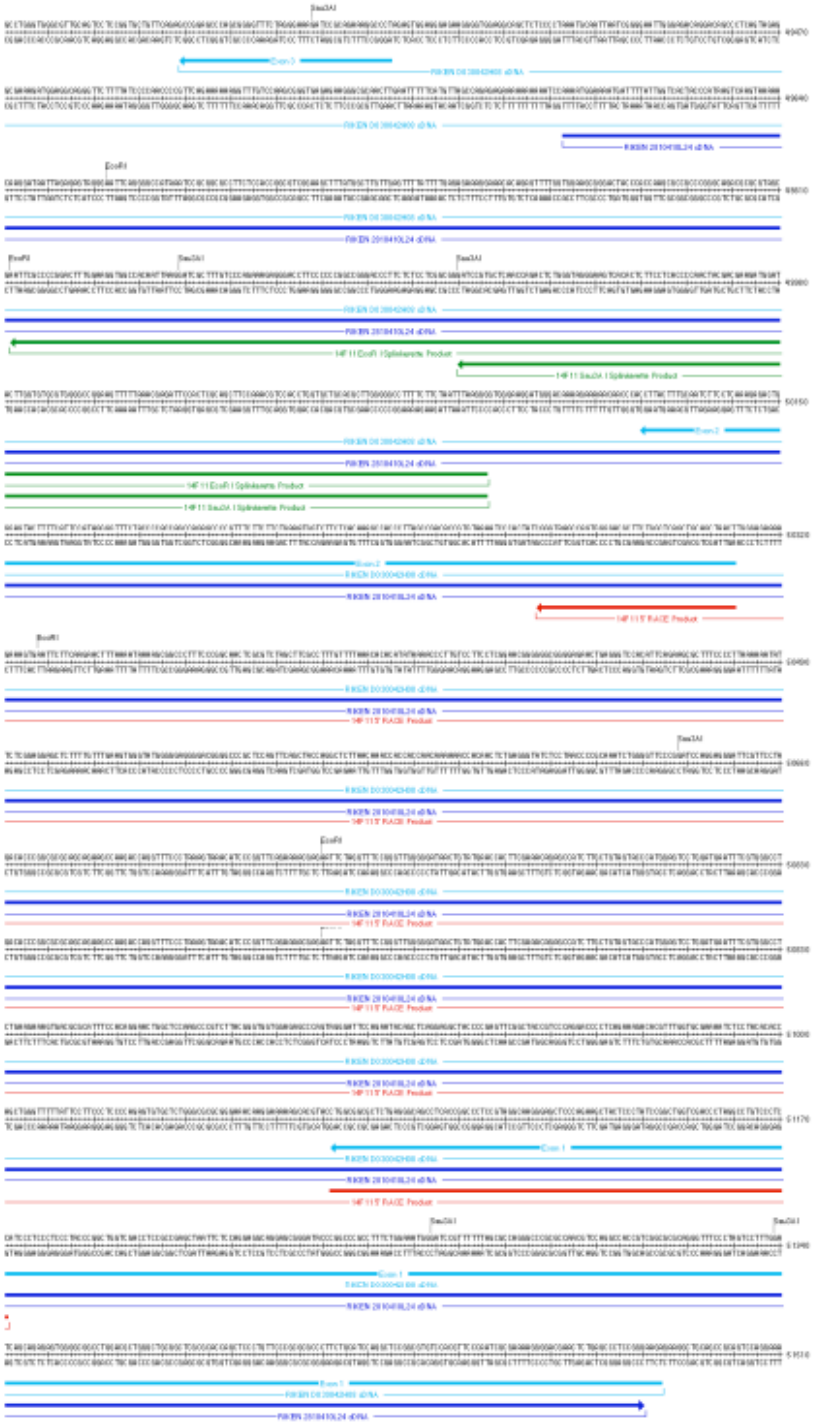
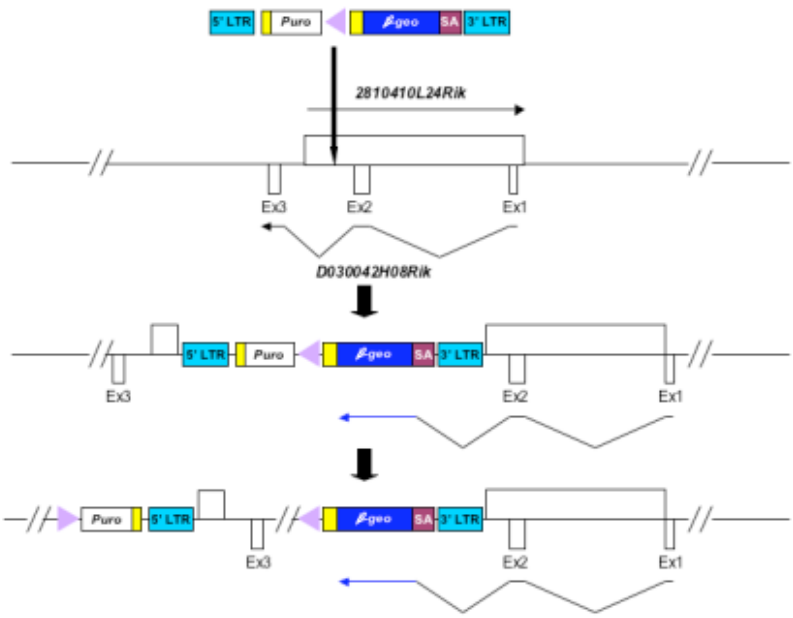

Trapping
Inversion
Mitotic recombination
8E6
Mitotic recombination
8G9

Fig. 5-7 WW103-8E6 and 8G9 are both chromosome 11 trisomies. a. Southern analysis of clones, WW103-8E6 and WW103-8G9. A *Pecam* specific probe was used to detect a 10.6 kb *Bgl*II wild type fragment and a 9 kb *Bgl*I targeted fragment. The same probe was also used to detect a 30 kb *Kpn*I wild type fragment and a 19 kb targeted fragment. Both WW103-8E6 and WW103-8G show a wild type fragment for both digestions. But the ratio between the targeted restriction fragment and the wild type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The Southern result suggested that both clones might be trisomic. **b.** A schematic illustration of possible recombination in WW103-8E6 and WW103-8G. The original trisomy probably contained two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild type chromosome. After regional trapping, the puromycin resistant trisomy has one 3' *Hprt* chromosome with the inversion, one 3' *Hprt* chromosome with the end point cassette and one 5' *Hprt* wild type chromosome. Induced mitotic recombination can generate two different products, depending on which 3' *Hprt* chromosome participates in the recombination process. Clones can be generated with two inversion chromosomes and one chromosome with the targeted end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the targeted end point cassette (WW103-8G9).

a



b


5' LTR
Puro
βgeo
SA
3' LTR
2810410L24Rik
Ex3
Ex2
Ex1
D030042H08Rik

c

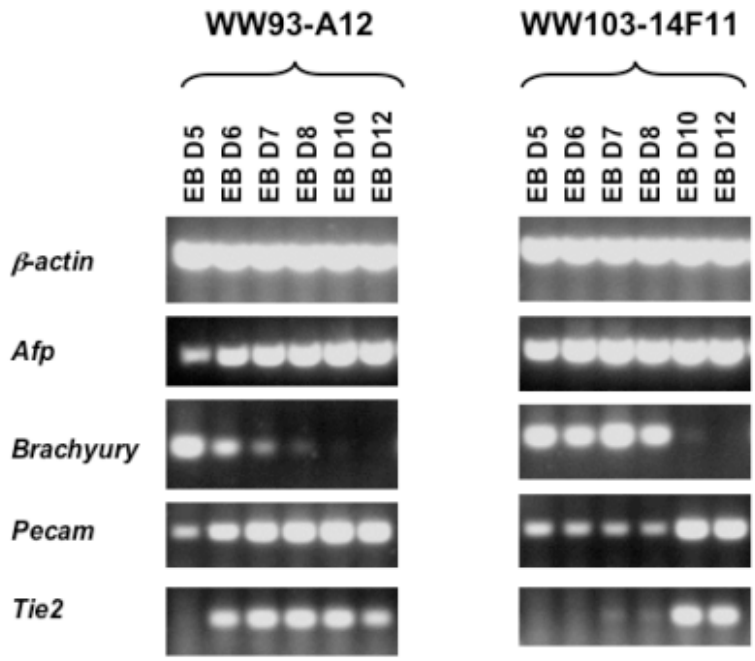

WW93-A12
WW103-14F11
EB D5
EB D6
EB D7
EB D8
EB D10
EB D12
β-actin
Afp
Brachyury
Pecam
Tie2

Fig. 5-8 WW103-14F11. a. Splinkerette PCR and 5' RACE.results. *Eco*RI and *Sau*3AI Splinkerette PCR products have both mapped the proviral insertion site to the *2810410L24Rik* locus. The 5' RACE result matched another transcript *D030042H08Rik*.The *2810410L24Rik* transcript is marked in dark blue, the *D030042H08Rik* transcript is marked in light blue, the splinkerette PCR fragments are marked in green and the 5'RACE product is marked in red. **b.** Schematic illustration of the structure of proviral insertion and the subsequent inversion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note that *Pecam* and *Tie2* RT-PCR results showed that the up-regulation of these two markers in the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed.

The splinkerette results mapped the proviral insertion site between the second and third exons of *D030042H08Rik*. However, the 5' RACE result did not match the *D030042H08Rik* cDNA sequence perfectly, although the transcript structure is similar. It is possible that the 5' RACE result and the *D030042H08Rik* cDNA sequence represent two different alternative splice forms of the same gene.

Nevertheless, the gene-trap retrovirus insertion and the subsequent inversion will disrupt the transcripts from both strands (Fig. 5-8b). The *in vitro* differentiation results showed that EBs derived from WW103-14F11 have impaired potential to develop into endothelial cells. RT-PCR using *Pecam* and *Tie2* primers has shown that the up-regulation of the expression of these two markers during the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed (Fig. 5-8c).

Further confirmation of this subclone is still undergoing. One way to directly confirm the defective endothelial cell differentiation is to use collagen IV coated dishes to induce undifferentiated ES cells to first differentiate into $Flk1^+$ cells (Yamashita, Itoh et al. 2000). When FACS sorted $Flk1^+$ cells were cultured with the addition of VEGF, these cells will further differentiate into PECAM1+ sheets of endothelial cells, which also express other endothelial cell-specific markers, such as *VE-cadherin* and *CD34*. By comparing the endothelia cell differentiation of the WW103-14F11 cells and the wild-type control cells, it will be possible to identify the molecular mechanism underlying the defective phenotype and determine at which stage the differentiation into endothelial lineage is blocked. However, it will be difficult to distinguish the phenotypes of the two genes transcribed from the opposite directions.

5.2.5 WW103-13D10 (*LOC217071*)

*Sau*3A1 and SpeI/*Xba*I/*Nhe*I Splinkerette PCR products mapped the proviral insertion site in the WW103-13D10 clones to the second intron of a hypothetical mouse gene, *LOC217071* (Fig. 5-9a and b). This gene is transcribed from the sense strand (from the centromere to telomere), and

Southern analysis using an *E_2DH* 3' probe has confirmed that this clone is homozygous for the targeted *E_2DH* allele. Southern analysis using a *LacZ* probe has shown that it only carried a 6.9 kb *Kpn*I restriction fragment which suggested that WW103-13D10 contains an intact proviral insertion. As discussed in the previous chapter, this might be caused by a G2 *trans* recombination event. The duplication chromosome has both a functional *Puro* and a functional *Bsd* cassette, and it can become homozygous after induced mitotic recombination because it has not lost any genetic material.

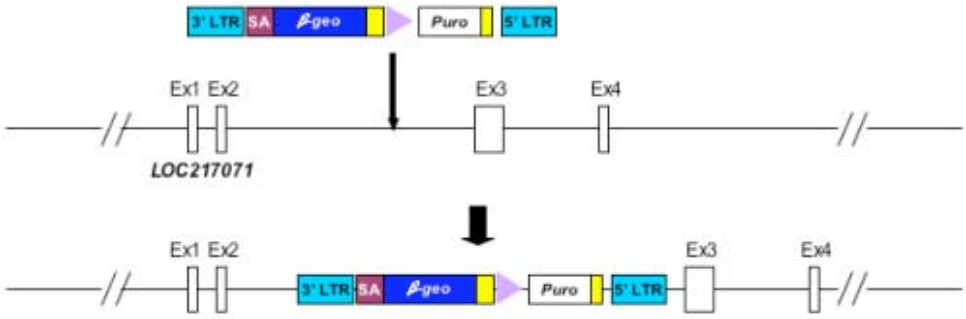
The homozygous duplication clone showed an obvious abnormality in *in vitro* differentiation. The undifferentiated WW103-13D10 ES cells expressed high levels of markers for differentiated cell types, such as *Afp*, *Gata4* and *Hnf4*. The expression of undifferentiated ES cell markers, like *Nodal* and *Oct3/4*, was significantly down-regulated, compared to the WW93-A12 control (Fig. 5-9c).

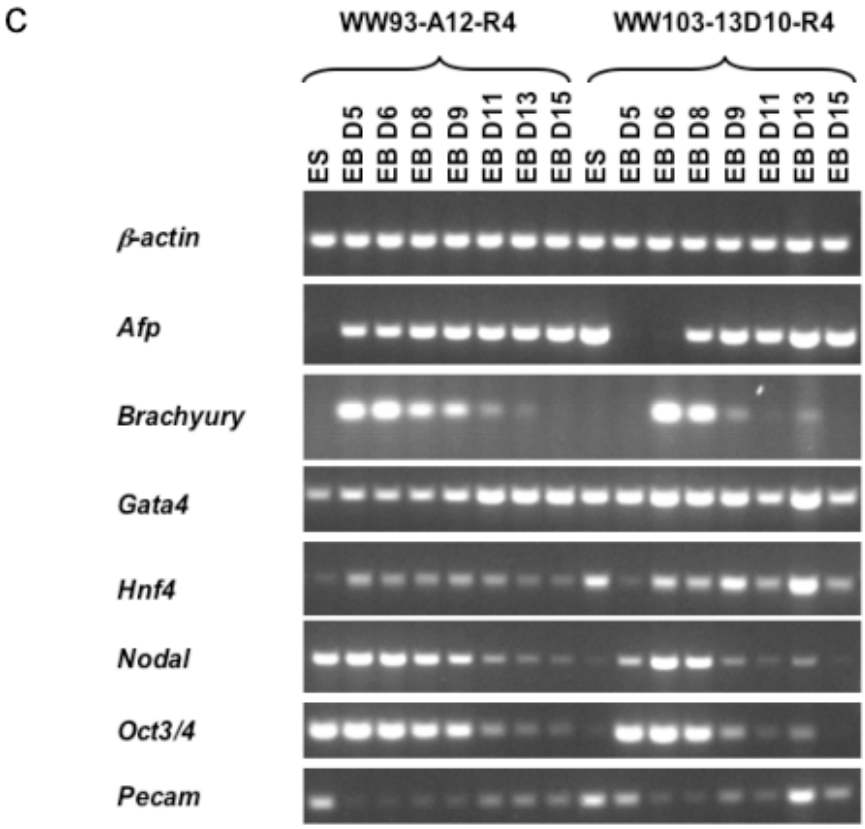
Interestingly, during the process of differentiation, the EBs made from WW103-13D10 ES cells seemed to differentiate normally. At day 5, they lost the expression of *Afp*, but regained the expression of *Nodal* and *Oct3/4*. After this, various markers showed expression patterns similar to those which were observed in the WW93-A12 control. But *Hnf4* and *Gata4* expression were still significantly up-regulated compared to the control.

Sib-selection was carried out on two subclones each from WW93-A12 and WW103-13D10. The same number of undifferentiated ES cells were plated into the wells of a gelatinized 24-well plate and selected in M15, M15+G418, M14+puromycin, M15+blasticidin and M15+HAT, respectively. WW103-13D10 was resistant to both puromycin and blasticidin, which suggested that this clone have two duplication chromosomes, instead of two inversion chromosomes (Fig. 5-9d). Most likely, the phenotype observed in WW103-13D10 was caused by the duplication, instead of the disruption of the *LOC217071* locus.

a

b


3' LTR
SA
β-geo
Puro
5' LTR
Ex1 Ex2
Ex3
Ex4
LOC217071
Ex1 Ex2
3' LTR
SA
β-geo
Puro
5' LTR
Ex3
Ex4


c
WW93-A12-R4
WW103-13D10-R4
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Afp
Brachyury
Gata4
Hnf4
Nodal
Oct3/4
Pecam

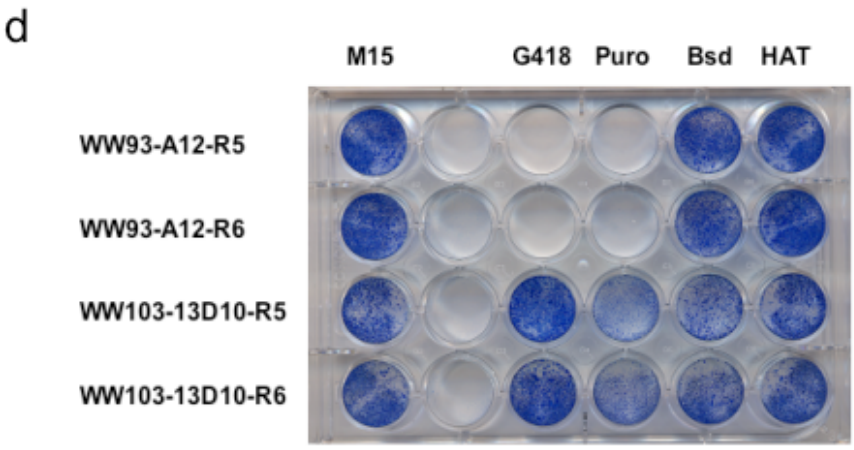

d
M15
G418
Puro
Bsd
HAT
WW93-A12-R5
WW93-A12-R6
WW103-13D10-R5
WW103-13D10-R6

Fig. 5-9 WW103-13D10 a. Splinkerette PCR results. *Spe*I/*Xba*I/*Nhe*I and *Sau*3AI Splinkerette PCR products have mapped the proviral insertion site to the *LOC217071* locus. The splinkerette PCR fragments are marked in green. **b.** Schematic illustration of the structure of proviral insertion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note undifferentiated WW103-13D10 ES cells express high amounts of *Afp*, *Hnf4* and *Gata4,* but low amount of *Oct3/4* and *Nodal*. Despite the altered expression of these markers,the EBs derived from WW103-13D10 ES cells seem to differentiate normally. **d.** Sib-selection of WW93-A12 and WW103-13D10 clones. Two subclones of each cell line were used for the sib selection. Same amount of undifferentiated ES cells were plated onto a gelatinized 24-well plate and selected with M15, M15+G418, M14+Puromycin, M15+Blasticidine and M15+HAT, respectively. WW103-13D10 was resistant to both Puromycin and Blasticidine, which suggested that this clone is a homozygous duplication, instead of a homozygous inversion.

5.2.6 WW103-18F11 (*Acly*)

One of the mitotic recombination clones, WW103-18F11, showed impaired *in vitro* differentiation potential. After EBs made of WW103-18F11 ES cells were plated on the gelatinized tissue culture plates at Day 5, the EBs did not form cystic three-dimensional structures.

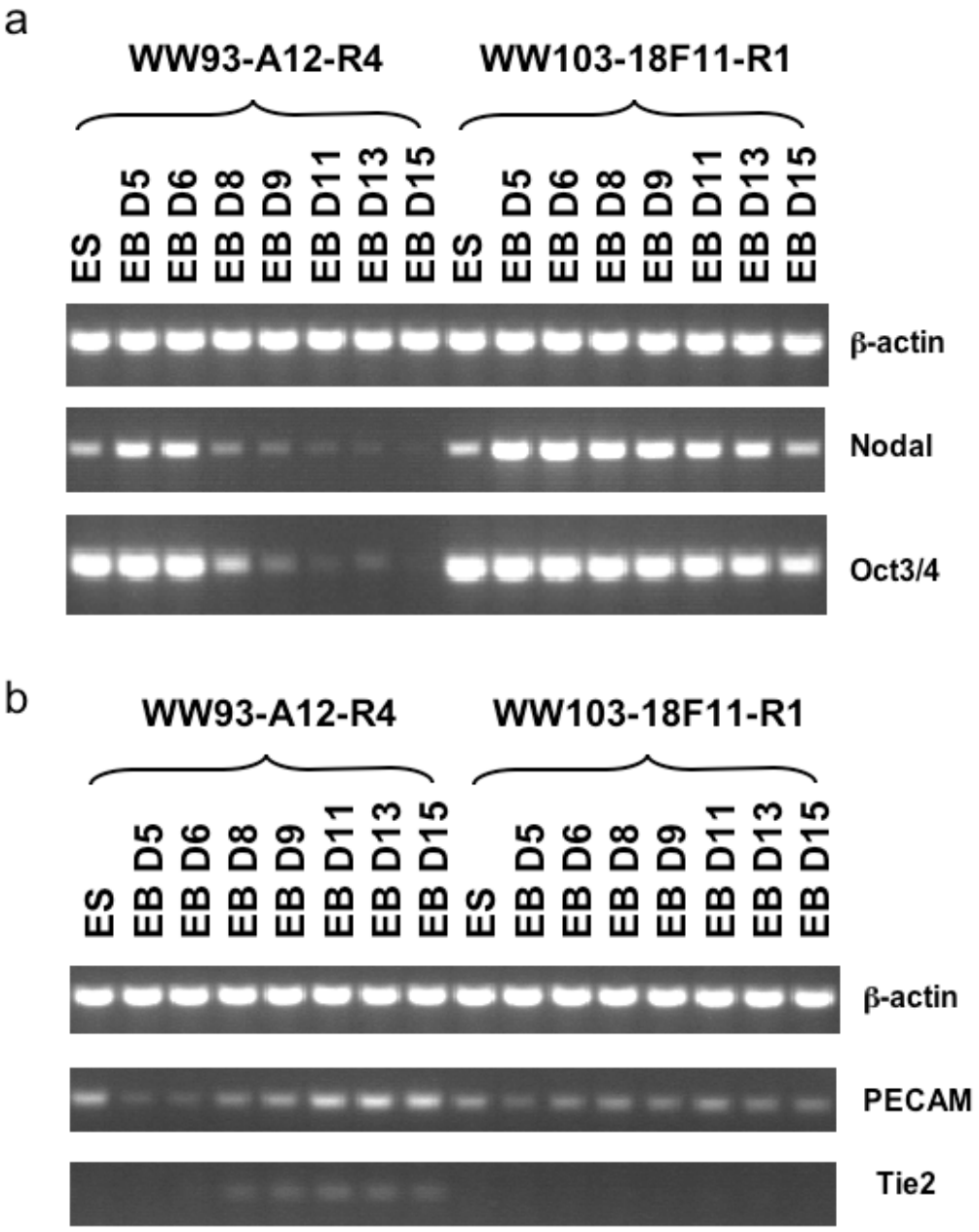
When RT-PCR was performed on RNA extracted from WW103-18F11 embryoid bodies collected at different time points, these EBs were found to express high levels of the undifferentiated ES cell markers, *Oct3/4* and *Nodal*, as late as Day 18 of the *in vitro* differentiation protocol. The expression of *Oct3/4* and *Nodal* still decreased a little during the differentiation process, but down-regulation was not as rapid as that in the control cell line (Fig. 5-10a and e).

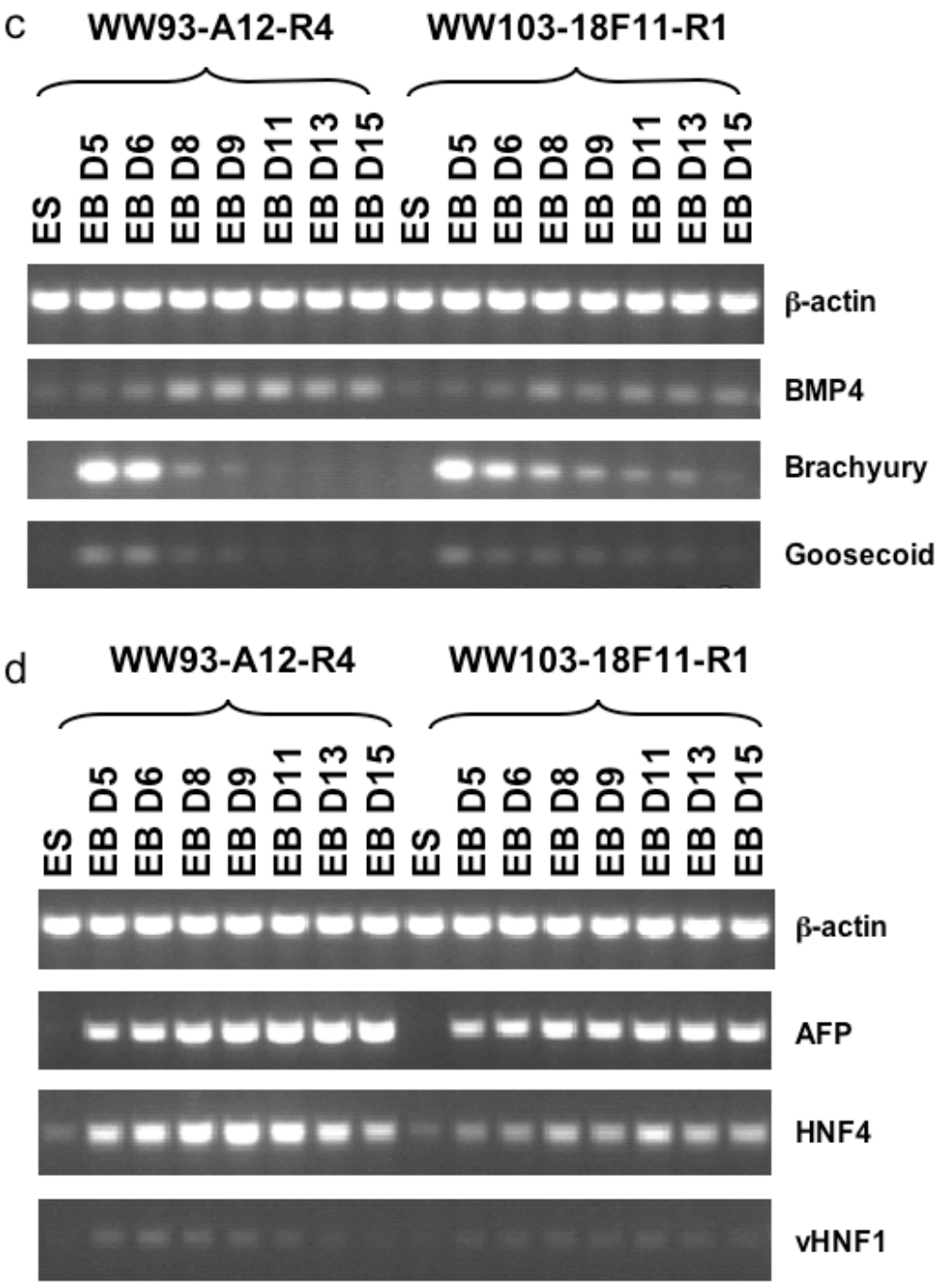
Tie2 expression was not detected during the whole process of differentiation of WW103-18F11 cells. The expression of *Pecam* was maintained at a constant basal level, instead of being up-regulated, as was observed in the WW93-A12 control cell line (Fig. 5-10b). Both of the markers are endothelial cell-specific proteins expressed during the formation of vascular structures in ES-derived EBs. The *Tie2* gene encodes a growth factor receptor, while the *Pecam* protein is an endothelial cell specific antigen. Vittet *et al*. (1996) has shown that both genes are expressed at low levels in undifferentiated ES cells. Normally, in the process of *in vitro* differentiation, the expression of both genes is absent at day 0-3 and is detected again from day 4. After this, the expression level of both genes is consistently up-regulated, as detected by Northern blotting and/or Immunofluorescence. However, in that experiment, only EBs from Day 3 to Day 7 were checked (Vittet, Prandini et al. 1996). In my experiment, I have observed the expression of *Tie2* and *Pecam* in the control line throughout the 15-day differentiation process. Thus, my observation suggested that the differentiation of endothelial cells in the mutant cell line was significantly impaired over the entire 15-day differentiation process.

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury*, *Goosecoid*) is similar between WW103-18F11 and WW93-A12 (Fig. 5-10c). However, low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. The expression of one of the endodermal markers, *Hnf4*, in WW103-18F11 was much lower than that in the control. Apart from these changes, no major differences were observed in the levels of expression of the other markers (Fig. 5-10d).

5' RACE results revealed that the gene-trap retrovirus trapped Exon 1 of ATP-citrate lyase (*Acly*) (Fig. 5-11a). *Acly* is one of two cytosolic enzymes in eukaryotes that synthesize acetyl-coenzyme A (acetyl-CoA), the other enzyme is acetyl-CoA synthetase 1. *Acly* catalyzes the formation of acetyl-coenzyme A (CoA) from citrate and CoA, and hydrolyzes ATP to ADP and phosphate. Because acetyl-CoA is an essential component for cholesterol and triglycerides synthesis, *Acly* is believed to be a potential therapeutic target for hyperlipidemias ad obesity (Beigneux, Kosinski et al. 2004).

