CHAPTER 1:
INTRODUCTION

1.1 Mouse as a genetic tool

1.1.1 A brief history

The inbred laboratory mouse has become an indispensable tool in modern biological and medical research. Among mammals, the mouse is ideally suited for genetic analysis because of its small size (average adult weight of 25-40 g) which allows mice to be housed in high density, its short generation time (about 10 weeks from being born to giving birth) and prolificacy in breeding (5-10 pups per litter and an immediate postpartum estrus). In addition, the docile nature of the mouse makes handling easy and the deposition of a vaginal plug upon mating with a male allows pregnancies to be timed accurately. William Ernest Castle was the pioneer mouse geneticist who carried out the first systematic analysis of Mendelian inheritance and genetic variation in mice at the Bussey Institute, Harvard University (Snell and Reed, 1993). One important milestone in the history of mouse genetics was the establishment of inbred mouse strains. An inbred strain is defined as one that has been maintained for more than 20 generations by brother-to-sister mating and is essentially homozygous at all genetic loci, except for mutations arising spontaneously. After 20 generations of inbreeding, approximately 98.7% of the genomic loci in each animal will be homozygous (Silver, 1985). Subsequently, further inbreeding will result in decreasing heterozygosity at the rate of 19.1% with each generation, and by the 40^{th} generation of inbreeding, essentially 99.98% of the genome will be homozygous. To date, all the commonly used inbred strains have been inbred for at least 60 generations; hence all siblings are essentially 100% identical. The inbred mouse strains have revolutionized studies in cancer research, tissue transplantation, and immunology by eliminating genetic variability and making it possible to compare and analyze data from different laboratories worldwide.

1.1.2 Using the mouse to model human disease

Similarities in both the anatomy and physiology between the mouse and human make the mouse the most popular model organism used to study basic biological

processes, development, immunology and behaviour. Sequencing and analysis of the mouse and human genomes have led to the identification of ~25,000 genes as well as hundreds of conserved non-coding regions. It was found that about 99% of mouse genes have at least one orthologue in the human genome (Lander et al., 2001; Venter et al., 2001; Waterston et al., 2002). Mice and humans diverged from a common ancestor about 65 million years ago. The sequenced mouse genome is about 14% smaller than the human genome (2.5×10^9 bases compared to 2.9×10^9 bases). Even though evolutionary divergence between the two genomes has occurred over time, comparative genomics analysis has identified syntenic regions between the human and mouse genomes (Waterston et al., 2002). As such, more than 90% of the human and mouse genomes can be clustered into segments of conserved synteny, indicating the high degrees of conservation of genomic organisation (Waterston et al., 2002). Thus the mouse represents a good model to study human gene function *in vivo*.

1.2 Methodologies to genetically manipulate the mouse genome

1.2.1 Pronuclear injections and transgenic mice

The development of many genetic and genomic tools has played a critical role in the widespread use of the mouse for biomedical research. In 1966, Teh Ping Lin first demonstrated that following the direct injection of macromolecules into the pronuclei of mouse zygotes, embryos developed and were viable, thus establishing the technical basis for generation of transgenic mice (Lin, 1966). Following this seminal discovery, several groups showed that pronuclear injection of DNA into mouse zygotes resulted in expression of the introduced 'transgenes' in the mouse; a process called transgenic mouse production (Brinster et al., 1981; Wagner et al., 1981a; Wagner et al., 1981b). In general, transgenic DNA injected into mouse pronuclei will concatamerize and integrate into the genome. Consequently, the integration sites of the transgene are random and this can be a problem because the insertion sites may drastically influence the expression of the transgene. Additionally, copy number variation can also influence the expression of the transgene. In some cases, transgenes are used for expression studies, where it is essential to include all the regulatory elements of the gene to faithfully recapitulate expression of the endogenous locus. It is estimated that in 5-10% of transgenic lines, homozygosity for

particular transgenes may cause developmental anomalies, including lethality. These recessive phenotypes are likely to be the result of disruption of vital genes at the insertion sites of the transgenes. In addition, pronuclear injection of DNA can lead to genomic DNA rearrangements which could complicate the identification of the mutant loci. Despite these drawbacks, pronuclear injection of DNA was one of the most significant breakthroughs in mouse genetics as it demonstrated that experimental modification of the animal germ line was possible and has provided fundamental insights into many biological processes.

1.2.2 Development of mouse embryonic stem (ES) cells

The most commonly used platform in genetic engineering came with the isolation of embryonic stem cells and the development of gene targeting. One of the landmarks in mouse genetics was the isolation of pluripotent mouse embryonic stem cells (ES) from mouse blastocysts (Evans and Kaufman, 1981; Martin, 1981). The concept of pluripotent ES cells was established when teratocarcinomas were found to contain cells (embryonic carcinoma, EC cells) that gave rise to new tumours (Solter et al., 1970). Teratomas contain a mixture of differentiated cell types derived from all three germ layers and these are present in the tumour mass in a disorganised manner. Malignant teratomas (teratocarcinomas) were shown to contain undifferentiated stem cells that formed tumours upon transplantation into secondary recipients (Solter et al., 1970). These teratocarcinoma stem cells formed secondary teratocarcinomas upon the transfer of individual undifferentiated cells, demonstrating their clonal potency (Kleinsmith and Pierce, 1964). When EC cells were injected into a blastocyst, these cells became incorporated into the embryo and contributed to the tissues of the developing fetus (Brinster, 1974). This indicates that the proliferation of undifferentiated EC cells can be controlled in response to appropriate *in vivo* cues. Importantly, EC cells can be cultured and maintained *in vitro* on a feeder layer of fibroblasts (Martin and Evans, 1975). Based on these initial observations, pluripotent mouse ES cells were isolated successfully from the inner cell mass of 3.5 days post-coitum (dpc) wild-type embryos (Evans and Kaufman, 1981; Martin, 1981). These ES cells were initially characterized as teratocarcinoma stem cell lines due to their similarities in morphology, culture conditions

and pluripotency. However, it became apparent that these ES cells were distinct from EC cells because they were more stable and could be controlled. The most important difference between ES cells and EC cells is that ES cells are euploid and are able to repopulate somatic tissues and also transmit through the mouse germ-line during embryogenesis (Bradley et al., 1984). A major advance in ES cell technology is that these cultured ES cells maintain their pluripotency even after modification of their genomes by the use of retroviral vectors (Robertson et al., 1986). Therefore, the desired modifications can be introduced into the ES cells before they are injected into mouse blastocysts. Subsequently, the development of homologous recombination technology permitted targeted mutagenesis in ES cells, thus facilitating the potential disruption of any gene in the mouse (Capecchi, 1989; Koller and Smithies, 1989; Smithies et al., 1985; Thomas et al., 1986). These recent technological advances have allowed us to study the consequences of individual gene loss to deduce the gene's function in the mouse and likely, its role in humans. These approaches, together with the transgenic technique of zygote injection, are classified as reverse genetics. With the development of ES cell technology and the discovery of homologous recombination, reverse genetics became the main approach to study gene function in the mouse and in subsequent years, numerous gene functions were elucidated and mouse models of human diseases were created, advancing the understanding of gene function and human diseases. In recognition of their pioneering contributions to the field of mouse genetics, Mario R. Capecchi, Sir Martin J. Evans and Oliver Smithies were awarded the Nobel Prize in Physiology or Medicine in 2007 'for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells'.

1.2.3 Homologous recombination and gene targeting

Homologous recombination is the ability of the cellular machinery to combine DNA fragments that share stretches of similar or identical sequences. Early studies, which focused on extrachromosomal recombination between transfected DNA molecules, showed that mammalian cells possess effective machinery for mediating the homologous recombination of introduced DNA fragments. Although the mechanism of homologous recombination is not fully understood, recombination between introduced DNA

molecules and homologous sequences in the chromosome can be exploited to precisely alter genes of interest in mouse ES cells. Using homologous recombination, Smithies *et al*. introduced an exogenous DNA sequence into the human chromosomal *β-globin* locus (Smithies et al., 1985) and shortly afterwards, Thomas *et al*. corrected a mutant neomycin resistance gene (*neo*) in the host genome with an injected *neo* gene fragment (Thomas et al., 1986). These key experiments paved the way for the current gene targeting technology, thus enabling the analysis of the effects of loss- or gain-of-function of a specific gene.

A myriad of sophisticated gene targeting strategies in ES cells have since been developed to manipulate the mouse genome (Bradley and Liu, 1996). In addition to the conventional use of ES cells to generate knockout mice, it can also be used for transgene expression analysis. Using ES cells as a genetic vehicle facilitated experiments that were not possible by pronuclear injections; for example where a dominant lethal phenotype is caused by transgene expression (Warren et al., 1994). Furthermore using gene knock-in strategies, Hanks *et al*. showed that the heterologous genes *engrailed 1/2* were capable of functional compensation by targeting one gene to the locus of its homologue and observing rescue of the knockout phenotype (Hanks et al., 1995). In the last two decades, gene targeting has been widely utilized in ES cells to make a variety of genetic mutations in many loci, allowing the phenotypic analyses of the genetic modifications to be determined. The combination of gene targeting and ES cell technology provides two vital components of the current mouse gene knockout technology.

1.2.4 Conventional gene knockout technology

The procedure for generating mice that have been genetically modified using gene targeting strategies is essentially the same regardless of the specific targeting strategy used (Figure 1.1). Firstly, the targeting construct which contains DNA fragments homologous to the targeted gene and a selectable marker is generated. The construct is then linearized and electroporated into wild-type ES cells which are cultured in the presence of a selection agent to select for transfectants which have stably integrated the construct into their genome. The surviving ES cell colonies are then isolated and examined for presence of the targeted allele using long range PCR or Southern blot

analysis to confirm that the desired recombined event has occurred. ES cell clones containing the correctly targeted allele are then expanded and injected into 3.5 dpc blastocysts and transplanted into the uteri of pseudopregnant surrogate females. The resulting pups are then examined for their degree of chimerism (percentage of genetic makeup of the mouse that was contributed by the ES cell). Male chimera showing high levels of chimerism are then mated with wild-type females to check for germ-line transmission of the targeted allele in the F1 offspring. The F1 heterozygotes can then be inter-crossed to breed to homozygosity.

Introducing a loss-of-function mutation is the most common method for determining the function of a gene. In order to ablate the function of a gene of interest (to generate a null allele), the most common experimental strategy is to replace all or part of the coding sequence of the gene with a selectable marker. There are two main types of targeting vectors for mutating a gene through homologous recombination in ES cells: (1) replacement vectors and (2) insertion vectors (Thomas and Capecchi, 1986). A replacement vector consists of a positive selection marker such as Neomycin (*Neo*), Puromycin (*Puro*) or Blasticidin (*Bsd*) resistance markers, which is flanked by isogenic DNA homologous arms (5-8 kb), and a linearization site outside of the homologous arms of the vector (Figure 1.2A). Usually, a negative selectable marker Herpes Simplex Virus thymidine kinase gene (*HSVtk*) is also used to enrich for correctly targeted events from random integrations. The *HSVtk* is inserted next to the homology arms. Random insertion events are likely to retain the functional *HSVtk* gene which would encode for thymidine kinase (Thomas and Capecchi, 1986). Negative selection is then performed by adding chemical compounds, such as gancyclovir or FIAU [1-2-deoxy-2-fluoro-β-D-arabinofuranosyl)-5-iodouracil]. These drugs are transformed into cytotoxic substances by the HSVtk. Thus, the cells that have the random integration events and thus harbour the *HSVtk* gene are eliminated. Those cells in which homologous recombination have occurred do not harbour the *HSVtk* gene and thus survive the selection. The final post-recombination product is the replacement of the targeted region with all the components of the vector. The basic elements of insertion vectors are a genomic DNA fragment and a selection marker. Unlike the replacement vector where the linearization site is outside the homology arms, the site of linearization of the insertion vector is within the homologous

DNA sequence. Homologous recombination between the insertion vector and its target locus results in insertion of the entire vector into the locus, leading to a duplication of the genomic sequences (Figure 1.2B).

Several parameters are known to affect the efficiency of the targeting events in ES cells. Firstly, the length of homologous sequence on the targeting vector; generally, the longer the homologous sequence, the higher the efficiency and typically, 7-8 kb in total (or 3-4 kb on each side of the replacement vector) is sufficient for successful targeting (Deng and Capecchi, 1992). Secondly, the use of isogenic DNA from the strain of mouse from which the ES cells are derived (such as 129/Sv for AB2.2 ES cells) has been shown to enhance the recombination efficiencies during targeting (van Deursen and Wieringa, 1992). Finally, the employment of promoter or poly-adenylation (polyA) trapping has also been shown to enhance targeting efficiencies (Donehower et al., 1992). In promoter trapping, the positive selection marker is promoter-less which allows for the selection of insertions that occur next to an endogenous promoter. However, this limits the targeting events to genes that are expressed in ES cells. To overcome this problem, polyA trapping which uses a positive selection marker that does not have a polyA signal and allows for the selection of integrations into genes, regardless of whether or not they are expressed in ES cells, was developed. However in our lab, neither promoter nor polyA trapping strategies are required for successful gene targeting in mouse ES cells.

The genetic changes created by conventional gene targeting vectors as discussed above are usually null mutations. This is usually accomplished by the deletion of exons and/or the introduction of a selection marker into the coding sequences which disrupts the reading frame of the gene, resulting in premature termination of the transcript or frame-shifts. Analyses of these null mutant mice have provided fundamental insights into mammalian gene function. Modification of the mouse genome can also produce other types of mutations; for example, hypomorphic alleles of a gene are also desirable to complement the null allele. In addition, subtle mutations such as point mutations in a gene are necessary in order to study functional domains of proteins and to accurately mirror some human diseases. Importantly, for genes that are critical for embryonic development, null mutants usually die *in uteri*, thereby precluding the study of their functions in late development or in specific tissues in the adult mice. To circumvent this

problem and to enable the investigation of gene function in a temporal and spatial manner, conditional knockout (cko) approaches have been developed (Glaser et al., 2005; Rajewsky et al., 1996). The cko approach allows a more accurate mouse model of human disease and sporadic cancer initiation and progression to be created.

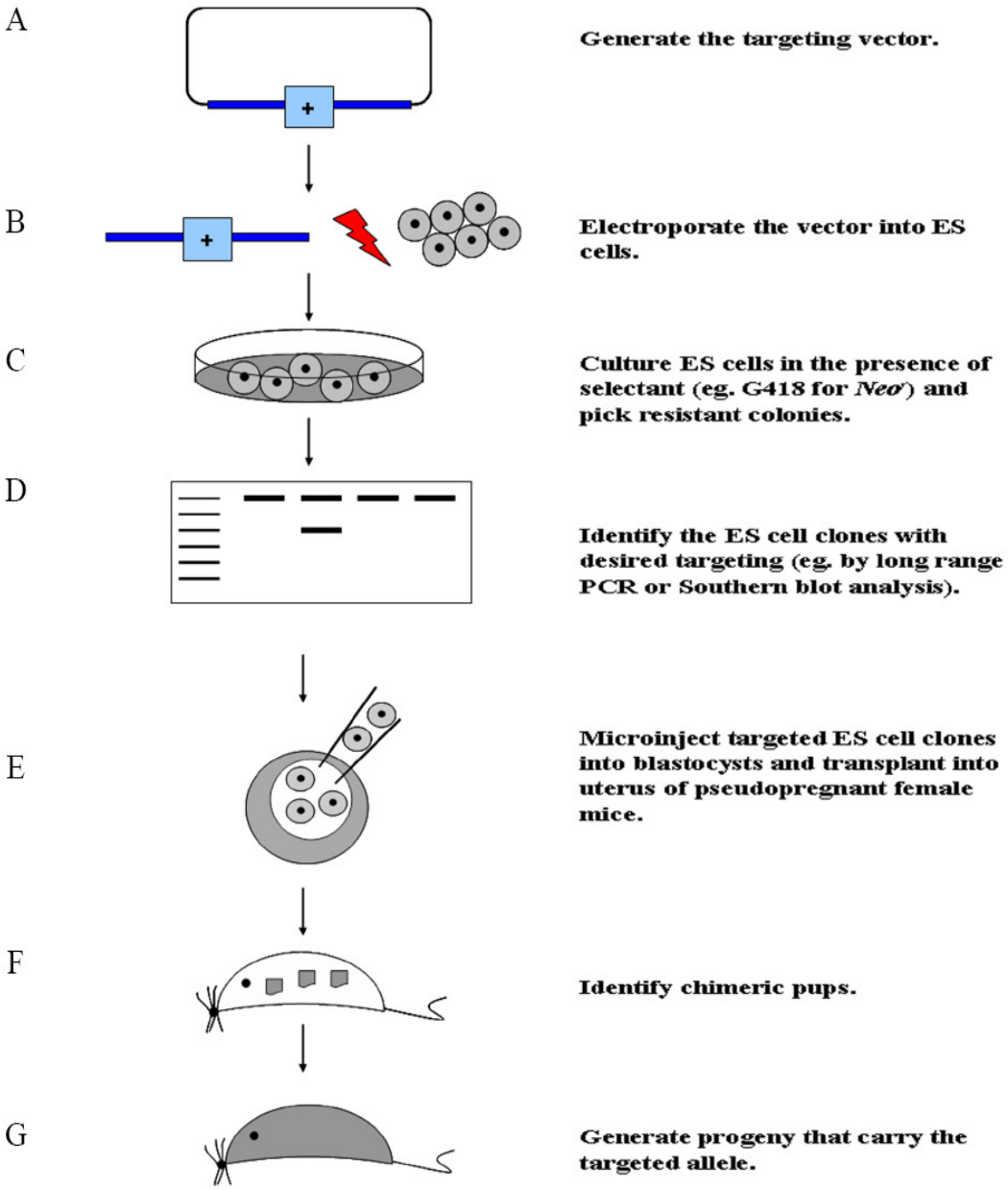


Figure 1.1. General procedure for generation of genetically modified mice by gene targeting strategies. (A) Generation of a targeting vector containing a positive (+) selection cassette and sequences of homology with the target locus (blue line). **(B)** The vector is linearized and electroporated into ES cells. **(C)** Correct transformants are selected for in the presence of a selectant (eg. G418 if a neomycin resistance cassette is present in the targeting vector). **(D)** Correctly targeted ES cell clones are then identified and genetically characterized using long range PCR or Southern blot analysis. **(E)** The selected ES cell clones are then microinjected into 3.5 days post-coitum blastocysts and transplanted into the uteri of pseudopregnant females. **(F)** Chimera obtained from the microinjections are mated with wild-type mice to establish germ-line transmission of the modified allele. **(G)** Progeny derived from the chimeras are characterized using long range PCR or Southern blot analysis, and a mutant mouse line that carries the desired targeted allele is established.

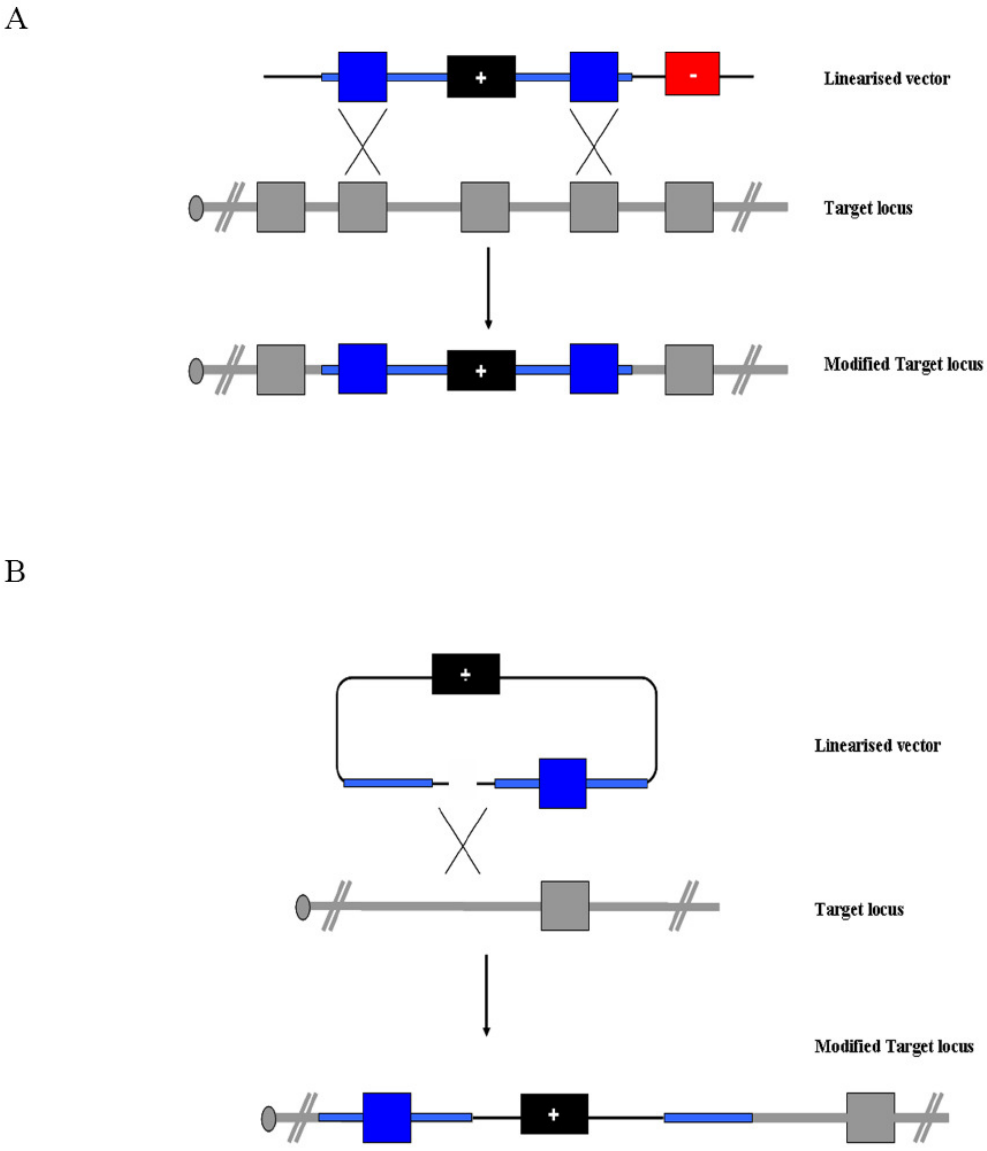


Figure 1.2. Replacement and insertion type vectors. (A) Replacement vectors target the locus by double reciprocal recombination, and they insert only the genetic sequences contained within the homologous sequences (blue). The positive selection markers (+) can be used to replace an exon or disrupt an exon. A negative selection marker (-) can be used to select against the random insertion of vectors. **(B)** Insertion vectors target the locus by single reciprocal recombination and insert the entire vector sequence. The vector contains regions of DNA (blue lines) homologous to the target locus (grey line) and a positive selection marker (+). X represents recombination between vector and the genomic loci.

1.2.5 Conditional knockout and Cre-*lox*P recombination system

Site-specific recombinases catalyze the recombination between two recombinant sites. By placing these sites strategically within the gene loci, deletion, insertion, inversion or translocation of specific regions of DNA can be produced after expression of the specific recombinase. There are two commonly used site-specific recombinase system in the mice: (1) the Cre/*lox*P system from the bacteriophage P1 (Figure 1.3A) and (2) the Flp/*FRT* system from the budding yeast *S. cerevisiae* (Figure 1.3B). Cre recombinase appears more effective in recombining its substrates *in vitro* and *in vivo* than Flp and it is the more widely used site-specific DNA system in both ES cells and mice (Nagy, 2000). This prokaryotic protein was first shown to work in mammalian cells by Sauer *et al.* (Sauer and Henderson, 1988) and subsequently in the mouse by Lakso *et al* (Lakso et al., 1992). Cre recombinase cleaves DNA between two 34-bp-long specific consensus sites known as *lox*P sites (Figure 1.3A). The *lox*P site has a unique structure that contains two palindromic 13-bp repeats flanking an 8-bp core sequence. The asymmetry of the core confers the orientation of the *lox*P site which is important for gene targeting as the directionality of the *lox*P sites determines the final configuration of the DNA sequence after expression of Cre recombinase. For example, if the two *lox*P sites are in the same orientation, expression of Cre recombinase would result in deletion of the intervening sequences (Figure 1.3C). In contrast, if the two *lox*P sites are in opposite orientations, Cre recombinase expression would result in an inversion (Figure 1.3D). Additionally, if the *lox*P sites are in *trans* (on different chromosomes), expression of Cre recombinase will lead to reciprocal exchange of the two chromosomes resulting in a translocation (Ramirez-Solis et al., 1995) (Figure 1.3E).

The Cre/*lox*P system has several advantages that make it suitable as a DNA recombination system in the mouse. Firstly, the *lox*P site is short enough to be targeted to intronic regions without disrupting endogenous transcription of the gene, but long enough to avoid the random occurrence of intrinsic *lox*P sites in the mouse genome. Analysis of the genomes of the mouse, rat and zebrafish revealed no perfectly matched intrinsic *lox*P sites. However, it has been noted that cryptic *lox*P sites are present in the mouse genome. Recombination between the cryptic *lox*P sites and *lox*P sites is expected to occur at a much lower efficiency, but the efficiency has not been comprehensively investigated. Cre

recombinase has also been shown to be functional in a wide range of cell types and can act on various forms of DNA substrates (supercoiled, relaxed or linear). *In vitro* Cre-mediated excision has been shown to facilitate genomic deletion of up to 400 kb, and recombinants can be identified without selection (Nagy, 2000). *In vivo* Cre-mediated excision is also extremely efficient and many Cre transgenic mouse lines have been generated to facilitate Cre-mediated excision in a wide range of tissues and over numerous developmental time points (Nagy, 2000).

One of the main uses of the Cre/*lox*P system in the mouse is to generate cko alleles. The first step to generating a cko allele is to identify the critical exon(s) of the gene which encodes for the essential functional domains of the protein. The critical exon(s) is then flanked by *lox*P sites which are targeted to the intronic regions flanking the critical exon(s). Mice containing the cko alleles should be phenotypically normal unless they are bred to a Cre-expressing transgenic line. Depending on the nature of the promoter driving Cre recombinase expression (ubiquitous or tissue-specific), deletion of the floxed region can occur in all cells or in specific cells/tissues or at certain developmental stages (Nagy, 2000).

To design a successful cko allele, there are several considerations that need to be addressed. Firstly, the critical exon(s) which encodes the essential functional domains has to be selected and flanked by *lox*P sites such that Cre-mediated excision would result in a null allele. The ideal scenario would be to flank the entire gene with *lox*P sites but this is usually not technically feasible as the efficiency of Cre-mediated excision decreases dramatically with increasing distance (Nagy, 2000). Another method would be to target the first exon of the gene. However there are pitfalls regarding this strategy such as placing a *lox*P site upstream of the ATG start site might disrupt endogeneous promoter sequences and there might also be alternative start sites downstream of the first exon to generate alternative transcripts of the gene. In addition, one also has to consider the strategy of targeting *lox*P sites to the gene locus and keeping the modified allele functional. The selection cassette used in the targeting step is usually removed because the presence of the selection cassette can result in a hypomorphic allele or affect the expression of neighbouring genes (Fiering et al., 1995). Such interference can generate a

phenotype that is normally not associated with loss-of-function of the targeted allele and can cause misinterpretation of the results.

The more sophisticated techniques of generating cko alleles use both Cre/*lox*P and Flp/*FRT* systems. In this method, the selection marker is flanked by two *FRT* sites and expression of Flp can be carried out either *in vitro* (ES cells using a Flp-expression plasmid) or *in vivo* (Flp-expressing mice) to remove the selection cassette (Figure 1.4). Following expression of Flp recombinase, the selection cassette is removed, leaving behind two *lox*P sites flanking the critical exon, generating a cko allele. The disadvantage of this system is that Flp recombinase is not as efficient as Cre recombinase; however with the generation of the enhanced Flp recombinase (Flpe) and mouse codon-optimised Flp recombinase (Flpo), the recombination efficiency of Flp has improved and is closer to that of Cre recombinase (Buchholz et al., 1998; Raymond and Soriano, 2007).

The cko alleles would not be possible unless there were well-characterized Cre transgenic lines which permitted the Cre-mediated excision of cko alleles. Thus, a major challenge is to identify promoters that are characteristic for different cell types, lineages and developmental events (Nagy, 2000). The Cre transgenic lines generated from these promoters will have to be validated to ensure their efficacy as well as specificity. This is because Cre-mediated excision does not usually reach 100% efficiency and partial excision creates mosaicism that can complicate interpretation of the observed phenotype. Nevertheless, these tissue- and cell-lineage-specific Cre transgenic lines are invaluable resources for the cko technology. A further improvement to Cre transgenic technology was achieved with the creation of inducible Cre systems. The inducible Cre system allows for the regulated control of the induction of Cre expression, thus permitting temporal deletion of the conditional alleles. One such system is the tamoxifen (TAM) or 4-hydroxytamoxifen (4-OHT) inducible Cre-ERT (Metzger and Chambon, 2001). The Cre recombinase has been fused to a mutated ligand binding domain of the human estrogen receptor (ER), resulting in a tamoxifen-dependent Cre recombinase, Cre-ERT, that is activated by TAM or 4-OHT, but not by endogenous estrogen or progesterone. In the absence of TAM or 4-OHT, the Cre-ERT protein resides in the cytoplasm, whereas upon treatment, binding of the ligand to the Cre-ERT protein results in a transformational change, causing the Cre-ERT protein to translocate into the nucleus where it executes its

function (Metzger and Chambon, 2001). Another widely used inducible Cre technology is the tetracycline (Tet)-dependent regulatory system (Tet-on or Tet-off) (Gossen and Bujard, 1992). The Tet systems use a chimeric transactivator to control transcription of the gene of interest from a silent promoter and are based on two regulatory elements derived from the *E. coli* Tet resistance operon: the Tet repressor protein (TetR) and the Tet operator sequence (*tetO*). The Tet-off system uses a plasmid that expresses a fusion protein known as the tetracycline-controlled transactivator (tTA), which is composed of TetR and the VP16 activation domain of the herpes simplex virus. The tTA binds to the tetracycline response element (TRE) and activates transcription of the target gene in the absence of the inducer, doxycyclin (Dox) (Gossen and Bujard, 1992). In contrast, the Tet-on system uses a form of TetR containing four amino acid changes that results in altered binding characteristics and creates the reverse TetR (rTetR). The rTetR binds the TRE and activates transcription of the target gene in the presence of Dox (Gossen and Bujard, 1992). The temporal control of Cre activity makes it an attractive system to combine inducibility with cell- or lineage-specific expression of Cre recombinase, thus allowing the temporal and spatial control of the deletion of the cko allele.

While expression of Cre recombinase is not normally associated with lethality, it is important to note that there are reported Cre toxicity effects. Cre expression in cultured mammalian cells results in a markedly reduced rate of proliferation and this effect is dependent on the endonuclease activity of Cre (Loonstra et al., 2001). *In vivo* titration experiments revealed that toxicity is due to the level of Cre activity (Loonstra et al., 2001). Hence proper controls are required to ensure that the observed phenotypes in cko mice are due to deletion of the targeted allele and not due to side effects of Cre toxicity.

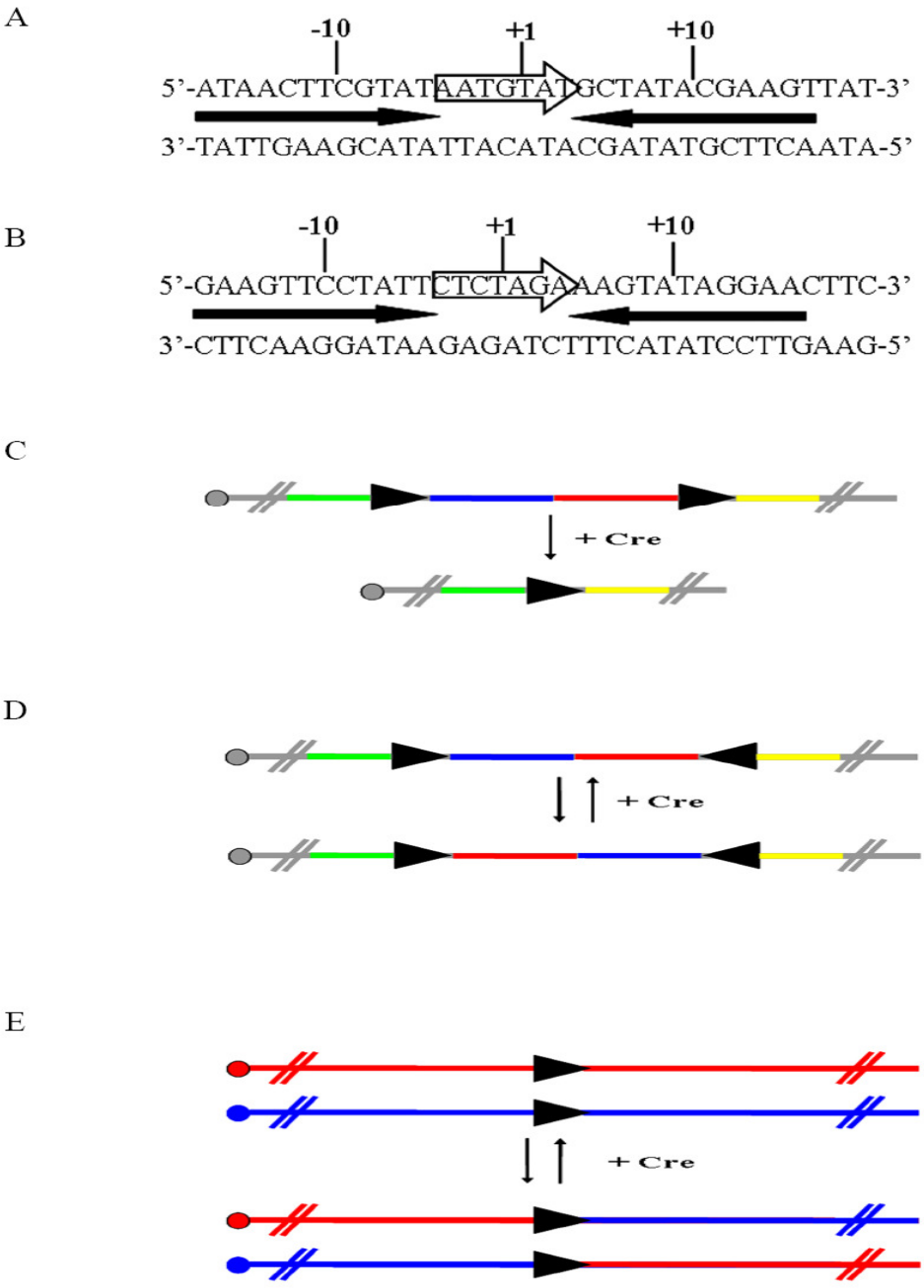


Figure 1.3. Properties of Cre recombinase. The consensus DNA sequence is recognized by **(A)** Cre recombinase and **(B)** Flp recombinase. Target sites (*lox*P of Cre; *FRT* of Flp) contain inverted 13-bp symmetry elements (indicated by the bold arrows) flanking an 8-bp A:T rich non-palindromic core (indicated by the open arrow). One recombinase monomer binds to each symmetry element, while the core sequence provides the site of strand cleavage, exchange, and ligation. The asymmetry of the core region (open arrow) imparts directionality on the reaction. Depending on the directionality of the *lox*P sites, expression of Cre recombinase can result in **(C)** deletion of the intervening sequences if *lox*P sites are in *cis*, **(D)** an inversion if the *lox*P sites are oriented in opposite directions or **(E)** reciprocal exchange of the regions that flank the *lox*P sites if the two sites are in *trans*. The arrows between the recombinase substrates and the products indicate the reversible nature of each reaction.

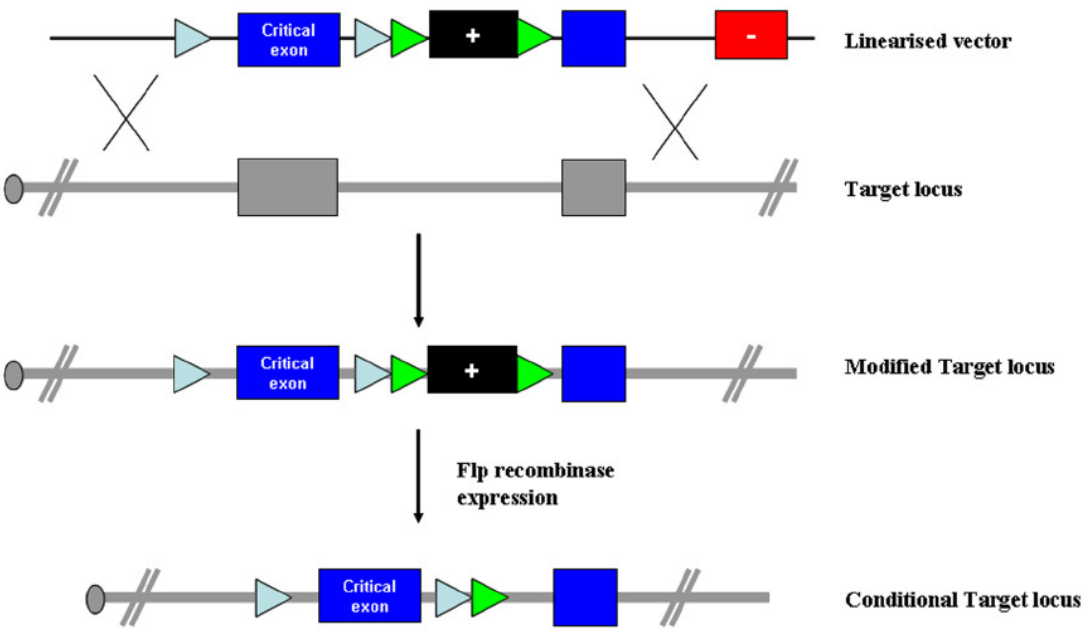


Figure 1.4. Strategy for generation of a conditional knockout allele. This strategy employs the use of both Cre/*lox*P and Flp/*FRT* systems. The selection marker (+) is flanked by two *FRT* sites (green arrow heads) and expression of Flp recombinase (*in vitro* or *in vivo*) results in the excision of the selection marker, leaving behind a single *FRT* site and two *lox*P sites (blue arrow heads) flanking the critical exon(s). (-) indicates negative selection marker. X represents recombination between vector and the genomic loci.