To characterize this mutant cell line further, pure subclones of WW103-18F11 were derived by low density plating to form single colonies. Six subclones were picked and expanded. To confirm chromosomal structure of these subclones, sib-selection was performed on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of the control cell line, WW93-A12-R4 and WW93-A12-R5. An equal number of ES cells from each subclone were plated onto multiple 6-well plates and selected with M15, M15+puromycin, M15+blasticidin, M15+G418 and M15+HAT, respectively. As expected, the two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggests that WW103-18F11 have undergone correct recombination (Fig. 5-11b).


a
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
b
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
PECAM
Tie2


c
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
BMP4
Brachyury
Goosecoid
d
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
AFP
HNF4
vHNF1

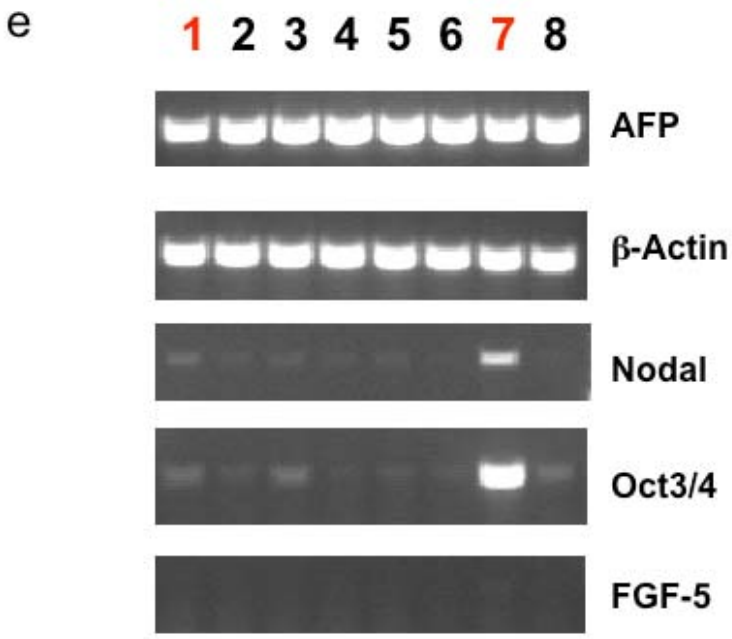
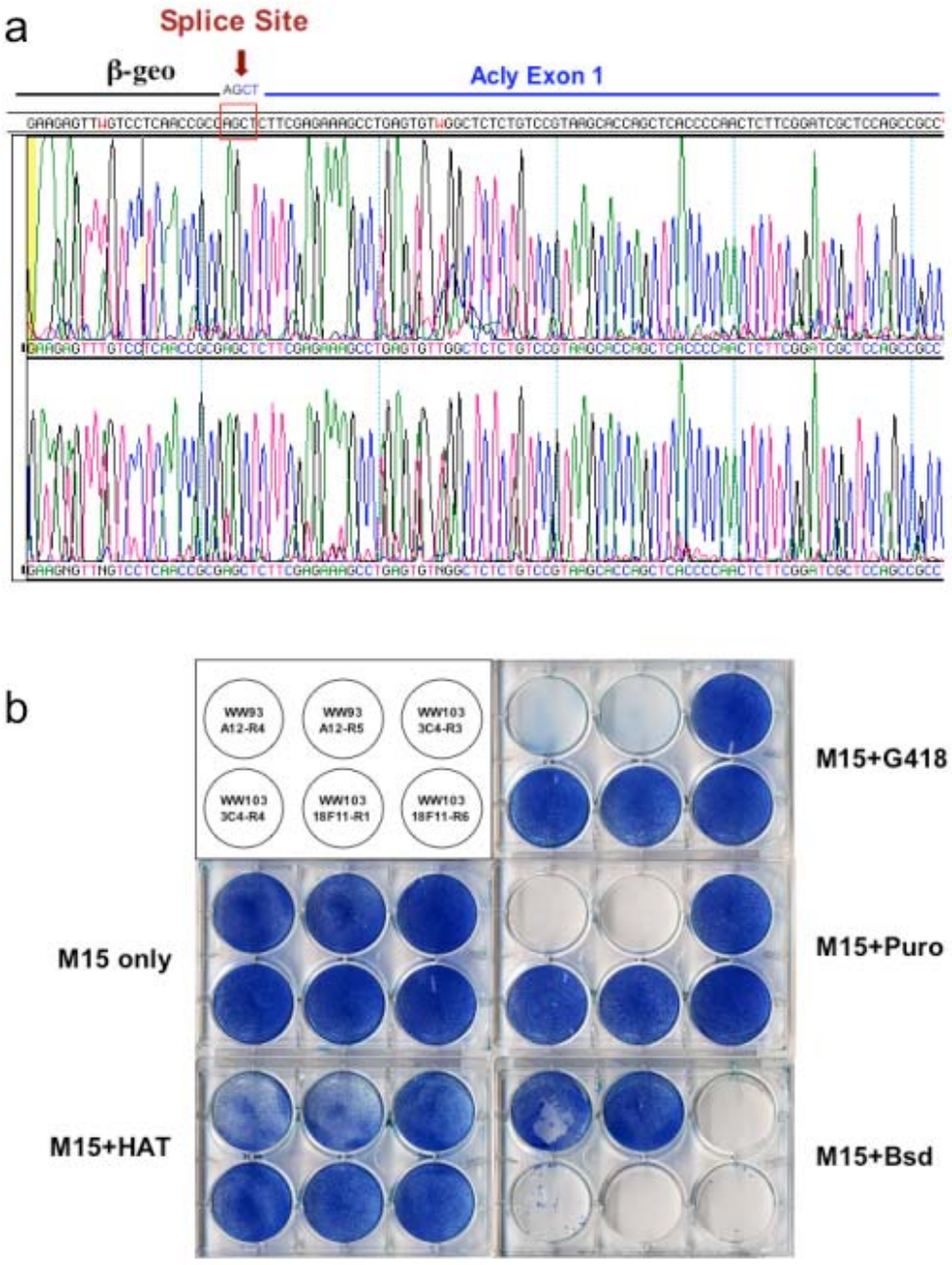
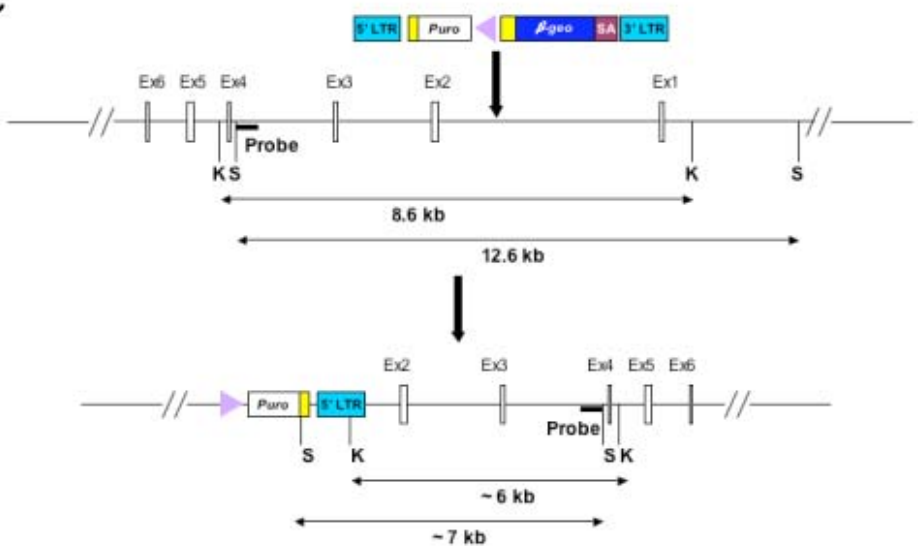


Fig. 5-10 RT-PCR results of WW103-18F11. a. Primitive ectoderm markers. The expression of *Oct3/4* and *Nodal* decreased slightly in the differentiation process. **b.** Endothelial markers. *Tie2* expression was not detected throughout the differentiation process of WW103-18F11 cells. The expression of *Pecam* was not up-regulated as observed in WW93-A12 control cell line. **c.** Early mesoderm markers. Low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. **d.** Endoderm markers. The expression of most endoderm markers appears quite similar between the two cell lines. However, the expression of *Hnf4* in WW103-18F11 was much lower than that in the control. **e.** Day 18 embryoid bodies RT-PCR results. Note that WW103-18F11 EBs still express high amount of *Nodal* and *Oct3/4* at day 18. Lane 1, WW93-A12 control cell line; lane 7, WW103-18F11. Lanes 2-6 and 8 are day 18 EBs derived from other irrelevant mutant cell lines.


a
Splice Site
β-geo
AGCT
Acly Exon 1
b
WW93 A12-R4
WW93 A12-R5
WW103 3C4-R3
WW103 3C4-R4
WW103 18F11-R1
WW103 18F11-R6
M15+G418
M15 only
M15+Puro
M15+HAT
M15+Bsd

c


5' LTR
Puro
βgeo
SA
3' LTR
Ex6
Ex5
Ex4
Ex3
Ex2
Ex1
Probe
K S
K
S
8.6 kb
12.6 kb
Puro
5' LTR
Ex2
Ex3
Ex4
Ex5
Ex6
Probe
S
K
S K
~ 6 kb
~ 7 kb

d

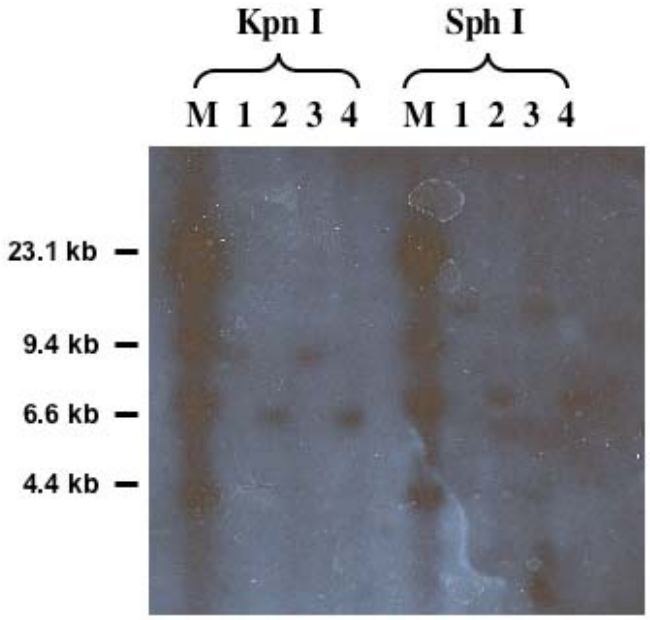

Kpn I
Sph I
M 1 2 3 4
M 1 2 3 4
23.1 kb
9.4 kb
6.6 kb
4.4 kb

Fig. 5-11 Identification of WW103-18F11. a. 5' RACE result. The sequence of the 5' RACE product matched the first exon of *Acly* gene. **b.** Sib-selection of WW103-18F11 and WW93-A12. Sib-selection was carried out on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of control cell line, WW93-A12-R4 and WW93-A12-R5. The same number of ES cells from two subclones each of WW93-A12 and WW103-18F11 were plated onto gelatinized 6-well plates and selected with M15, M15+Puromycin, M15+Blasticidine, M15+G418 and M15+HAT, respectively. The two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggested that it is a homozygous inversion clone. **c.** Schematic illustration of the structure of the proviral insertion and the subsequent inversion in WW103-18F11. **d.** Southern analysis of WW103-18F11. Southern analysis has been carried out using an *Acly* gene specific probe. This probe can detect an 8.6 kb *Kpn*I restriction fragment for the wild type allele. For WW103-18F11, the probe only detected an 6 kb *Kpn*I restriction fragment for the gene trap allele. Also, the probe only detected an 7 kb *Sph*I restriction fragment for the gene trap allele, instead of the 12.6 kb *Sph*I restriction fragment for the wild type allele. This Southern result has confirmed that both *Acly* alleles have been disrupted by the gene trap insertion. M, λ *Hin*dIII marker (New England Biolabs); lane 1, WW93-A12-R4; lane 2, WW103-18F11-R1; lane 3, WW93-A12-R5; lane 4, WW103-18F11-R6.

Southern analysis was carried out using an *Acly* gene specific probe. When this probe was hybridized to *Kpn*I digested genomic DNA, it detects an 8.6 kb wild-type fragment and an approximately 7 kb gene-trap fragment. When this probe was hybridized to *Sph*I digested genomic DNA, it detects a 12.6 kb wild-type fragment and an approximately 8 kb gene-trap fragment. As expected, only the gene-trap fragment was detected in the WW103-18F11 subclones. This Southern result confirms that both alleles of the *Acly* gene have been disrupted by the gene-trap insertion.

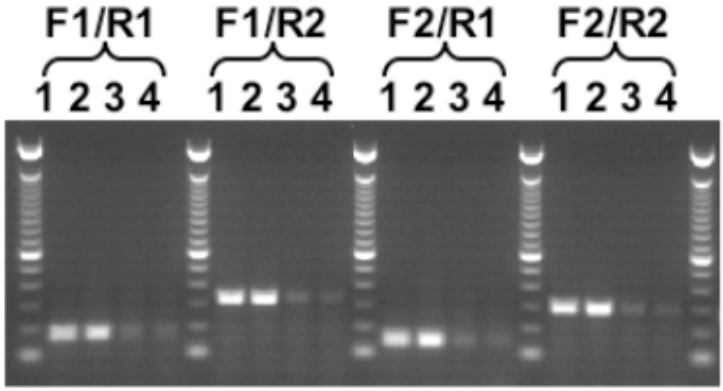
To see whether the gene-trap insertion and the subsequent inversion has disrupted transcription of the locus, PCR primers were used to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). First strand cDNA was synthesised using total RNA extracted from WW103-18F11 and WW93-A12 ES cells. The RT-PCR results showed that transcription from Exon 1 to downstream exons was blocked. Weak PCR bands were detected for the *Acly*-deficient cell lines, which are likely to be contamination from feeder cells (Fig. 5-12a). Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the Exon 1/β-geo fusion transcript from the trapped allele. As expected, specific bands were only detected for WW103-18F11 subclones, but not for WW93-A12 control (Fig. 5-12c).

To see whether the gene-trap insertion and the subsequent inversion has affected the transcription of downstream exons, PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). Specific PCR bands were detected in the WW103-18F11 mutant cell line and the WW93-A12 control cell line. Therefore it is likely that there is an alternative transcription start point between the retroviral insertion point and Exon 24, but the precise location of the mutant transcript start in WW103-18F11 is not known (Fig. 5-12b).

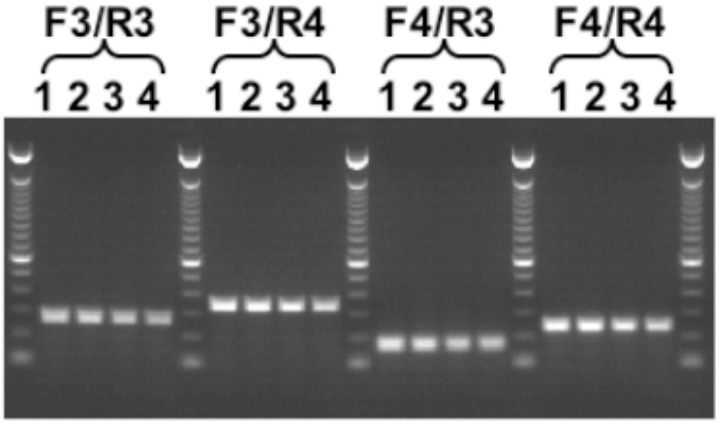
These RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation. In the WW93-A12 control cell line, *Acly* was highly expressed in undifferentiated ES cells, as well as throughout the whole differentiation process. In the WW103-18F11 mutant cell line, the F1/R1 and F2/R2 primer pairs did not detect the expression of the *Acly* upstream exons during *in vitro* differentiation, but the F3/R2 and F4/R4 primer pairs did detect expression of the *Acly* downstream exons (Fig. 5-12d).

To identify a causal link between the gene-trap insertion and inversion at the *Acly* locus and the severely impaired differentiation potential, a BAC rescue experiment was carried out to reverse the phenotype of the WW103-18F11 ES cell clone. A 129 S7 BAC clone, BMQ-290J5 was identified in Ensembl and confirmed to contain the complete *Acly* gene by PCR (Fig. 5-13a and data not shown). A *PGK-EM7-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination (Liu, Jenkins et al. 2003). The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe (Fig. 5-13b and c).

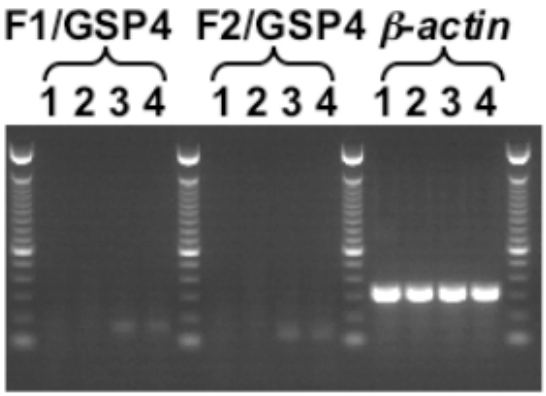
a


F1/R1
F1/R2
F2/R1
F2/R2
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

b


F3/R3
F3/R4
F4/R3
F4/R4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

c


F1/GSP4
F2/GSP4
β-actin
1 2 3 4
1 2 3 4
1 2 3 4

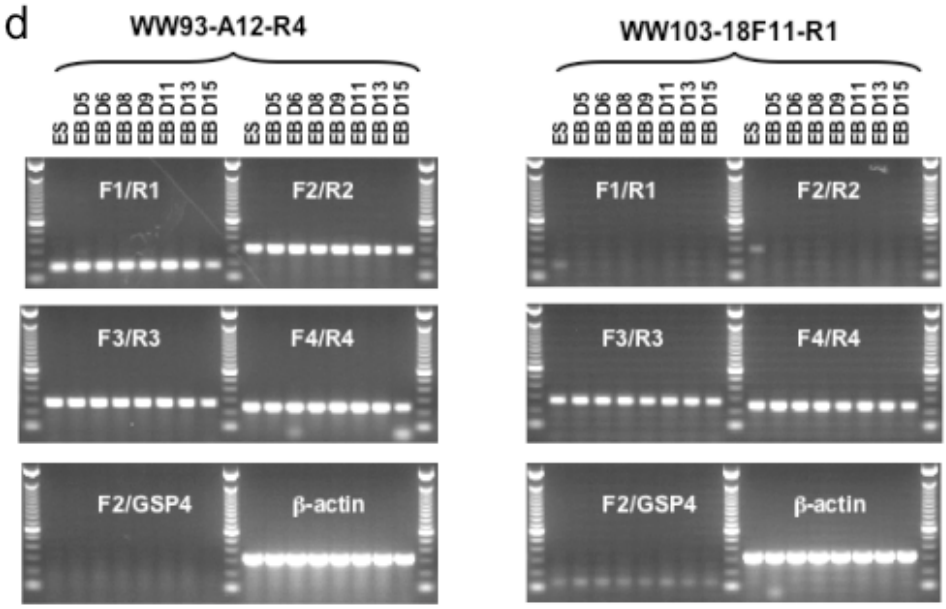


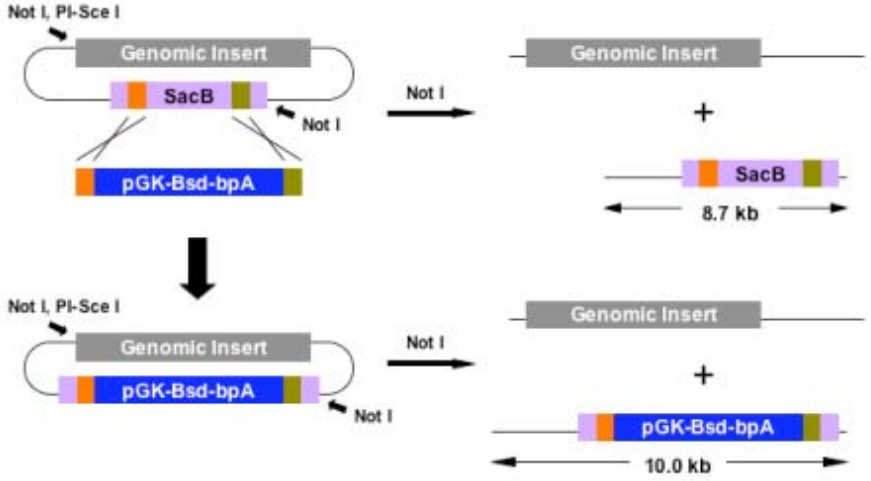
Fig. 5-12 Confirmation of WW103-18F11. a. RT-PCR to detect transcripts across the inversion breakpoint. PCR primers were designed to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). RNA extracted from undifferentiated ES cells was used as template. lane 1, WW93-A12-R4; lane 2, WW93-A12-R5; lane 3, WW103-18F11-R1; lane 4, WW103-18F11-R6. RT-PCR results showed that *Acly* transcription across the inversion breakpoint was blocked. **b.** RT-PCR to detect the transcription of downstream exons. PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). RT-PCR results showed that the transcription of downstream exons was not affected by the inversion. This suggests the existence of alternative transcript start points. **c.** RT-PCR to detect the Exon 1/β-geo fusion transcript. Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the gene trap fusion transcript. **d.** RT-PCR to detect *Acly* expression during ES cell *in vitro* differentiation. High *Acly* expression was detected in the whole process of differentiation of WW93-A12 control cell line, but not in WW103-18F11.

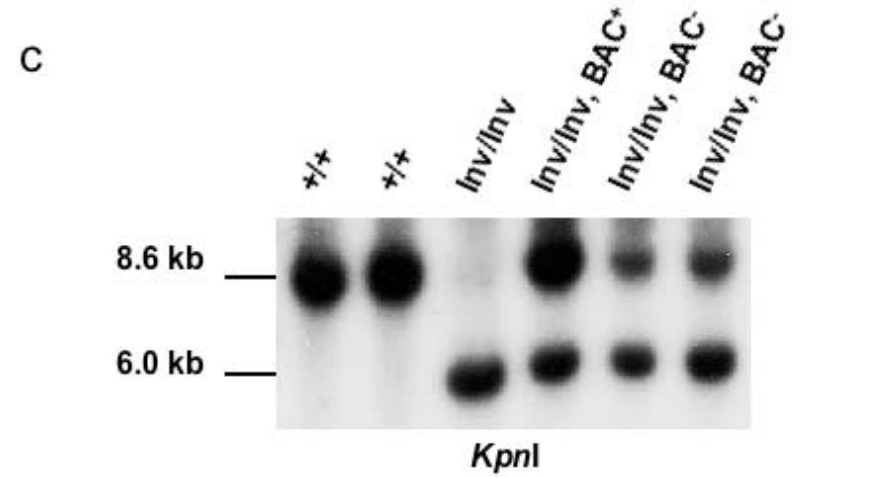
The modified BAC clone was linearized by I-*Sce*I and electroporated into WW103-18F11 ES cells (HAT^R, Neo^R, $Puro^R$, Bsd^S). 12 blasticidin resistant clones were picked and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild-type restriction fragment (Fig. 5-13c). One of the clones, WW113-2-8 has the wild-type restriction fragment and the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild-type copies of *Acly* gene. Another two clones, WW113-2-10 and WW113-2-11 also have the wild-type restriction fragment. But the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single copy of the BAC DNA, which can be randomly truncated and are likely to be incomplete. Western analysis was performed on whole-cell lysates extracted from undifferentiated wild-type control, *Acly*-deficient and BAC-rescued ES cells using a polyclonal rabbit anti-*Acly* antibody. *Acly* protein was not detected in the lysates from the WW103-18F11 cells. However, one of the BAC-rescued clones (WW113-8) expressed similar level of the *Acly* protein as the WW93-A12 wild-type control cells, indicating that this clone (WW113-2-8) is a rescued clone (Fig. 5-13d). The other two clones, WW113-2-10 and 2-11, which did not express *Acly* protein, are likely to only contain a truncated form of the BAC DNA and thus they were used as negative controls.

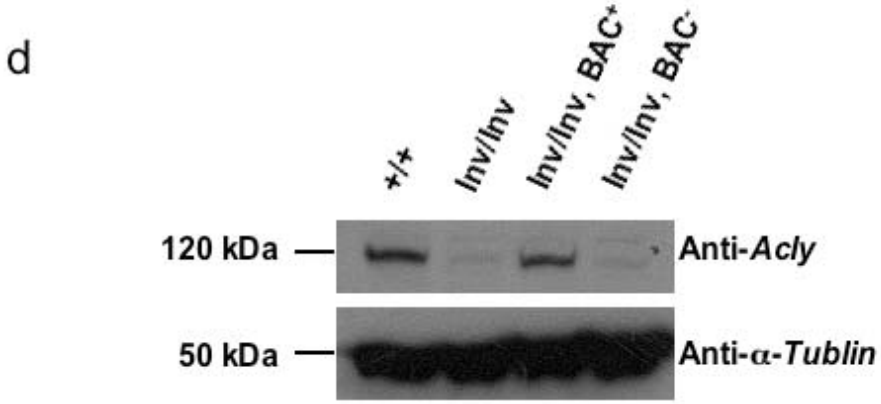
These three clones were expanded and induced to differentiate *in vitro*. After EBs were plated on the gelatinized tissue culture plates at Day 5, the EBs derived from WW113-2-8 ES cells could form cystic three-dimensional structures, while the EBs derived from WW113-2-10 and 2-11 ES cells could not (Fig. 5-13e).

a

b


Not I, PI-Sce I
Genomic Insert
SacB
Not I
pGK-Bsd-bpA
Not I
Genomic Insert
+
SacB
8.7 kb
Not I, PI-Sce I
Genomic Insert
pGK-Bsd-bpA
Not I
Not I
Genomic Insert
+
pGK-Bsd-bpA
10.0 kb


c
+/+
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
Inv/Inv, BAC-
8.6 kb
6.0 kb
KpnI


d
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
120 kDa
Anti-Acly
50 kDa
Anti-α-Tublin

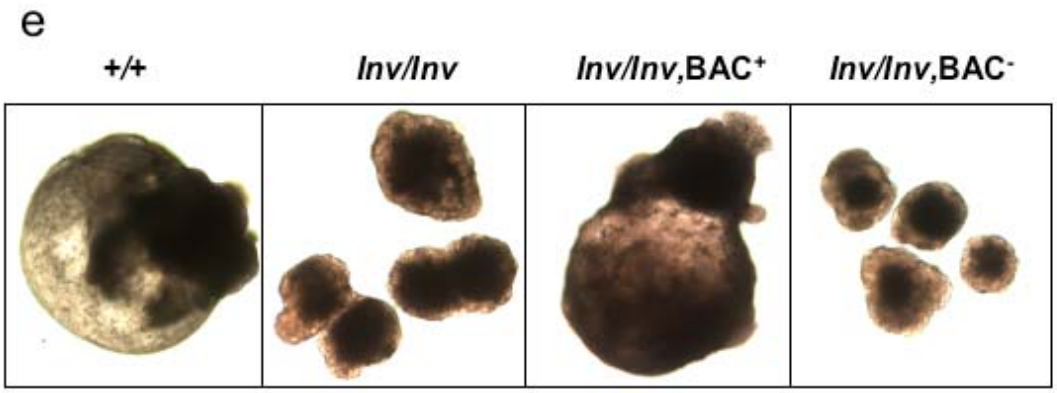

e
+/+
Inv/Inv
Inv/Inv,BAC+
Inv/Inv,BAC-

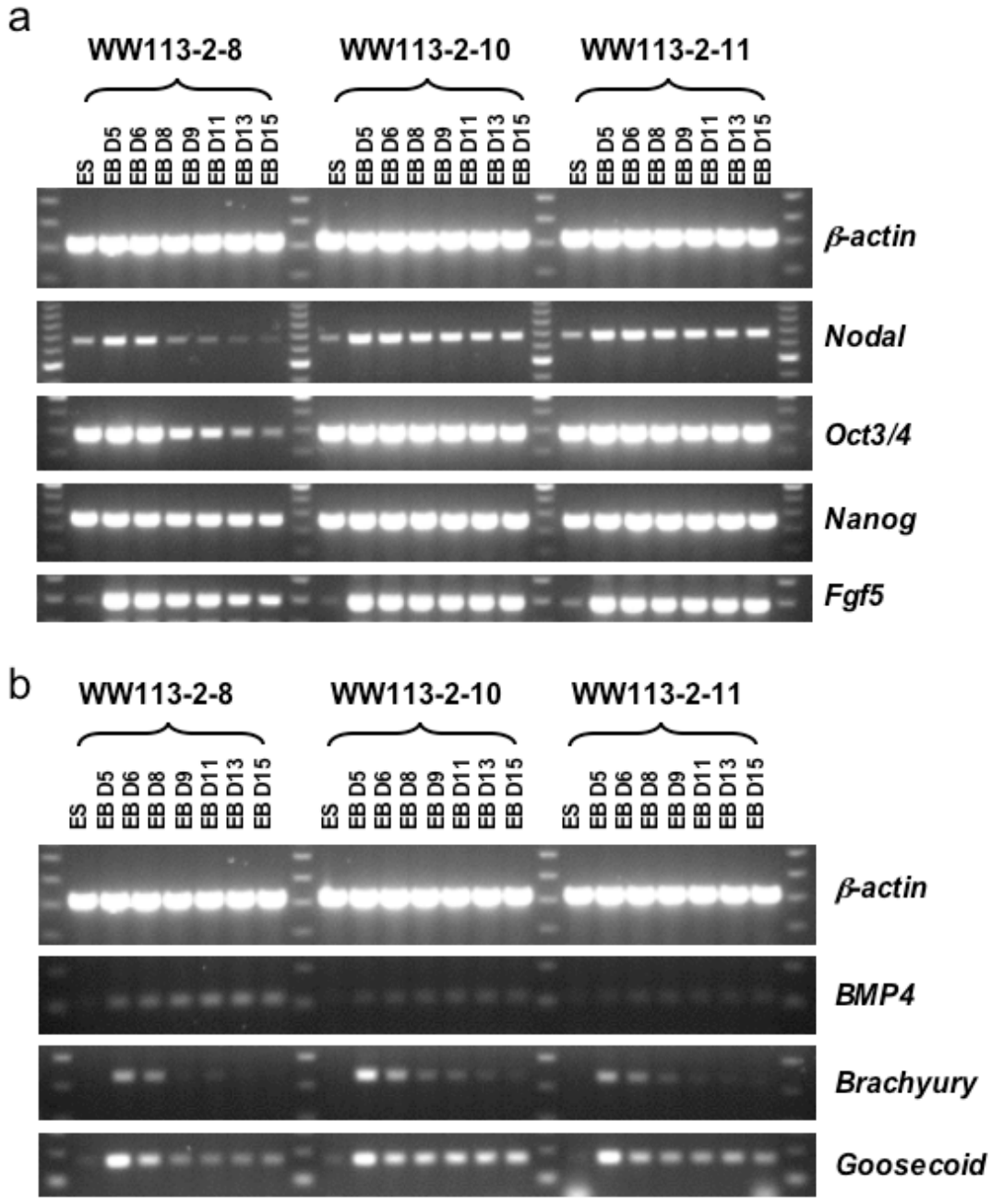
Fig. 5-13 BAC rescue. a. BAC clone BMQ-290J5. An 129 S7 BAC clone, BMQ-290J5 was picked according to the BAC end mapping information of this clone in the Ensembl database. This clone was confirmed to contain the complete *Acly* gene by PCR. BAC clone BMQ-290J5 is highlighted in blue. **b.** Insertion of a *PGK-Bsd-bpA* cassette into the BAC backbone. To facilitate the drug selection of BAC integration in ES cells, a *PGK-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination. **c.** Southern analysis of modified BAC clones. The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe. BAC DNA from the modified clones was digested with *Not*I. The *SacB* probe detected a 8.7 kb restriction restriction fragment for the unmodified BAC clone, and it detected a 10 kb fragment for the modified clones. **d.** Southern analysis of BAC inserted ES cell clones. Genomic DNA from 12 blasticidin resistant clones were digested with *Kpn*I, and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild type restriction fragment. WW113-2-8 (red) has the wild type restriction fragment and the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild type copies of Acly gene. Another two clones, WW113-2-10 and WW113-2-11 (blue) have the wild type restriction fragment, but the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single BAC clone which is likely to be incomplete. These two clones were used as control for the in vitro differentiation experiment.

When RT-PCR was performed on RNA extracted from WW113-2-8 embryoid bodies collected at different time points, the expression of *Oct3/4*, *Nodal* and *Nanog* were down-regulated rapidly as observed in WW93-A12 derived EBs. However, WW113-2-10 and 2-11 still expressed high level of these primitive ectoderm markers at late stages of their differentiation process (Fig. 5-14a).

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury* and *Goosecoid*) also became normal in the EBs derived from WW113-2-8 ES cells. The expression of *Brachyury* and *Goosecoid* was down-regulated much quicker in the WW113-2-8 derived EBs than in the WW113-2-10 or 2-11 derived EBs. The expression pattern of these markers in WW113-2-8 derived EBs was similar to that observed in the control WW93-A12 derived EBs (Fig. 5-14b).

Acly gene-specific RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation of the BAC rescue cell line, WW113-2-8. In the WW113-2-8 cell line, significantly higher *Acly* expression than the other two control cell lines was observed in undifferentiated ES cells, as well as throughout the whole differentiation process. RT-PCR using the F1/GSP4 primer pair confirmed that gene-trap transcripts were still present in the rescued cell line, WW113-2-8. So the phenotype observed in the WW103-18F11 mutant cell line was caused by the loss of normal *Acly* transcription, instead of the dominant-negative effects, as the phenotypes could be reversed by re-introducing a wild-type copy of *Acly* gene (Fig. 5-14c).

All these data suggest a direct link between the reduction in *Acly* expression and the impaired *in vitro* differentiation potential observed in WW103-18F11 derived EBs.


a
WW113-2-8
WW113-2-10
WW113-2-11
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
Nanog
Fgf5
b
WW113-2-8
WW113-2-10
WW113-2-11
β-actin
BMP4
Brachyury
Goosecoid

5.3 Discussion

In this chapter, I have described how I have used ES cell *in vitro* differentiation to screen a set of homozygous mutant ES clones. A panel of 16 markers was used to carry out the primary screen. To increase the throughput of the screen, I only took samples at three time points. If a homozygous mutant ES cell clone showed abnormal expression for one or more markers, the clone was subsequently tested in the second round screen. In the second round screen, more samples were taken at different time points, and additional markers were checked by RT-PCR to confirm the authenticity of the phenotype and also try to explain the phenotype at the molecular level by the gain or loss of specific differentiation markers.

5.3.1 Throughput of the screen

In this experiment, only a limited number of homozygous mutant ES clones were used for the *in vitro* differentiation screen. Therefore, it is possible to make a large number of EBs for each cell line and take samples at multiple time points. However, if the number of cell lines for screening increases to several hundred or several thousand, it would be necessary to make tens of thousands of plates of EBs. To make hundreds of thousands of “hanging drops”, transfer them to gelatinized tissue culture plates and change media regularly will be a labour-intensive work.

The RT-PCR method is not sensitive enough for the high-throughput analysis either. Approximately 20 μg of total RNA can be extracted from a plate of 40 EBs after Day 10. But for the early EBs (Day 5 to Day 10), sometimes two or three plates of EBs need to be combined together to get enough RNA. The cDNA synthesized from 5 μg of total RNA is only enough for about 20 RT-PCR reactions. If 10 cell lines are checked at the same time, 8 time points are taken for each cell line, and 16 markers are screened, this will require 1,280 PCR reactions. Any clones that do not show an obvious abnormality in these 16 markers will be discarded which is not a thorough analysis of the differentiation potential. Also, RT-PCR is a semi-quantitative approach to

assess gene expression, which makes it difficult to detect minor changes in expression.

An alternative approach to RT-PCR is to use cDNA and oligonucleotide microarray technology, which has been well characterized and proven to be a powerful tool for large-scale screens. This technology enables one to check the expression of all the genes in the mouse genome simultaneously. The development of array technology has made it possible to use very small amounts of starting RNA template. However, the downside of this technology is that it is still very expensive and the high cost makes it impractical to screen a lot of samples. Another problem of using microarray analysis to study ES cell *in vitro* differentiation is the complexity of the input material. It would be necessary to perform many control experiments to define the normal ranges of expression levels during differentiation, before comparisons can be made with samples from the mutant lines. Fluorescent reporters and FACS can also be used to screen the mutants in a high-throughput manner. It will be further discussed in the final chapter.

Considerable data has accumulated on the expression pattern of various markers characterizing the development of the three germ layers and other differentiated cell types during the ES cell *in vitro* differentiation process. However, this data is scattered throughout the literature and is far from being systematic or comprehensive. The results in these publications were generated by various methods, including RT-PCR, Northern, *in situ* hybridization or immunohistochemistry. Different ES cell lines (feeder-free or feeder-dependent), different differentiation protocols, and different lengths of observation periods make the data generated from these different experiments difficult to compare.

So before ES cell *in vitro* differentiation is used for a large-scale *in vitro* recessive screen, a systematic, quantitative study should be performed to determine the expression pattern of important developmental and differentiation markers in the differentiation process of the wildly used ES cell lines (AB2.2, D3, R1 and E14.1, etc.). Ideally, this data needs to be compared

to the expression pattern of these markers *in vivo* to link the *in vitro* differentiation with its *in vivo* counterpart.

5.3.2 Alternative recombination

Interestingly, two clones that have shown an abnormality during *in vitro* differentiation both contain either an extra chromosome 11 (WW103-8E6) or two partial duplication chromosome 11s (WW103-13D10). As discussed in the previous chapter, some clones can undergo a G2 *trans* recombination event and the resulting duplication chromosome can become homozygous by induced mitotic recombination. These homozygous duplication clones have as many as four copies of all the genes in the chromosomal region between the gene-trap locus and the end point targeting locus (*E_2DH*, 100.7 Mb). For WW103-13D10 (*LOC217071*, 88.7Mb), the duplication region is 12 Mb. It is reasonable to expect that such a big chromosomal rearrangement will cause an abnormality in differentiation. The WW103-8E6 clone has accumulated an extra chromosome before regional trapping. The subsequent mitotic recombination has duplicated the inversion chromosome, but a wild-type chromosome with the end point targeting cassette is still present. So the phenotypes of these clones with alternative recombination events are not related to the gene-trap loci, and are caused by the duplication of a part of or the whole chromosome.