1.2.6 Recombineering technology

The development of recombinogenic engineering or recombineering has greatly reduced the time required to generate DNA constructs for making knockout mice. Recombineering is based on the *E. coli* phage homologous recombination systems which enable modification of genomic DNA such as those carried on bacterial artificial chromosomes (BACs) or P1 artificial chromosomes (PACs) without any restriction enzymes or ligases. Traditionally, *E. coli* has been the organism of choice for carrying out the genetic modifications required for making targeting vectors. Numerous restriction digestion and ligation steps are required to piece different DNA fragments together in order to complete the targeting vectors. This is usually tedious and requires the use of specific compatible restriction sites. Using homologous recombination for DNA modification was first widely used in *S. cerevisiae* to combine transformed, linear, double-stranded DNA (dsDNA) with homologous regions in the yeast genome (Lafontaine and Tollervey, 1996). However, unlike in the yeast, linear dsDNA are highly unstable in *E. coli* due to the presence of the ATP-dependent, linear-dsDNA exonuclease RecBCD (Copeland et al., 2001). RecBCD unwinds and degrades DNA to generate 3' single-stranded DNA (ssDNA) tails, which are used by RecA to initiate recombination. Therefore transformation of linear dsDNA into wild-type *E. coli* would result in degradation of the exogenous DNA, making homologous recombination unfeasible. To overcome this, *E. coli* strains that contain the *RecBC* (hence lacking RecBCD), *sbcB* and *sbcC* mutations (which restore recombination capabilities) were initially used for genetic manipulation (Copeland et al., 2001). However, loss of the RecBCD complex generally impairs cellular integrity and as the recombination pathway is constitutively active in these strains, rearrangements and deletions between the repeat sequences found within BACs and PACs usually occur. To overcome this deficiency, a temperature-sensitive shuttle-vector-based system based on the RecA pathway was developed (O'Connor et al., 1989). The *E. coli RecA* gene was cloned into a temperature-sensitive shuttle vector that has the temperature-sensitive origin of replication which functions at 30^{o}C but not at the higher restrictive temperature of 42-44^{o}C. Hence, transformation of this plasmid, which also contains cloned sequences homologous to the target genomic DNA, into the *RecA*$^{-}$ *E. coli* containing the BAC confers recombination-competence to the *E. coli*, allowing

homologous recombination to occur (O'Connor et al., 1989). As loss of *RecA* in *E coli* generally impairs general cellular integrity, an improved recombineering approach by transient expression of *RecA* has been developed to generate targeting constructs (Wang et al., 2006).

In 1998, an important advancement to the field of homologous recombination was made by Francis Steward and colleagues who developed the Rac-encoded RecET system (also known as Red recombination and lambda-mediated cloning). They showed that PCR-amplified fragment of linear dsDNA, flanked by short homology arms (42 bp homologous sequence to target plasmid or BAC) can be efficiently targeted to a plasmid or BAC by electroporating the dsDNA into *recBC sbcA E. coli* strains (Zhang et al., 1998). Instead of using *RecBC* mutants as discussed above, RecBC can be inactivated either by the *sbcA* mutation, which removes a repressor for the endogenous Rac prophage to induce expression of *recE* and *recT* (Lloyd, 1974) or the *gam* protein of λ bacteriophage (Murphy, 1998). Recombination functions encoded by *recE* and *recT* genes enabled genomic DNA to be modified using PCR products with short homology arms (Zhang et al., 1998). The overview of the RecET system is depicted in Figure 1.5. Firstly, PCR amplification is used to generate the homology arms (to the target plasmid or BAC DNA) flanking the insert DNA. Secondly, phage-derived recombineering competence was conferred to the BAC-containing *E. coli* host strain. Finally, the PCR-generated cassette is introduced into the host by electroporation to initiate homologous recombination and generate the recombinant cassette (Figure 1.5). Bacteriophage λ contains an efficient recombination system known as the Red system. The key genes in this system are the *redα/exo* and *redβ/bet*. Exo is a 5'-3' exonuclease that acts on linear dsDNA while Bet binds to the 3' ssDNA (single-stranded DNA) overhangs generated by Exo and promotes annealing to a complementary strand. The Red system is assisted by the bacteriophage-encoded Gam protein that inhibits host RecBCD activity (Murphy, 1998).

After the initial publication by Steward, several labs developed other methods based on expression of the Exo, Bet and Gam proteins to maximise efficient homologous recombination efficiency (Datsenko and Wanner, 2000; Yu et al., 2000). In the defective prophage-based system, *exo*, *bet* and *gam* are expressed from a defective prophage that

has integrated into the *E. coli* genome and their expression is regulated under the tight control of the temperature-sensitive λ-CI857 repressor (this strain is known as DY380) (Yu et al., 2000). At 32^{o}C, the repressor is active and there is no expression of the three genes, thus the *E. coli* cells do not have recombineering activity. In contrast, upon shifting the culture to 42^{o}C for 15 min, the CI repressor is inactivated and the *red* genes are co-ordinately expressed from a λ-*pL* promoter at high levels, conferring recombineering-competence to the cells. The λ prophage-based recombineering system proved to be extremely efficient because all three recombineering genes (*exo*, *bet* and *gam*) are present as a single copy and each is expressed from their natural operon and from a strong promoter *pL*, which means that they are present in molar ratio suitable for forming a complex *in vivo*. This tightly regulated expression ensures that the Exo and Bet proteins are only present after induction, thus minimizing the deleterious effects of rearrangement or deletion of the BAC. To facilitate *in vitro* expression of Cre and Flpe recombinase, arabinose-inducible Cre and Flpe recombinases have also been introduced into DY380 strain to generate EL350 and EL250 strains respectively (Lee et al., 2001). Transient expression of Cre or Flpe recombinases after induction with arabinose would mediate recombination between *lox*P and *FRT* sites respectively, allowing the removal of undesired selection markers from targeting cassettes. Using these bacterial strains, a highly efficient strategy for generation of conditional targeting vectors was developed (Figure 1.6) (Liu et al., 2003a). In this method, longer homology arms (200-500 bp for retrieving and 100-300 bp for targeting) were generated by PCR amplification using BAC DNA as a template. These homology arms were used to subclone BAC DNA into high-copy plasmid backbone (pBluescript) by gap repair using recombineering in the bacterial strain EL350. Similarly, long homology arms were also generated by PCR and used to generate the mini-targeting vector (containing selection markers and *lox*P sites) and used for the first targeting to the retrieved BAC DNA in pBluescript backbone. Subsequently, the selection marker was excised by the expression of Cre recombinase in EL350 and a final targeting step was carried out to complete the cko targeting construct. By using pBluescript as a backbone for the vector construct, the problems caused by *lox*P sites present in the BAC backbone can be eliminated. In addition, the long homology arms increase the recombineering efficiency when compared to that obtained with

conventional 45 bp arms, and typically >95% of the colonies obtained after recombineering contain the correct DNA constructs (Liu et al., 2003a). A drawback of this method is that high-copy plasmids are used for retrieving BAC DNA and they can be unstable when harbouring large pieces of DNA. Since cloning steps are still involved, this method is not suitable for high throughput vector construction.

Subsequently, our lab developed a new set of recombineering reagents and protocols that would facilitate the high-throughput generation of cko targeting constructs in a rapid and highly efficient manner (Chan et al., 2007). Two new mobile recombineering reagents were generated, the first of which is a complete λ phage that is replication defective in BAC-harbouring DH10B *E. coli* but still retains its full temperature-inducible homologous recombination functions. In addition, a set of low-copy plasmids (pSim), which contains the three recombineering genes in their native operon, *pL*, under the control of the λ CI repressor. BAC-harbouring cells can be made recombineering-competent by a simple infection (with λ phage) or transformation (with pSim), allowing genetic manipulation to be carried out to facilitate the generation of targeting constructs in a 96-well plate. The overall strategy for using mobile recombineering reagents to generate cko targeting constructs is depicted in Figure 1.7. Briefly, the critical exons to be deleted are identified and the BAC clone containing the region of interest is then made recombineering-competent either by λ phage infection or by transformation with pSim. Next, 70-80 bp of homology sequence (homologous to the target BAC DNA) is included in the primers used to amplify the 5' and 3' targeting cassettes (5' – *Bsd* cassette; 3' – *Neo* cassette). Two sequential targeting steps are then carried out to target the 5' and 3' cassettes to the BAC DNA (Figure 1.7A) and the targeted BAC region is then retrieved into a pBR322-based backbone (PL611) (Figure 1.7B). Finally, the 5' *Bsd* cassette is replaced by a reporter cassette (*lacZ*) to generate the final targeting vector (Figure 1.7C). This type of targeting vector would generate a multi-purpose allele that can serve as a reporter, null and conditional allele (Figure 1.7D). Our new recombineering system enables numerous cko targeting constructs to be generated simultaneously in a 96-well format and has solved a key technical bottleneck in genome-wide targeted mutagenesis programmes of the mouse (Collins et al., 2007).

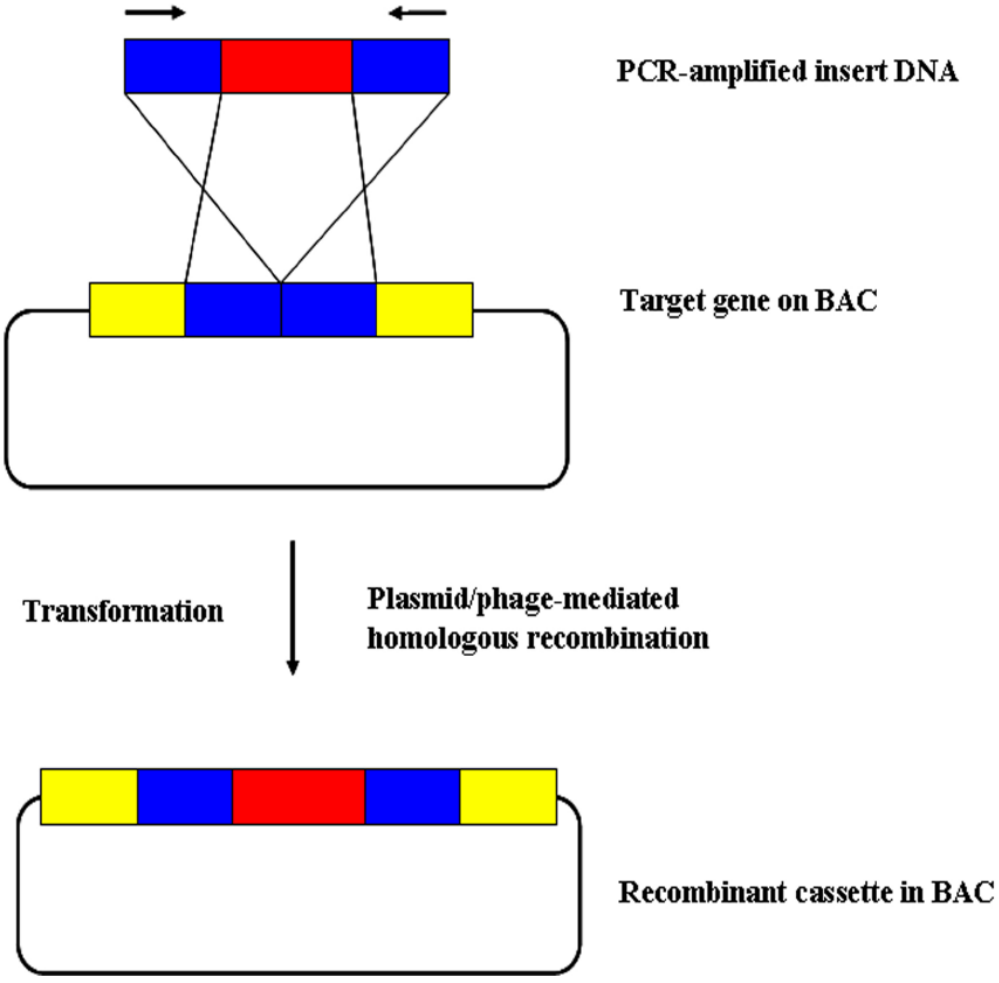


Figure 1.5. Overview of recombineering using Rac-encoded RecET system. PCR amplification is first carried out to generate homology arms (to the target plasmid or BAC DNA) flanking the cassette of interest (such as a selection marker). Next, phage-derived recombination functions are introduced into the target host bacterial strain or the BAC can also be transformed into a bacterial strain that is recombination competent. Finally, the PCR products are electroporated into the target host bacterial strain and recombinant products can be detected by selection or counter-selection.

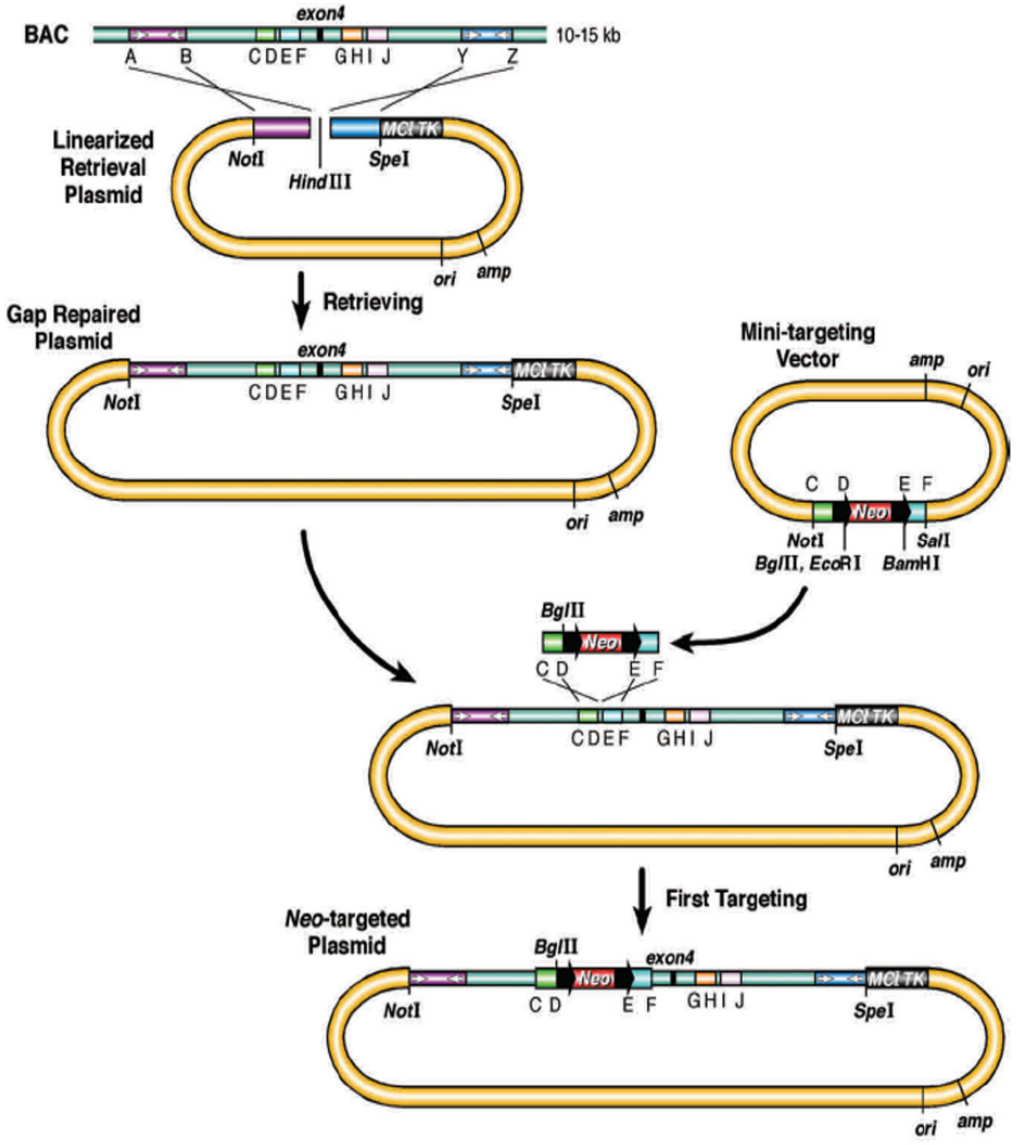
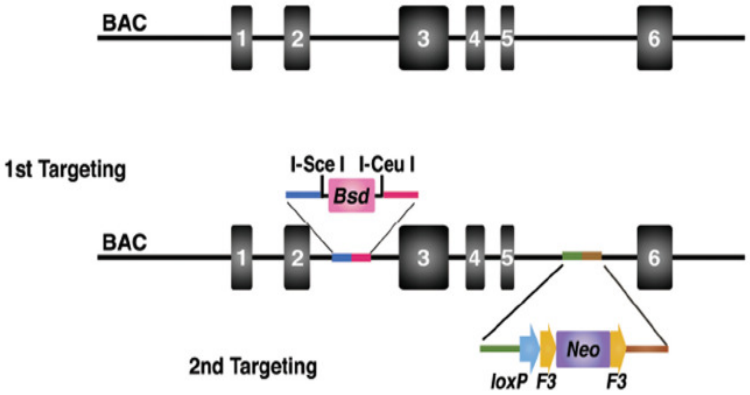
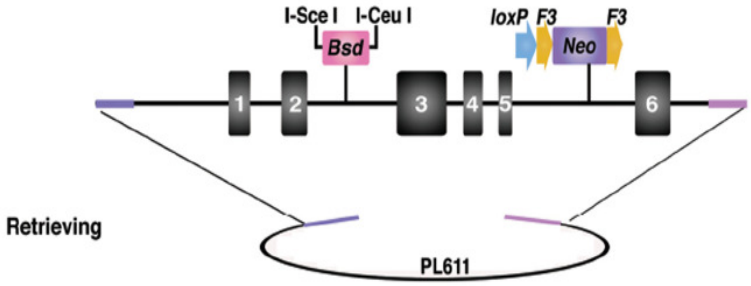


Figure 1.6. Generation of conditional targeting vectors using the Red system. DNA from a BAC containing the region of interest is sub-cloned from the BAC and is used for the construction of conditional targeting vectors. The homology arms used for subcloning and for targeting are generated by PCR using BAC DNA as a template. The two homology arms (purple or dark blue), which are amplified using primers A and B or primers Y and Z, are cloned into an *MC1-TK*-containing plasmid. The gap repair plasmid is then linearized with *Hind*III to create a double-strand break for gap repair. A mini-targeting vector is constructed by ligation of the two PCR products generated by amplification of BAC DNA with primers C and D (light green) or primers E and F (blue), with a floxed selection marker and pBluescript. A *Bgl*II restriction site is included in the mini targeting vector for diagnosing gene targeting in ES cells. The black arrows denote *lox*P sites. The targeting cassette is excised by *Not*I and *Sal*I digestion, or by PCR amplification, using primers C and F. The gap-repaired plasmid and the excised targeting cassette are cotransformed into recombineering-competent DY380 or EL350 cells. The recombinants have a floxed *Neo* cassette inserted between primers D and E and can be selected on kanamycin plates. The *Neo* cassette is excised with Cre recombinase, leaving a single *lox*P site at the targeted locus. Figure obtained from (Liu et al., 2003a).