5.3.3 *Acly* deficiency and the impaired differentiation potential

Acly is an important enzyme involved in fatty acid biosynthesis. Its product, acetyl-CoA, is the key building block for *de novo* lipogenesis (Beigneux, Kosinski et al. 2004). There are at least three principal sources of acetyl-CoA: 1) amino acid degradation produces cytosolic acetyl-CoA, 2) fatty acid oxidation produces mitochondrial acetyl-CoA, 3) Glycolysis produces pyruvate, which is converted to mitochondrial acetyl-CoA by pyruvate dehydrogenase (Garrett and Grisham 1999). The acetyl-CoA from amino acid degradation is not sufficient for fatty acid biosynthesis, and the acetyl-CoA produced by fatty acid oxidation and by pyruvate dehydrogenase can not cross the mitochondrial membrane. So cytosolic acetyl-CoA is mainly generated from citrate which is transported from the mitochondria to the

cytosol. ATP-citrate lyase converts the citrate to acetyl-CoA and oxaloacetate. Acetyl-CoA provides the substrate for cytosolic fatty acid synthesis, while the oxaloacetate is converted to malate which is transported back into the mitochondria where it can be converted back into citrate (Fig. 5-15).

5.3.3.1 *Acly* deficiency in the mouse

To investigate the phenotype of *Acly* deficiency in the mouse, an *Acly* knockout has been examined. This mouse line was generated from the Bay Genomics gene-trap resource (Stryke, Kawamoto et al. 2003). In this clone, a β-galactosidase marker is expressed from *Acly* regulatory sequences. Beigneux *et al*. (2004) have found that *Acly* is required for embryonic development, because no viable homozygous embryos were identified after 8.5 dpc. The early embryonic lethality suggested that the alternative pathways to produce acetyl-CoA in the cytosol are not sufficient to support development in the absence of *Acly* during development (Beigneux, Kosinski et al. 2004).

Northern and Western analysis of *Acly* mRNA and protein showed that in all the tissues examined (liver, heart, kidney, brain, and white adipose tissue), heterozygous mice expressed half of the normal amount of *Acly* mRNA and protein. But the heterozygous mice were healthy, fertile, and normolipidemic on both normal and high fat diets. The expression of another acetyl-CoA enzyme, Acetyl-CoA synthetase 1, was not up-regulated. Thus it seems that *Acly* is synthesized in adequate quantities and half-normal amount of the enzyme is enough for providing sufficient acetyl-CoA (Beigneux, Kosinski et al. 2004).

One interesting finding is that *Acly* is expressed at high levels in the neural tube at 8.5 dpc. The fact that *Acly* is not expressed in other foetal tissues suggests that *Acly* might not function as a house-keeping gene during development. Otherwise, widespread expression of *Acly* will be detected in all the cell lineages. Instead, it might have a tissue-specific function in embryogenesis, apart from producing Acetyl-CoA for lipogenesis.

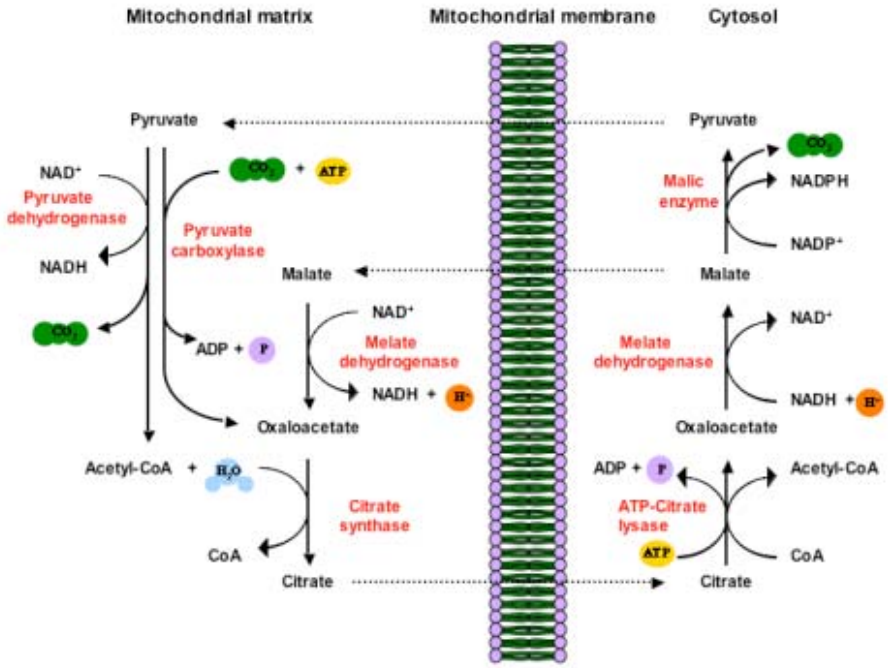


Fig. 5-15 The function of ATP-citrate lyase gene. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This is the first step for the *de novo* biosynthesis of sterol and fatty acid.

5.3.3.2 *Acly* and cell differentiation during sexual development

Acly has been shown to be involved in the sexual development of the fungus *Sordaria macrospora*. The fruiting body formation of filamentous ascomycetes involves the formation of the outer structures, as well as the development of mature ascospores within the fruiting body itself. Since this process requires the differentiation of several specialized tissues and some dramatic morphological and physiological changes, fruiting body maturation has been used as a model system to study multicellular development in eukaryotes (Nowrousian, Masloff et al. 1999).

Norwrousian *et al*. (1999) has used UV mutagenesis to screen for mutants with defects in fruiting body formation. One of the sterile mutants, per5, showed normal vegetative growth. But the fruiting body neck of the mutant strain was much shorter than that of the wild-type control. Most importantly, the fruiting bodies of the mutant strain only contain immature asci with no ascospores. DAPI staining showed that the immature asci still have eight nuclei within them, which suggests that there is no impairment in karyogamy or meiotic and postmeiotic divisions (Nowrousian, Masloff et al. 1999).

An indexed cosmid library was used to rescue the phenotype. A single complementing cosmid was isolated and sequence analysis has identified an ORF which has significant homology with higher eukaryotic *Acly*s. Analysis of the mutant *Acly* gene has identified a single nucleotide exchange (T to A), which altered a codon for aspartic acid into one for glutamic acid. The cloned mutant *Acly* gene can not rescue the phenotype of the mutant strain. Therefore, the mutation in *Acly* gene is responsible for the sterile phenotype (Nowrousian, Masloff et al. 1999).

As the mutant strain showed normal vegetative growth, it seems that sufficient acetyl-CoA is still produced for lipogenesis, either by residual *Acly* activity and/or expression of other acetyl-CoA-producing enzymes. But the attenuated *Acly* production can not satisfy the demand of acetyl-CoA during sexual development. So, the house-keeping functions of *Acly* can be circumvented to

a certain degree, but are essential under specific physiological conditions, such as sexual differentiation (Nowrousian, Masloff et al. 1999).

5.3.3.3 *Acly* as an *Brachyury* downstream notochord gene

Acly appears to be a downstream target of *Brachyury* in *Ciona intestinalis*. It is expressed specifically in the notochord in the embryogenesis process in *Ciona intestinalis*. The notochord has two major functions during chordate embryogenesis, providing inductive signals for the patterning of the neural tube and paraxial mesoderm and supporting the larval tail. The *Brachyury* gene encodes a transcription factor which contains a T DNA-binding domain. In vertebrates, *Brachyury* is first expressed in the presumptive mesoderm, and its expression is gradually restricted to the developing notochord and tailbud. *Brachyury* is believed to be one of the determinants for posterior mesoderm formation and notochord differentiation (Hotta, Takahashi et al. 2000).

By expressing the *Ciona intestinalis Brachyury* gene, *Ci-Bra*, in endoderm cells, Hotta *et al*. (1999) have isolated cDNA clones for 501 independent genes that were activated by *Ci-Bra* mis- and/or overexpression. By *in situ* hybridization, nearly 40 genes were found to be specifically or predominantly expressed in notochord, and therefore suggested to be *Brachyury*-downstream genes involved in notochord formation and function (Hotta, Takahashi et al. 1999). One of these genes, *Ci-Acl*, was found to share a high degree of homology with human ATP-citrate lyase (*ACLY*). The expression of this gene was first detected at the neural plate stage by *in situ* hybridization and its expression is restricted to notochord cells (Hotta, Takahashi et al. 2000). The fact that *Ci-Acl* only begins to express at the neural plate stage suggests that this gene might not be the immediate or direct target of *Ci-Bra*. Instead, it might be regulated by transcription factors, which in turn are regulated by *Ci-Bra* (Hotta, Takahashi et al. 2000).

It is interesting to notice that during embryogenesis of *Ciona intestinalis*, the expression of *Acly* is also highly restricted, similar to its expression pattern in murine embryogenesis. Considering the function of notochord in the

patterning of neural tube and paraxial mesoderm, it is likely that in both organisms, *Acly* plays some roles in neural tube and mesoderm differentiation.

5.3.3.4 Radicicol binds and inhibits mammalian *Acly*

Radicicol was first isolated from *Monosporium bonorden* as an antifungal antibiotic. But recently, this chemical was found to be able to reverse the transformed phenotype in *src*, *ras*, *mos*, *raf*, *fos*, and SV40-transformed cell lines. It can also cause cell cycle arrest and inhibit *in vivo* angiogenesis. So radicicol and its derivative are considered to be potential anti-cancer drugs (Ki, Ishigami et al. 2000).

To identify the *in vivo* target molecule of radicicol, Ki *et al*. (2000) used an affinity matrix to isolate radicicol-binding protein. Radicicol was biotinylated at various positions, and these variant compounds were then tested for their activity of morphological reversion of *src*-transformed phenotype. Two of the compounds, BR-1 and BR-6 were found to retain the activity. BR-6 was found to bind a 90-kDa protein, which was identified to be *Hsp90* by immunoblotting. BR-1 was shown to bind another 120-kDa protein, whose internal amino acid sequence was identical to human and rat ATP-citrate lyase. The identity of this 120-kDa protein was then confirmed by immunoblotting. Kinetic analysis showed that the activity of rat ATP-citrate lyase was inhibited by radicicol and BR-1, but not by BR-6. Radicicol was also found to be a non-competitive inhibitor of ATP-citrate lyase (Ki, Ishigami et al. 2000).

The fact that two radicicol derivatives, BR-1 and BR-6, bind two different proteins *in vivo* suggests that radicicol can bind different targets through different portions of its molecular structure. But the K_i value for ATP-citrate lyase was higher than the effective concentration of radicicol to reverse the transformed phenotype in *src*-tranformed cells, which suggests that this enzyme might not be directly involved in this process (Ki, Ishigami et al. 2000). Ki *et al*. (2000) hypothesized that BR-1 might be not stable and may be cleaved *in vivo* to generate free radicicol (Ki, Ishigami et al. 2000). But it could also be possible that radicicol is modified or cleaved *in vivo* to generate more

potent molecules to inhibit the enzyme activity of ATP-citrate lyase. Thus the phenotypes of radicicol, especially the ability to reverse the transformed phenotype in the cancer cell lines, might be partially associated with its binding and subsequent inhibition of *Acly* protein.

5.3.3.5 *Acly* is an important component of cell growth and transformation

Stable knockdown of *Acly* leads to impaired glucose-dependent lipid synthesis and also impaired *Akt*-mediated tumorigenesis (Bauer, Hatzivassiliou et al. 2005). Mammalian cells can not autonomously utilize the environmental nutrients to sustain their growth. Instead, constant extracellular signalling is needed to regulate the cellular metabolism of nutrients. However, cancer cells gain the autonomous ability to utilize nutrients by constitutively activating the normal signalling pathway without extracellular signals.

PI3K/*Akt* signalling pathway is critical for the cytokine-stimulated glucose metabolism, and its constitutive activation is commonly observed in cancer cells. In mammalian cells, glucose can either be oxidized to generate bioenergy, or be converted into other macromolecules to support biosynthesis. PI3K/*Akt* pathway can regulate the conversion of glucose to lipid and thus is essential for channeling the glucose into biosynthesis pathways. *Acly* is the main enzyme for producing cytosolic Acetyl-CoA for lipogenesis, and it is phosphorylated by *Akt in vivo* (Berwick, Hers et al. 2002). So it is possible that *Akt*-dependent cell transformation depends on *Acly* for *de novo* lipogenesis.

Bauer et al. (2005) used a shRNA construct to stably knock down the expression of *Acly* in a *Akt*-transformed cell line, FL5.12. *Akt*-expressing cells with or without *Acly* knockdown were injected into nude mice intravenously, and the mice were monitored for Akt-dependent leukemogenesis. Mice administrated *Acly* knockdown cells exhibited a significant delay or even a complete resistance to leukemogenesis.

5.3.3.6 A possible explanation of the phenotype of *Acly* deficient ES cells

Our *in vitro* differentiation results and works published before (Hotta, Takahashi et al. 1999; Nowrousian, Masloff et al. 1999; Hotta, Takahashi et al. 2000; Ki, Ishigami et al. 2000; Bauer, Hatzivassiliou et al. 2005) all suggested a pivotal function of *Acly* in cell differentiation and transformation. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This serves as the first step for the *de novo* biosynthesis of sterol and fatty acid. So the housekeeping function of the gene should be important for cell survival. But our observation and other published works (Nowrousian, Masloff et al. 1999; Beigneux, Kosinski et al. 2004) suggested that the house-keeping function of this gene can be circumvented to some degree either by the residual *Acly* activity or other alternative acetyl-CoA producing pathways.

In this study, there is no apparent difference in the growth rate, colony formation ability or ES cell/colony morphology in *Acly*-deficient ES cells compared to the wild-type control (data not shown). Microarray analysis using RNA extracted from undifferentiated WW103-18F11 and WW93-A12 ES cells showed that the expression levels of most mouse genes are similar (this work is still ongoing), which suggests that *Acly* deficiency does not cause observable phenotype in ES cells and the acetyl-CoA production in *Acly*-deficient ES cells seems to be sufficient to sustain the normal growth and division of ES cells.

However, when the *Acly*-deficient ES cells were differentiated *in vitro*, they could not form the typical three-dimensional cystic structures. In addition, the expression of some germ layer and cell type specific markers had changed. The RT-PCR results suggested that most cells in the cell aggregates were still undifferentiated ES cells. It is possible that the transition from the normal ES cell growth/division to the drastic re-programming and cell fate determination in the differentiation process demands higher than normal amounts of acetyl-CoA. A similar situation might accompany the transition from vegetative to sexual development in *S. macrospora* (Nowrousian, Masloff et al. 1999). It is

possible that this energetic demand can not be fulfilled by the alternative metabolic pathways, which might be partly due to different metabolic costs of lipogenesis.

Another possible explanation is that in order for some ES cells in an EB to be differentiated into a certain cell type, these cells must gain “competence” before the differentiation process is induced. The differentiation competence might involve as one component a threshold in acetyl-CoA concentration, which might be much higher than the level that is necessary for the ES cell growth and division.

The exact mechanism by which the acetyl-CoA production can influence the potential of ES cells to differentiate *in vitro* is unknown. Acetyl-CoA can be used to produce fatty acids, sterols and other important molecules which need the acetyl base, such as acetylcholine (Beigneux, Kosinski et al. 2004). Therefore, ATP-citrate lyase may either control the overall cytosolic acetyl-CoA concentration to indirectly regulate the pathways that need acetyl-CoA, or it could directly interact with various acetyltransferases or lipid/sterol synthetases to form an enzyme complex to provide acetyl-CoA. Nevertheless, *Acly* seems to play an important role in development and differentiation of certain cell types.

The difficulty to determine the primary locus of action of *Acly* make it hard to link this gene directly with any known genetic pathways controlling ES cell differentiation. It is not unexpected for mutants identified by such a genetic screen. However, if more homozygous ES cell mutants are generated in the future and screened using the same strategy, it will be possible to group the mutants by their apparent defects and study the relationships between the mutants with similar phenotypes. The importance of a genetic screen is that it can not only fill in the gaps in a known pathway, but also identify new pathways that are not necessarily overlapping with the known ones.

5.3.3.7 Future experiments to identify the function of *Acly* in ES cell *in vitro* differentiation

The RT-PCR results detected transcription of *Acly* downstream exons in the mutant line. Since *Acly* is a large gene (51.54 kb), it is possible that there are other alternative transcription start points. The proviral insertion and the subsequent inversion might not completely block all the *Acly* transcripts, so the mutation generated in the homozygous mutant cell line might not a null allele. To resolve this, a homozygous *Acly* gene targeted ES cell line can be constructed and these ES cells can be differentiated to confirm the function of the gene in ES cell *in vitro* differentiation.

To investigate the *in vivo* differentiation potential of the WW103-18F11 ES cell line, 1X 10^7 undifferentiated WW103-18F11 and control WW93A12 ES cells were injected subcutaneously into both flanks of 8-week old F1 hybrid mice (129 S7/SvEv$^{\text{Brd-}Hprtb\text{-}m2}$ X C57$^{\text{TyrBrdC1}}$ female). The animals were examined periodically over 4 weeks for the appearance and growth of tumours. 4 weeks after injection of ES cells, the mice were sacrificed, and the size of each tumor was measured after dissection. Tumor samples were cut into two halves, one half was fixed in 10% formalin for histopathological analysis, and the other half was dissected into several pieces (depending on its size) and snap-frozen in liquid nitrogen for subsequent RNA and DNA extraction.

For the WW93-A12 ES cell line, tumours were found at every ES cell injection site (8/8). Though the size of the tumours varied, all the tumours collected were dark red and highly vascular. When the tumours were bisected, a fluid filled central cavity was always found in the centre of the tumour. In contrast, for the WW103-18F11 ES cell line, only 3 tumours were found 4 weeks after the injection (3/8). All three tumours were very small and pale. No blood vessels were found on their surface. When the tumours were bisected, no fluid filled cavities were present.

Histopathology results of three tumours generated from WW103-18F11 cells and eight cases of WW93-A12 teratocarcinomas have confirmed that the differentiation potential of WW103-18F11 clones were greatly impaired

(pathology analysis was performed by Dr. Madhuri Warren). The WW103-18F11 tumours are circumscribed mixed ganglier/neuroepithelial tumours plus embryonal carcinoma composed predominantly of nests of mature glial cells with scarce neuroepithelial differentiation in the form of Homer-Wright rosettes. Nests of undifferentiated embryonal carcinoma (ES cells) are also seen. There was no evidence of differentiation into other germ cell lineages.

All the WW93-A12 tumour are circumscribed immature teratocarcinomas composed predominantly of immature glial tissue and tissues from all three germ layers: simple cuboidal epithelium, columnar epithelium, ciliated respiratory type epithelium, mucin secreting gastrointestinal epithelium; cartilage, osteoid, immature neuroepithelium, smooth muscle; and stratified squamous epithelium. In some samples, nests of immature embryonal carcinoma (undifferentiated ES cells) and isolated syncytiotrophoblast cells were also found.

We have also injected the BAC rescued ES cells into the F1 hybrid mice and are now waiting for the pathology results of the teratocarcinomas generated by subcutaneous injection. For the rescued cell line, tumours were found at every ES cell injection site (8/8). All of the tumours were dark red and highly vascular. Some of these tumours have a fluid filled central cavity in the centre of the tumour.

From the initial result, we can conclude that the differentiation potential of the *Acly*-deficient ES cells is also impaired *in vivo*. But complete pathology results of the tumours derived from the rescued cells are needed to confirm that the *in vivo* differentiation potential is fully recovered in these cells.

Because the phenotype of the *Acly*-deficient ES cells may depend on some mutations or silencing in a second gene, inactivation of *Acly* in an independent ES cell clone is necessary to prove that the *Acly* gene is solely responsible for the differentiation defect we observed in WW103-18F11 deficient cell line. Although BAC rescue experiment can make a causal link between the mutation in *Acly* gene and the defective phenotypes, it is

possible that other genes or transcriptional elements also play some roles in the differentiation defects. If over-expression of *Acly* cDNA can also rescue the defective phenotypes, it will effectively exclude the involvement of other genes or transcriptional elements.

5.3.4 Summary

In this chapter, I have described the strategy used to screen for an *in vitro* differentiation phenotype in homozygous mutant ES cell lines. Restricted by the detection method, I checked the expression of a limited number of markers in the differentiation process. In spite of this limitation, I successfully identified several clones with a reproducible *in vitro* phenotype. By Southern analysis and sib-selection using different drugs, I found some of these clones are the products of alternative recombination events. But two of the clones, WW103-14F11 and WW103-18F11, are products of regional trapping and subsequent inversion. Detailed expression analysis and functional studies have been carried out on WW103-18F11. The impaired *in vitro* differentiation potential observed in this clones was caused by the disruption of the ATP-Citrate lyase (*Acly*) gene. Therefore, this strategy has proved to be able to identify *in vitro* differentiation mutants and facilitate regional screens for genes involved in the early embryogenesis in the mouse genome.

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.

References

(1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science **282**(5396): 2012-8.
Abuin, A. and A. Bradley (1996). "Recycling selectable markers in mouse embryonic stem cells." Mol Cell Biol **16**(4): 1851-6.
Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-95.
Ankeny, R. A. (2001). "The natural history of Caenorhabditis elegans research." Nat Rev Genet **2**(6): 474-9.
Araki, K., M. Araki, et al. (1995). "Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase." Proc Natl Acad Sci U S A **92**(1): 160-4.
Austin, C. P., J. F. Battey, et al. (2004). "The knockout mouse project." Nat Genet **36**(9): 921-4.
Auwerx, J., P. Avner, et al. (2004). "The European dimension for the mouse genome mutagenesis program." Nat Genet **36**(9): 925-7.
Bauer, D. E., G. Hatzivassiliou, et al. (2005). "ATP citrate lyase is an important component of cell growth and transformation." Oncogene.
Beigneux, A. P., C. Kosinski, et al. (2004). "ATP-citrate lyase deficiency in the mouse." J Biol Chem **279**(10): 9557-64.
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.
Berwick, D. C., I. Hers, et al. (2002). "The identification of ATP-citrate lyase as a protein kinase B (Akt) substrate in primary adipocytes." J Biol Chem **277**(37): 33895-900.
Bessereau, J. L., A. Wright, et al. (2001). "Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line." Nature **413**(6851): 70-4.
Bonaldo, P., K. Chowdhury, et al. (1998). "Efficient gene trap screening for novel developmental genes using IRES beta geo vector and in vitro preselection." Exp Cell Res **244**(1): 125-36.
Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-6.
Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.
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.
Carthew, R. W. (2001). "Gene silencing by double-stranded RNA." Curr Opin Cell Biol **13**(2): 244-8.
Cattanach, B. M. and J. Jones (1994). "Genetic imprinting in the mouse: implications for gene regulation." J Inherit Metab Dis **17**(4): 403-20.
Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chambers, I. and A. Smith (2004). "Self-renewal of teratocarcinoma and embryonic stem cells." Oncogene **23**(43): 7150-60.
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.
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.
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.
Czyz, J., C. Wiese, et al. (2003). "Potential of embryonic and adult stem cells in vitro." Biol Chem **384**(10-11): 1391-409.
Doetschman, T. C., H. Eistetter, et al. (1985). "The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium." J Embryol Exp Morphol **87**: 27-45.
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.
Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.
Farley, F. W., P. Soriano, et al. (2000). "Widespread recombinase expression using FLPeR (flipper) mice." Genesis **28**(3-4): 106-10.
Forsburg, S. L. (2001). "The art and design of genetic screens: yeast." Nat Rev Genet **2**(9): 659-68.
Friedman, R. and A. L. Hughes (2001). "Pattern and timing of gene duplication in animal genomes." Genome Res **11**(11): 1842-7.
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.
Gajovic, S., K. Chowdhury, et al. (1998). "Genes expressed after retinoic acid-mediated differentiation of embryoid bodies are likely to be expressed during embryo development." Exp Cell Res **242**(1): 138-43.
Garrett, H. G. and C. M. Grisham (1999). "Biochemistry (Second edition)." **Chapter 25**: 802-805.
Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.
Goldstein, J. L. (2001). "Laskers for 2001: knockout mice and test-tube babies." Nat Med **7**(10): 1079-80.
Golic, K. G. (1991). "Site-specific recombination between homologous chromosomes in Drosophila." Science **252**(5008): 958-61.
Golic, K. G. and S. Lindquist (1989). "The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome." Cell **59**(3): 499-509.

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.
Greber, B., H. Lehrach, et al. (2005). "Mouse splice mutant generation from ENU-treated ES cells--a gene-driven approach." Genomics **85**(5): 557-62.
Grignani, F., T. Kinsella, et al. (1998). "High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein." Cancer Res **58**(1): 14-9.
Grimm, S. (2004). "The art and design of genetic screens: mammalian culture cells." Nat Rev Genet **5**(3): 179-89.
Gu, H., Y. R. Zou, et al. (1993). "Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting." Cell **73**(6): 1155-64.
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.
Hamilton, D. L. and K. Abremski (1984). "Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites." J Mol Biol **178**(2): 481-6.
Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.
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.
Hartwell, L. H., J. Culotti, et al. (1970). "Genetic control of the cell-division cycle in yeast. I. Detection of mutants." Proc Natl Acad Sci U S A **66**(2): 352-9.
Hotta, K., H. Takahashi, et al. (2000). "Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo." Dev Biol **224**(1): 69-80.
Hotta, K., H. Takahashi, et al. (1999). "Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the Ciona intestinalis embryo." Dev Growth Differ **41**(6): 657-64.
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.
Humpherys, D., K. Eggan, et al. (2001). "Epigenetic instability in ES cells and cloned mice." Science **293**(5527): 95-7.
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.
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.
Jansen, G., E. Hazendonk, et al. (1997). "Reverse genetics by chemical mutagenesis in Caenorhabditis elegans." Nat Genet **17**(1): 119-21.

Jorgensen, E. M. and S. E. Mango (2002). "The art and design of genetic screens: caenorhabditis elegans." Nat Rev Genet **3**(5): 356-69.
Justice, M. J. (2000). "Capitalizing on large-scale mouse mutagenesis screens." Nat Rev Genet **1**(2): 109-15.
Kamath, R. S., A. G. Fraser, et al. (2003). "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi." Nature **421**(6920): 231-7.
Kawasaki, H., K. Mizuseki, et al. (2000). "Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity." Neuron **28**(1): 31-40.
Ki, S. W., K. Ishigami, et al. (2000). "Radicicol binds and inhibits mammalian ATP citrate lyase." J Biol Chem **275**(50): 39231-6.
Kile, B. T., K. E. Hentges, et al. (2003). "Functional genetic analysis of mouse chromosome 11." Nature **425**(6953): 81-6.
Koller, B. H., P. Marrack, et al. (1990). "Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells." Science **248**(4960): 1227-30.
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.
Kumar, A. and M. Snyder (2001). "Emerging technologies in yeast genomics." Nat Rev Genet **2**(4): 302-12.
Lakso, M., B. Sauer, et al. (1992). "Targeted oncogene activation by site-specific recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(14): 6232-6.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Leahy, A., J. W. Xiong, et al. (1999). "Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation." J Exp Zool **284**(1): 67-81.
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.
Li, J., H. Shen, et al. (1999). "Leukaemia disease genes: large-scale cloning and pathway predictions." Nat Genet **23**(3): 348-53.
Lindsay, E. A., F. Vitelli, et al. (2001). "Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice." Nature **410**(6824): 97-101.
Lindsley, D. L., L. Sandler, et al. (1972). "Segmental aneuploidy and the genetic gross structure of the Drosophila genome." Genetics **71**(1): 157-84.
Liu, L. X., J. M. Spoerke, et al. (1999). "High-throughput isolation of Caenorhabditis elegans deletion mutants." Genome Res **9**(9): 859-67.
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.

Liu, P., H. Zhang, et al. (1998). "Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11." Genetics **150**(3): 1155-68.

Liu, X., H. Wu, et al. (1997). "Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission." Dev Dyn **209**(1): 85-91.

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.

Maine, E. M. (2001). "RNAi As a tool for understanding germline development in Caenorhabditis elegans: uses and cautions." Dev Biol **239**(2): 177-89.

Mann, J. R., I. Gadi, et al. (1990). "Androgenetic mouse embryonic stem cells are pluripotent and cause skeletal defects in chimeras: implications for genetic imprinting." Cell **62**(2): 251-60.

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.

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.

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.

Mitchell, K. J., K. I. Pinson, et al. (2001). "Functional analysis of secreted and transmembrane proteins critical to mouse development." Nat Genet **28**(3): 241-9.

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.

Munroe, R. J., S. L. Ackerman, et al. (2004). "Genomewide two-generation screens for recessive mutations by ES cell mutagenesis." Mamm Genome **15**(12): 960-5.

Munroe, R. J., R. A. Bergstrom, et al. (2000). "Mouse mutants from chemically mutagenized embryonic stem cells." Nat Genet **24**(3): 318-21.

Nagy, A. (2000). "Cre recombinase: the universal reagent for genome tailoring." Genesis **26**(2): 99-109.

Nagy, A., C. Moens, et al. (1998). "Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles." Curr Biol **8**(11): 661-4.

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.

Nakano, T., H. Kodama, et al. (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science **265**(5175): 1098-101.

Nichols, J., E. P. Evans, et al. (1990). "Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity." Development **110**(4): 1341-8.

Ning, Z., A. J. Cox, et al. (2001). "SSAHA: a fast search method for large DNA databases." Genome Res **11**(10): 1725-9.

Nishikawa, S. I., S. Nishikawa, et al. (1998). "Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages." Development **125**(9): 1747-57.

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.

Nowrousian, M., S. Masloff, et al. (1999). "Cell differentiation during sexual development of the fungus Sordaria macrospora requires ATP citrate lyase activity." Mol Cell Biol **19**(1): 450-60.

Nurse, P. (1975). "Genetic control of cell size at cell division in yeast." Nature **256**(5518): 547-51.

Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." Nature **287**(5785): 795-801.

Orban, P. C., D. Chui, et al. (1992). "Tissue- and site-specific DNA recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(15): 6861-5.

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.

Ramalho-Santos, M., S. Yoon, et al. (2002). ""Stemness": transcriptional profiling of embryonic and adult stem cells." Science **298**(5593): 597-600.

Ramirez-Solis, R., A. C. Davis, et al. (1993). "Gene targeting in embryonic stem cells." Methods Enzymol **225**: 855-78.

Ramirez-Solis, R., P. Liu, et al. (1995). "Chromosome engineering in mice." Nature **378**(6558): 720-4.

Rohwedel, J., K. Guan, et al. (2001). "Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects." Toxicol In Vitro **15**(6): 741-53.

Rorth, P. (1996). "A modular misexpression screen in Drosophila detecting tissue-specific phenotypes." Proc Natl Acad Sci U S A **93**(22): 12418-22.

Rorth, P., K. Szabo, et al. (1998). "Systematic gain-of-function genetics in Drosophila." Development **125**(6): 1049-57.

Rossant, J. and A. Spence (1998). "Chimeras and mosaics in mouse mutant analysis." Trends Genet **14**(9): 358-63.

Rubin, G. M. and E. B. Lewis (2000). "A brief history of Drosophila's contributions to genome research." Science **287**(5461): 2216-8.

Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proc Natl Acad Sci U S A **85**(14): 5166-70.

Schwartzberg, P. L., E. J. Robertson, et al. (1990). "Targeted gene disruption of the endogenous c-abl locus by homologous recombination with DNA encoding a selectable fusion protein." Proc Natl Acad Sci U S A **87**(8): 3210-4.

Sharov, A. A., Y. Piao, et al. (2003). "Transcriptome analysis of mouse stem cells and early embryos." PLoS Biol **1**(3): E74.
Shuman, H. A. and T. J. Silhavy (2003). "The art and design of genetic screens: Escherichia coli." Nat Rev Genet **4**(6): 419-31.
Simon, M. A. (1994). "Signal transduction during the development of the Drosophila R7 photoreceptor." Dev Biol **166**(2): 431-42.
Skarnes, W. C., J. E. Moss, et al. (1995). "Capturing genes encoding membrane and secreted proteins important for mouse development." Proc Natl Acad Sci U S A **92**(14): 6592-6.
Skarnes, W. C., H. von Melchner, et al. (2004). "A public gene trap resource for mouse functional genomics." Nat Genet **36**(6): 543-4.
Smith, A. J., M. A. De Sousa, et al. (1995). "A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination." Nat Genet **9**(4): 376-85.
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.
Stoykova, A., K. Chowdhury, et al. (1998). "Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system." Dev Dyn **212**(2): 198-213.
Stryke, D., M. Kawamoto, et al. (2003). "BayGenomics: a resource of insertional mutations in mouse embryonic stem cells." Nucleic Acids Res **31**(1): 278-81.
Su, H., A. A. Mills, et al. (2002). "A targeted X-linked CMV-Cre line." Genesis **32**(2): 187-8.
Su, H., X. Wang, et al. (2000). "Nested chromosomal deletions induced with retroviral vectors in mice." Nat Genet **24**(1): 92-5.
Sulston, J. E. (1976). "Post-embryonic development in the ventral cord of Caenorhabditis elegans." Philos Trans R Soc Lond B Biol Sci **275**(938): 287-97.
Sulston, J. E. and H. R. Horvitz (1977). "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans." Dev Biol **56**(1): 110-56.
Sulston, J. E., E. Schierenberg, et al. (1983). "The embryonic cell lineage of the nematode Caenorhabditis elegans." Dev Biol **100**(1): 64-119.
Suzuki, T., H. Shen, et al. (2002). "New genes involved in cancer identified by retroviral tagging." Nat Genet **32**(1): 166-74.
Tada, S., T. Era, et al. (2005). "Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture." Development **132**(19): 4363-74.
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.
Tessier, C. R., G. A. Doyle, et al. (2004). "Mammary tumor induction in transgenic mice expressing an RNA-binding protein." Cancer Res **64**(1): 209-14.
Theodosiou, N. A. and T. Xu (1998). "Use of FLP/FRT system to study Drosophila development." Methods **14**(4): 355-65.