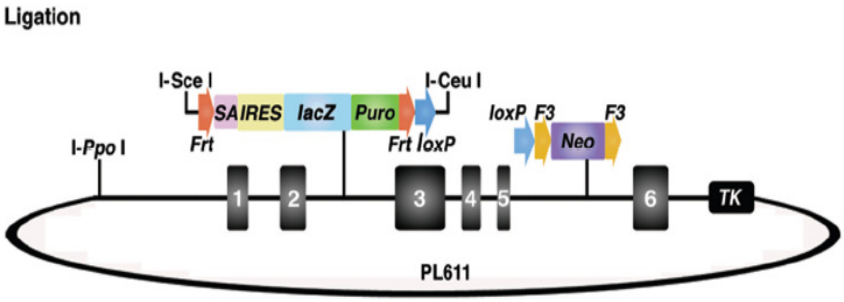
A


BAC
1
2
3
4
5
6
1st Targeting
I-Sce I
I-Ceu I
Bsd
BAC
2nd Targeting
Neo
loxP
F3
F3

B


I-Sce I
I-Ceu I
Bsd
loxP
F3
F3
Neo
1
2
3
4
5
6
Retrieving
PL611

C


Ligation
I-Sce I
I-Ceu I
SA
IRES
lacZ
Puro
loxP
F3
F3
Neo
Frt
Frt
loxP
I-Ppo I
1
2
3
4
5
6
TK
PL611

D

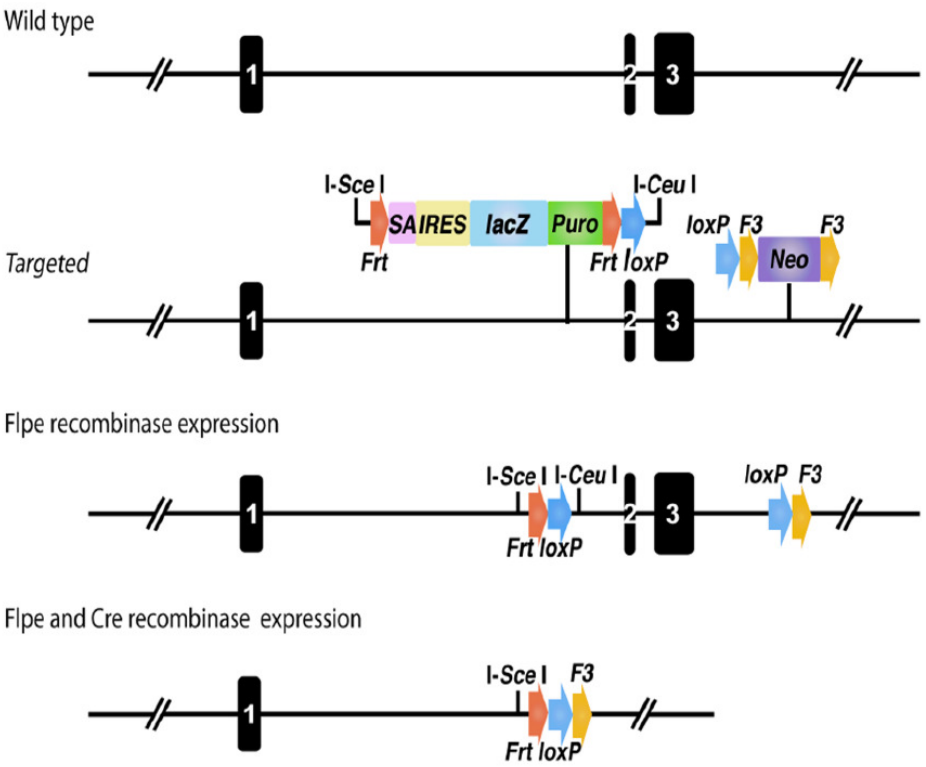


Figure 1.7. Generation of conditional targeting vectors using mobile reagents. (A) The critical exon(s) to be deleted in the cko allele is identified and in this example, exons 3-5 are to be deleted. The *Bsd* cassette, which is flanked by the two rare cutter sites, *I-Sce*I and *I-Ceu*I, is targeted to the 5' side of the intended deletion region. Subsequently, the *lox*P-*F3-PGK-EM7-Neo-F3* (*Neo*) cassette is targeted to the 3' side of the deletion region. Coloured lines represent the short homology arms used for recombineering. **(B)** The genomic DNA fragment is then retrieved from the BAC to PL611, which has the Amp^R gene. **(C)** The *Bsd* cassette can be conveniently replaced by a reporter, i.e. *lacZ*, in a simple ligation reaction. The final targeting vector has the reporter flanked by two *FRT* sites followed by a *lox*P site at the 5' side of the intended deletion region, and a *F3* flanked *Neo* cassette that provides for positive selection in ES cells. The negative selection marker *TK* is added to the vector backbone by recombineering. The vector is then linearized with the rare-cutter *I-Ppo*I before targeting to ES cells. **(D)** The targeted allele is a reporter allele in which endogenous regulatory elements control the expression of *lacZ* and staining for X-gal reveals the spatial expression patterns of the gene. The conditional knockout allele can be obtained by expression of Flpe recombinase either *in vitro* or *in vivo* to delete the intervening sequences between *FRT*/*F3* sites, leaving behind two *lox*P sites flanking critical exons of the gene. The conditional allele can then be crossed to a cell or lineage-specific Cre recombinase-expressing mice. Figure obtained from Chan *et al.* (Chan et al., 2007).

1.3 B-cell lymphoma/leukaemia 11(*Bcl11*) gene family

1.3.1 Discovery of the *Bcl11* genes

Loss-of-function analyses using knockout mice generated using the techniques discussed above have provided great insights into the specific roles of genes at different stages of development and in different tissues. One such example is the *Bcl11* family. The *Bcl11* family consists of two members in the mouse and human, *Bcl11a* and *Bcl11b*, which encode for Krüppel-like transcription factors that contain both single and double C_2H_2 zinc finger motifs. C_2H_2 zinc fingers are modular protein domains that generally confer sequence-specific DNA binding activity, and transcriptional regulatory domains. Structurally, the C_2H_2 zinc finger motif is comprised of a β-hairpin followed by an α-helix that folds around a single zinc ion (Suzuki et al., 1994). Sequence-specific recognition of DNA is mediated primarily by interactions between the variable amino acids within and around the α-helix and nucleotides within the major groove of DNA (Suzuki et al., 1994). Outside of these zinc finger domains, the rest of the coding sequences of Bcl11a and Bcl11b showed no homology to known protein domains in the database, suggesting that the *Bcl11* genes encode for two novel proteins.

Bcl11a and Bcl11b proteins were shown to interact with all members of the chicken ovalbumin upstream promoter transcription factor (COUP-TF) subfamily of orphan nuclear receptors and were termed COUP-TF-interacting proteins 1 and 2, respectively (CTIP1 and CTIP2) (Avram et al., 2000). These studies identified *Bcl11a* (*Evi9*/*Ctip1*) and *Bcl11b* (*Rit1*/*Ctip2*) as C_2H_2 zinc finger transcription factors and preliminary analyses suggest that they may function as transcription repressors.

Bcl11a was first identified as a common site of retroviral integration in BXH2 murine myeloid leukemias and was initially called Ectropic viral integration site 9 (*Evi9*) (Nakamura et al., 2000). Retroviral integration into the *Evi9* locus leads to the development of murine leukaemia either by enhancing the expression of proto-oncogenes or by disrupting the expression of tumour suppressor genes. Because proviral integration into the genomic loci leaves behind a genetic tag, the integration site can be readily determined and the affected gene(s) identified. Retroviral mutagenesis screens in the mouse have been instrumental in identifying numerous leukaemia disease genes (Jonkers and Berns, 1996). Sequence analysis showed that proviral integrations were located

within the first intron of *Evi9* in the reverse transcriptional orientation. This location and orientation is frequently found when retroviral integration activates gene transcription through an enhancer mechanism (Jonkers and Berns, 1996). Northern blots of cell lines with proviral integration at *Evi9* showed elevated levels of *Evi9* expression, consistent with the hypothesis that proviral integration results in the up-regulation of *Evi9* expression, suggesting that *Evi9* may be a putative proto-oncogene (Nakamura et al., 2000). Subsequently, *Evi9* was shown to possess oncogenic properties as transformation of NIH 3T3 cells with *Evi9* resulted in the growth of anchorage-independent colonies in soft agar (Nakamura et al., 2000).

In addition to retroviral screens, radiation-induced malignancies in the mouse, such as thymic lymphomas, can be studied to define the regions of allelic loss and delineate the putative tumour suppressor(s) within the deletion region. In a γ-ray induced mouse thymic lymphoma, *Bcl11b* (also known as Radiation-induced tumour suppressor gene 1, *Rit1*) was identified as a novel tumour suppressor gene located on mouse chromosome 12 (Shinbo et al., 1999; Wakabayashi et al., 2003a). Homozygous deletions of exons 2 and 3 of *Rit1* and point mutations of *Rit1* were detected in these thymic lymphomas and resulted in reduced or no expression of Rit1 protein (Wakabayashi et al., 2003a). In the same study, Rit1 proteins were found to suppress tumour cell growth and there was also a preferential inactivation of *Rit1* in *p53* wild-type lymphomas, suggesting that *Rit1* may be a potential tumour suppressor.

1.3.2 Properties of *Bcl11* genes

The human *BCL11A* gene is located on chromosome 2 while its mouse orthologue is found on chromosome 11. There is a high level of conservation between the human and mouse Bcl11a proteins (99% identity), suggesting that there is conservation of function in both species. There are at least three known isoforms of Bcl11a in both the human and mouse (XL – 835 aa; L - 773 aa and S - 243 aa) (Nakamura et al., 2000; Satterwhite et al., 2001; Weniger et al., 2006), though bioinformatical prediction and Northern blot analysis suggest there could be three more isoforms (Figure 1.8A). The three common Bcl11a isoforms are derived from five exons and all three isoforms contain the first three exons. The longest isoform (XL) contains the sequences from exons 1 to 4 only and is

predicted to have six zinc finger domains. Alternative splicing within exon 4 to exon 5 is thought to result in formation of the other two common isoforms. *BCL11B*, the paralogue of *BCL11A*, is found on chromosome 14 while its mouse orthologue is located on chromosome 12. Similarly, a high level of conservation between the human and mouse Bcl11b proteins is observed (93% identity), again suggesting conservation of function in both species. There are at least two isoforms of Bcl11b in both the human and mouse (α – 894/884 aa; β – 823/812 aa in human/mouse) (Figure 1.8A), which are derived from four exons; the longest isoform (α) contains exons 1 to 4 while the shorter isoform (β) only contains exons 1, 2 and 4.

Bcl11a and Bcl11b share 67% similarity at the nucleotide and 63% similarity at the amino acid levels, but the level of similarity increases to 95% between the zinc finger domains (Figure 1.8B and 1.8C). Exon 4 of both genes, which encodes 75% of the protein sequence, constitutes the main functional domain of these proteins. Bcl11a (XL) contains six zinc fingers, a proline-rich domain between zinc fingers 1 and 2 and an acidic domain between zinc fingers 3 and 4 that contains a run of 21 consecutive residues (Figure 1.8B-D). Zinc fingers 1 and 6 are different from zinc finger 2, 3, 4 and 5 in that they have 4 amino acids separating the two zinc-binding histidines, whereas the others have 3 amino acids (Satterwhite et al., 2001). The internal zinc fingers (2, 3, 4 and 5) are arranged in pairs, with each pair being separated by a canonical 'linker' sequence. The Bcl11a (L) isoform contains only four zinc fingers as it lacks the two C_2H_2 zinc fingers present at the C-terminus of Bcl11a (XL), and the shortest isoform (S), has only one zinc finger domain. The two Bcl11b isoforms contain six zinc fingers domains, a proline-rich domain between zinc fingers 1 and 2, and an acidic domain between zinc fingers 3 and 4 that contains a run of 13 consecutive residues (Figure 1.8B-D). Thus, the Bcl11 transcription factors have similar sequences, structures and physical properties.

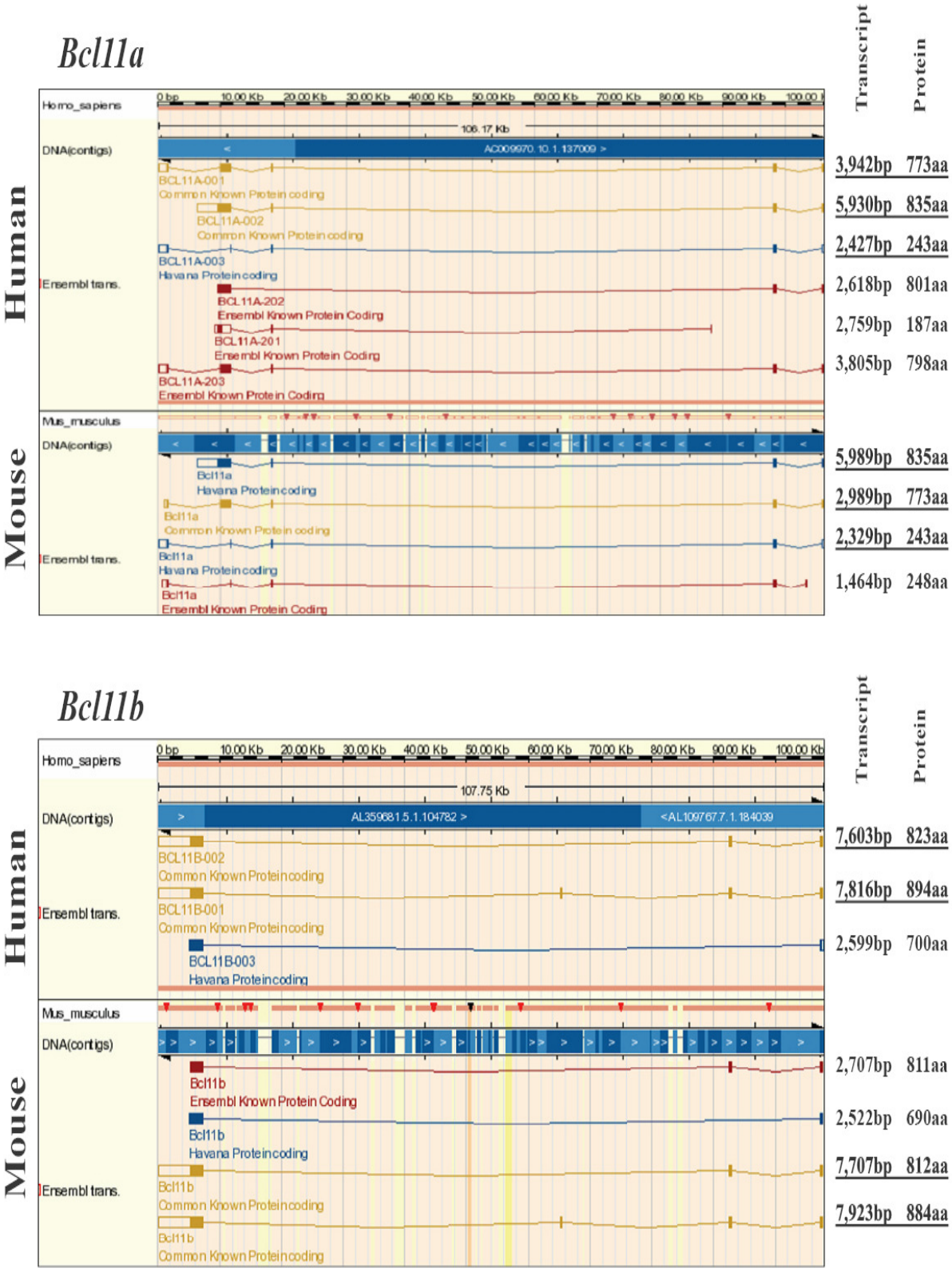
Different murine Bcl11a isoforms were shown to have different subcellular localization patterns. For example, one isoform (L) was shown to be located exclusively in the nucleus and was distributed within spherical nuclear structures, 50 to 500 nm in diameter (Nakamura et al., 2000). In contrast, the shortest isoform (S) was detected predominantly in the cytoplasm. In addition, Bcl11a (L) colocalized with BCL6 but not with PML or Sp100 (Nakamura et al., 2000). Further analysis by coimmunoprecipitation

and GST pull-down assay showed that Bcl11a and BCL6 interact directly, suggesting that Bcl11a may function in part through its interaction with BCL6 (Nakamura et al., 2000). Similarly, different human isoforms of BCL11A also show varying subcellular localization: the human BCL11A (XL) was shown to localize exclusively within nuclear dots (paraspeckles), and colocalized with the major fraction of BCL6 found inside the nucleus, while BCL11A (L) showed fewer nuclear dots and more diffuse nuclear staining, but nevertheless, BCL11A (L) colocalized with BCL6 (Liu et al., 2006; Pulford et al., 2006). In contrast, the shortest BCL11A (S) isoform is only relocalized from the cytoplasm to the nucleus by interaction with the other isoforms.

Putative target DNA binding sites of Bcl11a proteins have also been determined by reiterative oligonucleotide selection. The human BCL11A (XL) isoform binds to the core consensus 5'C-C-C/T-A/G-C-3' (Liu et al., 2006), which differs from the consensus binding sequence (5'-GGCCGGAGG-3') determined for the mouse Bcl11a (L) isoform (Avram et al., 2002). However, neither of these binding sites has been validated *in vivo*. As discussed above, both Bcl11 proteins were identified as COUP-TF interaction partners (Avram et al., 2000). COUP-TF family members generally mediate transcriptional repression by recruiting nuclear receptor co-repressor (NCoR) and/or silencing mediator for retinoid and thyroid hormone receptor (SMRT) to the template (Avram et al., 2000). NCoR and SMRT are components of a larger repressor complex that also includes mSin3A/B and a trichostatin-sensitive histone deacetylase (Heinzel et al., 1997). Avram *et al*. investigated the nature of the interaction between Bcl11a and ARP1, a COUP-TF member in yeast, and found that Bcl11a contained two independent ARP1 interaction domains, ID1 and ID2 that bound to the putative AF-2 domain of ARP1 *in vitro* (Avram et al., 2000). Bcl11a exhibited a punctuate distribution in the nucleus and recruited co-transfected ARP1 to these foci. Interestingly, Bcl11a potentiated the transcriptional repression activity of ARP1 in HEK293 cells independently of trichostatin-sensitive histone deacetylation, suggesting that Bcl11a does not mediate transcriptional repression by acting through recruitment of trichostatin-sensitive class I or II histone deacetylase(s). These results demonstrate that Bcl11a and Bcl11b transcription factors may either bind directly to the target sites in a sequence-specific manner or be recruited to the target sites by a COUP-TF family member. Recent studies demonstrated that Bcl11a and Bcl11b

recruited sirtuin 1 (SIRT1), a trichostatin-insensitive, nicotinamide-sensitive class III histone deacetylase, to the promoter region of a reporter gene template in HEK293 cells (Senawong et al., 2003; Senawong et al., 2005). Additionally, SIRT1 was shown to catalyze the deacetylation of histones H3 and/or H4 on the reporter gene template, suggesting that SIRT1 contributes at least partially to the transcriptional repression activities of both Bcl11 proteins. Interestingly, Bcl11b was also shown to be associated with the nucleosome remodelling and deacetylase (NuRD) complex in T lymphocytes (Cismasiu et al., 2005). Further analysis showed that the endogenous Bcl11b complexes from $CD4^+$ T lymphocytes harbour histone deacetylase activity sensitive to trichostatin A and that both metastasis-associated proteins MTA1 and MTA2 interact directly with BCL11B. In summary, the current data suggests that the transcriptional repression activities of Bcl11 proteins could be mediated by recruitment of SIRT1 (trichostatin-insensitive) and/or NuRD complex (trichostatin-sensitive) depending on the cellular context.