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.

Tsien, J. Z., D. F. Chen, et al. (1996). "Subregion- and cell type-restricted gene knockout in mouse brain." Cell **87**(7): 1317-26.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Vittet, D., M. H. Prandini, et al. (1996). "Embryonic stem cells differentiate in vitro to endothelial cells through successive maturation steps." Blood **88**(9): 3424-31.

Vivian, J. L., Y. Chen, et al. (2002). "An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells." Proc Natl Acad Sci U S A **99**(24): 15542-7.

von Melchner, H. and H. E. Ruley (1989). "Identification of cellular promoters by using a retrovirus promoter trap." J Virol **63**(8): 3227-33.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Wiles, M. V. and G. Keller (1991). "Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture." Development **111**(2): 259-67.

Wobus, A. M. (2001). "Potential of embryonic stem cells." Mol Aspects Med **22**(3): 149-64.

Wobus, A. M., K. Guan, et al. (2002). "Embryonic stem cells as a model to study cardiac, skeletal muscle, and vascular smooth muscle cell differentiation." Methods Mol Biol **185**: 127-56.

Wobus, A. M., G. Wallukat, et al. (1991). "Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers." Differentiation **48**(3): 173-82.

Wood, V., R. Gwilliam, et al. (2002). "The genome sequence of Schizosaccharomyces pombe." Nature **415**(6874): 871-80.

Xu, T. and S. D. Harrison (1994). "Mosaic analysis using FLP recombinase." Methods Cell Biol **44**: 655-81.

Xu, T. and G. M. Rubin (1993). "Analysis of genetic mosaics in developing and adult Drosophila tissues." Development **117**(4): 1223-37.

Xu, T., W. Wang, et al. (1995). "Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase." Development **121**(4): 1053-63.

Yamashita, J., H. Itoh, et al. (2000). "Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors." Nature **408**(6808): 92-6.

Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nat Biotechnol **21**(2): 183-6.

Yu, Y. and A. Bradley (2001). "Engineering chromosomal rearrangements in mice." Nat Rev Genet **2**(10): 780-90.

5 Genetic screen on homozygous gene traps

5.1 Introduction

During *in vitro* differentiation, ES cells can form cystic embryo-like aggregates, embryoid bodies (EB), that contain cells of endodermal, ectodermal and mesodermal lineages, which can further differentiate into more specialized cell types. The morphological changes of embryoid bodies are accompanied, at the molecular level, by the changes in the expression of a set of lineage-specific and tissue-specific markers. By comparing the dynamic changes in the expression of these markers *in vivo* and *in vitro*, different stages of EB differentiation *in vitro* can be linked to different stages of embryogenesis *in vivo* (Leahy, Xiong et al. 1999). These properties allow us to use ES cell *in vitro* differentiation as an *in vitro* model to study early embryogenesis and this facilitates genetic approaches.

5.1.1 *In vitro* differentiation protocols

There are three main protocols for ES cell *in vitro* differentiation: the hanging drop method (Wobus, Wallukat et al. 1991); the mass culture method (Doetschman, Eistetter et al. 1985); and the methylcellulose method (Wiles and Keller 1991). All three of these have been widely used for making embryoid bodies (EB) for different purposes.

The advantage of the hanging drop method is that the starting number of ES cells in an embryoid body is defined, so the size and the differentiation pattern of the EBs generated by this method is more consistent than with the other two methods. This characteristic is particularly important for developmental studies, which require the comparisons between EBs under different culture conditions and/or with different mutations (Wobus, Guan et al. 2002). However, this method is also more complicated than the other two methods.

On the other hand, the mass culture method is useful for differentiating a large number of ES cells. By plating undifferentiated ES cells onto bacteriological Petri dishes, the cells automatically form cell aggregates, and the aggregates can differentiate into a variety of different cell types. However, the size and the differentiation pattern can vary significantly between plates or between

experiments, even when the same ES cell line is used. The methylcellulose method is used specifically for the differentiation of haematopoietic lineages, and is not suitable for other purposes.

In this project, it was necessary to compare the *in vitro* differentiation potential of a number of homozygous mutant ES cell clones. Therefore the hanging drop method was the most appropriate *in vitro* differentiation protocol to use.

5.1.2 Parameters influencing *in vitro* differentiation of ES cells

The developmental potency of ES cells in culture is dependent on a number of intrinsic and extrinsic parameters. These include the number of ES cells used to make the EBs; the composition of the differentiation medium; cellular growth factors and differentiation inducers added to the culture medium; the ES cell lines; as well as the genetic changes in the ES cell genome.

Compared to *in vivo* differentiation in the mouse, the parameters for *in vitro* differentiation are more controllable. Whichever differentiation protocol is chosen, extrinsic parameters can be effectively controlled by using defined medium and culture conditions. Variations caused by intrinsic parameters can be eliminated by choosing an appropriate control ES cell line. Thus loss-of-function or gain-of-function studies using *in vitro* differentiation can be an ideal alternative to study the phenotypes of mutations on embryogenesis and early development (Wobus, Guan et al. 2002).

5.1.3 Recessive genetic screens using ES cell *in vitro* differentiation

Genetic analysis of recessive mutations in ES cells is informative on possible functions *in vivo*, especially for mutations that result in embryonic lethality. A recessive genetic screen using ES cell *in vitro* differentiation can be used to identify important genes in the differentiation process.

The bottleneck of recessive genetic screens in ES cells is the difficulty of obtaining enough homozygous mutant ES cells. If a genetic screen is performed to identify genes involved in ES cell *in vitro* differentiation, pure homozygous mutant ES cell clones need to be differentiated individually to

check for their differentiation potential. Existing methods to generate homozygous mutations in ES cells are not ideal for this purpose. In the previous chapters, I have demonstrated that a strategy combining regional trapping and inducible mitotic recombination can be used to generate homozygous mutations in a genomic region of interest. By Splinkerette PCR and 5' RACE, proviral/host flanking genomic sequences and/or cDNA sequence were isolated to identify the proviral insertion sites and inversion breakpoints of these mutant clones.

A total of 30 different gene-trap loci on chromosome 11 that are homozygously mutated were isolated. These homozygous gene-trap clones can be used to perform a small-scale genetic screen to identify the mutations that will disrupt the normal ES *in vitro* differentiation process. Each gene-trap clone has been differentiated individually, and a set of important lineage-specific and tissue-specific markers have been checked to determine the differentiation potential of each of the homozygous mutant ES cell lines. Mutant cell lines that show an abnormal differentiation pattern have been confirmed using independent methods.

5.2 Results

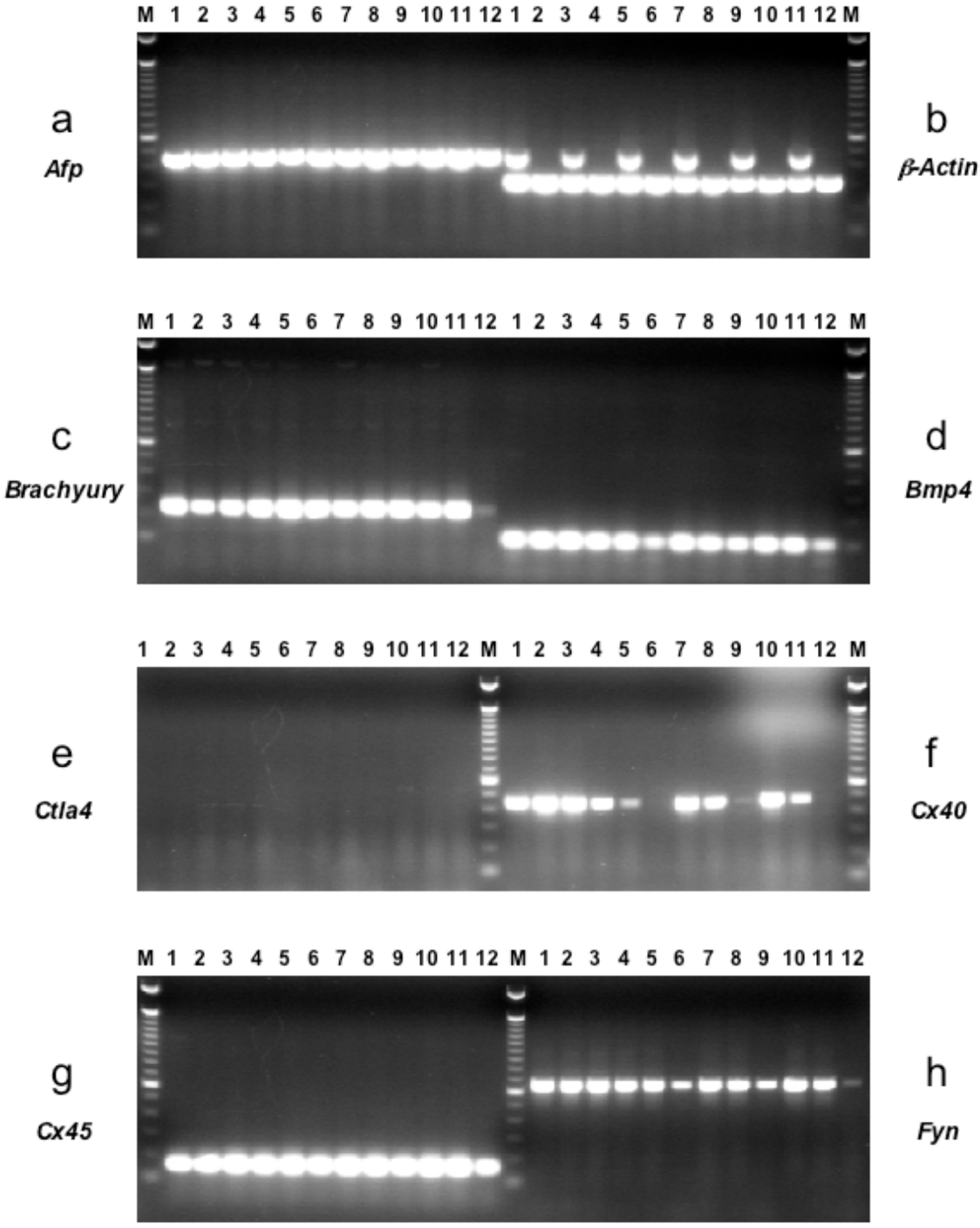
5.2.1 Primary screen

For each of the 33 mapped gene-trap loci, at least one subclone was chosen for the primary *in vitro* differentiation screen. Embryoid bodies were made and cultured as described before (Wobus, Guan et al. 2002). In brief, undifferentiated ES cells were maintained on feeder layers until they were used for *in vitro* differentiation. To setup the assay, ES cells were trypsinized and diluted to a final concentration of approximately 600 cells in 20 μl Differentiation Medium (see material and methods). 20 μl drops of the ES cell suspension were laid onto the bottom of 100-mm bacteriological Petri dishes. The Petri dishes were inverted and the ES cell aggregates were cultured in the resulting hanging drops for two days. After this, the Petri dishes were turned the right way up and Differentiation Medium was added into each dish to rinse the aggregates. The aggregates were cultured in suspension for

another three days. The sample was harvested at Day 5. At the same time, the EBs on the remaining dishes were plated out onto gelatinized 90-mm tissue culture plates. The plated EBs were subsequently cultured in Differentiation Medium supplemented with 10^{-8} M retinoic acid (RA) and the medium was changed every two days. Subsequent samples were taken at Day 8 and Day 11.

When all the samples were taken, RNA was extracted from each sample and quantified. 5 μg total RNA was used for first strand cDNA synthesis. The resulting cDNA was used as a template for RT-PCR. In the primary of screen, 16 pairs of primers were used (*Afp*, *β-Actin*, *Brachyury*, *Bmp4*, *Ctla4*, *Cx40*, *Cx45*, *Fyn*, *Gata4*, *Goosecoid*, *Hnf4*, *Nodal*, *Oct3/4*, *Pecam*, *Tie2* and *vHNF1*). All the homozygous mutant cell lines that showed abnormal expression (significant up-regulation or down-regulation compared to the WW93A12 control line) for one or more markers in the primary screen were selected for the second round screen (Fig. 5-1 and Fig. 5-2).

All the mitotic recombination clones (WW103) used in the screen carry two homologs of chromosome 11 from the same parent, either bi-paternal or bi-maternal. It is possible that because of the imprinting, the *in vitro* differentiation pattern of ES cells carrying bi-paternal or bi-maternal homologs of chromosome 11 will be different from that of the wild-type ES cells that have one paternal and one maternal homologs of chromosome 11. Also, the *in vitro* differentiation potential of ES cells homozygous for the targeted E_2DH allele has not been assessed. So an ideal control cell line for this experiment will carry two homologs of chromosome 11 from the same parent as the WW103 clones, and this control cell line is also homozygous for the targeted E_2DH allele.


M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 1 2 3 4 5 6 7 8 9 10 11 12 M
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn

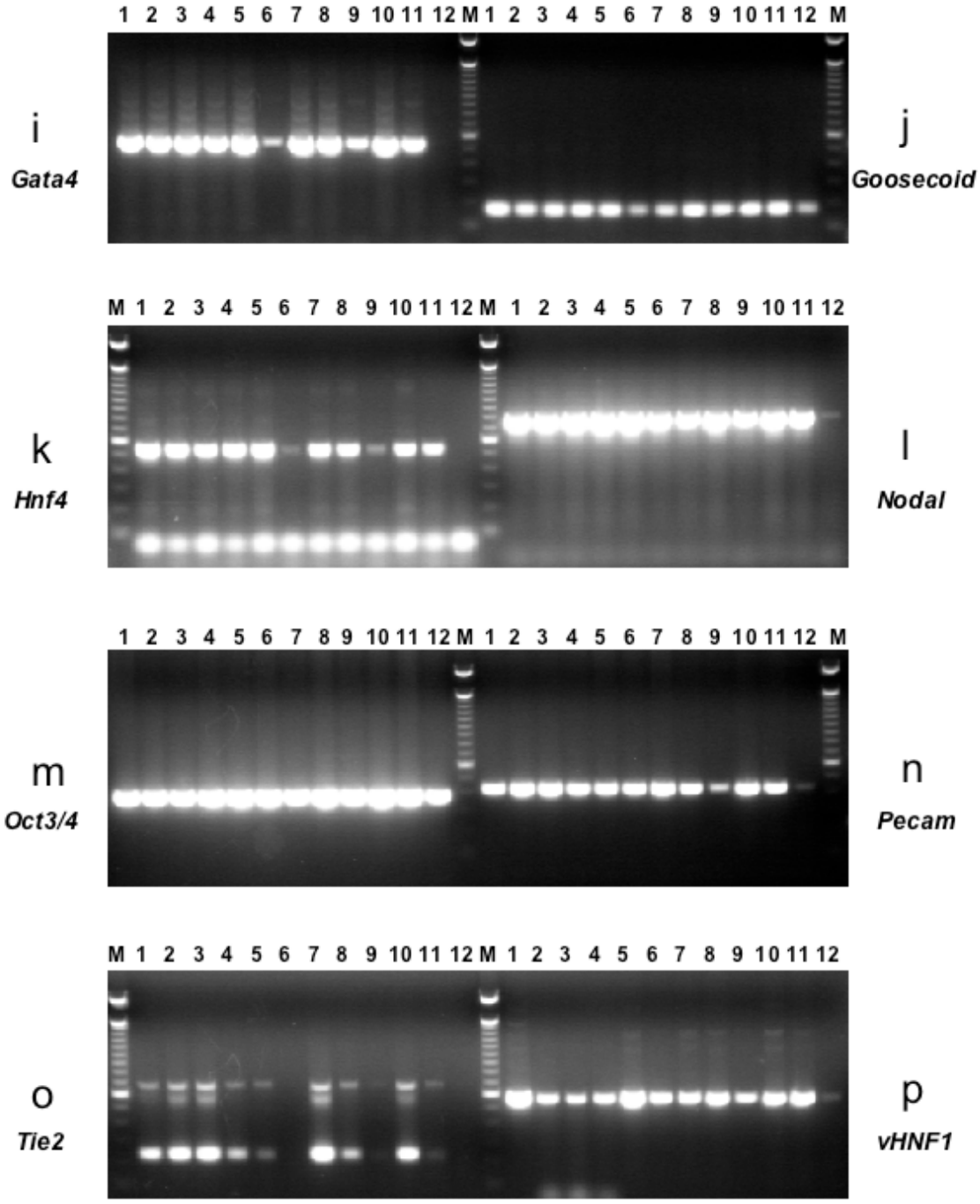
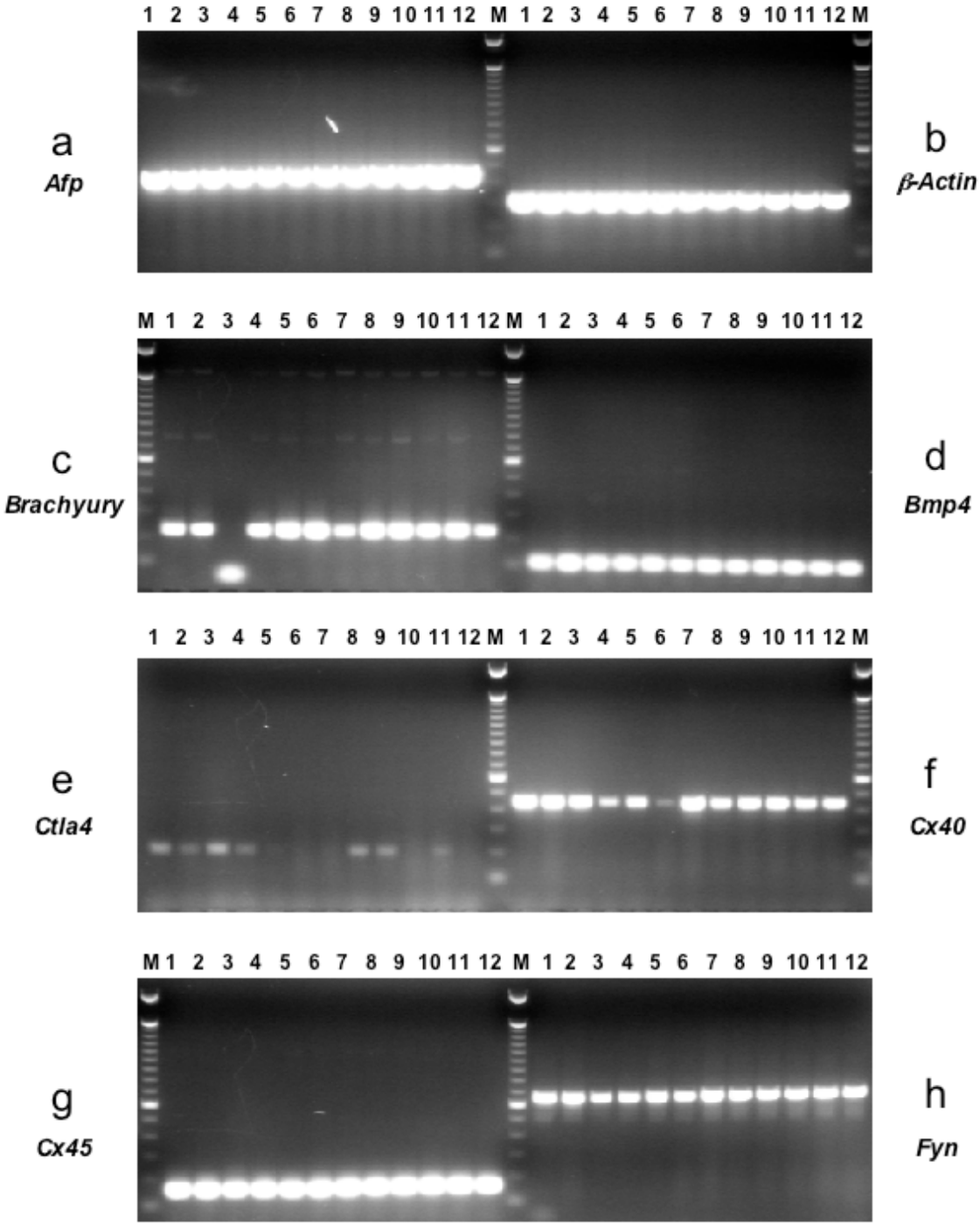
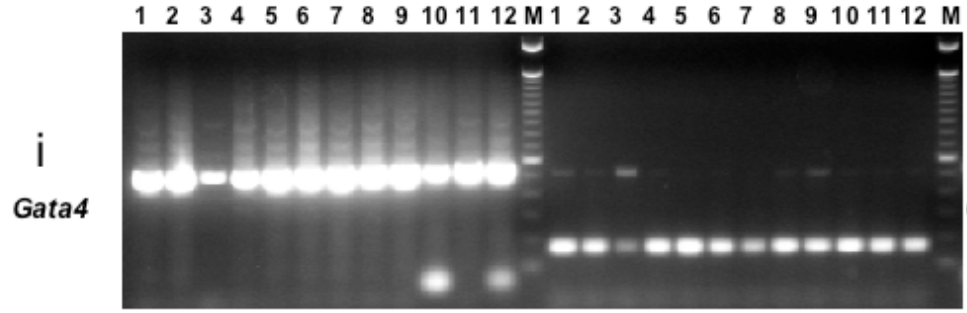

1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
i
Gata4
j
Goosecoid
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
k
Hnf4
l
Nodal
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
m
Oct3/4
n
Pecam
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
o
Tie2
p
vHNF1

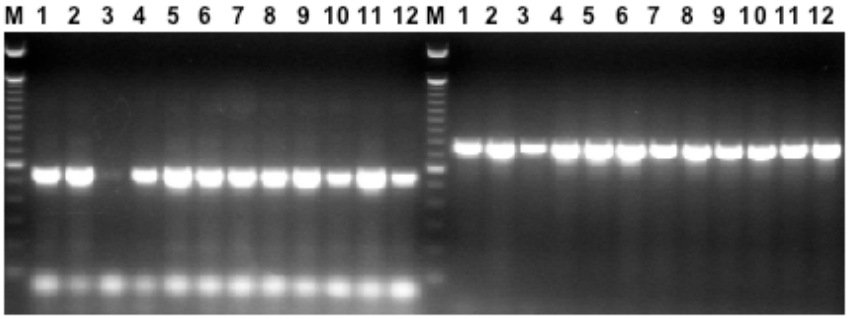
Fig. 5-1 RT-PCR results of Day 5 embryoid bodies. a. *Afp*. **b.** *β-Actin*. Note that line 1, 3, 5, 7, 9 and 11 were mistakenly mixed with PCR products from **a**. *β-Actin* was used as a loading control. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)


1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
a
Afp
b
β-Actin
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
c
Brachyury
d
Bmp4
1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12 M
e
Ctla4
f
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 M 1 2 3 4 5 6 7 8 9 10 11 12
g
Cx45
h
Fyn



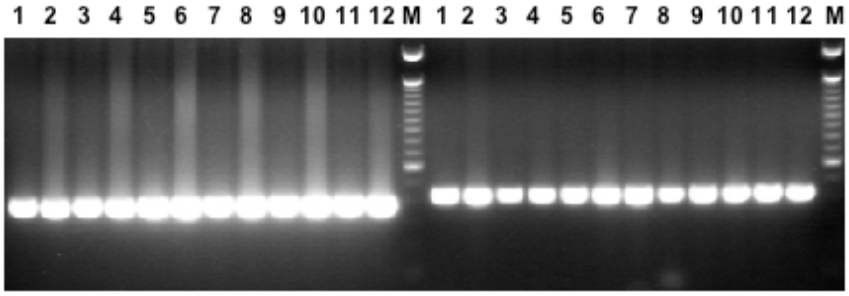
i
Gata4

j
Goosecoid



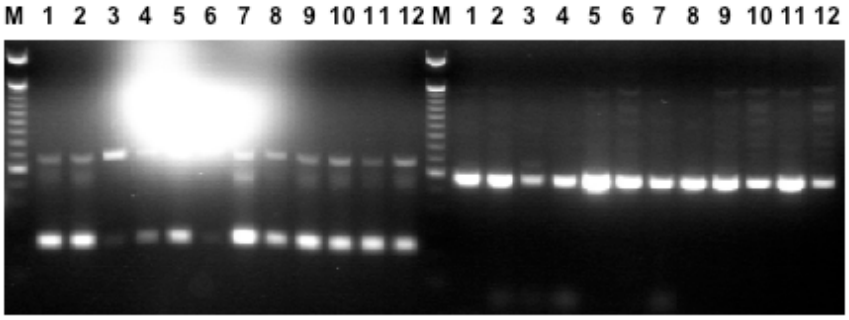
k
Hnf4

l
Nodal



m
Oct3/4

n
Pecam



o
Tie2

p
vHNF1

Fig. 5-2 RT-PCR result of Day 8 embryoid bodies. a. *Afp*. **b.** *β-Actin*. **c.** *Brachyury*. **d.** *Bmp4*. **e.** *Ctla4*. **f.** *Cx40*. **g.** *Cx45*. **h.** *Fyn*. **i.** *Gata4*. **j.** *Goosecoid*. **k.** *Hnf4*. **l.** *Nodal*. **m.** *Oct3/4*. **n.** *Pecam*. **o.** *Tie2*. **p.** *vHNF1*. Lane 1, WW103-16B3; 2, WW103-16D2; 3, WW103-17F2; 4, WW103-17G6; 5, WW103-18E8; 6, WW103-18F11; 7, WW103-18G10; 8, WW103-19A1; 9, WW103-19A2; 10, WW103-19A6; 11, WW103-19A8; 12, WW103-19D3; M, 100 bp ladder (Invitrogen)

To generate this control ES cell line, a Cre expression plasmid was electroporated into the WW69-D6 cell line and mitotic recombination clones were selected in M15 supplemented with HAT. The clones with the desired phenotype were identified both by sib-selection and by Southern analysis using an *E_2DH* 3' probe. The G2-X recombinants are resistant to HAT and blasticidin, but sensitive to G418 and puromycin. Southern analysis of *Nde*I digested genomic DNA will generate a 9.6 kb targeted fragment instead of the 13.1 kb wild-type fragment. One clone with the desired genotype, WW93-A12 and its subclones were used as controls in the ES cell *in vitro* differentiation screen.

5.2.2 Secondary screen

Mutant cell lines that showed an abnormal expression pattern for the markers checked in the primary screen were subcloned and single colonies were picked to avoid cross-contamination by ES cells that did not have the correct genotype. The control cell line, WW93-A12 was also subcloned. Southern analysis was performed on all the subclones to confirm their identities (Fig. 5-3).

The *in vitro* differentiation protocol for the second round screen is essentially the same as that of the first round. But more time points were taken and more molecular markers were checked using RT-PCR. The clones that still showed abnormal expression for the markers checked were characterized individually.

5.2.3 WW103-8E6 (*Pecam*)

One of the mitotic recombination clones, WW103-8E6, have overtly impaired *in vitro* differentiation potential. When the EBs were plated onto the gelatinized tissue culture plates, the EBs could not form cystic three-dimensional structures. When RT-PCR was performed using a series of molecular markers, the expression of some markers in the day 8 EBs was significantly down-regulated compared to the wild-type control, WW93-A12 (Fig. 5-4).

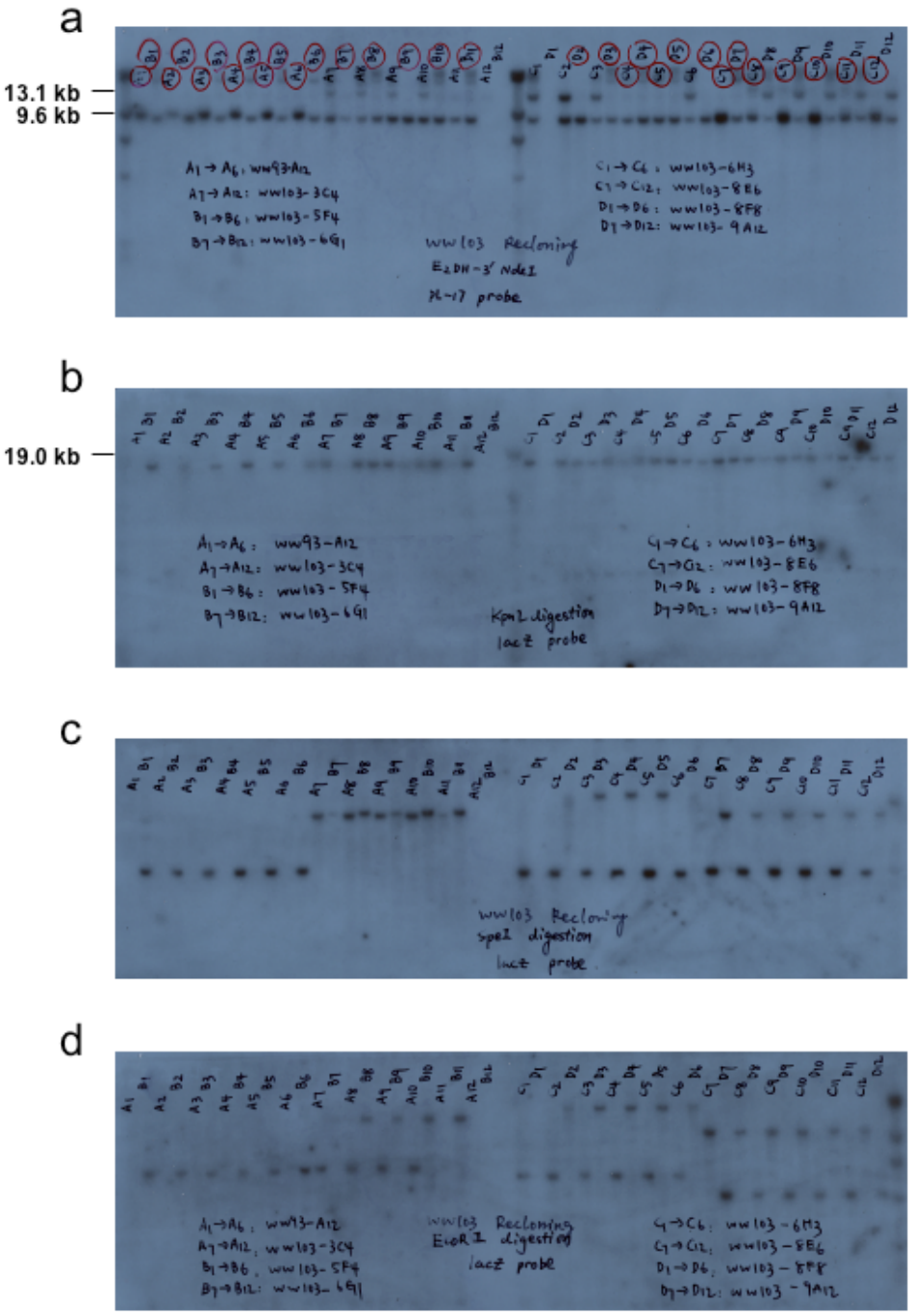
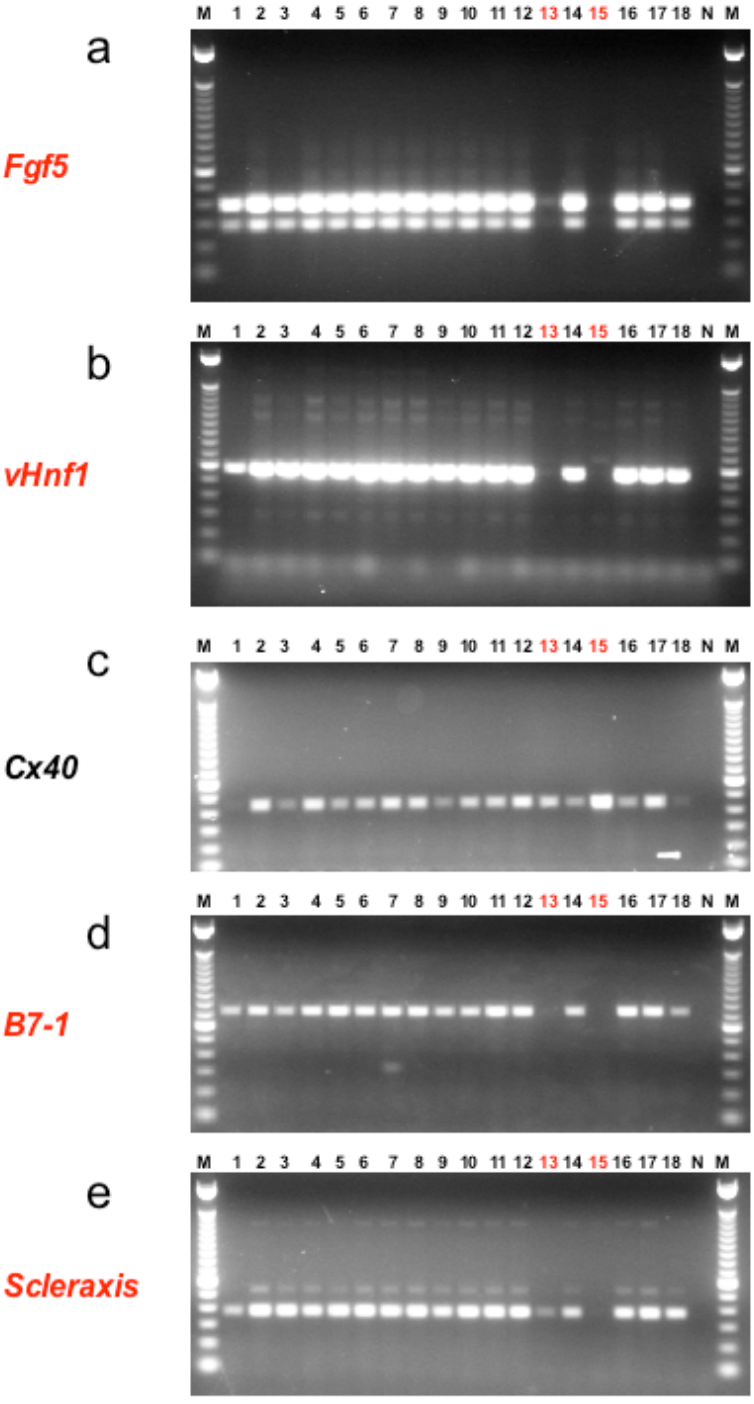
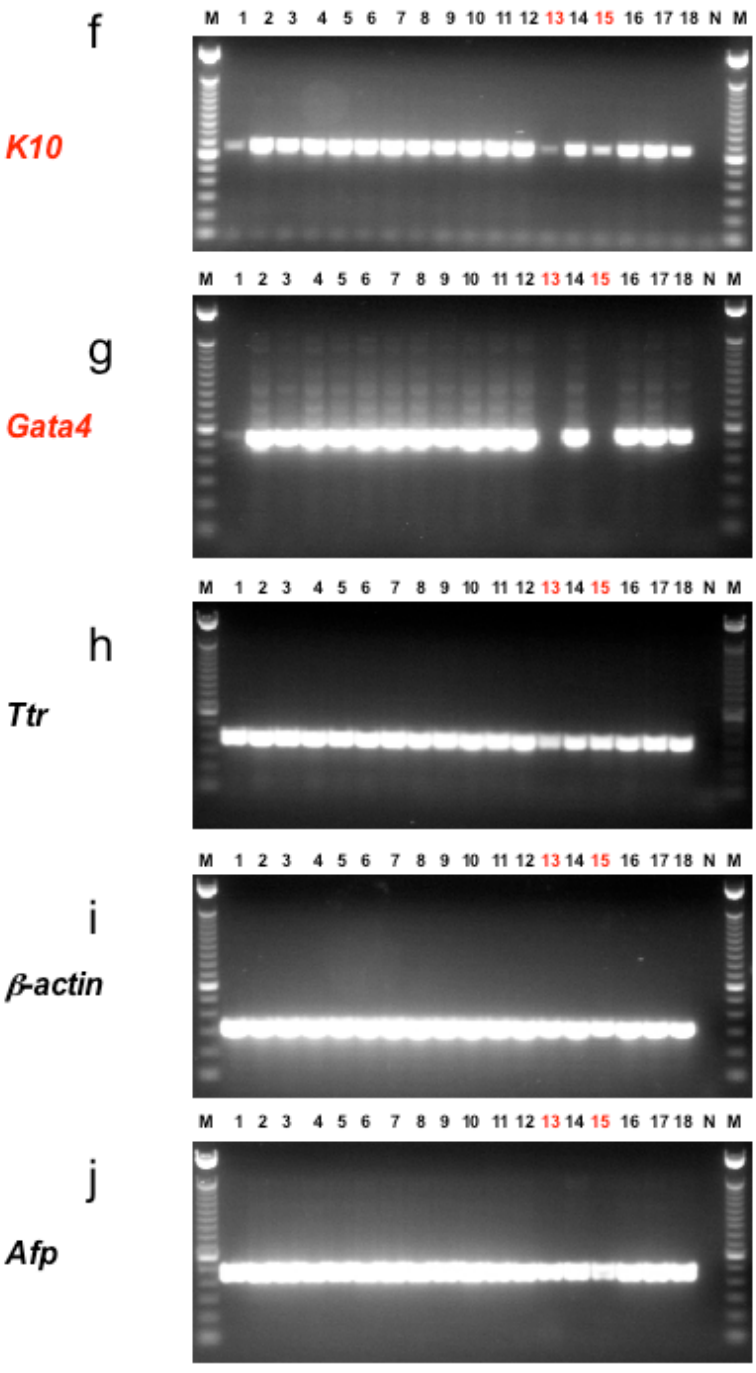

a
13.1 kb
9.6 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
E2DH-3' NdeI
PL-17 probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
b
19.0 kb
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
KpnI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12
c
ww103 Recloning
SpeI digestion
lacZ probe
d
A1→A6: ww93-A12
A7→A12: ww103-3C4
B1→B6: ww103-5F4
B7→B12: ww103-6G1
ww103 Recloning
EcoRI digestion
lacZ probe
C1→C6: ww103-6H3
C7→C12: ww103-8E6
D1→D6: ww103-8F8
D7→D12: ww103-9A12