A


Bcl11a
Human
Mouse
Transcript
Protein
Homo_sapiens
DNA(contigs)
Ensembl trans.
106.17 Kb
AC009970.10.1.137009 >
BCL11A-001
Common Known Protein coding
3,942bp 773aa
BCL11A-002
Common Known Protein coding
5,930bp 835aa
BCL11A-003
Havana Protein coding
2,427bp 243aa
BCL11A-202
Ensembl Known Protein Coding
2,618bp 801aa
BCL11A-201
Ensembl Known Protein Coding
2,759bp 187aa
BCL11A-203
Ensembl Known Protein Coding
3,805bp 798aa
Mus_musculus
Bcl11a
Havana Protein coding
5,989bp 835aa
Bcl11a
Common Known Protein coding
2,989bp 773aa
Bcl11a
Havana Protein coding
2,329bp 243aa
Bcl11a
Ensembl Known Protein Coding
1,464bp 248aa
Bcl11b
Human
Mouse
Homo_sapiens
DNA(contigs)
Ensembl trans.
107.75 Kb
AL359681.5.1.104782 >
<AL109767.7.1.184039
BCL11B-002
Common Known Protein coding
7,603bp 823aa
BCL11B-001
Common Known Protein coding
7,816bp 894aa
BCL11B-003
Havana Protein coding
2,599bp 700aa
Mus_musculus
Bcl11b
Ensembl Known Protein Coding
2,707bp 811aa
Bcl11b
Havana Protein coding
2,522bp 690aa
Bcl11b
Common Known Protein coding
7,707bp 812aa
Bcl11b
Common Known Protein coding
7,923bp 884aa

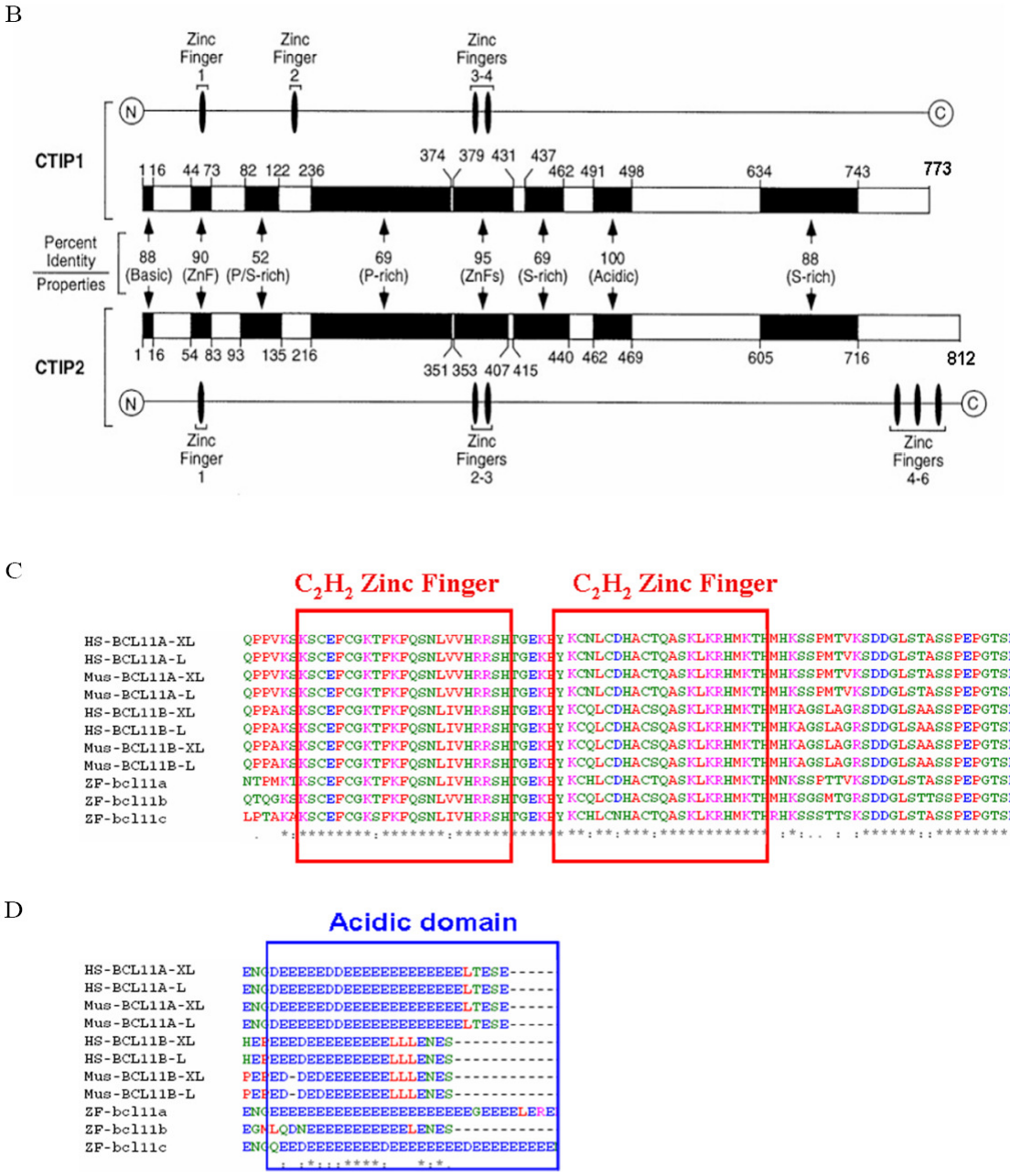


Figure 1.8. Properties of *Bcl11* genes. (A) The human *BCL11A* gene is located on chromosome 2 while the mouse *Bcl11a* gene is found on chromosome 11. The human *BCL11B* gene is located on chromosome 14 while the mouse *Bcl11b* gene is found on chromosome 12. Each gene is predicted to have multiple transcripts and proteins; the length and size of the transcripts and proteins are shown beside the diagrammatic representation of each gene. (Underlined numbers – transcripts and proteins that have been detected experimentally). **(B)** Schematic diagram of Bcl11a and Bcl11b amino acid alignment. The homologous regions are represented by black boxes and the percentage of identity between each region of Bcl11a and Bcl11b is indicated. Bcl11a is predicted to have three C_2H_2 zinc finger domains while Bcl11b is predicted to have an extra three zinc finger domains at its C-terminus. (Figure adapted from Avram *et al.*) (Avram et al., 2000). Clustal w aligment of **(C)** C_2H_2 zinc finger domains and **(D)** acidic domains of human, mouse and zebrafish Bcl11 proteins. These alignments show that the C_2H_2 zinc finger and acidic domains are highly conserved between the three species.

1.3.3 Bcl11 and tumorigenesis

The discovery of the *Bcl11* genes has brought about great interest because their expression has been linked to the etiology of disease in both mice and humans. Over-expression of Bcl11a following proviral integration resulted in the development of myeloid leukaemia in mice (Nakamura et al., 2000). In addition, over-expression of Bcl11a transformed NIH 3T3 cells and this transformation event is thought to be partly facilitated by the physical interaction of Bcl11a with BCL6. Importantly, deregulated expression of *BCL11A* has been consistently found in rare but clinically aggressive cases of B-cell malignancy with chromosomal translocation t(2; 14) (p13; q32.3) (Satterwhite et al., 2001). These data suggest that *Bcl11a* is a proto-oncogene in both the mouse and human. Moreover, recent studies have implicated *BCL11A* in other human diseases, for example, two mutations in *BCL11A* were identified and validated in human breast cancers (Wood et al., 2007) and a quantitative trait locus (QTL) that influences F-cell production (that measures the presence of fetal haemoglobin) was mapped to *BCL11A* in human thalassemia patients (Menzel et al., 2007).

The majority of studies have suggested that, in contrast to *Bcl11a*, *Bcl11b* may be a tumour suppressor. p53 is an important transcription factor that functions as a tumour suppressor and is critical in preventing the genomic instability that leads to cancer. An association between p53 and Bcl11b in mice has been implicated in the development of thymic lymphomas. It was reported that inactivation of *Bcl11b* in normal thymocytes triggered cell proliferation and this was accompanied by a profound p53-dependent apoptosis, whereas *Bcl11b* deficiency appeared to activate cell growth in transformed lymphocytes (Wakabayashi et al., 2003b). In a separate study, preferential inactivation of *Bcl11b* was found in *p53* wild-type lymphomas, suggesting that loss of *Bcl11b* may contribute to oncogenesis only in *p53*-proficient lymphocytes or that it may not be required in p53-null lymphomas (Wakabayashi et al., 2003a). Another functional link between p53 and Bcl11b was demonstrated by Okazuka *et al.* who showed that inactivation of p53 is sufficient for $CD4^-CD8^-$ double-negative ($CD44^-CD25^+$) DN3 stage thymocytes to differentiate into the $CD4^-CD8^+$ immature single-positive, but not to the $CD4^+CD8^+$ double-positive stage of thymocyte development in $Bcl11b^{-/-}$ mice (Okazuka et al., 2005). In addition, $Bcl11b^{+/-}p53^{+/-}$ mice exhibited greater susceptibility to

lymphomas than *Bcl11b*$^{+/+}$*p53*$^{+/-}$ mice, suggesting that functional loss of one *Bcl11b* allele confers a selective advantage for tumour growth (Kamimura et al., 2007). *Bcl11b* has also been shown to be involved in other human disorders, for example BCL11B was shown to repress Tat-mediated transcriptional activation and to inhibit HIV-1 replication in human microglial cells (Rohr et al., 2003). *BCL11B* was also disrupted in a novel chromosomal aberration, inv(14)(q11.2q32.31), in T-cell acute lymphoblastic leukaemia (T-ALL), which resulted in the absence of wild-type BCL11B (Przybylski et al., 2005). Taken together, these studies underline the importance of *Bcl11* genes in human diseases, and underscore the need for further studies on their functional and expression patterns during development.

1.3.4 Hematopoiesis and the roles of *Bcl11* genes in hematopoietic development

As discussed above, hematopoietic malignancies account for most of the abnormalities observed in humans and mice when *Bcl11* genes are dysregulated. As such, the majority of the research to-date has focused on dissecting the roles of *Bcl11* genes in hematopoietic lineages. All the hematopoietic cells, including the red blood cells, platelets and white blood cells, are derived from the same progenitor or precursor cells – the hematopoietic stem cell (HSC) (Figure 1.9A). The self-renewal HSCs are mostly quiescent but can be rapidly mobilized from their bone marrow (BM) niche to proliferate and differentiate into lineages of the innate and adaptive immune systems, as well as into red blood cells and platelets. In the mouse, the contribution of the yolk-sac to the formation of adult HSCs remains controversial and it is thought that most HSCs arise from the aorta-gonand-mesonephros (AGM) region of the embryo (Godin et al., 1995). These HSCs then seed the fetal liver where embryonic hematopoiesis occurs and the stem cells proliferate and differentiate into myeloid and lymphoid lineages (Godin and Cumano, 2002). Subsequently, the HSCs migrate out of the fetal liver and populate the spleen and eventually the BM, where a specialized microenvironment forms – the endosteal niche - where HSCs become quiescent (Godin and Cumano, 2002).

The continual generation and replenishment of the hematopoietic system involves the sequential commitment of the HSCs to the more restricted multi-potent progenitor

cells (lineage-restricted progenitor cells such as common lymphoid progenitors and myeloid progenitors) and eventually to the functionally distinct mature blood cells (Cantor and Orkin, 2001). The myeloid progenitor is the precursor of the granulocytes, macrophages and some dendritic cells of the immune system (Figure 1.9A). Granulocytes (neutrophils, eosinophils and basophils) are involved in the clearing of bacterial and parasitic infections and as such, their production is increased during the immune response. Macrophages are the mature form of monocytes which play a critical role in innate immunity. Dendritic cells are specialized antigen-presenting cells that take up antigens and display them to lymphocytes.

The common lymphoid progenitor gives rise to the lymphocytes: T lymphocytes (T cells), B lymphocytes (B cells) and natural killer cells. T cells can be classified into two main classes: Cytotoxic T cells, which kill cells infected with viruses, and Helper T cells, which differentiate into cells that activate other cells such as B cells and macrophages. T cell progenitors are produced in the BM but migrate to the thymus where they undergo maturation (Rothenberg, 2007a). T cell development proceeds through a series of stages that are well defined by the expression of cell surface makers (Rothenberg, 2007a) (Figure 1.9B). There are two types of thymocytes that are normally present within the thymus: those expressing αβ T cell receptor (αβ T cells) and those expressing γδ (γδ T cells). The αβ T cells later develop into two distinct functional subsets, CD4 and CD8 single positive mature T cells. Development of the prospective αβ T cells proceeds through several stages, one of which is the double-negative (DN) stage, when both CD4 and CD8 are not expressed. This DN stage can be further subdivided into at least four stages (DN1-4) depending on the expression of the adhesion molecules CD44 and CD25 (the α chain of the IL-2 receptor) (Figure 1.9B): the DN1 stage is characterized by expression of CD44 but not CD25; the DN2 stage by expression of both CD44 and CD25; the DN3 stage by expression of CD25 and down-regulation of CD44 and finally, the DN4 stage by down-regulation of both CD44 and CD25 (Rothenberg, 2007a). An important checkpoint occurs at the transition between DN3 to DN4 where thymocytes that failed to produce a functional TCRβ protein undergo apoptosis. Subsequently, thymocytes begin to express both CD4 and CD8 (double-positive, DP) and

eventually, these DP cells mature and express either CD4 or CD8; becoming single-positive thymocytes.

B cell development occurs in the BM, and after immature B cells migrate to the spleen, they further develop into antibody-secreting memory B cells and plasma cells (Hardy and Hayakawa, 2001). Similar to T cell development, B cell maturation occurs through several distinct stages that can be identified by the expression of cell surface markers (Figure 1.9C). The earliest B-lineage cells are known as pre/pro-B cells and can be characterized by the expression of B220 and CD43 cell surface markers. These cells develop into pro-B cells and up-regulate the expression of CD19 (Figure 1.9C). The pro-B cells can be further divided into 'early' or 'late' pro-B cells depending on their expression of BP-1: early Pro-B cells do not express BP-1 but late Pro-B cells do. Next, the pro-B cell develops into a pre-B cell which no longer expresses CD43. Pre-B cells will have successfully undergone immunoglobulin (Ig) loci rearrangement to express IgM. These immature B cells begin to express IgD in addition to IgM and are now known as mature (naïve) B cells. Naïve B cells circulate in the blood and peripheral lymphoid tissues such as the spleen, where upon encountering antigens, these cells differentiate into antibody-producing plasma cells or memory B cells.

Since both *Bcl11a* and *Bcl11b* are involved in hematopoietic disorders, each of these genes was knocked out in the mouse to characterize their roles in hematopoietic development (Liu et al., 2003b; Wakabayashi et al., 2003b). Interestingly, even though Bcl11a and Bcl11b are highly similar in terms of their sequence, structure and physical properties, the knockout mice have different phenotypes. As shown by *Liu* et al., loss of *Bcl11a* in the mouse resulted in a complete block of B cell development, which occurred at the earliest stage of B cell development (Liu et al., 2003b). This phenotype, together with the observation in humans that over-expression of *BCL11A* causes B cell tumourigenesis, demonstrates that *Bcl11a* is essential for B cell development (Satterwhite et al., 2001; Weniger et al., 2006). Although proviral insertion in the *Bcl11a* locus resulted in over-expression of *Bcl11a*, leading to myeloid leukaemia (Nakamura et al., 2000); there was no obvious defect in myeloid development in the knockout mice. In addition to the B cell defects, T cell development was also affected in the absence of *Bcl11a*. The proportion of αβ T cells to γδ T cells was perturbed and there was a three- to

four-fold increase in the number of γδ T cells concomitant with a loss of αβ T cells (Liu et al., 2003b). Loss of *Bcl11a* also affected subsets of αβ T cells as there was an increase in the population of $CD4^-CD8^-$ and $CD4^{lo}CD8^+$ cells and an absence of $CD4^{hi}CD8^-$ thymocytes. Intriguingly, a 'non-cell autonomous tumour suppressor function' for *Bcl11a* was identified in the same study. When mutant *Bcl11a* fetal liver cells were transplanted into lethally-irradiated host mice, the transplanted mice developed $CD4^+CD8^+$ T cell leukaemia three months after transplantation. Further analysis of these tumours showed that most of the cells originated from the wild-type host and not from *Bcl11a* mutant cells (Liu et al., 2003b). This suggests that mutant *Bcl11a* behaves in a cell non-autonomous manner in the development of T cell leukaemia in these mice. In addition, in these T cell leukemic cells, there was a dramatic up-regulation of *Notch1* transcripts. As Notch1 signalling has been shown to inhibit B cell development and promote T cell development (Radtke et al., 2004), a possible explanation for the *Bcl11a* knockout phenotype would be that *Bcl11a* normally inhibits Notch1 to permit B cell development; loss of *Bcl11a* would release the brake on *Notch1*, which would result in a block in B cell development. However, formal proof of this hypothesis is still lacking.

On the other hand, knockout of *Bcl11b* in the mouse resulted in a block in thymocyte development (Wakabayashi et al., 2003b). Only αβ T cell development was affected while γδ T cells and B cells were unaffected in the *Bcl11b* mutant mice. Further analysis showed that rearrangement of the TCRβ locus in *Bcl11b* mutant double negative (DN) thymocytes is incomplete – while they successfully rearranged from D to J, they did not rearrange from V to D. T cell development was predominantly arrested at the DN3 stage ($CD44^-CD25^+$) and some at the immature single positive stage. Host mice transplanted with mutant *Bcl11b* fetal liver cells showed reconstitution of B cells and γδ T cells as well as immature (DN1 and DN2), but not mature T cells, suggesting that the block in T cell development occurred after the γδ T cell lineage commitment. In addition, extensive apoptosis was observed in *Bcl11b* mutant thymocytes. These results suggest that *Bcl11b* is a critical regulator of both development and survival of thymocytes.

Collectively, the knockouts of *Bcl11a* and *Bcl11b* clearly establish their essential roles at different stages of lymphopoietic development and illustrate the importance of transcription factors in hematopoiesis.