Fig. 5-3 Southern analysis of subclones of homozygous mutant ES cell lines. a. Genomic DNA from the subclones was cut with *Nde*I and hybridized to an E_2DH 3' probe. The targeted restriction fragment is 9.6 kb and the wild type fragment is 13.1 kb. Note that some subclones of WW103-6H3 (c1, c2, c3 and c6) are heterozygous. **b.** Genomic DNA from the subclones was cut with *Kpn*I and hybridized to a *lacZ* probe. A 19.0 kb inversion restriction fragment was detected for all the subclones except those of the control cell line, WW93-A12. **c.** Genomic DNA from the subclones was cut with *Spe*I and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12. **d.** Genomic DNA from the subclones was cut with *Eco*RI and hybridized to a *lacZ* probe. Proviral/host junction fragments of different sizes were detected for all the subclones except those of the control cell line, WW93-A12.

The *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product from this clone mapped the proviral insertion site to the first intron of *Pecam* (Platelet endothelial cell Adhesion Molecule Precursor, CD31) (Fig. 5-5a). The 5' RACE product matched an alternative spliced exon (Exon 1b) (Fig. 5-5b). In the Ensembl browser, there are at least three different spliced forms at the 5' end of this gene. *Pecam* transcripts can start from Exon 1a, Exon 1b or a site just 5' to Exon2 (Fig. 5-5c). The open reading frame (ORF) of *PECAM* starts from Exon 2. So the breakpoint in intron 1 created by the inversion would disrupt the transcripts starting from Exon1a and 1b, but it may not affect the transcripts starting from Exon 2. RT-PCR primers were used to determine the expression of different alternative spliced forms of *Pecam* in undifferentiated WW93-A12 and WW103-8E6 ES cells. This analysis revealed that none of the transcripts in undifferentiated ES cells started from Exon 1a (data not shown). In undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. However, in undifferentiated WW103-8E6 ES cells, *Pecam* transcripts starting from Exon 2 and Exon 1b were both detected, implying that the inversion did not completely block the transcription across the breakpoint (Fig. 5-5d).

The *in vitro* differentiation of another *Pecam* gene-trap clone, WW103-4A6, showed that the differentiation of this clone was not impaired by the proviral insertion and the breakage caused by inversion. The *Sau*3A1 Splinkerette PCR product fro this clone has mapped the proviral insertion site to the third intron of *Pecam* (Fig. 5-6a and b). RT-PCR analysis of WW103-4A6 during the process of differentiation showed the expression of all the molecular markers during differentiation which was the same as the control cell line, WW93-A12 (data not shown). RT-PCR using a pair of primers specifically designed to amplify Exons 6, 7 and 8 of *Pecam* showed that the Pecam expression was completely blocked in WW103-4A6. On the other hand, WW103-8E6 and WW103-8G9, subclones in the same group as 8E6, showed normal *Pecam* expression (Fig. 5-6c).


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
a
Fgf5
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
b
vHnf1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
c
Cx40
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
d
B7-1
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
e
Scleraxis


f
K10
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
g
Gata4
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
h
Ttr
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
i
β-actin
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M
j
Afp
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 N M

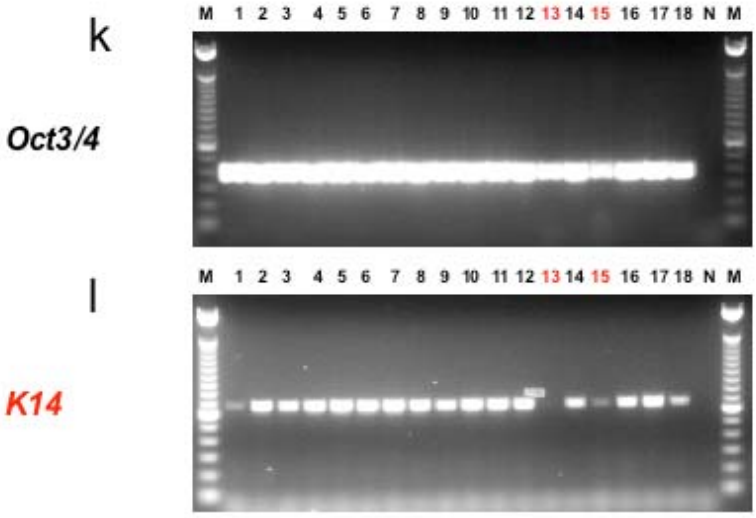
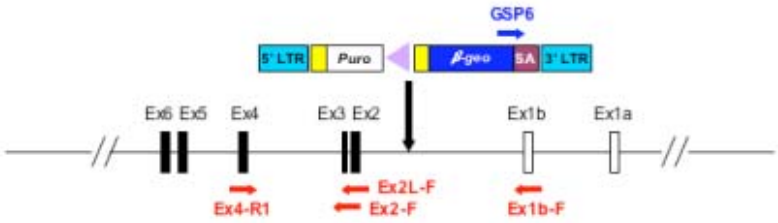


Fig. 5-4 RT-PCR results of Day 8 embryoid bodies. a. *Fgf5*. **b.** *vHnf1*. **c.** *Cx40*. **d.** *B7-1*. **e.** *Scleraxis*. **f.** *K10*. **g.** *Gata4*. **h.** *Ttr*. **i.** *β-actin*. **j.** *Afp*. **k.** *Oct3/4*. **l.** *K14*. Note that lane 13 is WW103-8E6 (marked in red). Lane 1 is the AB2.2 wild type ES cell line. Lane 2 is the WW93-A12 control line, all the other lanes are irrelevant mutant cell lines. At day 8, EBs derived from WW103-8E6 didn't express *Fgf5, vHnf, B7-1, Gata4 and K14*. The expression of *Scleraxis* and *K10* was also down-regulated compared to WW93-A12 control. Markers that are down-regulated are marked in red. Interestingly, the expression of various markers differs between EBs from AB2.2 (lane 1) and those from WW93A12 (lane 2).The expression pattern of the mutant cell lines is more similar to WW93-A12. In the whole screen, WW93A12 and its subclones were used as control.

a

b

c


GSP6
5' LTR
Puro
β-geo
SA
3' LTR
Ex6 Ex5
Ex4
Ex3 Ex2
Ex1b
Ex1a
Ex4-R1
Ex2L-F
Ex2-F
Ex1b-F

d

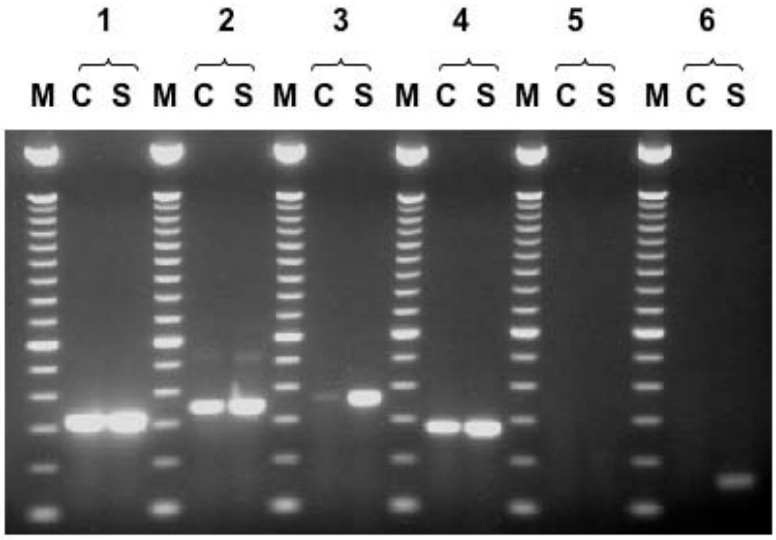
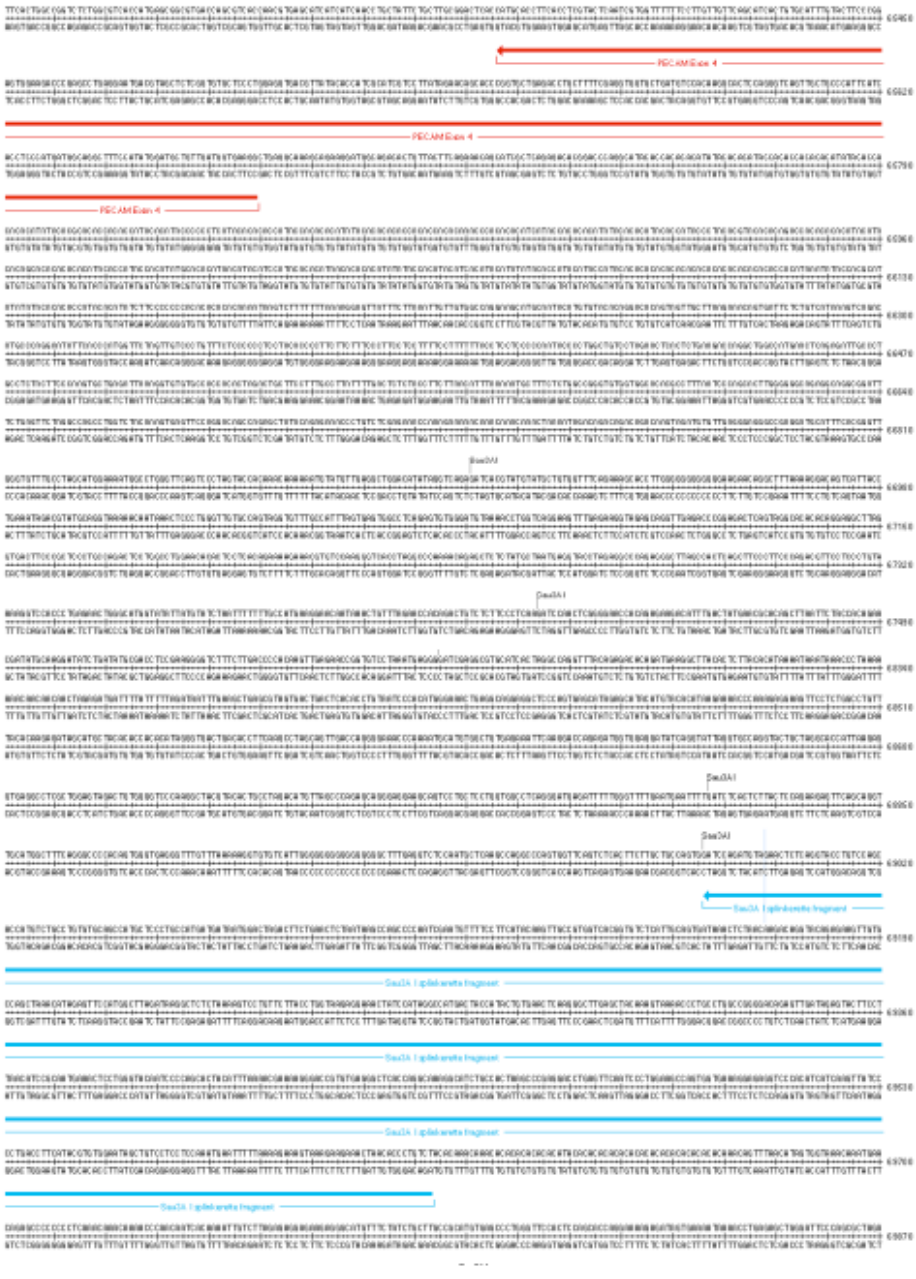

1
2
3
4
5
6
M C S M C S M C S M C S M C S M C S

Fig. 5-5 WW103-8E6. a. Splinkerette PCR result. *Spe*I/*Xba*I/*Nhe*I Splinkerette PCR product mapped the proviral insertion site to the first intron of the *Pecam* gene. The second and the third exons of *Pecam a*re marked in red, and the splinkerette PCR fragment is marked in blue. **b.** 5'RACE result. The 5' RACE product was mapped to an alternative exon of *Pecam*.The splice acceptor of the SA-βgeo cassette is marked in black, the *Pecam* alternative exon 1b is marked in red, the polyC tail is marked in green. **c.** Schematic illustration of the proviral insertion in WW103-8E6. *Pecam* gene specific primers were designed to decide the structure of the clone. PECAM-Ex1b-F, PECAM-Ex2L-F, PECAM-Ex2-F and PECAM-Ex4-R1 primers are shown as red arrows, SA-βgeo cassette specific primer GSP6 is shown as a blue arrow, the coding exons are marked in black. **d.** RT-PCR result. 1: positive control, *β-actin* RT primers; 2: PECAM-Ex1b-F/PECAM-Ex4-R1; 3: PECAM-Ex2L-F/PECAM-Ex4-R1; 4: PECAM-Ex2-F/PECAM-Ex4-R1; 5: negative control, no primers were added. 6: PECAM-Ex1b-F/GSP6. M, 100 bp ladder (Invitrogen); C: RNA extracted from undifferentiated WW93-A12 ES cells; S: RNA extracted from undifferentiated WW103-8E6 ES cells. Note that in undifferentiated WW93-A12 ES cells, most *Pecam* transcripts start from Exon 1b. In undifferentiated WW103-8E6 ES cells, a large portion of Pecam transcripts start from Exon 2. But significant amount of transcripts started from Exon 1b were still detected.

a


PECAM Exon 4
PECAM Exon 4
PECAM Exon 4
Sau3AI
Sau3AI
Sau3AI
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment
Sau3A I gDNA enrich fragment

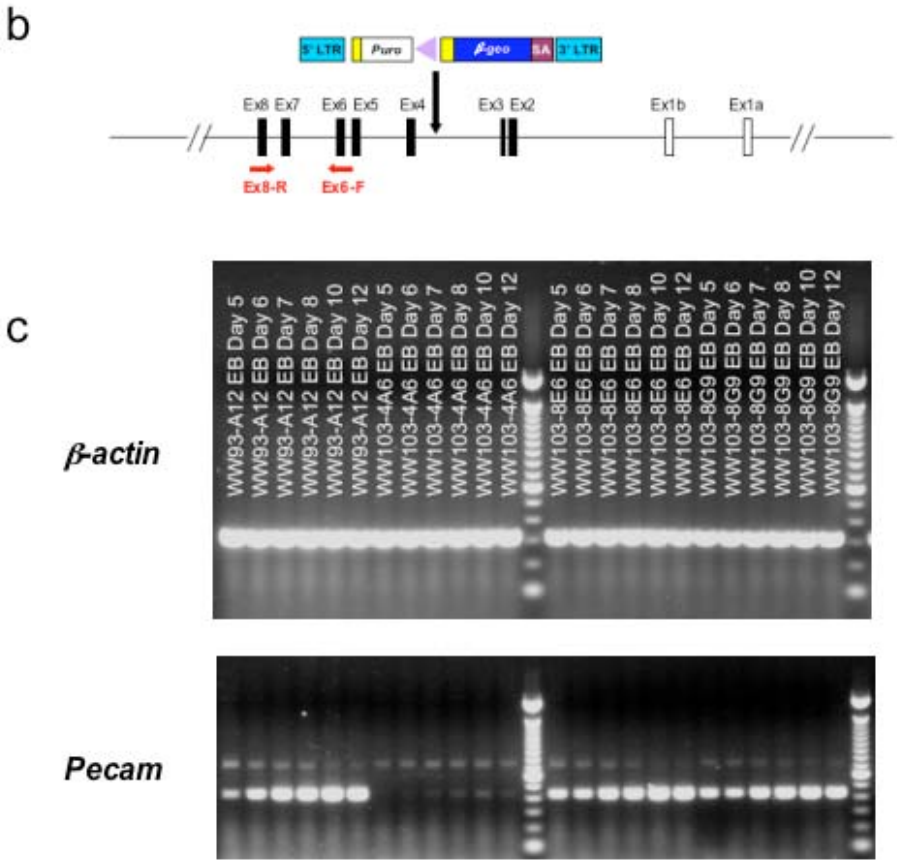


Fig. 5-6 WW103-4A6. a. Splinkerette PCR result. *Sau*3AI Splinkerette PCR product mapped the proviral insertion site to the third intron of the *Pecam* gene. The forth exon of *Pecam* is marked in red, and the splinkerette PCR fragment is marked in blue. **b.** Schematic illustration of the proviral insertion in WW103-4A6. *Pecam* gene specific primers were designed to confirm the structure of the allele. PECAM-Ex6-F and PECAM-Ex8-R primers are shown as red arrows, the coding exons are marked in black. **c.** RT-PCR result. The expression of *Pecam* during *in vitro* differentiation of WW93-A12, WW103-4A6, WW103-8E6 and WW103-8G9 was examined by RT-PCR using PECAM-Ex6-F and PECAM-Ex8-R primers. Note that the *Pecam* expression was completely blocked in WW103-4A6, but not in WW103-8E6 and WW103-8G9.

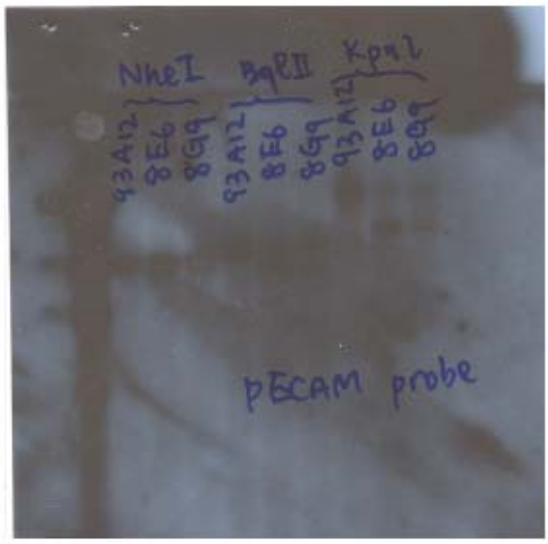
In an attempt to resolve how this situation could have occurred, Southern analysis was performed using a *Pecam* specific probe (Fig. 5-7). This revealed that both WW103-8E6 and WW103-8G9 were heterozygous for *Pecam* locus. But interestingly, the ratio between the targeted restriction fragment and the wild-type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The unexpected Southern result suggested that both clones might be trisomic. If so, it is most likely that the trisomy appeared after the end point cassette targeting and before the retrovirus infection. In this case, the original trisomy would contain two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild-type chromosome. After regional trapping, the puromycin resistant trisomy will have one 3' *Hprt* chromosome with an inversion, one 3' *Hprt* chromosome with targeted end point cassette and one 5' *Hprt* wild-type chromosome. Induced mitotic recombination can generate two different products: clones with two inversion chromosomes and one chromosome with the end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the end point cassette (WW103-8G9). In both cases, the clones will carry three targeted E_2DH alleles (end point targeting), thus Southern analysis using E_2DH probe can not distinguish these trisomies from the homozygous inversion clones.

Therefore the impaired differentiation potential of WW103-8E6 does not have any direct connection with the *Pecam* trapping and the subsequent inversion. This may be the result of the up-regulation of the chromosome 11 genes caused by the extra chromosome.

5.2.4 WW103-14F11 (*2810410L24Rik*)

As described in the previous chapter, the WW103-14F11 subclone has a proviral insertion at the *2810410L24Rik* locus (119.9 Mb) (Fig. 5-8a), which is close to the telomere of the chromosome 11. But instead of trapping the *2810410L24Rik* gene, which is transcribed from the sense strand (from centromere to telomere), the retrovirus trapped another transcript transcribed from the anti-sense strand (from telomere to centromere), *D030042H08Rik*.

a


NheI
BglII
KpnI
93A12
8E6
8G9
93A12
8E6
8G9
93A12
8E6
8G9
pECAM probe

b

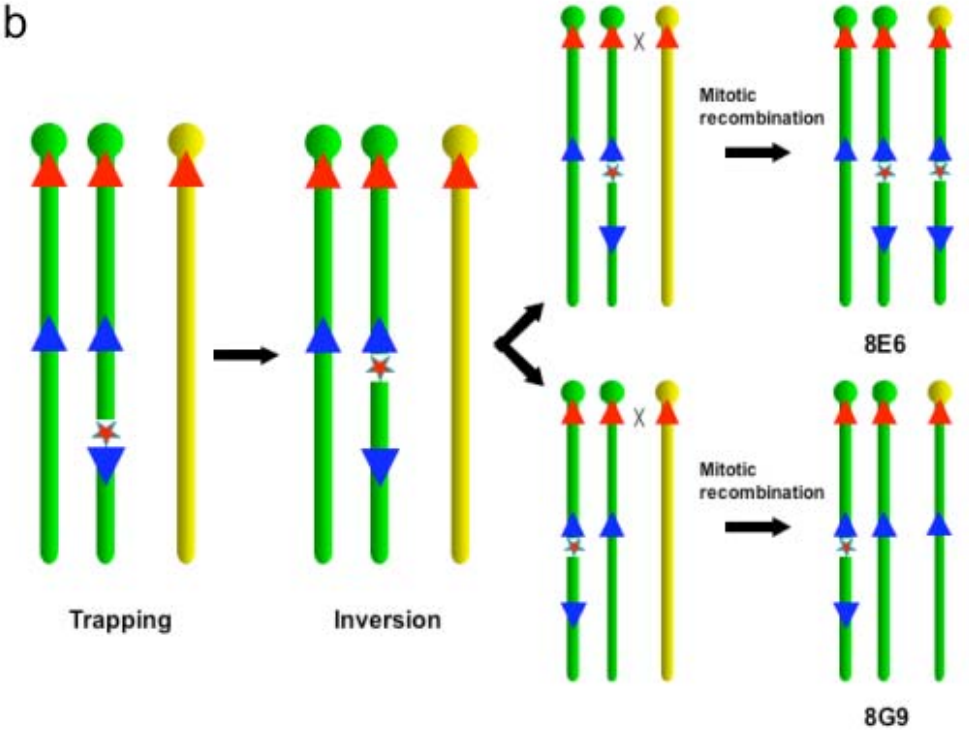
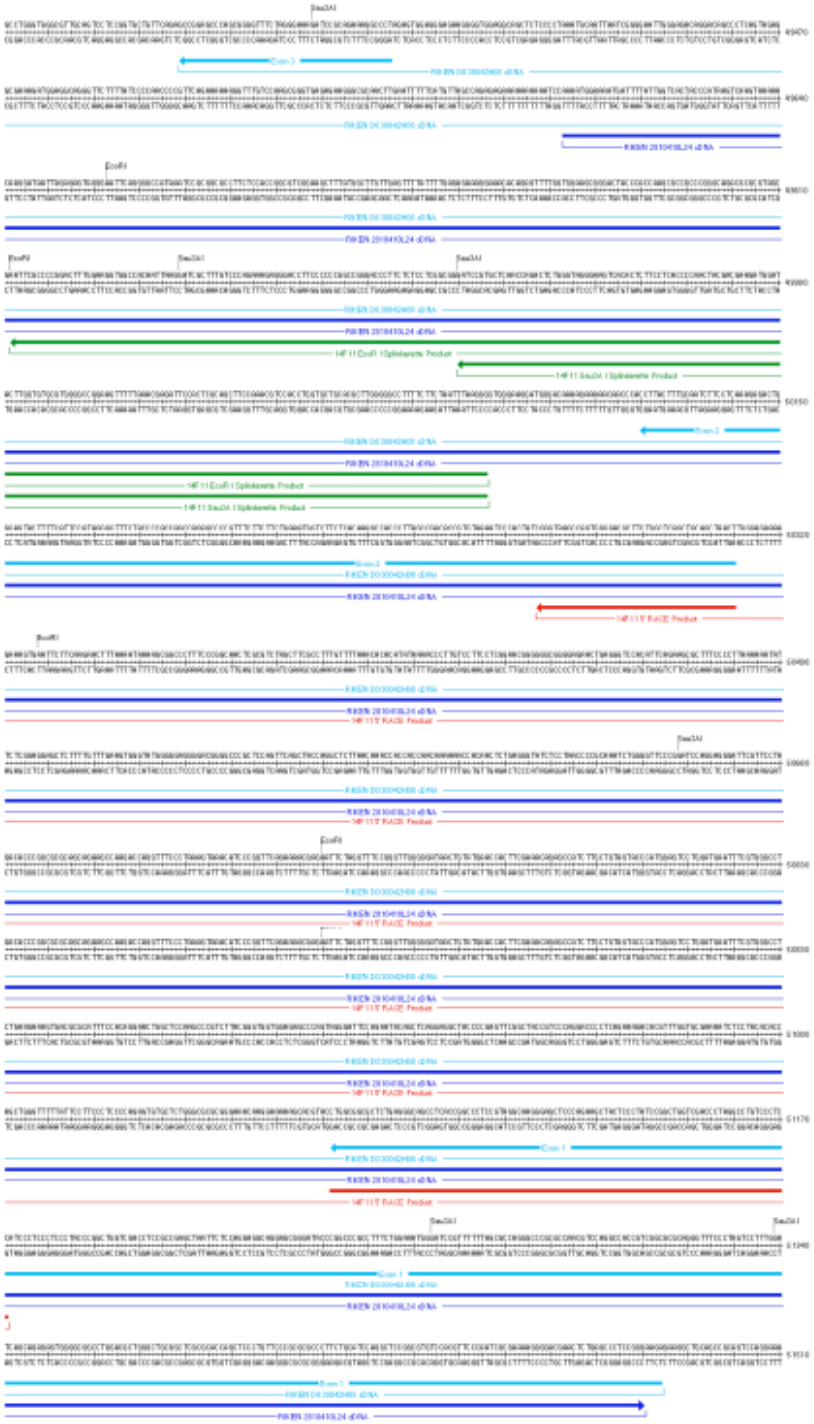
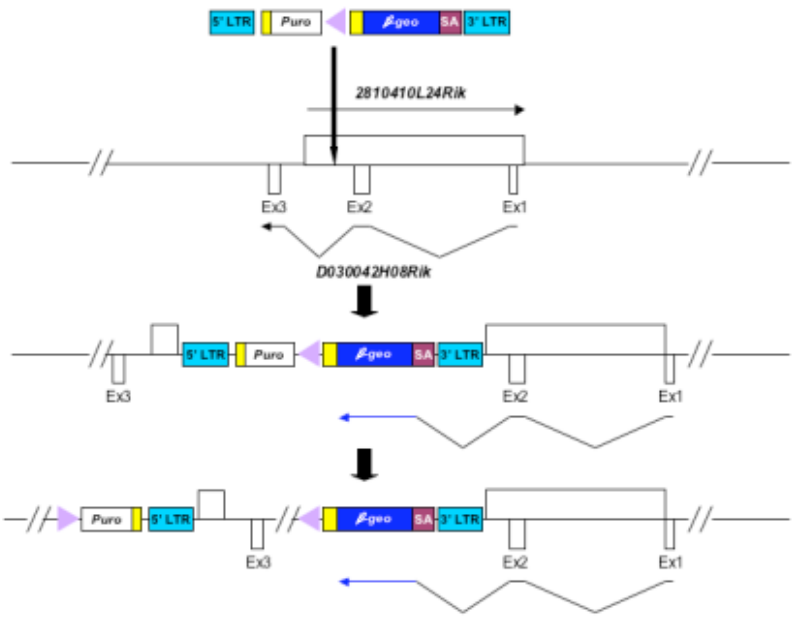

Trapping
Inversion
Mitotic recombination
8E6
Mitotic recombination
8G9

Fig. 5-7 WW103-8E6 and 8G9 are both chromosome 11 trisomies. a. Southern analysis of clones, WW103-8E6 and WW103-8G9. A *Pecam* specific probe was used to detect a 10.6 kb *Bgl*II wild type fragment and a 9 kb *Bgl*I targeted fragment. The same probe was also used to detect a 30 kb *Kpn*I wild type fragment and a 19 kb targeted fragment. Both WW103-8E6 and WW103-8G show a wild type fragment for both digestions. But the ratio between the targeted restriction fragment and the wild type restriction fragment is not 1:1. For WW103-8E6, the ratio is around 2:1, while for WW103-8G9 the ratio is around 1:2. The Southern result suggested that both clones might be trisomic. **b.** A schematic illustration of possible recombination in WW103-8E6 and WW103-8G. The original trisomy probably contained two 3' *Hprt* chromosomes targeted with the end point cassette, and one 5' *Hprt* wild type chromosome. After regional trapping, the puromycin resistant trisomy has one 3' *Hprt* chromosome with the inversion, one 3' *Hprt* chromosome with the end point cassette and one 5' *Hprt* wild type chromosome. Induced mitotic recombination can generate two different products, depending on which 3' *Hprt* chromosome participates in the recombination process. Clones can be generated with two inversion chromosomes and one chromosome with the targeted end point cassette (WW103-8E6), or clones with one inversion chromosome and two chromosomes with the targeted end point cassette (WW103-8G9).

a



b


5' LTR
Puro
βgeo
SA
3' LTR
2810410L24Rik
Ex3
Ex2
Ex1
D030042H08Rik

c

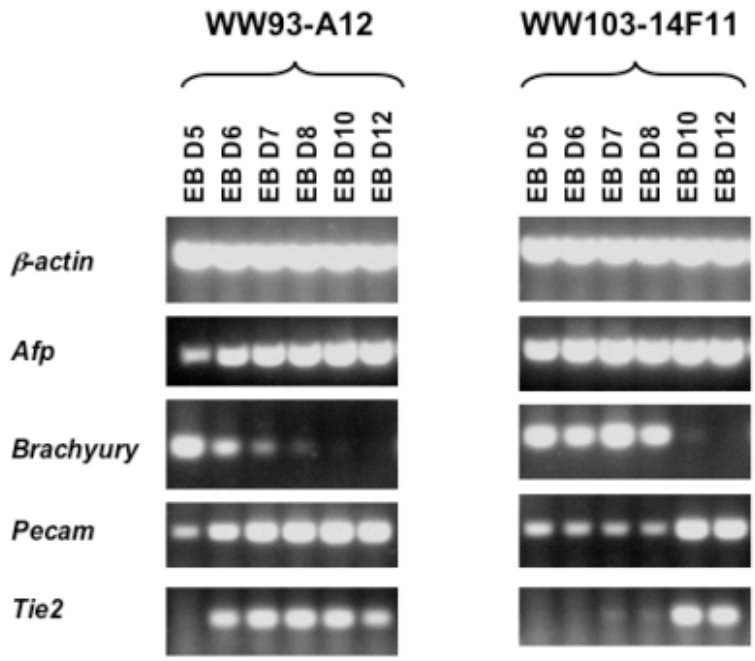

WW93-A12
WW103-14F11
EB D5
EB D6
EB D7
EB D8
EB D10
EB D12
β-actin
Afp
Brachyury
Pecam
Tie2

Fig. 5-8 WW103-14F11. a. Splinkerette PCR and 5' RACE.results. *Eco*RI and *Sau*3AI Splinkerette PCR products have both mapped the proviral insertion site to the *2810410L24Rik* locus. The 5' RACE result matched another transcript *D030042H08Rik*.The *2810410L24Rik* transcript is marked in dark blue, the *D030042H08Rik* transcript is marked in light blue, the splinkerette PCR fragments are marked in green and the 5'RACE product is marked in red. **b.** Schematic illustration of the structure of proviral insertion and the subsequent inversion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note that *Pecam* and *Tie2* RT-PCR results showed that the up-regulation of these two markers in the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed.

The splinkerette results mapped the proviral insertion site between the second and third exons of *D030042H08Rik*. However, the 5' RACE result did not match the *D030042H08Rik* cDNA sequence perfectly, although the transcript structure is similar. It is possible that the 5' RACE result and the *D030042H08Rik* cDNA sequence represent two different alternative splice forms of the same gene.

Nevertheless, the gene-trap retrovirus insertion and the subsequent inversion will disrupt the transcripts from both strands (Fig. 5-8b). The *in vitro* differentiation results showed that EBs derived from WW103-14F11 have impaired potential to develop into endothelial cells. RT-PCR using *Pecam* and *Tie2* primers has shown that the up-regulation of the expression of these two markers during the differentiation process was significantly delayed. On the other hand, the early mesoderm marker, *Brachyury*'s down-regulation was also delayed (Fig. 5-8c).

Further confirmation of this subclone is still undergoing. One way to directly confirm the defective endothelial cell differentiation is to use collagen IV coated dishes to induce undifferentiated ES cells to first differentiate into $Flk1^+$ cells (Yamashita, Itoh et al. 2000). When FACS sorted $Flk1^+$ cells were cultured with the addition of VEGF, these cells will further differentiate into PECAM1+ sheets of endothelial cells, which also express other endothelial cell-specific markers, such as *VE-cadherin* and *CD34*. By comparing the endothelia cell differentiation of the WW103-14F11 cells and the wild-type control cells, it will be possible to identify the molecular mechanism underlying the defective phenotype and determine at which stage the differentiation into endothelial lineage is blocked. However, it will be difficult to distinguish the phenotypes of the two genes transcribed from the opposite directions.

5.2.5 WW103-13D10 (*LOC217071*)

*Sau*3A1 and SpeI/*Xba*I/*Nhe*I Splinkerette PCR products mapped the proviral insertion site in the WW103-13D10 clones to the second intron of a hypothetical mouse gene, *LOC217071* (Fig. 5-9a and b). This gene is transcribed from the sense strand (from the centromere to telomere), and

Southern analysis using an *E_2DH* 3' probe has confirmed that this clone is homozygous for the targeted *E_2DH* allele. Southern analysis using a *LacZ* probe has shown that it only carried a 6.9 kb *Kpn*I restriction fragment which suggested that WW103-13D10 contains an intact proviral insertion. As discussed in the previous chapter, this might be caused by a G2 *trans* recombination event. The duplication chromosome has both a functional *Puro* and a functional *Bsd* cassette, and it can become homozygous after induced mitotic recombination because it has not lost any genetic material.

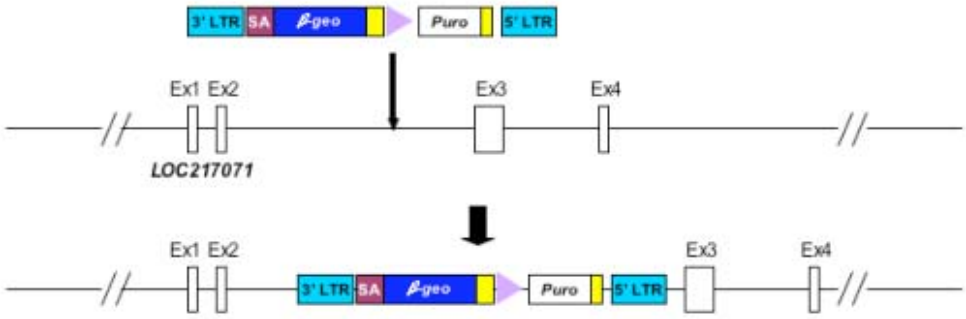
The homozygous duplication clone showed an obvious abnormality in *in vitro* differentiation. The undifferentiated WW103-13D10 ES cells expressed high levels of markers for differentiated cell types, such as *Afp*, *Gata4* and *Hnf4*. The expression of undifferentiated ES cell markers, like *Nodal* and *Oct3/4*, was significantly down-regulated, compared to the WW93-A12 control (Fig. 5-9c).

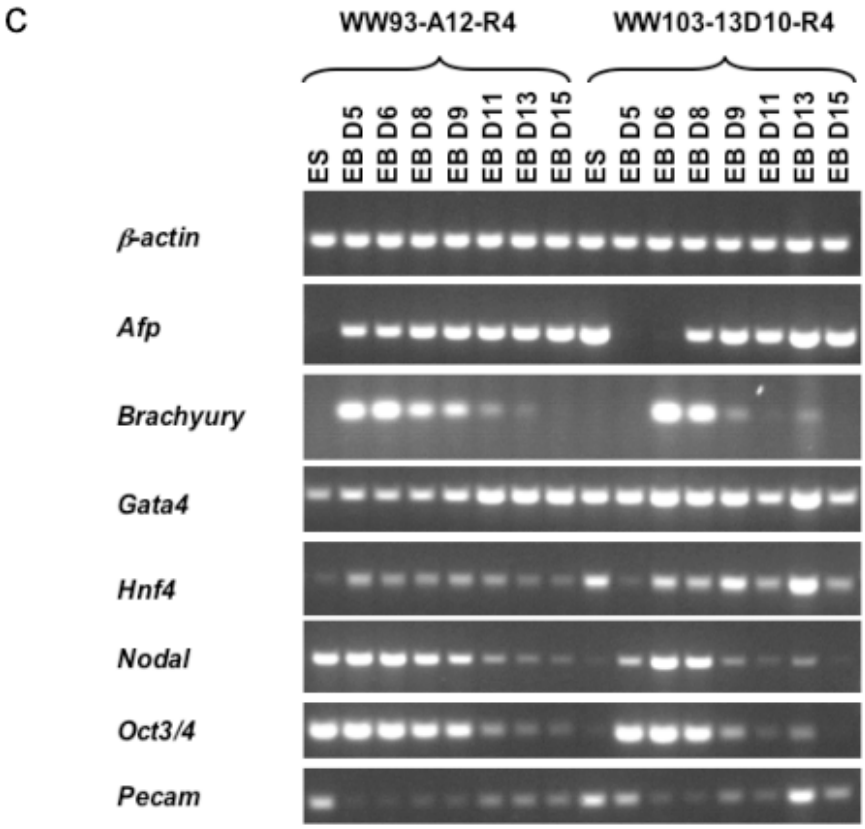
Interestingly, during the process of differentiation, the EBs made from WW103-13D10 ES cells seemed to differentiate normally. At day 5, they lost the expression of *Afp*, but regained the expression of *Nodal* and *Oct3/4*. After this, various markers showed expression patterns similar to those which were observed in the WW93-A12 control. But *Hnf4* and *Gata4* expression were still significantly up-regulated compared to the control.

Sib-selection was carried out on two subclones each from WW93-A12 and WW103-13D10. The same number of undifferentiated ES cells were plated into the wells of a gelatinized 24-well plate and selected in M15, M15+G418, M14+puromycin, M15+blasticidin and M15+HAT, respectively. WW103-13D10 was resistant to both puromycin and blasticidin, which suggested that this clone have two duplication chromosomes, instead of two inversion chromosomes (Fig. 5-9d). Most likely, the phenotype observed in WW103-13D10 was caused by the duplication, instead of the disruption of the *LOC217071* locus.

a

b


3' LTR
SA
β-geo
Puro
5' LTR
Ex1 Ex2
Ex3
Ex4
LOC217071
Ex1 Ex2
3' LTR
SA
β-geo
Puro
5' LTR
Ex3
Ex4


c
WW93-A12-R4
WW103-13D10-R4
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Afp
Brachyury
Gata4
Hnf4
Nodal
Oct3/4
Pecam

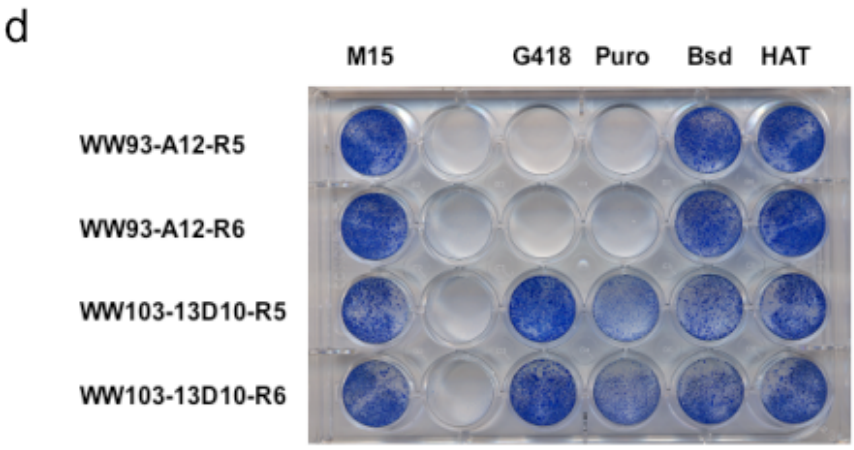

d
M15
G418
Puro
Bsd
HAT
WW93-A12-R5
WW93-A12-R6
WW103-13D10-R5
WW103-13D10-R6

Fig. 5-9 WW103-13D10 a. Splinkerette PCR results. *Spe*I/*Xba*I/*Nhe*I and *Sau*3AI Splinkerette PCR products have mapped the proviral insertion site to the *LOC217071* locus. The splinkerette PCR fragments are marked in green. **b.** Schematic illustration of the structure of proviral insertion in WW103-14F11. **c.** RT-PCR results of EBs derived from WW013-14F11 and WW93-A12. Note undifferentiated WW103-13D10 ES cells express high amounts of *Afp*, *Hnf4* and *Gata4,* but low amount of *Oct3/4* and *Nodal*. Despite the altered expression of these markers,the EBs derived from WW103-13D10 ES cells seem to differentiate normally. **d.** Sib-selection of WW93-A12 and WW103-13D10 clones. Two subclones of each cell line were used for the sib selection. Same amount of undifferentiated ES cells were plated onto a gelatinized 24-well plate and selected with M15, M15+G418, M14+Puromycin, M15+Blasticidine and M15+HAT, respectively. WW103-13D10 was resistant to both Puromycin and Blasticidine, which suggested that this clone is a homozygous duplication, instead of a homozygous inversion.

5.2.6 WW103-18F11 (*Acly*)

One of the mitotic recombination clones, WW103-18F11, showed impaired *in vitro* differentiation potential. After EBs made of WW103-18F11 ES cells were plated on the gelatinized tissue culture plates at Day 5, the EBs did not form cystic three-dimensional structures.

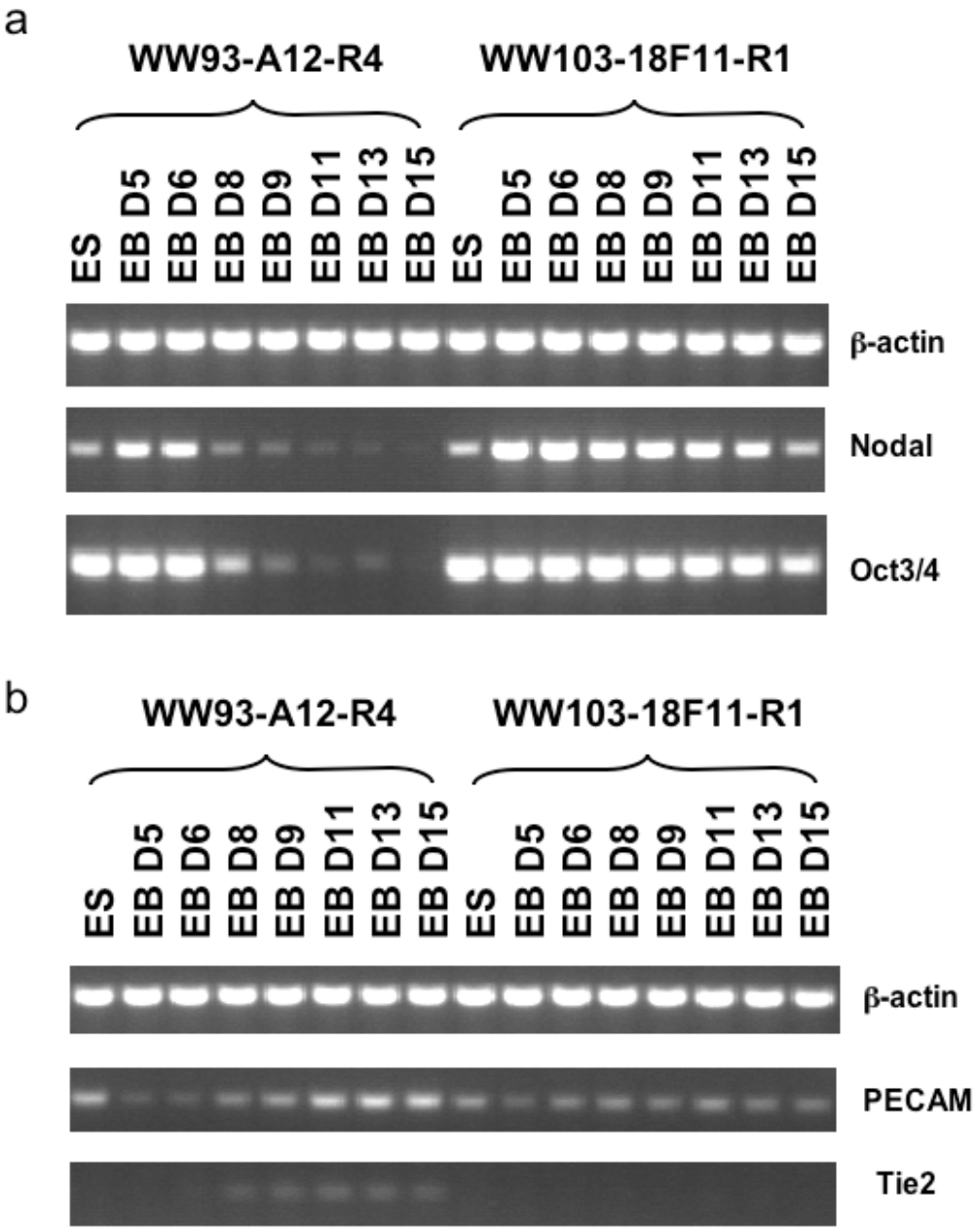
When RT-PCR was performed on RNA extracted from WW103-18F11 embryoid bodies collected at different time points, these EBs were found to express high levels of the undifferentiated ES cell markers, *Oct3/4* and *Nodal*, as late as Day 18 of the *in vitro* differentiation protocol. The expression of *Oct3/4* and *Nodal* still decreased a little during the differentiation process, but down-regulation was not as rapid as that in the control cell line (Fig. 5-10a and e).

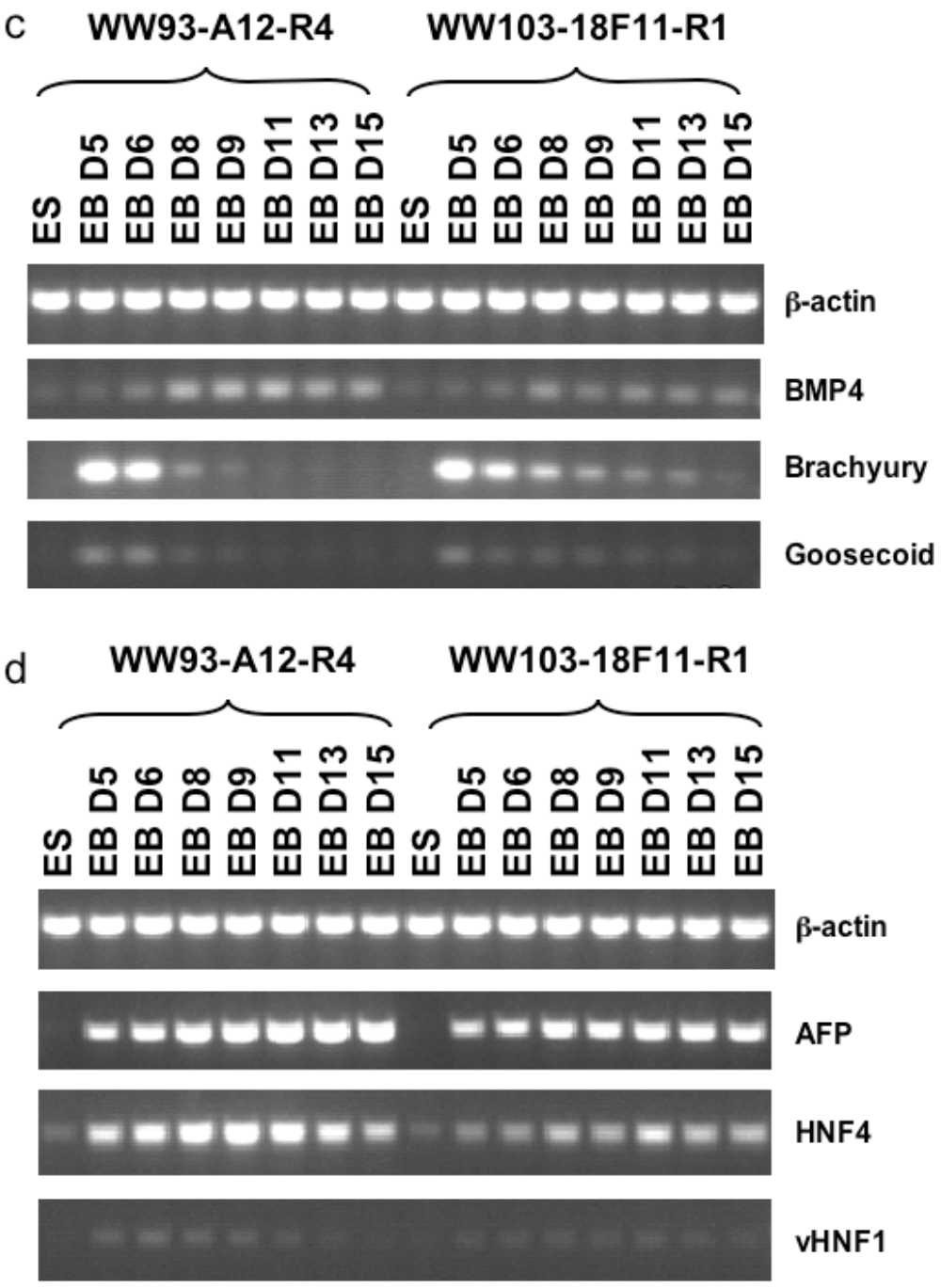
Tie2 expression was not detected during the whole process of differentiation of WW103-18F11 cells. The expression of *Pecam* was maintained at a constant basal level, instead of being up-regulated, as was observed in the WW93-A12 control cell line (Fig. 5-10b). Both of the markers are endothelial cell-specific proteins expressed during the formation of vascular structures in ES-derived EBs. The *Tie2* gene encodes a growth factor receptor, while the *Pecam* protein is an endothelial cell specific antigen. Vittet *et al*. (1996) has shown that both genes are expressed at low levels in undifferentiated ES cells. Normally, in the process of *in vitro* differentiation, the expression of both genes is absent at day 0-3 and is detected again from day 4. After this, the expression level of both genes is consistently up-regulated, as detected by Northern blotting and/or Immunofluorescence. However, in that experiment, only EBs from Day 3 to Day 7 were checked (Vittet, Prandini et al. 1996). In my experiment, I have observed the expression of *Tie2* and *Pecam* in the control line throughout the 15-day differentiation process. Thus, my observation suggested that the differentiation of endothelial cells in the mutant cell line was significantly impaired over the entire 15-day differentiation process.

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury*, *Goosecoid*) is similar between WW103-18F11 and WW93-A12 (Fig. 5-10c). However, low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. The expression of one of the endodermal markers, *Hnf4*, in WW103-18F11 was much lower than that in the control. Apart from these changes, no major differences were observed in the levels of expression of the other markers (Fig. 5-10d).

5' RACE results revealed that the gene-trap retrovirus trapped Exon 1 of ATP-citrate lyase (*Acly*) (Fig. 5-11a). *Acly* is one of two cytosolic enzymes in eukaryotes that synthesize acetyl-coenzyme A (acetyl-CoA), the other enzyme is acetyl-CoA synthetase 1. *Acly* catalyzes the formation of acetyl-coenzyme A (CoA) from citrate and CoA, and hydrolyzes ATP to ADP and phosphate. Because acetyl-CoA is an essential component for cholesterol and triglycerides synthesis, *Acly* is believed to be a potential therapeutic target for hyperlipidemias ad obesity (Beigneux, Kosinski et al. 2004).

To characterize this mutant cell line further, pure subclones of WW103-18F11 were derived by low density plating to form single colonies. Six subclones were picked and expanded. To confirm chromosomal structure of these subclones, sib-selection was performed on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of the control cell line, WW93-A12-R4 and WW93-A12-R5. An equal number of ES cells from each subclone were plated onto multiple 6-well plates and selected with M15, M15+puromycin, M15+blasticidin, M15+G418 and M15+HAT, respectively. As expected, the two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggests that WW103-18F11 have undergone correct recombination (Fig. 5-11b).


a
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
b
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
PECAM
Tie2


c
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
BMP4
Brachyury
Goosecoid
d
WW93-A12-R4
WW103-18F11-R1
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
AFP
HNF4
vHNF1

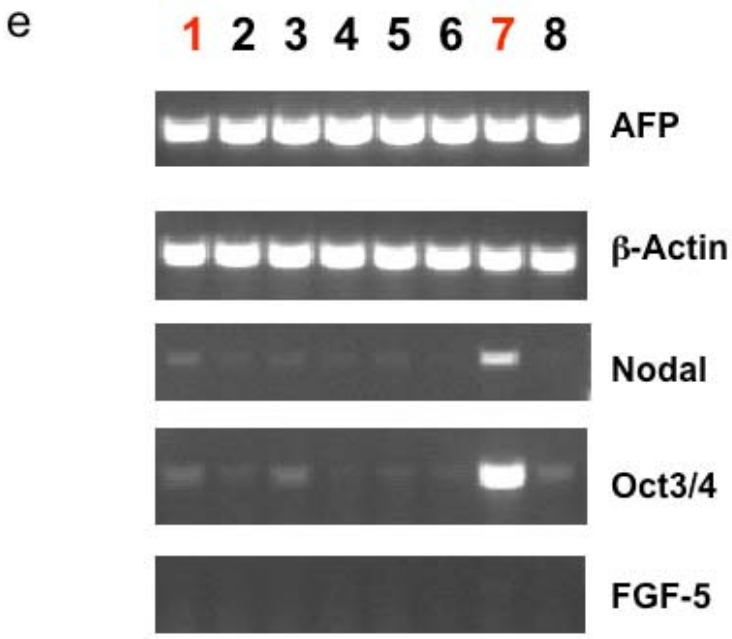
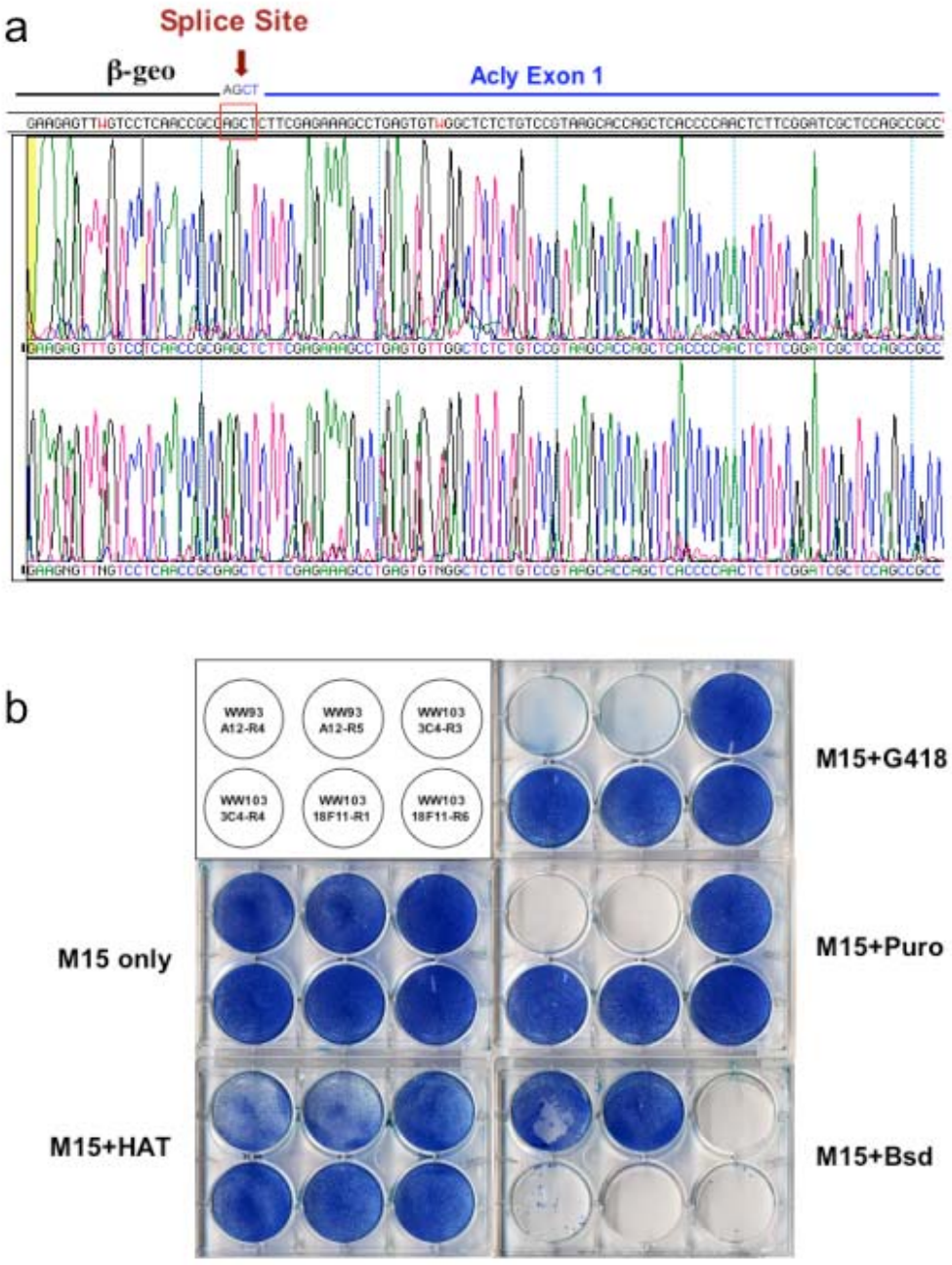
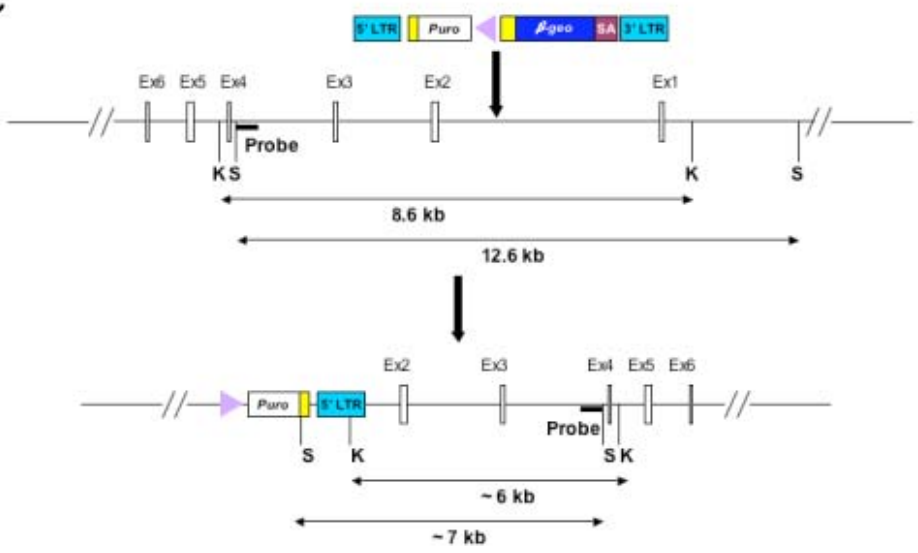


Fig. 5-10 RT-PCR results of WW103-18F11. a. Primitive ectoderm markers. The expression of *Oct3/4* and *Nodal* decreased slightly in the differentiation process. **b.** Endothelial markers. *Tie2* expression was not detected throughout the differentiation process of WW103-18F11 cells. The expression of *Pecam* was not up-regulated as observed in WW93-A12 control cell line. **c.** Early mesoderm markers. Low levels of expression of *Brachyury* and *Goosecoid* were still detected in WW103-18F11 derived EBs collected at later stages of the differentiation process, while no expression of these markers were detected in later stage EBs derived from WW93A12. **d.** Endoderm markers. The expression of most endoderm markers appears quite similar between the two cell lines. However, the expression of *Hnf4* in WW103-18F11 was much lower than that in the control. **e.** Day 18 embryoid bodies RT-PCR results. Note that WW103-18F11 EBs still express high amount of *Nodal* and *Oct3/4* at day 18. Lane 1, WW93-A12 control cell line; lane 7, WW103-18F11. Lanes 2-6 and 8 are day 18 EBs derived from other irrelevant mutant cell lines.


a
Splice Site
β-geo
AGCT
Acly Exon 1
b
WW93 A12-R4
WW93 A12-R5
WW103 3C4-R3
WW103 3C4-R4
WW103 18F11-R1
WW103 18F11-R6
M15+G418
M15 only
M15+Puro
M15+HAT
M15+Bsd

c


5' LTR
Puro
βgeo
SA
3' LTR
Ex6
Ex5
Ex4
Ex3
Ex2
Ex1
Probe
K S
K
S
8.6 kb
12.6 kb
Puro
5' LTR
Ex2
Ex3
Ex4
Ex5
Ex6
Probe
S
K
S K
~ 6 kb
~ 7 kb

d

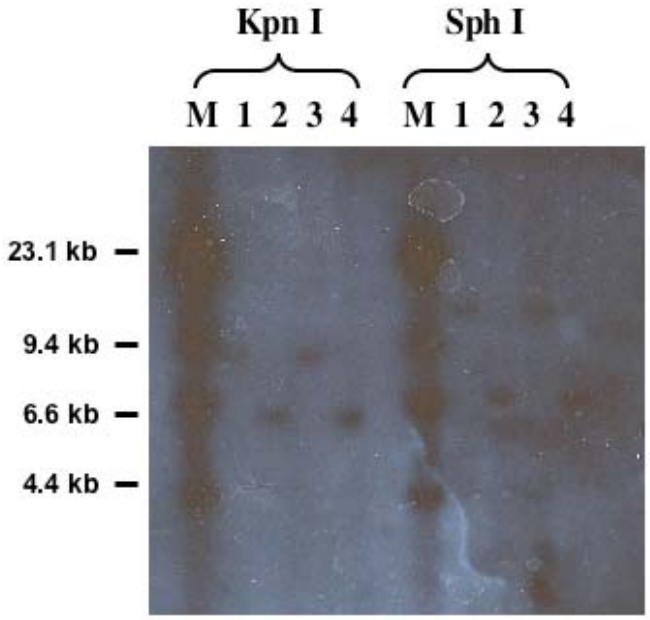

Kpn I
Sph I
M 1 2 3 4
M 1 2 3 4
23.1 kb
9.4 kb
6.6 kb
4.4 kb

Fig. 5-11 Identification of WW103-18F11. a. 5' RACE result. The sequence of the 5' RACE product matched the first exon of *Acly* gene. **b.** Sib-selection of WW103-18F11 and WW93-A12. Sib-selection was carried out on two of the WW103-18F11 subclones, WW103-18F11-R1 and WW103-18F11-R6, as well as two subclones of control cell line, WW93-A12-R4 and WW93-A12-R5. The same number of ES cells from two subclones each of WW93-A12 and WW103-18F11 were plated onto gelatinized 6-well plates and selected with M15, M15+Puromycin, M15+Blasticidine, M15+G418 and M15+HAT, respectively. The two WW103-18F11 subclones are $Puro^R$, Neo^R, Bsd^S and HAT^R, and the two WW93-A12 subclones are $Puro^S$, Neo^S, Bsd^R and HAT^R. The drug resistance pattern of WW103-18F11 suggested that it is a homozygous inversion clone. **c.** Schematic illustration of the structure of the proviral insertion and the subsequent inversion in WW103-18F11. **d.** Southern analysis of WW103-18F11. Southern analysis has been carried out using an *Acly* gene specific probe. This probe can detect an 8.6 kb *Kpn*I restriction fragment for the wild type allele. For WW103-18F11, the probe only detected an 6 kb *Kpn*I restriction fragment for the gene trap allele. Also, the probe only detected an 7 kb *Sph*I restriction fragment for the gene trap allele, instead of the 12.6 kb *Sph*I restriction fragment for the wild type allele. This Southern result has confirmed that both *Acly* alleles have been disrupted by the gene trap insertion. M, λ *Hin*dIII marker (New England Biolabs); lane 1, WW93-A12-R4; lane 2, WW103-18F11-R1; lane 3, WW93-A12-R5; lane 4, WW103-18F11-R6.