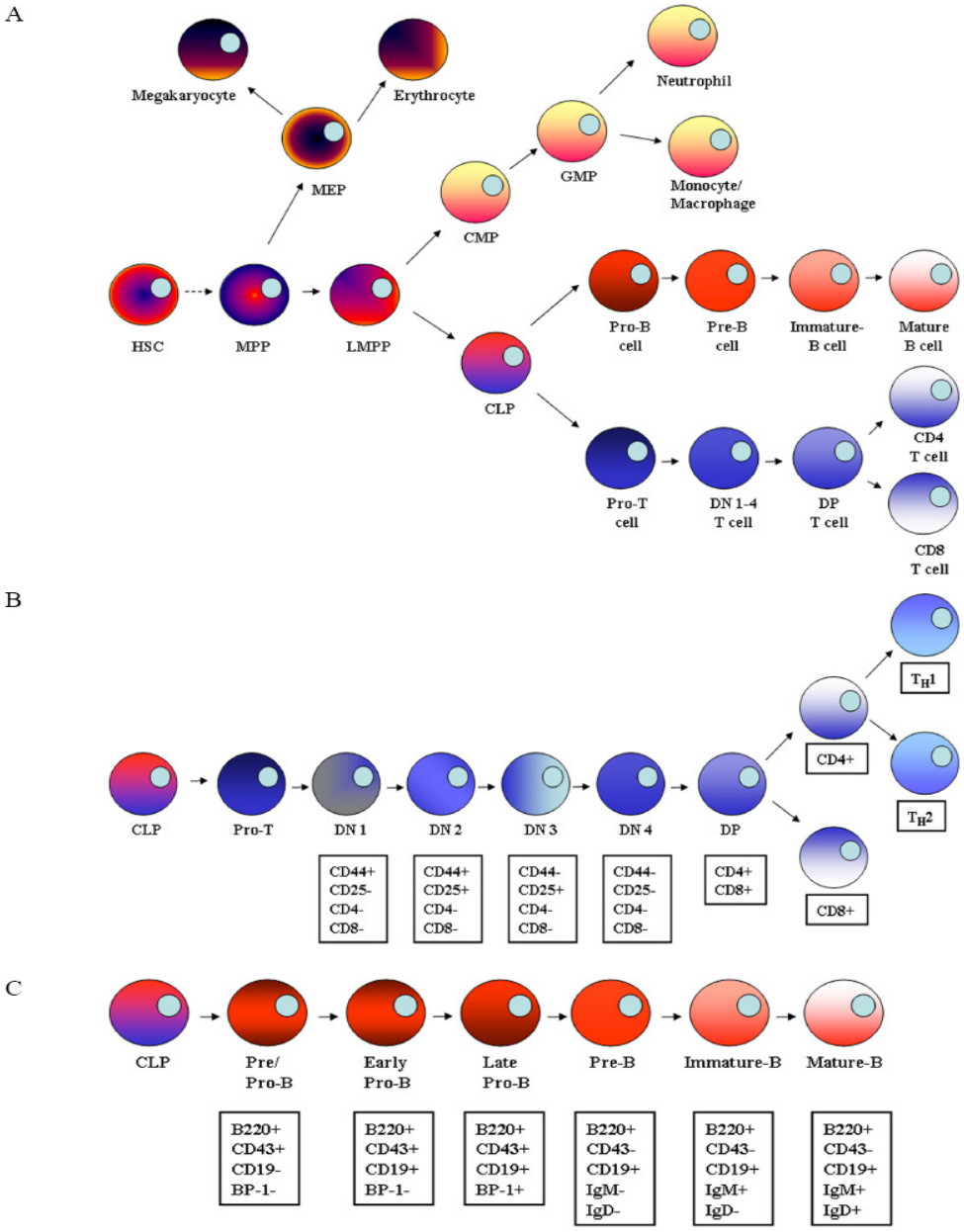


Figure 1.9. Schematic diagram showing hematopoietic development. (A) All lineages of the blood cells including red blood cells and platelets are believed to derive from the hematopoietic stem cells (HSCs). Proliferation and differentiation of the HSCs generate the more lineage-restricted multi-potent progenitors such as megakaryo-erythrocyte progenitor (MEP); common myeloid progenitor (CMP); common lymphoid progenitor (CLP). These lineage-restricted progenitors are responsible for setting up the entire repertoire of the hematopoietic system. **(B)** Schematic diagram depicting various stages of T lymphocyte development. From the CLP, thymocyte development can be defined into different stages characterized by the expression of cell surface markers CD4, CD8, CD44 and CD25. **(C)** Schematic diagram showing B lymphocyte development from the CLP. The various B cell development stages can be defined by expression of cell surface markers B220, CD43, BP-1, IgM and IgD. Abbreviation used: HSC: Hematopoietic stem cell; MPP: Multi-potent progenitor; LMPP: Lymphoid-primed MPP; CLP: Common lymphoid progenitor; DN: CD4/CD8 double-negative T cell; DP: CD4/CD8 double-positive T cell; CMP: Common myeloid progenitor; GMP: Granulocyte/macrophage progenitor; MEP: Megakaryocyte/erythroid progenitor.

1.3.5 Transcription factors in hematopoiesis

Understanding the molecular mechanisms that determine lineage specification of the entire repertoire of blood cells from HSCs is critical to the field of hematopoiesis. Extensive studies have shown that lineage-specific transcription factors play integral roles in hematopoietic lineage decisions (Cantor and Orkin, 2001; Orkin and Zon, 2008; Rothenberg, 2007a). A simplistic view of cell lineage commitment would be that certain lineage-specific transcription factors ('master regulators') act principally in a positive manner to specify lineage choice by activating unique programs of gene expression. Recent studies have shown that this interpretation is overly simplistic; instead, transcription factors act in a combinatorial fashion to direct cell lineage commitment and the establishment of certain cell fate (Cantor and Orkin, 2001; Orkin and Zon, 2008). Transcription factors are key components of complex gene regulatory networks. They orchestrate the entire developmental process through a series of co-ordinated hierarchical stages, allowing the determination of different cell fates from multi-potent stem or progenitor cells. Gene regulatory networks consist of many different yet interconnected hierarchical states and each of these different states is determined by a unique combinatorial code of transcription factors (Singh et al., 2005). Higher up the regulatory hierarchy, expression of certain transcription factors at a crucial time is critical to activate target genes that are the executioner of the lineage specification process. For example, Notch1 which is important for the cell-lineage specification of T cells from the common lymphoid progenitor (CLP), is not required to maintain γδ T cells past the DN2 stage nor for αβ T cell development after β-selection, except for a much later role in promoting T helper 2 (T_H2) subset differentiation in response to antigen (Artavanis-Tsakonas et al., 1999; Radtke et al., 2004). In both the initial T cell specification and T_H2 cell differentiation, Notch1 works in part by inducing expression of *Gata-3* (Amsen et al., 2007; Fang et al., 2007). However, in the CD4 versus CD8 T cell lineage decision, Notch1 is not required for the Gata-3 dependent determination of CD4 cell fate (Radtke et al., 2004; Rothenberg, 2007a). Taken together, these studies illustrate two pertinent points: (1) homeostasis of the level of transcription factors at each cell state is critical to initiate and define the specific cell fate; and (2) temporal and spatial regulation of the expression of these transcription factors is also vital to the cell-fate decisions.

In addition, it is becoming apparent that lineage-specific transcription factors exhibit cross-antagonism. Whilst specifying a certain lineage by activating lineage-specific genes, transcription factors simultaneously exert inhibitory effects on the alternative lineage gene programs by directly antagonizing the actions of opposing transcription factors. Thus, there appears to be a dynamic balance of forces that ultimately determine the phenotype of a cell. Such cross-antagonism of transcription factors in determining the cell fate has been reported in the hematopoietic system. For example, *Blimp1* and *Pax5* act in a mutually repressive opposing fashion to control the mature B cell versus plasma cell developmental switch (Davis, 2007); in T-helper cells, *Gata-3* and *T-bet* specify the T_H2 and T_H1 cell fates, respectively (Ho and Pai, 2007; Zheng and Flavell, 1997).

In summary, the gene regulatory networks play critical roles in establishing the hematopoietic system; precise, co-ordinated regulation of the levels of transcription factors facilitate the cell-fate decisions of progenitors and thus the generation of the entire repertoire of differentiated hematopoietic progenies from the HSC. However, the 'information content' of a transcription factor's activity depends on the cellular context and is not permanent or fixed for a particular lineage. This means a transcription factor can participate in the control of completely different sets of gene targets (Orkin and Zon, 2008; Rothenberg, 2007a). The functional flexibility of transcription factors depends on 'local factors' such as the chromatin status of a cell, which controls the accessibility of different target genes in different cell types, thereby allowing the same transcription factor to perform different functions at respective anatomical locations. Frequently, key transcription factors have been found to play essential roles in different tissue systems. For example, in addition to its role in T cell development, *Gata-3* is also important for skin, nervous system, adrenal and jaw development (Kaufman et al., 2003; Lim et al., 2000; Zheng and Flavell, 1997). Recently, *Gata-3* has also been shown to be a critical factor in epithelial cell-lineage determination in the mammary gland (Asselin-Labat et al., 2007; Kouros-Mehr et al., 2006). These studies also demonstrate that the genetic control between lymphocyte development and mammary development is likely to be conserved.

1.4 Mammary gland biology

The mammary gland serves to provide nourishment and passive immunity via milk to newborns until they are capable of independent feeding. In order to fulfil its function, the mammary gland has to develop an extensive network of milk-producing cells and interconnecting ducts to produce and deliver the milk to the young. Development of the mammary gland occurs in distinct phases that can be broadly classified into embryonic, postnatal (pre-pubertal and pubertal), pregnancy, lactation and involution stages. The entire developmental cycle involves specification of progenitors, proliferation and differentiation of these progenitors to generate the entire mammary gland and also apoptosis, de-differentiation and tissue remodelling during post-lactational regression. Hence the mammary gland is an excellent model system to study the gene regulatory networks that control specification, proliferation, differentiation, survival and death of cells. Understanding the molecular mechanisms that control these processes is important not only from a developmental biologist's point of view but also allows us to elucidate how dysregulation of these processes can lead to breast cancer. Eventually, this may facilitate the design of better therapeutics to treat breast cancer, which is the most commonly occuring cancer in women in the UK (Cancer Research UK). The mammary developmental process is tightly regulated by temporal and spatial stimuli, such as steroid and peptide hormones, through various signalling networks (Hennighausen and Robinson, 1998, 2005). Interestingly, recent studies have shown that transcription factors that are important in hematopoiesis also function in various aspects of mammary lineage commitment (Asselin-Labat et al., 2007; Khaled et al., 2007; Kouros-Mehr et al., 2006; Oakes et al., 2008). The relative contributions of hormonal regulation and transcription factors to mammary lineage commitment, differentiation and development of the mammary gland will be discussed in greater detail below.

1.4.1 Embryonic development

1.4.1.1 Overview of development

The initial stages of mammary development are independent of systemic cues and instead depend on reciprocal signalling between the mammary epithelium and the underlying mesenchyme. Embryonic mammary gland development in the mouse is

initiated around mid-gestation, at 10.5 days post-coitum (dpc), with the formation of mammary (milk) lines that run in an anteroposterior fashion along the ventral side of both the male and female embryos (Figure 1.10A). By 11.5 dpc, the mammary lines have given way to five pairs of mammary placodes that form at specific locations along the mammary line which eventually invaginate into the underlying mesenchymal layer to form mammary buds at 12.5 dpc (Figure 1.10B). Further development of the mammary bud and the surrounding mesenchymal cells continues between 13.5-15.5 dpc, when the mesenchymal cells around the mammary bud condense and differentiate to form the mammary mesenchyme and when the mammary bud begins to elongate and form a sprout (Figure 1.10C). Importantly, while the female embryonic mammary gland remains quiescent until 15.5 dpc, the dense mammary mesenchyme in the male embryos expresses androgen receptors that respond to testosterone and result in the disruption of the mammary anlage (Kratochwil and Schwartz, 1976). Between 15.5-18.5 dpc, the mammary sprout continues to penetrate into the underlying fat pad and undergoes arborization to eventually form a rudimentary ductal tree consisting of a primary duct and 15-20 ductules at birth (Hens and Wysolmerski, 2005).

1.4.1.2 Specification of the milk line

The milk line consists of an ectodermal thickening that arises in the skin and can be visualised using *in situ* staining for the wingless gene, *Wnt10b*, which is one of the earliest markers of mammary lineage (Veltmaat et al., 2004). Using the TOP-GAL Wnt reporter line (Chu et al., 2004), the dynamics of early mammary gland development was monitored and it was found that specification and formation of the milk line depends on canonical Wnt signalling. Forced activation of Wnt signalling promotes development of mammary placodes while inhibition of Wnt signalling blocks initiation of placode formation (Chu et al., 2004). Following the specification of the milk line, a cascade of Wnt expression was shown to be induced (includes *Wnt10b*, *Wnt10a* and *Wnt6*) within the milk line and eventually becomes localized to the mammary placodes (Chu et al., 2004). In addition to Wnt signalling, the fibroblast growth factor (Fgf) signalling pathway has also been suggested to be important for specification of the milk lines. Between 10.5-11.5 dpc, *Fgf10* is expressed in the most ventral-lateral reaches of the dermatomyotome

of the somites adjacent to the developing mammary line while *Fgfr2b* is expressed within the mammary epithelial placodes (Mailleux et al., 2002). Knocking out either *Fgf10* or its receptor *Fgfr2b* in the mouse disrupted four out of the five pairs of the mammary placodes (Numbers 1, 2, 3 and 5) (Mailleux et al., 2002). In addition, inhibition of Wnt signalling does not alter the expression of *Fgf10* or *Fgfr1* signalling (Mailleux et al., 2002), suggesting that Fgf signalling is important for the earliest stage of mammary development and probably acts in concert with Wnt signalling, rather than downstream of it. The T-box transcription factor, *Tbx3,* is also expressed in the mammary line from 10.25-10.5 dpc and formation of mammary placodes is abolished in *Tbx3* knockout mice (Eblaghie et al., 2004). In addition, *Tbx3* expression was shown to be induced by both Fgf and Wnt signalling within the mammary line of cultured mouse embryos (Eblaghie et al., 2004), thus establishing a link between *Tbx3* and the Fgf and Wnt pathways in specification of the milk line.

1.4.1.3 Formation of placodes

By 11.5 dpc, the mammary lines have given way to five pairs of mammary placodes that appear at specific positions along the ventral side of the embryos (Figure 1.10B). Placodes are visible as lens-shaped structures and consist of an ectodermal thickening. Even though the placodes appear at highly reproducible positions in the embryos, they are formed asynchronously in a specific order: the third pair is the first to appear, followed by the fourth pair, and then the first and fifth pairs, which appear simultaneously, and eventually the second pair appears. Using knockout mice, several genes were shown to be implicated in the formation of the mammary placodes as deletion of these genes resulted in mice missing all or some of the mammary placodes. As discussed above, loss of *Fgf10*/*Fgfr2* and *Tbx3* resulted in absence of certain mammary placodes (Eblaghie et al., 2004; Mailleux et al., 2002). Lymphoid enhancing factor 1 (*Lef1*) is a nuclear target of the canonical Wnt signalling pathway where it interacts with β-catenin (Behrens et al., 1996). It was found that *Lef1*-null mice form primitive mammary placodes that eventually degenerate, although the fourth pair of placodes is sometimes retained (van Genderen et al., 1994). This highlights the importance of *Lef1* as a survival factor in the mammary anlage. In the scaramanga (*ska*) mutant mouse, aberrant

bud formation was observed. Interestingly, in addition to the absence of the third pair of mammary buds, supernumerary buds were observed at the location normally occupied by the fourth pair of mammary buds (Howard et al., 2005). The '*ska'* gene was later identified as the neuregulin 3 (*Nrg3*) gene, which encodes the ligand for the *Erbb4* (*Her4*) tyrosine kinase receptor. It has been proposed that Nrg3 could function to transmit signals from Fgf10 and/or Tbx3 to the precursor mammary epithelial cells which appear to express Wnt10b (Veltmaat et al., 2004).

1.4.1.4 Bud formation

From 11.5-12.5 dpc, several morphogenic processes in the embryonic mammary gland begin to occur. Firstly, the mammary placodes began to invaginate into the underlying mesenchymal layer and develop into mammary buds. Secondly, the mesenchymal cells surrounding the mammary bud condense and differentiate into the dense mammary mesenchyme that surrounds the mammary bud. In addition, the mammary mesenchyme up-regulates the expression of androgen receptors which are important for the regression of male mammary buds in response to androgens (Kratochwil and Schwartz, 1976). As mentioned above, *Lef1* mutant embryos have primitive placodes but they eventually regress by 12.5 dpc, suggesting that Wnt signalling continues to play an important role in mammary bud development. Studies of TOP-GAL reporter mice showed that Wnt signalling is detectable in the mammary bud right up to 15.5 dpc: *Lef1* is expressed in the mammary placodes and bud until 15.5 dpc and in the mammary mesenchyme at 14.5 dpc. The homeodomain-containing transcription factors, *Msx1* and *Msx2* are both expressed in the mammary buds with the latter also expressed in the mammary mesenchyme (Satokata et al., 2000). *Msx* genes appear to have necessary but redundant functions in mammary development: knockout of a single *Msx* gene did not affect mammary bud formation (Satokata et al., 2000) but knocking-out both genes simultaneously disrupted mammary bud formation (Satokata et al., 2000). Recently, the hedgehog (Hh) signalling pathway has also been implicated in mammary bud formation (Hatsell and Cowin, 2006). *Gli3*, which encodes a microtubule-bound transcription factor, is an effector of the Hh signalling pathway (Hooper and Scott, 2005). Depending on its phosphorylation status, it can either become a transcription

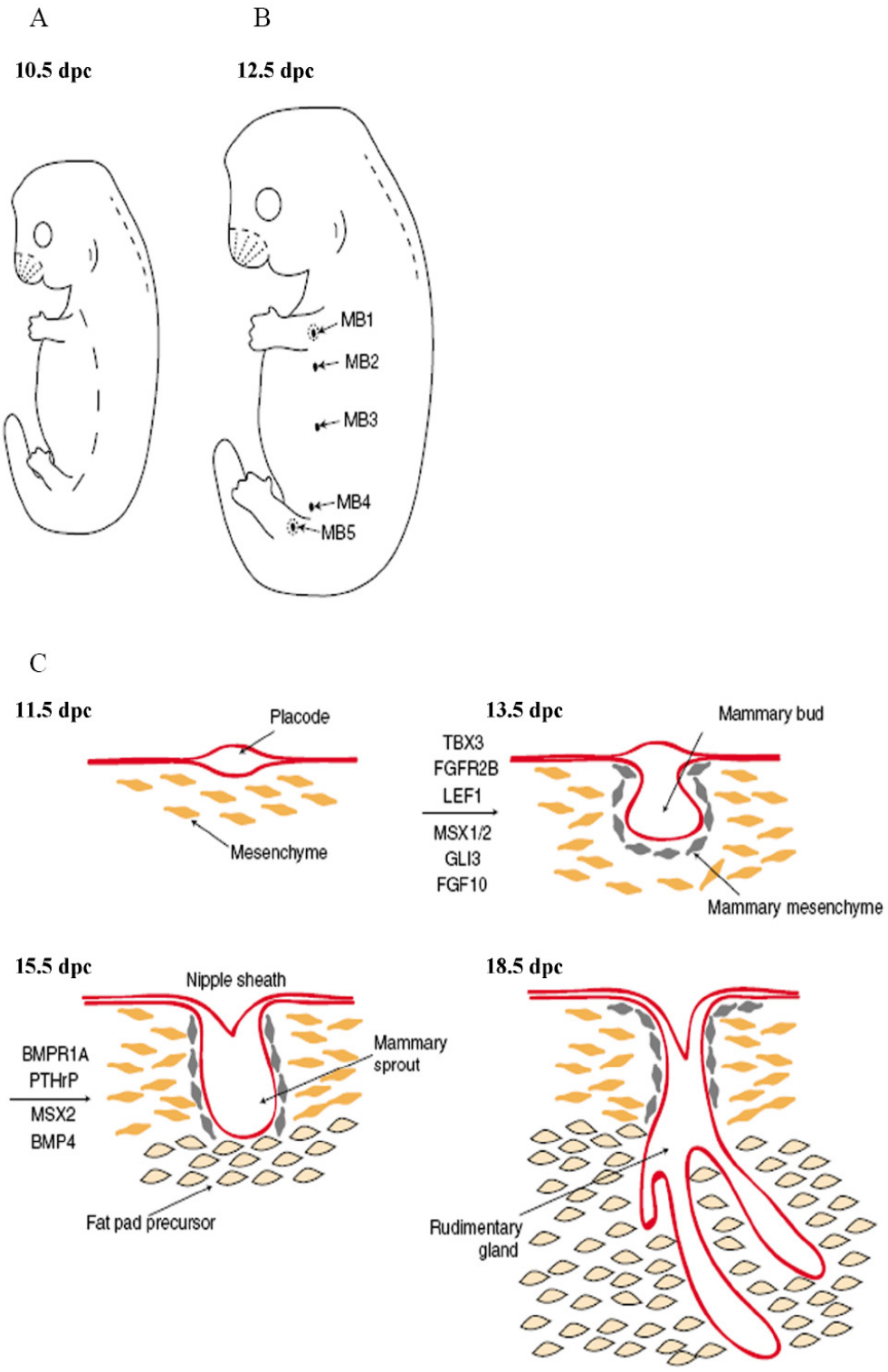
activator (Gli^A; phosphorylated form) or a transcription repressor (Gli^R; cleaved form) to mediate the signals from interactions between the secreted ligands (sonic hedgehog, Indian hedgehog or desert hedgehog) and their receptors (Patched1 and Patched2) (Hooper and Scott, 2005). The Gli^A/Gli^R ratio of Gli3 was found to be required to repress Hh target genes that are involved in patterning and bud formation for the third and fifth pairs of mammary buds (Hatsell and Cowin, 2006), suggesting that different inductive signals are required for different mammary buds. Another transcription factor, *Gata-3*, has also been shown to be essential for mammary bud formation. Expression of *Gata-3* was detected in the mammary buds from 12.5 dpc and conditional deletion of *Gata-3* using Cre recombinase driven by the keratin 14 (K14) promoter resulted in variable loss of mammary placodes and a failure to form the nipple sheath (Asselin-Labat et al., 2007).

1.4.1.5 Sprout formation and early ductal morphogenesis

Between 13.5-15.5 dpc, the mammary bud begins to elongate and penetrate into the underlying fat pad to form a sprout-like structure and nipple formation is initiated (Figure 1.10C). Epithelial-mesenchymal signalling through parathyroid hormone related peptide (PTHrP) and its receptor (PTHrPR) plays a critical role in mammary sprout and nipple sheath formation (Foley et al., 2001). PTHrP is expressed and secreted from the mammary epithelial bud which then binds to PTHrP receptors found in the mammary mesenchyme (Figure 1.10D) (Foley et al., 2001). This up-regulates the expression of bone morphogenic protein receptor 1A (*Bmpr1A*) in the mammary mesenchyme which makes the mesenchyme responsive to bone morphogenic protein 4 (Bmp4) signalling (Figure 1.10D). Binding of Bmp4 to Bmpr1A stimulates epithelial bud elongation and also up-regulates the expression of *Msx2*. This in turn inhibits the formation of hair follicles within the nipple sheath and permits the development of the nipple sheath (Foley et al., 2001). Thus, PTHrP is the first mammary epithelium-specific secreted signalling molecule that has been found to influence cell fate decisions in the surrounding mesenchyme to-date. From 15.5 dpc, the elongation and arborisation of the sprout continues to be stimulated by the interactions between PTHrP and Bmp4 signalling pathways. At birth, the mammary sprout has developed into a small rudimentary

mammary network with a primary duct and about 10-15 ductules. Mammary gland development is then arrested and does not recommence until puberty.

In summary, from the appearance of the milk line at 10.5 dpc to the formation of the rudimentary mammary tree structure at birth, embryonic mammary gland development is a highly co-ordinated developmental process that requires the interplay between several key signalling pathways such as the Wnt, Fgf, Bmp and PTHrP pathways. Studies have also shown that the interaction between the developing mammary bud and the surrounding mammary mesenchyme is also critical in ensuring proper mammary development. A more detailed functional analysis of genes that are expressed in the mammary buds and/or mammary mesenchyme is being studied in mouse knockouts and this would undoubtedly add new players to the mammary developmental process. Therefore, one major challenge would be to integrate these new genes into the existing signalling pathways in order to obtain a better understanding of the initiation, specification and differentiation process of formation of the embryonic mammary gland.


A
B
10.5 dpc
12.5 dpc
MB1
MB2
MB3
MB4
MB5
C
11.5 dpc
Placode
Mesenchyme
13.5 dpc
TBX3
FGFR2B
LEF1
MSX1/2
GLI3
FGF10
Mammary bud
Mammary mesenchyme
15.5 dpc
Nipple sheath
BMPR1A
PTHrP
MSX2
BMP4
Mammary sprout
Fat pad precursor
18.5 dpc
Rudimentary gland

D

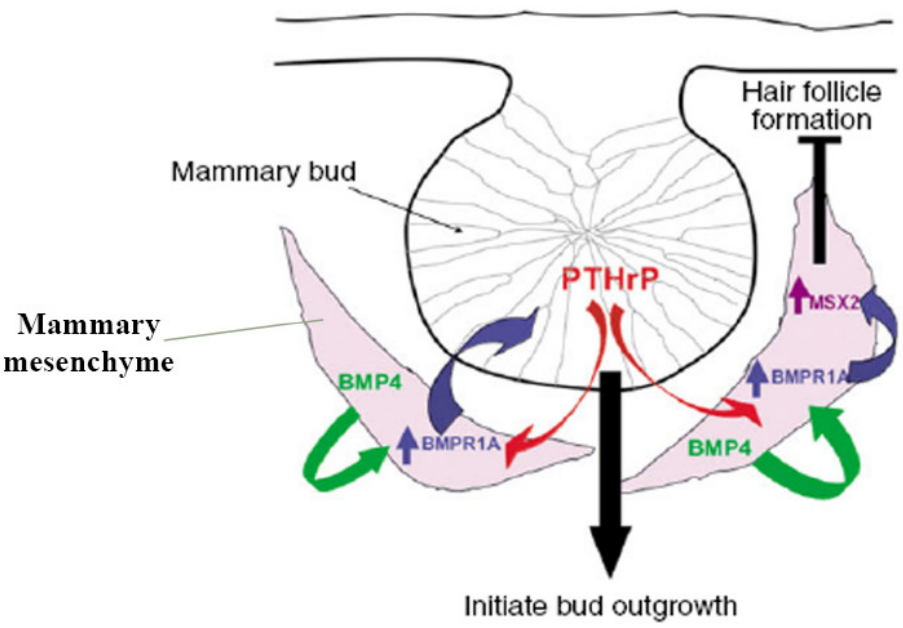


Figure 1.10. Overview of murine embryonic mammary gland development. (A) Diagrammatic view of the presumptive locations of the mammary milk lines which appear in an anteroposterior axis at around 10.5 dpc. **(B)** At 12.5 dpc, five pairs of mammary placodes which have become mammary buds (MB1-5) are visible at specific positions along the anteroposterior ventral side of the embryo. MB1 and MB5 are hidden behind the fore- and hind-limb respectively. **(C)** Overview of the murine embryonic development process from the milk line. At 11.5 dpc, the milk line has developed into an ectodermal thickening known as the placode. By 13.5 dpc, the placodes have invaginated into the underlying mesenchymal layer (orange) to form mammary buds. The mesenchymal cells (orange) that enclose the mammary buds begin to condense to form the mammary mesenchyme (grey). By 15.5 dpc, these buds have developed into sprouts which form a lumen with an opening to the skin, characterized by the formation of the nipple sheath. By 18.5 dpc, the mammary gland is present as a rudimentary arborized structure which has several small ducts that have invaded the mammary fat pad (buff). **(D)** Schematic cross-section of a mammary bud between 13.5-15.5 dpc. Within these time points, mammary buds begin to elongate and invaginate into the underlying fat pad stimulated by the reciprocal epithelial-mesenchymal signalling between PTHrP and its receptor. Binding of the PTHrP (secreted from the mammary bud) to its receptor (expressed on mammary mesenchyme) up-regulates expression of BMPR1A in the mammary mesenchyme which is now responsive to BMP4 signalling. This stimulates epithelial outgrowth, elevates *MSX2* expression, and inhibits hair follicle formation within the nipple sheath. Images obtained from (Watson and Khaled, 2008).

1.4.2 Postnatal morphogenesis

Unlike other organs, development of the mammary gland is completed during adolescent rather than during embryonic development. Mammary ductal elongation and branching develops predominantly after puberty, whereas alveolar proliferation occurs during pregnancy and functional differentiation of these alveolar cells is completed with parturition and lactation. Prior to the onset of puberty (which occurs at about four weeks after birth), the mammary gland enters an allometric growth phase where the mammary ducts elongate into the fat pad at a similar rate to the overall growth of the animal (Figure 1.11A). Accelerated ductal elongation and side branching occurs at the start of puberty, about four weeks after birth. Terminal end buds (TEBs), which are club-shaped-like structures found at the leading edges of the mammary ducts, are prominent throughout this stage of development; they lead to the elongation and penetration of the ducts through the fat pad. The TEB is a specialized structure that consists of two histologically distinct cell types: (1) cap cells, which are precursors of myoepithelial cells and are thought to contain mammary stem/progenitor cells, and (2) body cells which give rise to mammary epithelial cells (Figure 1.11A) (Humphreys et al., 1996). The TEB plays an important role in the generation of the mature mammary duct. A highly regulated process of proliferation and apoptosis in the TEB generates a hollow lumen surrounded by a single layer of luminal epithelial cells that serves as a channel for the transport of milk during lactation and the surrounding myoepithelial layers, which serve to provide contractile forces to facilitate transport of milk to the nipple (Humphreys et al., 1996). Additionally, ductal arborization is initiated from the highly proliferative TEBs which form secondary and tertiary branches (Daniel et al., 1987). At the end of puberty, the TEBs have disappeared and the entire fat pad is now invaginated by a network of mammary ducts (Figure 1.11B).

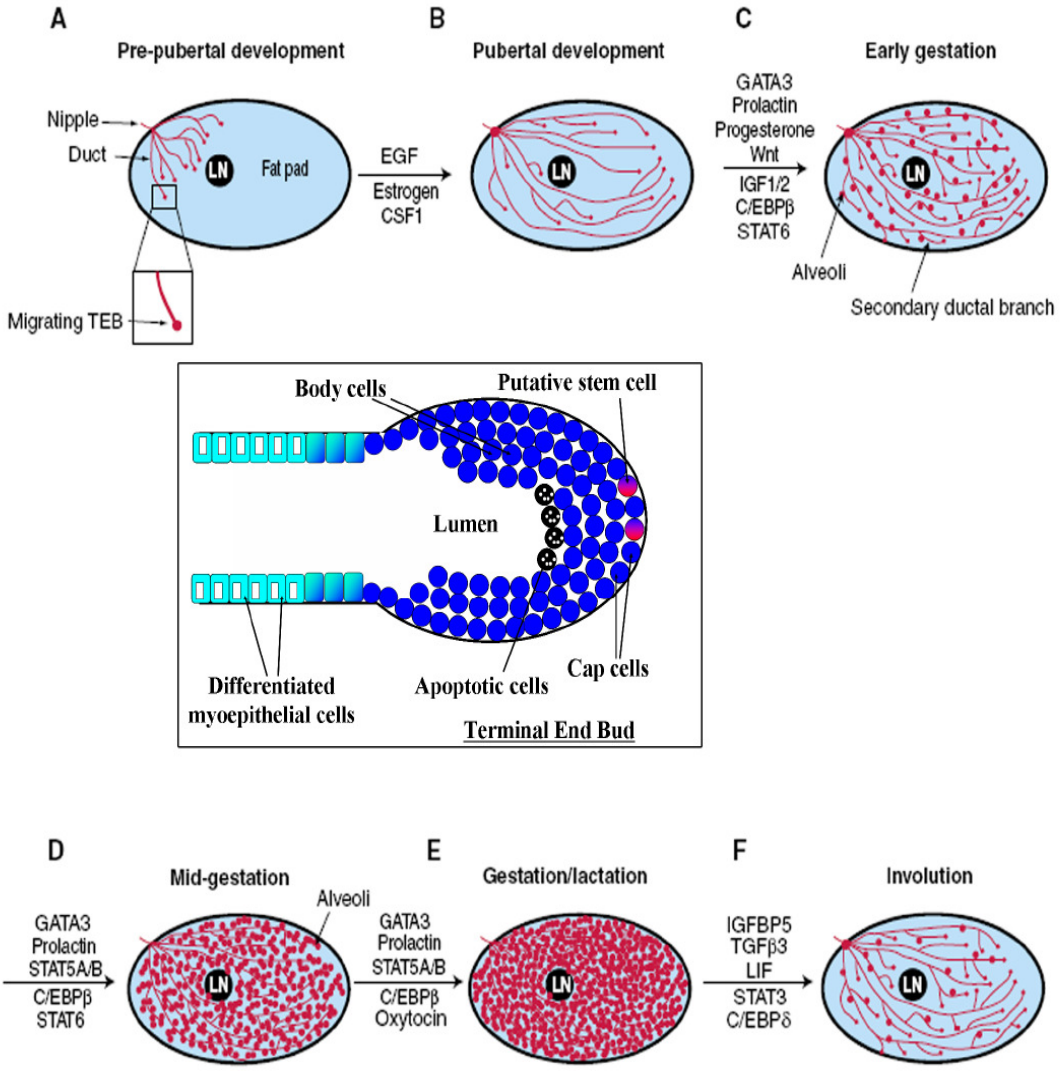


Figure 1.11. Overview of murine adult mammary development. Schematic diagrams depicting the major phases of adult mammary gland development: **(A)** Pre-pubertal, **(B)** Pubertal, **(C)** Early gestation, **(D)** Mid-gestation, **(E)** Gestation/Lactation and **(F)** Involution. Some of the important factors that are involved in the various phases are highlighted. Insert shows a diagrammatic representation of the cap and body cell layers of TEB. LN: lymph node; TEB: terminal end bud.

1.4.2.1 Hormonal regulation

From the onset of puberty, functional development of the mammary gland is essentially defined by the hormonal status of the animal. The combinatorial effects of systemic hormones such as growth hormone (GH), estrogen and progesterone, together with the stroma are major influences in the development of the postnatal mammary gland (Sakakura et al., 1987). Pituitary GH, which is already present before the pubertal up-regulation of estrogen, acts via its receptor on mammary stromal cells to elicit expression of insulin-like growth factor-1 (IGF-1). IGF-1 then stimulates TEB formation and epithelial branching in a paracrine manner (Sternlicht, 2006).

Estrogen plays an important role in branching morphogenesis in the virgin gland. There are two estrogen receptors in the mouse, ERα and ERβ, with ERα playing a more important role in ductal morphogenesis. Development of the murine mammary ductal network is completed by four months, whereas the mammary glands of the ERα knockout females displayed a rudimentary epithelium devoid of any TEBs, demonstrating that estrogen is required for ductal outgrowth (Bocchinfuso and Korach, 1997; Korach, 2000). Interestingly, embryonic tissue recombination studies suggest that only stromal ERα is required, whereas adult tissue transplants indicate that both epithelial and stromal ERα are necessary for ductal morphogenesis (Cunha and Hom, 1996; Mueller et al., 2002). However, this original ERα knockout was not a true null allele as alternative splicing generated a trans-activation protein and a failure of estrogen signalling through the hypothalamic/pituitary axis resulted in reduced levels of prolactin, which complicates analysis (Bocchinfuso and Korach, 1997). Subsequently, a complete ERα knockout mouse was generated and TEBs were completely absent and ducts failed to penetrate the mammary fat pad (Mallepell et al., 2006). Recently, conditional deletion of ERα at various stages of mammary development demonstrated that estrogen is required for both pubertal development and development of alveolar structures during late gestation and lactation (Feng et al., 2007).

Progesterone mediates its effects through two receptors, PR-A and PR-B, which are encoded by a single gene containing two distinct promoters. Deletion of both PR isoforms in the mouse resulted in severe defects in the reproductive tissues (Lydon et al., 1995). Mammary gland development was also affected in PR knockout mice. While

ductal growth was normal, side branching was nearly completely absent. The main effects of progesterone appear to be during pregnancy (See Chapter 1.4.3.1). Collectively, the combined effects of GH, estrogen and progesterone ensure the proper development of the ductal system - setting up a mature mammary network that is primed for further proliferation and differentiation during pregnancy.

1.4.3 Lobulo-alveolar development

The most dramatic changes in the mammary gland occur during pregnancy in preparation for the birth of the pups. From the onset of pregnancy, the mammary epithelial cells, stimulated by prolactin and placental lactogens, proliferate and undergo differentiation to form lobulo-alveolar structures that eventually fill up the entire fat pad (Figure 1.11C-F). The maximal rate of proliferation of the epithelial cells is observed between days six and ten of gestation. Extensive branching morphogenesis occurs from day six of gestation, producing secondary and tertiary side branches from the primary duct; this is followed by proliferation and differentiation of alveolar buds (Figure 1.11C). By mid-gestation, clusters of alveoli are evident throughout the entire fat pad and further differentiation of the alveoli results in the production of milk proteins such as β-casein (Figure 1.11D) (Hennighausen and Robinson, 2005; Watson and Khaled, 2008). By late gestation, the mammary epithelial cells account for about 90% of the cells in the mammary gland, most of which have began to produce other milk proteins such as whey acidic protein (WAP) (Watson and Khaled, 2008).

1.4.3.1 Progesterone signalling

In the PR knockout mouse, ductal growth was normal, however, pregnancy induced side branching and alveolar development were completely absent, suggesting that progesterone is necessary for side branching and alveolar differentiation (Lydon et al., 1995). Reciprocal transplantation experiments showed that the effects of progesterone are mediated via the mammary epithelium. Notably, Wnt4 was shown to be required for tertiary side-branching and is also regulated by progesterone, hence Wnt4 might function downstream of the progesterone receptor (Brisken et al., 2000).

1.4.3.2 Prolactin signalling

Prolactin signalling is essential for the proliferation and functional differentiation of lobulo-alveolar structures during pregnancy (Topper and Freeman, 1980). Prolactin is produced and secreted from the anterior pituitary glands of the mouse and binding of prolactin to its receptor results in receptor dimerization and activation of the Janus kinase 2 (JAK2), Fyn, and the mitogen-activated protein (MAP) kinase *in vitro* (Hennighausen and Robinson, 1998). The *in vivo* effects of prolactin signalling are mediated predominantly through the JAK-Stat (signal transducer and activator of transcription) pathway, which will be discussed in detail below (Chapter 1.4.3.3) (Liu et al., 1997). Deletion of *Prl* (prolactin) in the mouse curtailed ductal arborisation in adult virgins, suggesting that ductal morphogenesis is affected; this could however be due to indirect effects because *Prl*-deficient females do not ovulate (Horseman et al., 1997). However, the effects of *Prl* deletion during pregnancy and lactation cannot be examined because *Prl*-deficient mice are infertile. Using *PrlR* (prolactin receptor) knockout mice, Ormandy *et al.* showed that prolactin signalling is important for mammary gland development (Ormandy et al., 1997). Females with only one copy of *PrlR* allele failed to lactate after their first pregnancy, indicating that levels of prolactin are important for epithelial cell proliferation and differentiation (Ormandy et al., 1997). However, mammary gland development was normal after the second pregnancy or in older females, and dams were able to lactate and nurture their pups. This demonstrates that continued hormonal stimuli were sufficient to ensure complete functional development of the mammary gland. Collectively, these studies demonstrate that prolactin signalling is essential for reproductive functions and mammary gland development in the mouse. This was confirmed by studying various mouse knockouts of different components of the prolactin signalling pathway, in particular, components of the JAK-Stat pathway.

1.4.3.3 JAK/Stat signalling

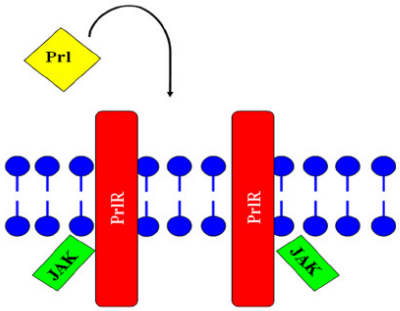
Prolactin signalling in the mammary gland is mediated largely through the JAK-Stat pathway (Liu et al., 1997). Binding of prolactin to its receptor results in receptor dimerization and activation of the receptor-associated kinase (JAK) which then phosphorylates specific tyrosine residues of the receptor (Figure 1.12A and 1.12B).