Southern analysis was carried out using an *Acly* gene specific probe. When this probe was hybridized to *Kpn*I digested genomic DNA, it detects an 8.6 kb wild-type fragment and an approximately 7 kb gene-trap fragment. When this probe was hybridized to *Sph*I digested genomic DNA, it detects a 12.6 kb wild-type fragment and an approximately 8 kb gene-trap fragment. As expected, only the gene-trap fragment was detected in the WW103-18F11 subclones. This Southern result confirms that both alleles of the *Acly* gene have been disrupted by the gene-trap insertion.

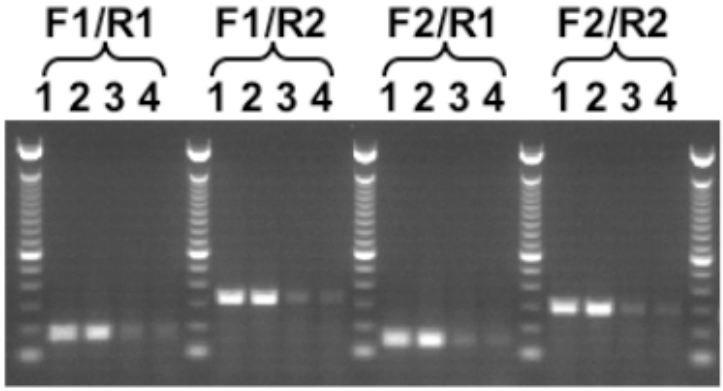
To see whether the gene-trap insertion and the subsequent inversion has disrupted transcription of the locus, PCR primers were used to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). First strand cDNA was synthesised using total RNA extracted from WW103-18F11 and WW93-A12 ES cells. The RT-PCR results showed that transcription from Exon 1 to downstream exons was blocked. Weak PCR bands were detected for the *Acly*-deficient cell lines, which are likely to be contamination from feeder cells (Fig. 5-12a). Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the Exon 1/β-geo fusion transcript from the trapped allele. As expected, specific bands were only detected for WW103-18F11 subclones, but not for WW93-A12 control (Fig. 5-12c).

To see whether the gene-trap insertion and the subsequent inversion has affected the transcription of downstream exons, PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). Specific PCR bands were detected in the WW103-18F11 mutant cell line and the WW93-A12 control cell line. Therefore it is likely that there is an alternative transcription start point between the retroviral insertion point and Exon 24, but the precise location of the mutant transcript start in WW103-18F11 is not known (Fig. 5-12b).

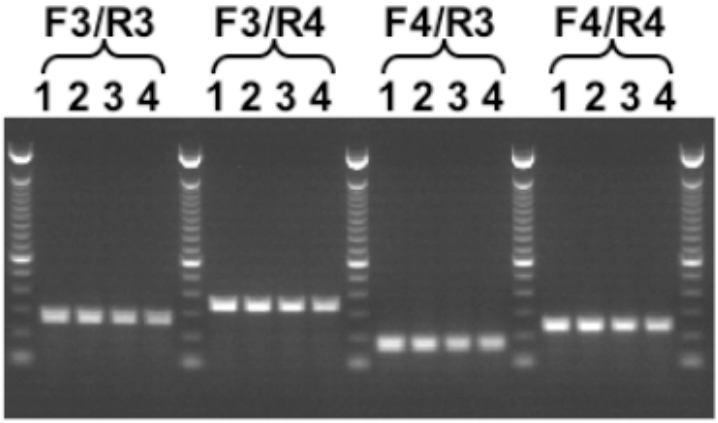
These RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation. In the WW93-A12 control cell line, *Acly* was highly expressed in undifferentiated ES cells, as well as throughout the whole differentiation process. In the WW103-18F11 mutant cell line, the F1/R1 and F2/R2 primer pairs did not detect the expression of the *Acly* upstream exons during *in vitro* differentiation, but the F3/R2 and F4/R4 primer pairs did detect expression of the *Acly* downstream exons (Fig. 5-12d).

To identify a causal link between the gene-trap insertion and inversion at the *Acly* locus and the severely impaired differentiation potential, a BAC rescue experiment was carried out to reverse the phenotype of the WW103-18F11 ES cell clone. A 129 S7 BAC clone, BMQ-290J5 was identified in Ensembl and confirmed to contain the complete *Acly* gene by PCR (Fig. 5-13a and data not shown). A *PGK-EM7-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination (Liu, Jenkins et al. 2003). The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe (Fig. 5-13b and c).

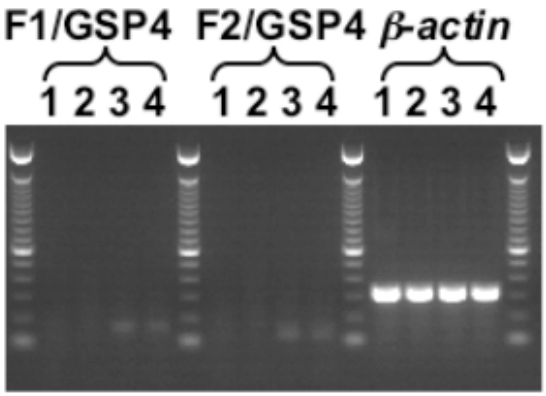
a


F1/R1
F1/R2
F2/R1
F2/R2
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

b


F3/R3
F3/R4
F4/R3
F4/R4
1 2 3 4
1 2 3 4
1 2 3 4
1 2 3 4

c


F1/GSP4
F2/GSP4
β-actin
1 2 3 4
1 2 3 4
1 2 3 4

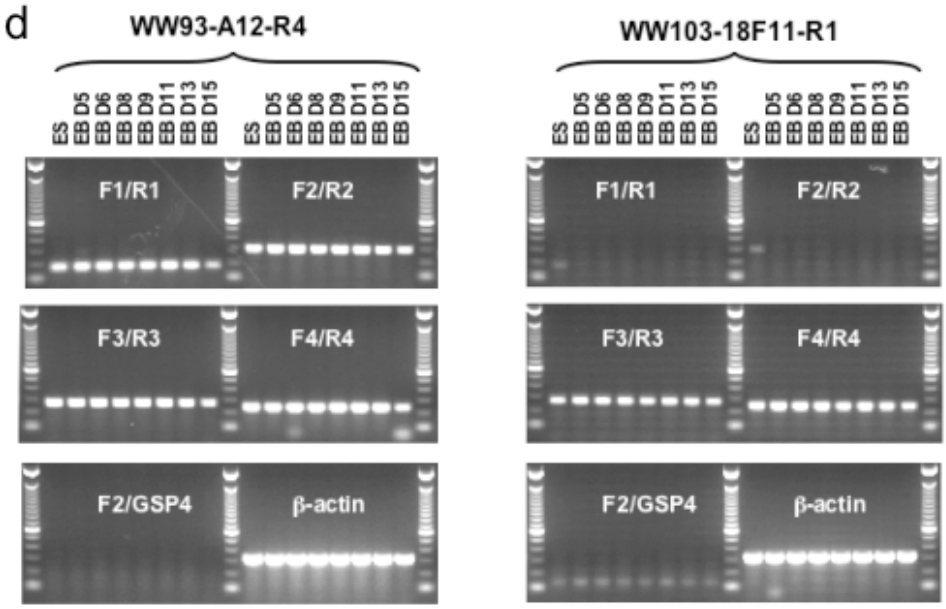


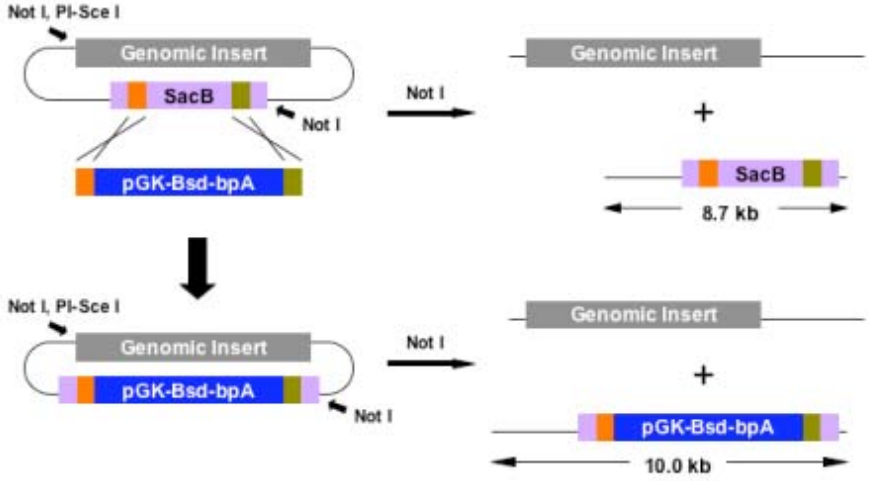
Fig. 5-12 Confirmation of WW103-18F11. a. RT-PCR to detect transcripts across the inversion breakpoint. PCR primers were designed to specifically amplify cDNA fragments from Exon 1 to Exon 2 (F1/R1 and F2/R1) and Exon 1 to Exon 3 (F1/R2 and F2/R2). RNA extracted from undifferentiated ES cells was used as template. lane 1, WW93-A12-R4; lane 2, WW93-A12-R5; lane 3, WW103-18F11-R1; lane 4, WW103-18F11-R6. RT-PCR results showed that *Acly* transcription across the inversion breakpoint was blocked. **b.** RT-PCR to detect the transcription of downstream exons. PCR primers were designed to specifically amplify cDNA fragments from Exon 24 to Exon 28 (F3/R3 and F3/R4) and Exon 25 to Exon 28 (F4/R3 and F4/R4). RT-PCR results showed that the transcription of downstream exons was not affected by the inversion. This suggests the existence of alternative transcript start points. **c.** RT-PCR to detect the Exon 1/β-geo fusion transcript. Primer pairs F1/GSP4 and F2/GSP4 were used to specifically amplify the gene trap fusion transcript. **d.** RT-PCR to detect *Acly* expression during ES cell *in vitro* differentiation. High *Acly* expression was detected in the whole process of differentiation of WW93-A12 control cell line, but not in WW103-18F11.

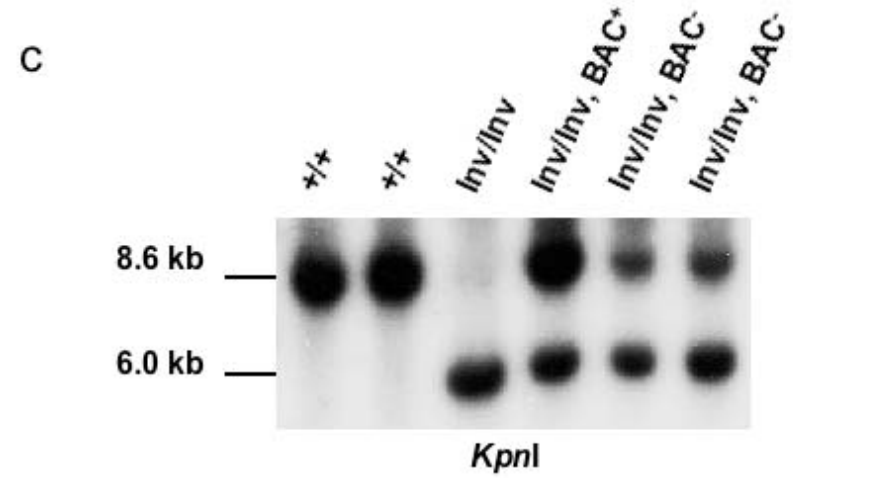
The modified BAC clone was linearized by I-*Sce*I and electroporated into WW103-18F11 ES cells (HAT^R, Neo^R, $Puro^R$, Bsd^S). 12 blasticidin resistant clones were picked and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild-type restriction fragment (Fig. 5-13c). One of the clones, WW113-2-8 has the wild-type restriction fragment and the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild-type copies of *Acly* gene. Another two clones, WW113-2-10 and WW113-2-11 also have the wild-type restriction fragment. But the ratio between the wild-type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single copy of the BAC DNA, which can be randomly truncated and are likely to be incomplete. Western analysis was performed on whole-cell lysates extracted from undifferentiated wild-type control, *Acly*-deficient and BAC-rescued ES cells using a polyclonal rabbit anti-*Acly* antibody. *Acly* protein was not detected in the lysates from the WW103-18F11 cells. However, one of the BAC-rescued clones (WW113-8) expressed similar level of the *Acly* protein as the WW93-A12 wild-type control cells, indicating that this clone (WW113-2-8) is a rescued clone (Fig. 5-13d). The other two clones, WW113-2-10 and 2-11, which did not express *Acly* protein, are likely to only contain a truncated form of the BAC DNA and thus they were used as negative controls.

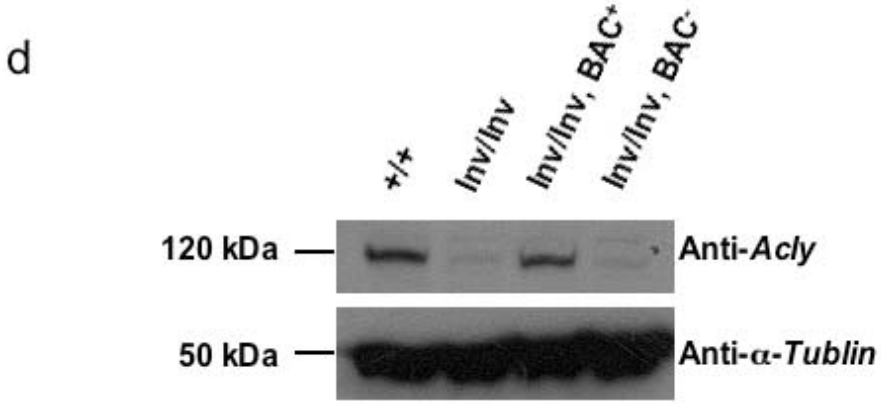
These three clones were expanded and induced to differentiate *in vitro*. After EBs were plated on the gelatinized tissue culture plates at Day 5, the EBs derived from WW113-2-8 ES cells could form cystic three-dimensional structures, while the EBs derived from WW113-2-10 and 2-11 ES cells could not (Fig. 5-13e).

a

b


Not I, PI-Sce I
Genomic Insert
SacB
Not I
pGK-Bsd-bpA
Not I
Genomic Insert
+
SacB
8.7 kb
Not I, PI-Sce I
Genomic Insert
pGK-Bsd-bpA
Not I
Not I
Genomic Insert
+
pGK-Bsd-bpA
10.0 kb


c
+/+
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
Inv/Inv, BAC-
8.6 kb
6.0 kb
KpnI


d
+/+
Inv/Inv
Inv/Inv, BAC+
Inv/Inv, BAC-
120 kDa
Anti-Acly
50 kDa
Anti-α-Tublin

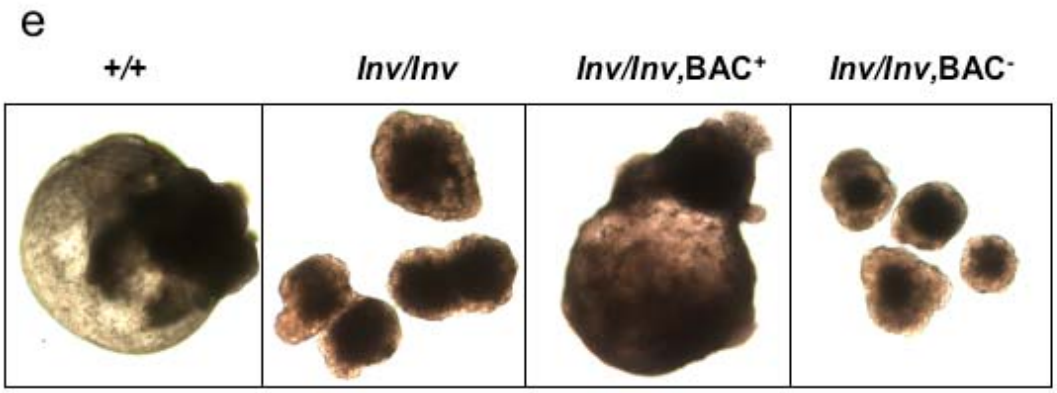

e
+/+
Inv/Inv
Inv/Inv,BAC+
Inv/Inv,BAC-

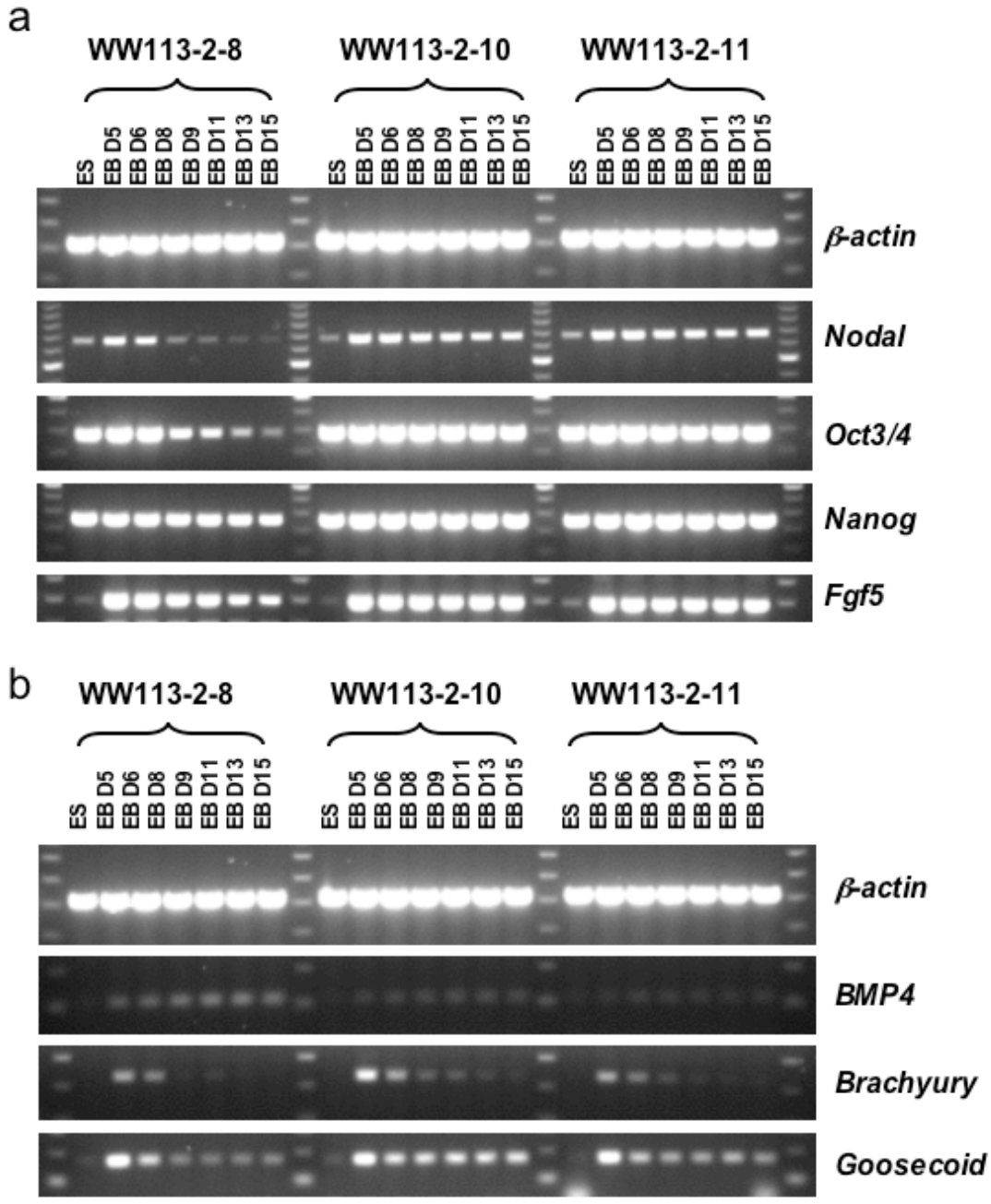
Fig. 5-13 BAC rescue. a. BAC clone BMQ-290J5. An 129 S7 BAC clone, BMQ-290J5 was picked according to the BAC end mapping information of this clone in the Ensembl database. This clone was confirmed to contain the complete *Acly* gene by PCR. BAC clone BMQ-290J5 is highlighted in blue. **b.** Insertion of a *PGK-Bsd-bpA* cassette into the BAC backbone. To facilitate the drug selection of BAC integration in ES cells, a *PGK-Bsd-bpA* cassette (pL313) was inserted into the *SacB* gene on the backbone of this BAC clone by *E. coli* recombination. **c.** Southern analysis of modified BAC clones. The correct insertion of the *Bsd* cassette into the BAC backbone was confirmed by Southern using a *SacB* specific probe. BAC DNA from the modified clones was digested with *Not*I. The *SacB* probe detected a 8.7 kb restriction restriction fragment for the unmodified BAC clone, and it detected a 10 kb fragment for the modified clones. **d.** Southern analysis of BAC inserted ES cell clones. Genomic DNA from 12 blasticidin resistant clones were digested with *Kpn*I, and Southern analysis was carried out using an *Acly* gene specific probe to identify ES cell clones with a wild type restriction fragment. WW113-2-8 (red) has the wild type restriction fragment and the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:1, suggesting that this is likely to be a complemented clone which contains two wild type copies of Acly gene. Another two clones, WW113-2-10 and WW113-2-11 (blue) have the wild type restriction fragment, but the ratio between the wild type restriction fragment and the targeted restriction fragment is about 1:2, which suggests that both clones might contain a single BAC clone which is likely to be incomplete. These two clones were used as control for the in vitro differentiation experiment.

When RT-PCR was performed on RNA extracted from WW113-2-8 embryoid bodies collected at different time points, the expression of *Oct3/4*, *Nodal* and *Nanog* were down-regulated rapidly as observed in WW93-A12 derived EBs. However, WW113-2-10 and 2-11 still expressed high level of these primitive ectoderm markers at late stages of their differentiation process (Fig. 5-14a).

The expression pattern of the early mesodermal markers (*Bmp4*, *Brachyury* and *Goosecoid*) also became normal in the EBs derived from WW113-2-8 ES cells. The expression of *Brachyury* and *Goosecoid* was down-regulated much quicker in the WW113-2-8 derived EBs than in the WW113-2-10 or 2-11 derived EBs. The expression pattern of these markers in WW113-2-8 derived EBs was similar to that observed in the control WW93-A12 derived EBs (Fig. 5-14b).

Acly gene-specific RT-PCR primer pairs have also been used to check *Acly* expression during *in vitro* differentiation of the BAC rescue cell line, WW113-2-8. In the WW113-2-8 cell line, significantly higher *Acly* expression than the other two control cell lines was observed in undifferentiated ES cells, as well as throughout the whole differentiation process. RT-PCR using the F1/GSP4 primer pair confirmed that gene-trap transcripts were still present in the rescued cell line, WW113-2-8. So the phenotype observed in the WW103-18F11 mutant cell line was caused by the loss of normal *Acly* transcription, instead of the dominant-negative effects, as the phenotypes could be reversed by re-introducing a wild-type copy of *Acly* gene (Fig. 5-14c).

All these data suggest a direct link between the reduction in *Acly* expression and the impaired *in vitro* differentiation potential observed in WW103-18F11 derived EBs.


a
WW113-2-8
WW113-2-10
WW113-2-11
ES
EB D5
EB D6
EB D8
EB D9
EB D11
EB D13
EB D15
β-actin
Nodal
Oct3/4
Nanog
Fgf5
b
WW113-2-8
WW113-2-10
WW113-2-11
β-actin
BMP4
Brachyury
Goosecoid

5.3 Discussion

In this chapter, I have described how I have used ES cell *in vitro* differentiation to screen a set of homozygous mutant ES clones. A panel of 16 markers was used to carry out the primary screen. To increase the throughput of the screen, I only took samples at three time points. If a homozygous mutant ES cell clone showed abnormal expression for one or more markers, the clone was subsequently tested in the second round screen. In the second round screen, more samples were taken at different time points, and additional markers were checked by RT-PCR to confirm the authenticity of the phenotype and also try to explain the phenotype at the molecular level by the gain or loss of specific differentiation markers.

5.3.1 Throughput of the screen

In this experiment, only a limited number of homozygous mutant ES clones were used for the *in vitro* differentiation screen. Therefore, it is possible to make a large number of EBs for each cell line and take samples at multiple time points. However, if the number of cell lines for screening increases to several hundred or several thousand, it would be necessary to make tens of thousands of plates of EBs. To make hundreds of thousands of “hanging drops”, transfer them to gelatinized tissue culture plates and change media regularly will be a labour-intensive work.

The RT-PCR method is not sensitive enough for the high-throughput analysis either. Approximately 20 μg of total RNA can be extracted from a plate of 40 EBs after Day 10. But for the early EBs (Day 5 to Day 10), sometimes two or three plates of EBs need to be combined together to get enough RNA. The cDNA synthesized from 5 μg of total RNA is only enough for about 20 RT-PCR reactions. If 10 cell lines are checked at the same time, 8 time points are taken for each cell line, and 16 markers are screened, this will require 1,280 PCR reactions. Any clones that do not show an obvious abnormality in these 16 markers will be discarded which is not a thorough analysis of the differentiation potential. Also, RT-PCR is a semi-quantitative approach to

assess gene expression, which makes it difficult to detect minor changes in expression.

An alternative approach to RT-PCR is to use cDNA and oligonucleotide microarray technology, which has been well characterized and proven to be a powerful tool for large-scale screens. This technology enables one to check the expression of all the genes in the mouse genome simultaneously. The development of array technology has made it possible to use very small amounts of starting RNA template. However, the downside of this technology is that it is still very expensive and the high cost makes it impractical to screen a lot of samples. Another problem of using microarray analysis to study ES cell *in vitro* differentiation is the complexity of the input material. It would be necessary to perform many control experiments to define the normal ranges of expression levels during differentiation, before comparisons can be made with samples from the mutant lines. Fluorescent reporters and FACS can also be used to screen the mutants in a high-throughput manner. It will be further discussed in the final chapter.

Considerable data has accumulated on the expression pattern of various markers characterizing the development of the three germ layers and other differentiated cell types during the ES cell *in vitro* differentiation process. However, this data is scattered throughout the literature and is far from being systematic or comprehensive. The results in these publications were generated by various methods, including RT-PCR, Northern, *in situ* hybridization or immunohistochemistry. Different ES cell lines (feeder-free or feeder-dependent), different differentiation protocols, and different lengths of observation periods make the data generated from these different experiments difficult to compare.

So before ES cell *in vitro* differentiation is used for a large-scale *in vitro* recessive screen, a systematic, quantitative study should be performed to determine the expression pattern of important developmental and differentiation markers in the differentiation process of the wildly used ES cell lines (AB2.2, D3, R1 and E14.1, etc.). Ideally, this data needs to be compared

to the expression pattern of these markers *in vivo* to link the *in vitro* differentiation with its *in vivo* counterpart.

5.3.2 Alternative recombination

Interestingly, two clones that have shown an abnormality during *in vitro* differentiation both contain either an extra chromosome 11 (WW103-8E6) or two partial duplication chromosome 11s (WW103-13D10). As discussed in the previous chapter, some clones can undergo a G2 *trans* recombination event and the resulting duplication chromosome can become homozygous by induced mitotic recombination. These homozygous duplication clones have as many as four copies of all the genes in the chromosomal region between the gene-trap locus and the end point targeting locus (*E_2DH*, 100.7 Mb). For WW103-13D10 (*LOC217071*, 88.7Mb), the duplication region is 12 Mb. It is reasonable to expect that such a big chromosomal rearrangement will cause an abnormality in differentiation. The WW103-8E6 clone has accumulated an extra chromosome before regional trapping. The subsequent mitotic recombination has duplicated the inversion chromosome, but a wild-type chromosome with the end point targeting cassette is still present. So the phenotypes of these clones with alternative recombination events are not related to the gene-trap loci, and are caused by the duplication of a part of or the whole chromosome.

5.3.3 *Acly* deficiency and the impaired differentiation potential

Acly is an important enzyme involved in fatty acid biosynthesis. Its product, acetyl-CoA, is the key building block for *de novo* lipogenesis (Beigneux, Kosinski et al. 2004). There are at least three principal sources of acetyl-CoA: 1) amino acid degradation produces cytosolic acetyl-CoA, 2) fatty acid oxidation produces mitochondrial acetyl-CoA, 3) Glycolysis produces pyruvate, which is converted to mitochondrial acetyl-CoA by pyruvate dehydrogenase (Garrett and Grisham 1999). The acetyl-CoA from amino acid degradation is not sufficient for fatty acid biosynthesis, and the acetyl-CoA produced by fatty acid oxidation and by pyruvate dehydrogenase can not cross the mitochondrial membrane. So cytosolic acetyl-CoA is mainly generated from citrate which is transported from the mitochondria to the

cytosol. ATP-citrate lyase converts the citrate to acetyl-CoA and oxaloacetate. Acetyl-CoA provides the substrate for cytosolic fatty acid synthesis, while the oxaloacetate is converted to malate which is transported back into the mitochondria where it can be converted back into citrate (Fig. 5-15).

5.3.3.1 *Acly* deficiency in the mouse

To investigate the phenotype of *Acly* deficiency in the mouse, an *Acly* knockout has been examined. This mouse line was generated from the Bay Genomics gene-trap resource (Stryke, Kawamoto et al. 2003). In this clone, a β-galactosidase marker is expressed from *Acly* regulatory sequences. Beigneux *et al*. (2004) have found that *Acly* is required for embryonic development, because no viable homozygous embryos were identified after 8.5 dpc. The early embryonic lethality suggested that the alternative pathways to produce acetyl-CoA in the cytosol are not sufficient to support development in the absence of *Acly* during development (Beigneux, Kosinski et al. 2004).

Northern and Western analysis of *Acly* mRNA and protein showed that in all the tissues examined (liver, heart, kidney, brain, and white adipose tissue), heterozygous mice expressed half of the normal amount of *Acly* mRNA and protein. But the heterozygous mice were healthy, fertile, and normolipidemic on both normal and high fat diets. The expression of another acetyl-CoA enzyme, Acetyl-CoA synthetase 1, was not up-regulated. Thus it seems that *Acly* is synthesized in adequate quantities and half-normal amount of the enzyme is enough for providing sufficient acetyl-CoA (Beigneux, Kosinski et al. 2004).

One interesting finding is that *Acly* is expressed at high levels in the neural tube at 8.5 dpc. The fact that *Acly* is not expressed in other foetal tissues suggests that *Acly* might not function as a house-keeping gene during development. Otherwise, widespread expression of *Acly* will be detected in all the cell lineages. Instead, it might have a tissue-specific function in embryogenesis, apart from producing Acetyl-CoA for lipogenesis.

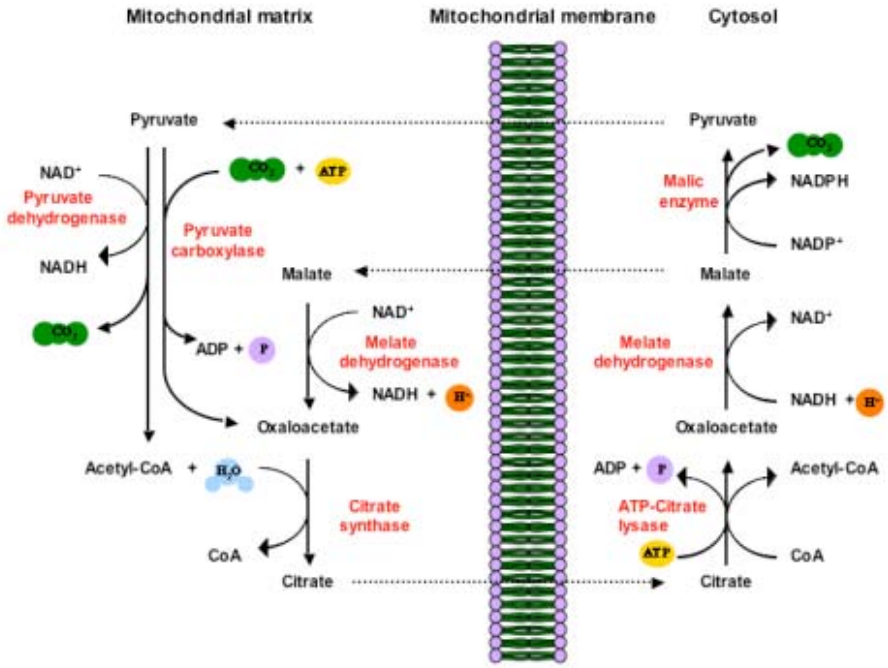


Fig. 5-15 The function of ATP-citrate lyase gene. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This is the first step for the *de novo* biosynthesis of sterol and fatty acid.

5.3.3.2 *Acly* and cell differentiation during sexual development

Acly has been shown to be involved in the sexual development of the fungus *Sordaria macrospora*. The fruiting body formation of filamentous ascomycetes involves the formation of the outer structures, as well as the development of mature ascospores within the fruiting body itself. Since this process requires the differentiation of several specialized tissues and some dramatic morphological and physiological changes, fruiting body maturation has been used as a model system to study multicellular development in eukaryotes (Nowrousian, Masloff et al. 1999).

Norwrousian *et al*. (1999) has used UV mutagenesis to screen for mutants with defects in fruiting body formation. One of the sterile mutants, per5, showed normal vegetative growth. But the fruiting body neck of the mutant strain was much shorter than that of the wild-type control. Most importantly, the fruiting bodies of the mutant strain only contain immature asci with no ascospores. DAPI staining showed that the immature asci still have eight nuclei within them, which suggests that there is no impairment in karyogamy or meiotic and postmeiotic divisions (Nowrousian, Masloff et al. 1999).

An indexed cosmid library was used to rescue the phenotype. A single complementing cosmid was isolated and sequence analysis has identified an ORF which has significant homology with higher eukaryotic *Acly*s. Analysis of the mutant *Acly* gene has identified a single nucleotide exchange (T to A), which altered a codon for aspartic acid into one for glutamic acid. The cloned mutant *Acly* gene can not rescue the phenotype of the mutant strain. Therefore, the mutation in *Acly* gene is responsible for the sterile phenotype (Nowrousian, Masloff et al. 1999).

As the mutant strain showed normal vegetative growth, it seems that sufficient acetyl-CoA is still produced for lipogenesis, either by residual *Acly* activity and/or expression of other acetyl-CoA-producing enzymes. But the attenuated *Acly* production can not satisfy the demand of acetyl-CoA during sexual development. So, the house-keeping functions of *Acly* can be circumvented to

a certain degree, but are essential under specific physiological conditions, such as sexual differentiation (Nowrousian, Masloff et al. 1999).

5.3.3.3 *Acly* as an *Brachyury* downstream notochord gene

Acly appears to be a downstream target of *Brachyury* in *Ciona intestinalis*. It is expressed specifically in the notochord in the embryogenesis process in *Ciona intestinalis*. The notochord has two major functions during chordate embryogenesis, providing inductive signals for the patterning of the neural tube and paraxial mesoderm and supporting the larval tail. The *Brachyury* gene encodes a transcription factor which contains a T DNA-binding domain. In vertebrates, *Brachyury* is first expressed in the presumptive mesoderm, and its expression is gradually restricted to the developing notochord and tailbud. *Brachyury* is believed to be one of the determinants for posterior mesoderm formation and notochord differentiation (Hotta, Takahashi et al. 2000).

By expressing the *Ciona intestinalis Brachyury* gene, *Ci-Bra*, in endoderm cells, Hotta *et al*. (1999) have isolated cDNA clones for 501 independent genes that were activated by *Ci-Bra* mis- and/or overexpression. By *in situ* hybridization, nearly 40 genes were found to be specifically or predominantly expressed in notochord, and therefore suggested to be *Brachyury*-downstream genes involved in notochord formation and function (Hotta, Takahashi et al. 1999). One of these genes, *Ci-Acl*, was found to share a high degree of homology with human ATP-citrate lyase (*ACLY*). The expression of this gene was first detected at the neural plate stage by *in situ* hybridization and its expression is restricted to notochord cells (Hotta, Takahashi et al. 2000). The fact that *Ci-Acl* only begins to express at the neural plate stage suggests that this gene might not be the immediate or direct target of *Ci-Bra*. Instead, it might be regulated by transcription factors, which in turn are regulated by *Ci-Bra* (Hotta, Takahashi et al. 2000).

It is interesting to notice that during embryogenesis of *Ciona intestinalis*, the expression of *Acly* is also highly restricted, similar to its expression pattern in murine embryogenesis. Considering the function of notochord in the

patterning of neural tube and paraxial mesoderm, it is likely that in both organisms, *Acly* plays some roles in neural tube and mesoderm differentiation.

5.3.3.4 Radicicol binds and inhibits mammalian *Acly*

Radicicol was first isolated from *Monosporium bonorden* as an antifungal antibiotic. But recently, this chemical was found to be able to reverse the transformed phenotype in *src*, *ras*, *mos*, *raf*, *fos*, and SV40-transformed cell lines. It can also cause cell cycle arrest and inhibit *in vivo* angiogenesis. So radicicol and its derivative are considered to be potential anti-cancer drugs (Ki, Ishigami et al. 2000).

To identify the *in vivo* target molecule of radicicol, Ki *et al*. (2000) used an affinity matrix to isolate radicicol-binding protein. Radicicol was biotinylated at various positions, and these variant compounds were then tested for their activity of morphological reversion of *src*-transformed phenotype. Two of the compounds, BR-1 and BR-6 were found to retain the activity. BR-6 was found to bind a 90-kDa protein, which was identified to be *Hsp90* by immunoblotting. BR-1 was shown to bind another 120-kDa protein, whose internal amino acid sequence was identical to human and rat ATP-citrate lyase. The identity of this 120-kDa protein was then confirmed by immunoblotting. Kinetic analysis showed that the activity of rat ATP-citrate lyase was inhibited by radicicol and BR-1, but not by BR-6. Radicicol was also found to be a non-competitive inhibitor of ATP-citrate lyase (Ki, Ishigami et al. 2000).

The fact that two radicicol derivatives, BR-1 and BR-6, bind two different proteins *in vivo* suggests that radicicol can bind different targets through different portions of its molecular structure. But the K_i value for ATP-citrate lyase was higher than the effective concentration of radicicol to reverse the transformed phenotype in *src*-tranformed cells, which suggests that this enzyme might not be directly involved in this process (Ki, Ishigami et al. 2000). Ki *et al*. (2000) hypothesized that BR-1 might be not stable and may be cleaved *in vivo* to generate free radicicol (Ki, Ishigami et al. 2000). But it could also be possible that radicicol is modified or cleaved *in vivo* to generate more

potent molecules to inhibit the enzyme activity of ATP-citrate lyase. Thus the phenotypes of radicicol, especially the ability to reverse the transformed phenotype in the cancer cell lines, might be partially associated with its binding and subsequent inhibition of *Acly* protein.

5.3.3.5 *Acly* is an important component of cell growth and transformation

Stable knockdown of *Acly* leads to impaired glucose-dependent lipid synthesis and also impaired *Akt*-mediated tumorigenesis (Bauer, Hatzivassiliou et al. 2005). Mammalian cells can not autonomously utilize the environmental nutrients to sustain their growth. Instead, constant extracellular signalling is needed to regulate the cellular metabolism of nutrients. However, cancer cells gain the autonomous ability to utilize nutrients by constitutively activating the normal signalling pathway without extracellular signals.

PI3K/*Akt* signalling pathway is critical for the cytokine-stimulated glucose metabolism, and its constitutive activation is commonly observed in cancer cells. In mammalian cells, glucose can either be oxidized to generate bioenergy, or be converted into other macromolecules to support biosynthesis. PI3K/*Akt* pathway can regulate the conversion of glucose to lipid and thus is essential for channeling the glucose into biosynthesis pathways. *Acly* is the main enzyme for producing cytosolic Acetyl-CoA for lipogenesis, and it is phosphorylated by *Akt in vivo* (Berwick, Hers et al. 2002). So it is possible that *Akt*-dependent cell transformation depends on *Acly* for *de novo* lipogenesis.

Bauer et al. (2005) used a shRNA construct to stably knock down the expression of *Acly* in a *Akt*-transformed cell line, FL5.12. *Akt*-expressing cells with or without *Acly* knockdown were injected into nude mice intravenously, and the mice were monitored for Akt-dependent leukemogenesis. Mice administrated *Acly* knockdown cells exhibited a significant delay or even a complete resistance to leukemogenesis.

5.3.3.6 A possible explanation of the phenotype of *Acly* deficient ES cells

Our *in vitro* differentiation results and works published before (Hotta, Takahashi et al. 1999; Nowrousian, Masloff et al. 1999; Hotta, Takahashi et al. 2000; Ki, Ishigami et al. 2000; Bauer, Hatzivassiliou et al. 2005) all suggested a pivotal function of *Acly* in cell differentiation and transformation. The only known function of *Acly in vivo* is to generate acetyl-CoA by the ATP-driven conversion of citrate and CoA into oxaloacetate and acetyl-CoA. This serves as the first step for the *de novo* biosynthesis of sterol and fatty acid. So the housekeeping function of the gene should be important for cell survival. But our observation and other published works (Nowrousian, Masloff et al. 1999; Beigneux, Kosinski et al. 2004) suggested that the house-keeping function of this gene can be circumvented to some degree either by the residual *Acly* activity or other alternative acetyl-CoA producing pathways.

In this study, there is no apparent difference in the growth rate, colony formation ability or ES cell/colony morphology in *Acly*-deficient ES cells compared to the wild-type control (data not shown). Microarray analysis using RNA extracted from undifferentiated WW103-18F11 and WW93-A12 ES cells showed that the expression levels of most mouse genes are similar (this work is still ongoing), which suggests that *Acly* deficiency does not cause observable phenotype in ES cells and the acetyl-CoA production in *Acly*-deficient ES cells seems to be sufficient to sustain the normal growth and division of ES cells.

However, when the *Acly*-deficient ES cells were differentiated *in vitro*, they could not form the typical three-dimensional cystic structures. In addition, the expression of some germ layer and cell type specific markers had changed. The RT-PCR results suggested that most cells in the cell aggregates were still undifferentiated ES cells. It is possible that the transition from the normal ES cell growth/division to the drastic re-programming and cell fate determination in the differentiation process demands higher than normal amounts of acetyl-CoA. A similar situation might accompany the transition from vegetative to sexual development in *S. macrospora* (Nowrousian, Masloff et al. 1999). It is

possible that this energetic demand can not be fulfilled by the alternative metabolic pathways, which might be partly due to different metabolic costs of lipogenesis.

Another possible explanation is that in order for some ES cells in an EB to be differentiated into a certain cell type, these cells must gain “competence” before the differentiation process is induced. The differentiation competence might involve as one component a threshold in acetyl-CoA concentration, which might be much higher than the level that is necessary for the ES cell growth and division.

The exact mechanism by which the acetyl-CoA production can influence the potential of ES cells to differentiate *in vitro* is unknown. Acetyl-CoA can be used to produce fatty acids, sterols and other important molecules which need the acetyl base, such as acetylcholine (Beigneux, Kosinski et al. 2004). Therefore, ATP-citrate lyase may either control the overall cytosolic acetyl-CoA concentration to indirectly regulate the pathways that need acetyl-CoA, or it could directly interact with various acetyltransferases or lipid/sterol synthetases to form an enzyme complex to provide acetyl-CoA. Nevertheless, *Acly* seems to play an important role in development and differentiation of certain cell types.

The difficulty to determine the primary locus of action of *Acly* make it hard to link this gene directly with any known genetic pathways controlling ES cell differentiation. It is not unexpected for mutants identified by such a genetic screen. However, if more homozygous ES cell mutants are generated in the future and screened using the same strategy, it will be possible to group the mutants by their apparent defects and study the relationships between the mutants with similar phenotypes. The importance of a genetic screen is that it can not only fill in the gaps in a known pathway, but also identify new pathways that are not necessarily overlapping with the known ones.

5.3.3.7 Future experiments to identify the function of *Acly* in ES cell *in vitro* differentiation

The RT-PCR results detected transcription of *Acly* downstream exons in the mutant line. Since *Acly* is a large gene (51.54 kb), it is possible that there are other alternative transcription start points. The proviral insertion and the subsequent inversion might not completely block all the *Acly* transcripts, so the mutation generated in the homozygous mutant cell line might not a null allele. To resolve this, a homozygous *Acly* gene targeted ES cell line can be constructed and these ES cells can be differentiated to confirm the function of the gene in ES cell *in vitro* differentiation.

To investigate the *in vivo* differentiation potential of the WW103-18F11 ES cell line, 1X 10^7 undifferentiated WW103-18F11 and control WW93A12 ES cells were injected subcutaneously into both flanks of 8-week old F1 hybrid mice (129 S7/SvEv$^{\text{Brd-}Hprtb\text{-}m2}$ X C57$^{\text{TyrBrdC1}}$ female). The animals were examined periodically over 4 weeks for the appearance and growth of tumours. 4 weeks after injection of ES cells, the mice were sacrificed, and the size of each tumor was measured after dissection. Tumor samples were cut into two halves, one half was fixed in 10% formalin for histopathological analysis, and the other half was dissected into several pieces (depending on its size) and snap-frozen in liquid nitrogen for subsequent RNA and DNA extraction.

For the WW93-A12 ES cell line, tumours were found at every ES cell injection site (8/8). Though the size of the tumours varied, all the tumours collected were dark red and highly vascular. When the tumours were bisected, a fluid filled central cavity was always found in the centre of the tumour. In contrast, for the WW103-18F11 ES cell line, only 3 tumours were found 4 weeks after the injection (3/8). All three tumours were very small and pale. No blood vessels were found on their surface. When the tumours were bisected, no fluid filled cavities were present.

Histopathology results of three tumours generated from WW103-18F11 cells and eight cases of WW93-A12 teratocarcinomas have confirmed that the differentiation potential of WW103-18F11 clones were greatly impaired

(pathology analysis was performed by Dr. Madhuri Warren). The WW103-18F11 tumours are circumscribed mixed ganglier/neuroepithelial tumours plus embryonal carcinoma composed predominantly of nests of mature glial cells with scarce neuroepithelial differentiation in the form of Homer-Wright rosettes. Nests of undifferentiated embryonal carcinoma (ES cells) are also seen. There was no evidence of differentiation into other germ cell lineages.

All the WW93-A12 tumour are circumscribed immature teratocarcinomas composed predominantly of immature glial tissue and tissues from all three germ layers: simple cuboidal epithelium, columnar epithelium, ciliated respiratory type epithelium, mucin secreting gastrointestinal epithelium; cartilage, osteoid, immature neuroepithelium, smooth muscle; and stratified squamous epithelium. In some samples, nests of immature embryonal carcinoma (undifferentiated ES cells) and isolated syncytiotrophoblast cells were also found.

We have also injected the BAC rescued ES cells into the F1 hybrid mice and are now waiting for the pathology results of the teratocarcinomas generated by subcutaneous injection. For the rescued cell line, tumours were found at every ES cell injection site (8/8). All of the tumours were dark red and highly vascular. Some of these tumours have a fluid filled central cavity in the centre of the tumour.

From the initial result, we can conclude that the differentiation potential of the *Acly*-deficient ES cells is also impaired *in vivo*. But complete pathology results of the tumours derived from the rescued cells are needed to confirm that the *in vivo* differentiation potential is fully recovered in these cells.

Because the phenotype of the *Acly*-deficient ES cells may depend on some mutations or silencing in a second gene, inactivation of *Acly* in an independent ES cell clone is necessary to prove that the *Acly* gene is solely responsible for the differentiation defect we observed in WW103-18F11 deficient cell line. Although BAC rescue experiment can make a causal link between the mutation in *Acly* gene and the defective phenotypes, it is

possible that other genes or transcriptional elements also play some roles in the differentiation defects. If over-expression of *Acly* cDNA can also rescue the defective phenotypes, it will effectively exclude the involvement of other genes or transcriptional elements.

5.3.4 Summary

In this chapter, I have described the strategy used to screen for an *in vitro* differentiation phenotype in homozygous mutant ES cell lines. Restricted by the detection method, I checked the expression of a limited number of markers in the differentiation process. In spite of this limitation, I successfully identified several clones with a reproducible *in vitro* phenotype. By Southern analysis and sib-selection using different drugs, I found some of these clones are the products of alternative recombination events. But two of the clones, WW103-14F11 and WW103-18F11, are products of regional trapping and subsequent inversion. Detailed expression analysis and functional studies have been carried out on WW103-18F11. The impaired *in vitro* differentiation potential observed in this clones was caused by the disruption of the ATP-Citrate lyase (*Acly*) gene. Therefore, this strategy has proved to be able to identify *in vitro* differentiation mutants and facilitate regional screens for genes involved in the early embryogenesis in the mouse genome.

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.

References

(1998). "Genome sequence of the nematode C. elegans: a platform for investigating biology." Science **282**(5396): 2012-8.
Abuin, A. and A. Bradley (1996). "Recycling selectable markers in mouse embryonic stem cells." Mol Cell Biol **16**(4): 1851-6.
Adams, M. D., S. E. Celniker, et al. (2000). "The genome sequence of Drosophila melanogaster." Science **287**(5461): 2185-95.
Ankeny, R. A. (2001). "The natural history of Caenorhabditis elegans research." Nat Rev Genet **2**(6): 474-9.
Araki, K., M. Araki, et al. (1995). "Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase." Proc Natl Acad Sci U S A **92**(1): 160-4.
Austin, C. P., J. F. Battey, et al. (2004). "The knockout mouse project." Nat Genet **36**(9): 921-4.
Auwerx, J., P. Avner, et al. (2004). "The European dimension for the mouse genome mutagenesis program." Nat Genet **36**(9): 925-7.
Bauer, D. E., G. Hatzivassiliou, et al. (2005). "ATP citrate lyase is an important component of cell growth and transformation." Oncogene.
Beigneux, A. P., C. Kosinski, et al. (2004). "ATP-citrate lyase deficiency in the mouse." J Biol Chem **279**(10): 9557-64.
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.
Berwick, D. C., I. Hers, et al. (2002). "The identification of ATP-citrate lyase as a protein kinase B (Akt) substrate in primary adipocytes." J Biol Chem **277**(37): 33895-900.
Bessereau, J. L., A. Wright, et al. (2001). "Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line." Nature **413**(6851): 70-4.
Bonaldo, P., K. Chowdhury, et al. (1998). "Efficient gene trap screening for novel developmental genes using IRES beta geo vector and in vitro preselection." Exp Cell Res **244**(1): 125-36.
Bradley, A., M. Evans, et al. (1984). "Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines." Nature **309**(5965): 255-6.
Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.
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.
Carthew, R. W. (2001). "Gene silencing by double-stranded RNA." Curr Opin Cell Biol **13**(2): 244-8.
Cattanach, B. M. and J. Jones (1994). "Genetic imprinting in the mouse: implications for gene regulation." J Inherit Metab Dis **17**(4): 403-20.
Chambers, I., D. Colby, et al. (2003). "Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells." Cell **113**(5): 643-55.

Chambers, I. and A. Smith (2004). "Self-renewal of teratocarcinoma and embryonic stem cells." Oncogene **23**(43): 7150-60.
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.
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.
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.
Czyz, J., C. Wiese, et al. (2003). "Potential of embryonic and adult stem cells in vitro." Biol Chem **384**(10-11): 1391-409.
Doetschman, T. C., H. Eistetter, et al. (1985). "The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium." J Embryol Exp Morphol **87**: 27-45.
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.
Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-8.
Evans, M. J. and M. H. Kaufman (1981). "Establishment in culture of pluripotential cells from mouse embryos." Nature **292**(5819): 154-6.
Farley, F. W., P. Soriano, et al. (2000). "Widespread recombinase expression using FLPeR (flipper) mice." Genesis **28**(3-4): 106-10.
Forsburg, S. L. (2001). "The art and design of genetic screens: yeast." Nat Rev Genet **2**(9): 659-68.
Friedman, R. and A. L. Hughes (2001). "Pattern and timing of gene duplication in animal genomes." Genome Res **11**(11): 1842-7.
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.
Gajovic, S., K. Chowdhury, et al. (1998). "Genes expressed after retinoic acid-mediated differentiation of embryoid bodies are likely to be expressed during embryo development." Exp Cell Res **242**(1): 138-43.
Garrett, H. G. and C. M. Grisham (1999). "Biochemistry (Second edition)." **Chapter 25**: 802-805.
Goffeau, A., B. G. Barrell, et al. (1996). "Life with 6000 genes." Science **274**(5287): 546, 563-7.
Goldstein, J. L. (2001). "Laskers for 2001: knockout mice and test-tube babies." Nat Med **7**(10): 1079-80.
Golic, K. G. (1991). "Site-specific recombination between homologous chromosomes in Drosophila." Science **252**(5008): 958-61.
Golic, K. G. and S. Lindquist (1989). "The FLP recombinase of yeast catalyzes site-specific recombination in the Drosophila genome." Cell **59**(3): 499-509.

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.
Greber, B., H. Lehrach, et al. (2005). "Mouse splice mutant generation from ENU-treated ES cells--a gene-driven approach." Genomics **85**(5): 557-62.
Grignani, F., T. Kinsella, et al. (1998). "High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein." Cancer Res **58**(1): 14-9.
Grimm, S. (2004). "The art and design of genetic screens: mammalian culture cells." Nat Rev Genet **5**(3): 179-89.
Gu, H., Y. R. Zou, et al. (1993). "Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting." Cell **73**(6): 1155-64.
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.
Hamilton, D. L. and K. Abremski (1984). "Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites." J Mol Biol **178**(2): 481-6.
Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.
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.
Hartwell, L. H., J. Culotti, et al. (1970). "Genetic control of the cell-division cycle in yeast. I. Detection of mutants." Proc Natl Acad Sci U S A **66**(2): 352-9.
Hotta, K., H. Takahashi, et al. (2000). "Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo." Dev Biol **224**(1): 69-80.
Hotta, K., H. Takahashi, et al. (1999). "Temporal expression patterns of 39 Brachyury-downstream genes associated with notochord formation in the Ciona intestinalis embryo." Dev Growth Differ **41**(6): 657-64.
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.
Humpherys, D., K. Eggan, et al. (2001). "Epigenetic instability in ES cells and cloned mice." Science **293**(5527): 95-7.
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.
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.
Jansen, G., E. Hazendonk, et al. (1997). "Reverse genetics by chemical mutagenesis in Caenorhabditis elegans." Nat Genet **17**(1): 119-21.

Jorgensen, E. M. and S. E. Mango (2002). "The art and design of genetic screens: caenorhabditis elegans." Nat Rev Genet **3**(5): 356-69.
Justice, M. J. (2000). "Capitalizing on large-scale mouse mutagenesis screens." Nat Rev Genet **1**(2): 109-15.
Kamath, R. S., A. G. Fraser, et al. (2003). "Systematic functional analysis of the Caenorhabditis elegans genome using RNAi." Nature **421**(6920): 231-7.
Kawasaki, H., K. Mizuseki, et al. (2000). "Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity." Neuron **28**(1): 31-40.
Ki, S. W., K. Ishigami, et al. (2000). "Radicicol binds and inhibits mammalian ATP citrate lyase." J Biol Chem **275**(50): 39231-6.
Kile, B. T., K. E. Hentges, et al. (2003). "Functional genetic analysis of mouse chromosome 11." Nature **425**(6953): 81-6.
Koller, B. H., P. Marrack, et al. (1990). "Normal development of mice deficient in beta 2M, MHC class I proteins, and CD8+ T cells." Science **248**(4960): 1227-30.
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.
Kumar, A. and M. Snyder (2001). "Emerging technologies in yeast genomics." Nat Rev Genet **2**(4): 302-12.
Lakso, M., B. Sauer, et al. (1992). "Targeted oncogene activation by site-specific recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(14): 6232-6.
Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Leahy, A., J. W. Xiong, et al. (1999). "Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation." J Exp Zool **284**(1): 67-81.
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.
Li, J., H. Shen, et al. (1999). "Leukaemia disease genes: large-scale cloning and pathway predictions." Nat Genet **23**(3): 348-53.
Lindsay, E. A., F. Vitelli, et al. (2001). "Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice." Nature **410**(6824): 97-101.
Lindsley, D. L., L. Sandler, et al. (1972). "Segmental aneuploidy and the genetic gross structure of the Drosophila genome." Genetics **71**(1): 157-84.
Liu, L. X., J. M. Spoerke, et al. (1999). "High-throughput isolation of Caenorhabditis elegans deletion mutants." Genome Res **9**(9): 859-67.
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.

Liu, P., H. Zhang, et al. (1998). "Embryonic lethality and tumorigenesis caused by segmental aneuploidy on mouse chromosome 11." Genetics **150**(3): 1155-68.

Liu, X., H. Wu, et al. (1997). "Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission." Dev Dyn **209**(1): 85-91.

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.

Maine, E. M. (2001). "RNAi As a tool for understanding germline development in Caenorhabditis elegans: uses and cautions." Dev Biol **239**(2): 177-89.

Mann, J. R., I. Gadi, et al. (1990). "Androgenetic mouse embryonic stem cells are pluripotent and cause skeletal defects in chimeras: implications for genetic imprinting." Cell **62**(2): 251-60.

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.

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.

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.

Mitchell, K. J., K. I. Pinson, et al. (2001). "Functional analysis of secreted and transmembrane proteins critical to mouse development." Nat Genet **28**(3): 241-9.

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.

Munroe, R. J., S. L. Ackerman, et al. (2004). "Genomewide two-generation screens for recessive mutations by ES cell mutagenesis." Mamm Genome **15**(12): 960-5.

Munroe, R. J., R. A. Bergstrom, et al. (2000). "Mouse mutants from chemically mutagenized embryonic stem cells." Nat Genet **24**(3): 318-21.

Nagy, A. (2000). "Cre recombinase: the universal reagent for genome tailoring." Genesis **26**(2): 99-109.

Nagy, A., C. Moens, et al. (1998). "Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles." Curr Biol **8**(11): 661-4.

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.

Nakano, T., H. Kodama, et al. (1994). "Generation of lymphohematopoietic cells from embryonic stem cells in culture." Science **265**(5175): 1098-101.

Nichols, J., E. P. Evans, et al. (1990). "Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity." Development **110**(4): 1341-8.

Ning, Z., A. J. Cox, et al. (2001). "SSAHA: a fast search method for large DNA databases." Genome Res **11**(10): 1725-9.

Nishikawa, S. I., S. Nishikawa, et al. (1998). "Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages." Development **125**(9): 1747-57.

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.

Nowrousian, M., S. Masloff, et al. (1999). "Cell differentiation during sexual development of the fungus Sordaria macrospora requires ATP citrate lyase activity." Mol Cell Biol **19**(1): 450-60.

Nurse, P. (1975). "Genetic control of cell size at cell division in yeast." Nature **256**(5518): 547-51.

Nusslein-Volhard, C. and E. Wieschaus (1980). "Mutations affecting segment number and polarity in Drosophila." Nature **287**(5785): 795-801.

Orban, P. C., D. Chui, et al. (1992). "Tissue- and site-specific DNA recombination in transgenic mice." Proc Natl Acad Sci U S A **89**(15): 6861-5.

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.

Ramalho-Santos, M., S. Yoon, et al. (2002). ""Stemness": transcriptional profiling of embryonic and adult stem cells." Science **298**(5593): 597-600.

Ramirez-Solis, R., A. C. Davis, et al. (1993). "Gene targeting in embryonic stem cells." Methods Enzymol **225**: 855-78.

Ramirez-Solis, R., P. Liu, et al. (1995). "Chromosome engineering in mice." Nature **378**(6558): 720-4.

Rohwedel, J., K. Guan, et al. (2001). "Embryonic stem cells as an in vitro model for mutagenicity, cytotoxicity and embryotoxicity studies: present state and future prospects." Toxicol In Vitro **15**(6): 741-53.

Rorth, P. (1996). "A modular misexpression screen in Drosophila detecting tissue-specific phenotypes." Proc Natl Acad Sci U S A **93**(22): 12418-22.

Rorth, P., K. Szabo, et al. (1998). "Systematic gain-of-function genetics in Drosophila." Development **125**(6): 1049-57.

Rossant, J. and A. Spence (1998). "Chimeras and mosaics in mouse mutant analysis." Trends Genet **14**(9): 358-63.

Rubin, G. M. and E. B. Lewis (2000). "A brief history of Drosophila's contributions to genome research." Science **287**(5461): 2216-8.

Sauer, B. and N. Henderson (1988). "Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1." Proc Natl Acad Sci U S A **85**(14): 5166-70.

Schwartzberg, P. L., E. J. Robertson, et al. (1990). "Targeted gene disruption of the endogenous c-abl locus by homologous recombination with DNA encoding a selectable fusion protein." Proc Natl Acad Sci U S A **87**(8): 3210-4.

Sharov, A. A., Y. Piao, et al. (2003). "Transcriptome analysis of mouse stem cells and early embryos." PLoS Biol **1**(3): E74.
Shuman, H. A. and T. J. Silhavy (2003). "The art and design of genetic screens: Escherichia coli." Nat Rev Genet **4**(6): 419-31.
Simon, M. A. (1994). "Signal transduction during the development of the Drosophila R7 photoreceptor." Dev Biol **166**(2): 431-42.
Skarnes, W. C., J. E. Moss, et al. (1995). "Capturing genes encoding membrane and secreted proteins important for mouse development." Proc Natl Acad Sci U S A **92**(14): 6592-6.
Skarnes, W. C., H. von Melchner, et al. (2004). "A public gene trap resource for mouse functional genomics." Nat Genet **36**(6): 543-4.
Smith, A. J., M. A. De Sousa, et al. (1995). "A site-directed chromosomal translocation induced in embryonic stem cells by Cre-loxP recombination." Nat Genet **9**(4): 376-85.
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.
Stoykova, A., K. Chowdhury, et al. (1998). "Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system." Dev Dyn **212**(2): 198-213.
Stryke, D., M. Kawamoto, et al. (2003). "BayGenomics: a resource of insertional mutations in mouse embryonic stem cells." Nucleic Acids Res **31**(1): 278-81.
Su, H., A. A. Mills, et al. (2002). "A targeted X-linked CMV-Cre line." Genesis **32**(2): 187-8.
Su, H., X. Wang, et al. (2000). "Nested chromosomal deletions induced with retroviral vectors in mice." Nat Genet **24**(1): 92-5.
Sulston, J. E. (1976). "Post-embryonic development in the ventral cord of Caenorhabditis elegans." Philos Trans R Soc Lond B Biol Sci **275**(938): 287-97.
Sulston, J. E. and H. R. Horvitz (1977). "Post-embryonic cell lineages of the nematode, Caenorhabditis elegans." Dev Biol **56**(1): 110-56.
Sulston, J. E., E. Schierenberg, et al. (1983). "The embryonic cell lineage of the nematode Caenorhabditis elegans." Dev Biol **100**(1): 64-119.
Suzuki, T., H. Shen, et al. (2002). "New genes involved in cancer identified by retroviral tagging." Nat Genet **32**(1): 166-74.
Tada, S., T. Era, et al. (2005). "Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture." Development **132**(19): 4363-74.
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.
Tessier, C. R., G. A. Doyle, et al. (2004). "Mammary tumor induction in transgenic mice expressing an RNA-binding protein." Cancer Res **64**(1): 209-14.
Theodosiou, N. A. and T. Xu (1998). "Use of FLP/FRT system to study Drosophila development." Methods **14**(4): 355-65.

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.

Tsien, J. Z., D. F. Chen, et al. (1996). "Subregion- and cell type-restricted gene knockout in mouse brain." Cell **87**(7): 1317-26.

Venter, J. C., M. D. Adams, et al. (2001). "The sequence of the human genome." Science **291**(5507): 1304-51.

Vittet, D., M. H. Prandini, et al. (1996). "Embryonic stem cells differentiate in vitro to endothelial cells through successive maturation steps." Blood **88**(9): 3424-31.

Vivian, J. L., Y. Chen, et al. (2002). "An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells." Proc Natl Acad Sci U S A **99**(24): 15542-7.

von Melchner, H. and H. E. Ruley (1989). "Identification of cellular promoters by using a retrovirus promoter trap." J Virol **63**(8): 3227-33.

Waterston, R. H., K. Lindblad-Toh, et al. (2002). "Initial sequencing and comparative analysis of the mouse genome." Nature **420**(6915): 520-62.

Wiles, M. V. and G. Keller (1991). "Multiple hematopoietic lineages develop from embryonic stem (ES) cells in culture." Development **111**(2): 259-67.

Wobus, A. M. (2001). "Potential of embryonic stem cells." Mol Aspects Med **22**(3): 149-64.

Wobus, A. M., K. Guan, et al. (2002). "Embryonic stem cells as a model to study cardiac, skeletal muscle, and vascular smooth muscle cell differentiation." Methods Mol Biol **185**: 127-56.

Wobus, A. M., G. Wallukat, et al. (1991). "Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers." Differentiation **48**(3): 173-82.

Wood, V., R. Gwilliam, et al. (2002). "The genome sequence of Schizosaccharomyces pombe." Nature **415**(6874): 871-80.

Xu, T. and S. D. Harrison (1994). "Mosaic analysis using FLP recombinase." Methods Cell Biol **44**: 655-81.

Xu, T. and G. M. Rubin (1993). "Analysis of genetic mosaics in developing and adult Drosophila tissues." Development **117**(4): 1223-37.

Xu, T., W. Wang, et al. (1995). "Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase." Development **121**(4): 1053-63.

Yamashita, J., H. Itoh, et al. (2000). "Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors." Nature **408**(6808): 92-6.

Ying, Q. L., M. Stavridis, et al. (2003). "Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture." Nat Biotechnol **21**(2): 183-6.

Yu, Y. and A. Bradley (2001). "Engineering chromosomal rearrangements in mice." Nat Rev Genet **2**(10): 780-90.

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.

Zambrowicz, B. P., A. Abuin, et al. (2003). "Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention." Proc Natl Acad Sci U S A **100**(24): 14109-14.

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.

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.

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.