Subsequently, Stats are recruited to the receptor through their Src-homology-2 (SH2) domains and become phosphorylated on tyrosine 694 by JAK (Figure 1.12C). Phosphorylated Stats form hetero- and/or homo-dimers and translocate into the nucleus where they activate transcription of target genes (Figure 1.12D) (Hennighausen and Robinson, 1998, 2005). There are seven different members of the Stat family (Stat1, Stat2, Stat3, Stat4, Stat5a, Stat5b and Stat6) and four known JAK proteins (JAK1, JAK2, JAK3 and JAK4) (Darnell, 1997). Most Stats bind to and induce transcription of genes containing γ-interferon activation sites (GAS), $TTCn_3GAA$ with only Stat6 binding to a unique consensus sequence $TTCn_4GAA$ (Darnell, 1997). Stat proteins have six conserved domains with the DNA binding domain in the centre and the SH2 domain at the C-terminus of the protein. Despite the common consensus binding sequence, Stats proteins achieve specificity in their activation of target genes by the different combinations of co-activators and co-repressors that are associated with Stats.

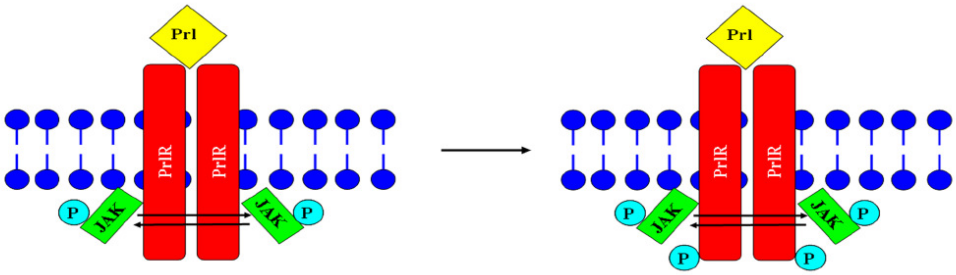
Two major mechanisms function to prevent precocious or sustained JAK-Stat signalling during pregnancy: (1) suppressor of cytokine signalling (SOCS) family of proteins and (2) caveolin-1 (Cav1). There are eight members of the SOCS family [SOCS1-7 and cytokine-inducible SH2 protein (CIS)] that interact with JAKs (Figure 1.12E1) and cytokine receptors (Figure 1.12E2) to control the activation of Stat proteins (Kubo et al., 2003). *In vitro* studies showed that SOCS proteins bind to tyrosine residues in cytokine receptors and block the binding and hence the activation of Stat proteins. SOCS1 has been shown to be a critical negative regulator of alveologenesis (Lindeman et al., 2001). Loss of *Socs1* in the mouse resulted in excessive alveologenesis during gestation and elevated Stat5 activity during lactation, suggesting that SOCS1 normally inhibits prolactin signalling. To demonstrate the direct modulation of prolactin signalling by SOCS1, Lindeman *et al.* showed that an equivalent loss of one *Socs1* allele in *PrlR* hemizygous mice rescued alveologenesis and Stat5 activity (Lindeman et al., 2001). Similarly, homozygous null mutants of *Socs2* rescued the alveologenesis defects in *PrlR* hemizygous mice, suggesting that SOCS2 is also a negative regulator of prolactin signalling (Harris et al., 2006). SOCS3 proteins function in a negative-feedback-loop that involves Stat5 and the gp130 receptor that is the shared subunit of cytokine receptors for interleukin 6 (IL6) and leukaemia inhibitory factor (Lif). Activation of gp130 signalling

occurs during involution which leads to activation of Stat3 (Humphreys et al., 2002). Expression of *Socs3* is up-regulated by Stat5 during pregnancy, resulting in the binding of SOCS3 to the tyrosine residues of gp130 to prevent the activation of Stat3 (Humphreys et al., 2002). *Socs5* and *Cav1* null mutants showed accelerated alveolar development and dramatic activation of Stat5 leading to precocious milk production in the pregnant gland (Khaled et al., 2007). Hence regulation of the activity of SOCS and Cav-1 proteins is also critical to the maintenance and duration of the JAK-Stat signalling pathway. Other negative regulators of JAK-Stat pathways in the mammary gland include SH2 containing phosphatases (SHP) and protein inhibitor of activated Stat (PIAS) family (Wormald and Hilton, 2004).

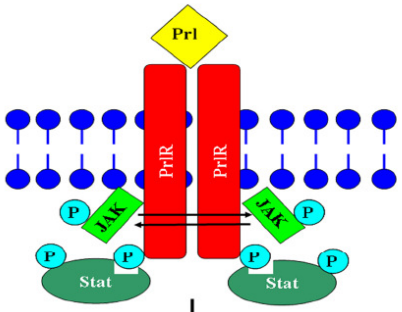
A


Prl
PrlR
PrlR
JAK
JAK

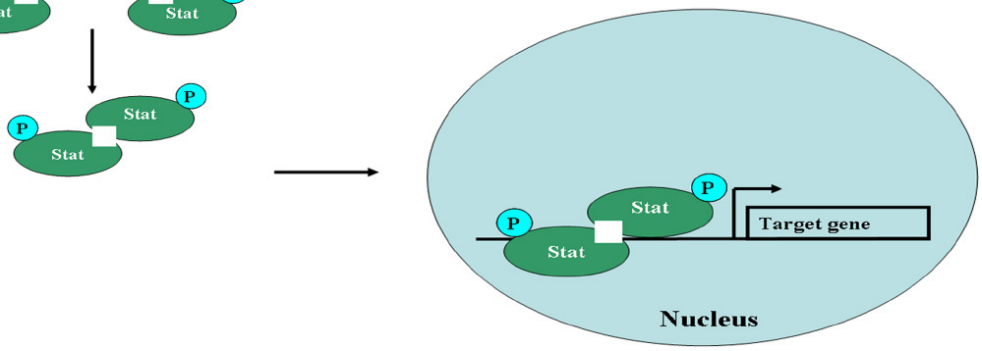
B


Prl
PrlR
PrlR
JAK
JAK
P
Prl
PrlR
PrlR
JAK
JAK
P

C


Prl
PrlR
PrlR
JAK
JAK
P
Stat
Stat

D


Stat
Stat
Stat
Stat
Target gene
Nucleus

E

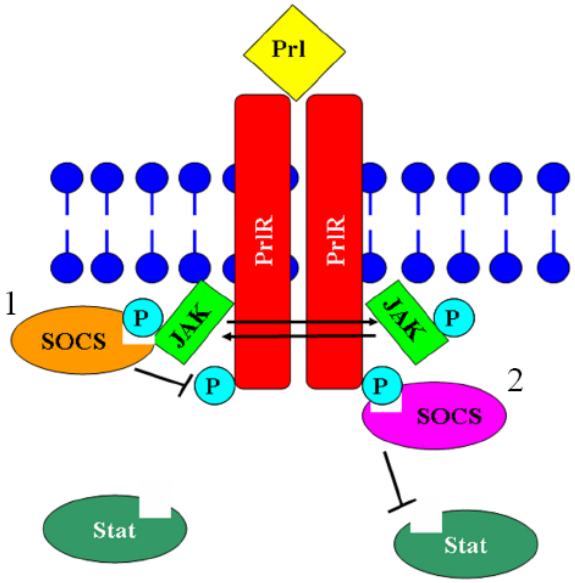


Figure 1.12. Stat signalling pathway. (A) PrlR with the associated tyrosine kinase JAK remains inactive in the absence of ligand (Prl). **(B)** Binding of Prl results in dimerization of the receptor and activation of JAK which in turn phosphorylates specific tyrosine residues on the receptor, creating docking sites for Stats. **(C)** Stat is recruited to the phosphorylated PrlR through its SH2 domains, where it becomes phosphorylated. **(D)** Phosphorylated Stat dimerizes and translocates into the nucleus where it activates the transcription of target genes. **(E)** SOCS proteins are induced in response to cytokine signalling. Different SOCS proteins negatively regulate Stat signalling via different methods: (1) SOCS can bind to phosphorylated JAKs to prevent activation of the receptor and subsequently phosphorylation of Stats; (2) SOCS can also bind directly to Stat binding sites on the activated receptor, preventing docking and hence activation of Stat proteins. Prl: Prolactin; PrlR: Prolactin receptor; JAK: Janus kinase; Stat: Signal transducer and activator of transcription; SOCS: Suppressor of cytokine signalling; P: phosphorylated residue.

1.4.3.4 Stat5 as a key mediator of lobulo-alveolar development

The effects of prolactin signalling during pregnancy are predominantly mediated via Stat5 (Hennighausen and Robinson, 2008). Stat5a and Stat5b are encoded by two juxtaposed genes and are highly similar; they show a 96% similarity at the amino acid level (Hennighausen and Robinson, 2008). Differential expression patterns of both genes were observed in the mouse: *Stat5a* is the predominantly expressed in mammary tissue while *Stat5b* is more abundant in the muscle and liver (Hennighausen and Robinson, 2008). Consistent with their expression patterns, knockout of *Stat5a* in mice resulted in the attenuation of alveolar development and failure to lactate (Liu et al., 1997) while knockout of *Stat5b* in mice did not affect mammary development but led to stunted body growth (Udy et al., 1997). These tissue-specific defects reflect either the tissue-specific expression of *Stat5* genes or the tissue-specific function of these genes. After multiple pregnancies, some functional mammary development was observed in *Stat5a* knockout mice, coinciding with an increase in the levels of Stat5b (Liu et al., 1998), which suggests that there might be partial compensation from *Stat5b* in the mutant gland. Deletion of *Stat5a* and *Stat5b* simultaneously in the mouse confirmed functional redundancy of both isoforms in the mammary gland (Miyoshi et al., 2001; Teglund et al., 1998). As the original *Stat5a/b* double knockout females were infertile due to non-functional corpora lutea, *Stat5a/b* conditional knockout mice were used to analyse the mammary phenotypes during pregnancy (Cui et al., 2004). Transplantation of *Stat5a/b* null epithelial cells into a cleared mammary fat pad showed that while ductal growth was unaffected, cell proliferation and alveologenesis during gestation were severely affected (Miyoshi et al., 2001). In addition, conditional deletion of *Stat5a/b* during late pregnancy led to premature cell death of the alveoli, suggesting a role for Stat5a/b in cell survival (Cui et al., 2004). Hence, *Stat5a/b* is a crucial mediator of proliferation and differentiation of mammary luminal cells and also a regulator of cell survival and function during lactation.

1.4.4 Lactation

At the onset of lactation, the mammary gland reaches maturity and produces and secretes milk to feed the newborn pups. Proliferation and expansion of the alveolar structures occurs at parturition and the entire mammy gland becomes populated by

alveoli with large lumens (Figure 1.11E). The mature alveoli acquire a secretory function and the luminal secretory cells are surrounded by basal myoepithelial cells, which contract and force the milk into the ducts that transport the milk to the nipple. Activation of Stat5 by prolactin signalling is required for β-casein production (Kazansky et al., 1995). Direct binding of activated Stat5 and recruitment of CCAAT/enhancer-binding protein (C/EBPβ) (Raught et al., 1995) to the β-casein promoter up-regulates β-casein transcript levels during pregnancy and lactation. In addition to milk protein production, the ability to expel the secreted milk into the mammary ducts is also critical to successful lactation. The systemic hormone, oxytocin (OT) plays an important role in this process by inducing the contraction of the myoepithelial layer surrounding the alveoli and ducts to facilitate delivery of the milk to the young (Hennighausen and Robinson, 1998). Mice lacking OT failed to lactate and undergo precocious involution most likely due to an accumulation of milk in the alveoli (Wagner et al., 1997). Therefore, successful lactation is dependent on the terminal differentiation of luminal epithelial cells, which produce and secrete milk, and on the contractile functionality of the myoepithelial layers that enables the milk to be transported to the nipple to nurse the young.

1.4.5 Involution

At the end of the lactation phase following the weaning of the pups, the milk-producing epithelial cells are no longer required and are removed (Figure 1.11F). This post-lactational mammary gland regression is known as involution and is mediated by apoptosis (Watson, 2006b). The main features of involution include removal of most of the secretory epithelial cells by apoptosis, clearing of the dead cells and milk, and re-differentiation of the adipocytes such that the mammary gland is remodelled to a pre-pregnant state in preparation for subsequent pregnancies. Failure to remove redundant lobulo-alveoli during involution could result in inflammation and tissue damage, therefore regulation of involution is an extremely important process (Watson, 2006a). Detailed analysis of the involution process has allowed the identification of morphological features and molecular events that occur in the first six days of involution. Studies from forced weaning (removal of pups from lactating females at day 10 lactation) and glucocorticoid administration in mice revealed two distinct phases of involution

(Watson, 2006a). In the first phase (first 48 hours of involution), no remodelling of the mammary gland is observed and only apoptosis is initiated, resulting in the detection of apoptotic cells in the alveolar lumens. The second phase of involution (72 hours after initiation of involution) is characterized by apoptosis together with remodelling of the mammary gland where the alveoli collapse and adipocytes re-differentiate and begin to refill the gland (Watson, 2006b). If pups are returned to the involuting female within 48 hours of initiation of involution, apoptosis is curtailed and lactation is re-initiated (Watson, 2006a). The first phase of involution is regulated by local factors such as milk accumulation within each gland and is not dependent on circulating hormones (Li et al., 1997; Marti et al., 1997). In contrast, the second phase of involution is dependent on circulating hormones and can be blocked by the administration of progesterone (Feng et al., 1995; Lund et al., 1996). In addition, the second phase is also dependent on the activity of matrix metalloproteinases whose function is inhibited by the expression of tissue inhibitors of metalloproteinases (Watson, 2006a). Immuno-infiltration of the mammary gland is also characteristic of the second phase where phagocytes clear the gland of cellular debris and milk components.

1.4.5.1 Lif/Stat3-mediated apoptosis during involution

Apoptosis mediated by the activation of the LIF/Stat3 pathway plays a critical role in the involution process (Watson, 2006a; Watson, 2006b). Apoptosis is a form of programmed cell death in multicellular organisms that was first coined by Kerr, Wyllie and Currie in 1972 (Kerr et al., 1972). Apoptosis has been extensively studied and the morphological changes that occur within an apoptotic cell are well-defined. The cells undergo cellular shrinking, membrane blebbing, chromatin condensation, nuclear fragmentation and chromosomal DNA fragmentation (Kerr et al., 1972). Caspases, which are cysteine proteases, are responsible for most of the morphological changes described in the apoptotic cells (Kerr et al., 1972). Several signalling pathways and apoptotic regulators have been implicated in the first phase of involution in the mammary gland such as members of the Bcl-2 family, Bcl-x and Bax. Conditional deletion of the anti-apoptotic *Bcl-x* in the mouse resulted in accelerated apoptosis (Walton et al., 2001), whereas inactivation of the pro-anti-apoptotic *Bax* delayed involution (Schorr et al.,

1999). However, the main signalling pathway that is involved in the first phase of involution is the JAK-Stat pathway, in particular, Stat3, which is the main player in the involution process (Chapman et al., 2000).

Stat3 is a member of the Stat family and activation of Stat3 occurs at the start of the first phase of involution (Liu et al., 1996). Conditional deletion of *Stat3* in the mammary glands of mice led to a dramatic reduction in epithelial apoptosis, resulting in delayed mammary gland involution (Chapman et al., 1999). This suggests that activation of Stat3 is a critical event in the initiation of apoptosis and involution. Subsequently, both the upstream and downstream targets of the Stat3 signalling pathway were identified (Kritikou et al., 2003; Thangaraju et al., 2005; Tonner et al., 2002). Leukemia inhibitory factor (Lif), a 40 kDa glycoprotein belonging to the IL6 family of cytokines, binds to the Lif receptor β (LifRβ) which forms a complex with gp130 to activate Stat3 *in vivo* (Figure 1.13A) (Kritikou et al., 2003). *Lif* is up-regulated dramatically within 12 hours of initiation of involution and deletion of *Lif* in the mouse resulted in the absence of Stat3 activation, decreased apoptosis and delayed involution (Kritikou et al., 2003). Interestingly in *Lif* knockout mice, there was precocious lobulo-alveolar development in the pregnant gland (when activation of Stat3 is normally undetected) (Kritikou et al., 2003). Activation of extracellular signal-regulated kinase 1/2 (ERK1/2) was dramatically reduced in the pregnant glands of *Lif* knockout mice, suggesting that Lif could act via the ERK pathway during lobulo-alveolar development (Kritikou et al., 2003). In addition to Lif, transforming growth factor (TGF) β3 has also been shown to be capable of activating Stat3 in the first phase of involution (Nguyen and Pollard, 2000).

Downstream targets of Stat3 have also been identified and they include insulin-like-growth-factor-binding protein 5 (IGFBP-5), C/EBPδ and regulatory subunits of the phosphatidylinositol-3-OH kinase (PI3K) pathway. Over-expression of *Igfbp-5* led to impaired lobulo-alveolar development in pregnant mice, an increase in expression of cleaved Caspase 3 (pro-apoptotic) and a decrease in expression of *Bcl-2* and *Bcl-xL* (Tonner et al., 2002). In contrast, deletion of *C/EBPδ* delayed involution due to the inactivation of pro-apoptotic genes and continued expression of anti-apoptotic genes (Thangaraju et al., 2005). Activation of Stat3 also up-regulates the expression of PI3K regulatory subunits *p55α* and *p50α* (Figure 1.13B) (Abell et al., 2005). PI3K is a family

of enzymes that catalyze the production of 3' phosphoinositide lipids in the cell membrane upon their recruitment in response to activation of growth factor receptors (Figure 1.13C) (Katso et al., 2001). Class I PI3Ks catalyze the phosphorylation of phosphatidylinositol-4,5-biphosphate (PtdIns-4,5-P_2) to phosphatidylinositol-3,4,5-triphosphate (PtdIns-3,4,5-P_3), which in turn acts as a docking site for adaptor proteins that contain the pleckstrin homology (PH) domain such as the serine-threonine kinase, protein kinase B (PKB)/Akt (Figure 1.13D) and phosphoinositide-dependent kinase 1 (PDK1) (Katso et al., 2001). After recruitment to the membrane, Akt is phosphorylated and consequently activated by PDK1 and is now able to mediate phosphorylation of target proteins on serine and threonine residues. Through phosphorylation of these target proteins, Akt functions as a key regulator of cellular proliferation and survival (Katso et al., 2001). PI3K consists of a p110 (α/β/δ isoforms) catalytic subunit associated with either of the p85 (α/β isoforms), p55 (α/γ isoforms) and p50α regulatory subunits (Funaki et al., 2000). Both p55α and p50α are alternative splice variants of the p85α subunit. Association of p110 with p85 isoforms render the enzyme active while p55α and p50α are negative regulators of p110 catalytic activity (Funaki et al., 2000). The levels of *p85α* remained constant throughout mammary gland development while expression of *p55α* and *p50α* were up-regulated at the onset of involution (Figure 1.13B) (Abell et al., 2005). Activation of Stat3 directly up-regulates the transcription of *p55α* and *p50α* via an internal Stat3 binding site in *p85α* (Abell et al., 2005). The increased levels of p55α and p50α correlate with a decrease in Akt phosphorylation during involution (Abell et al., 2005). Consistent with this, in the *Stat3* mutant mammary glands, levels of p55α and p50α were reduced, thus demonstrating that the Stat3-mediated apoptosis is regulated by a subunit switch in the PI3K enzyme (Abell et al., 2005).

In summary, mammary gland development consists of a tightly co-ordinated series of events. Embryonic development sets up the basis of a rudimentary mammary structure and at puberty, ductal morphogenesis stimulated by steroid hormones, generates the entire mature mammary network. Terminal differentiation of the mammary epithelium occurs during pregnancy and lactation. During involution, the redundant lobulo-alveolar structures regress by apoptosis to return the mammary epithelium to a pre-pregnant state, in preparation for the next pregnancy.

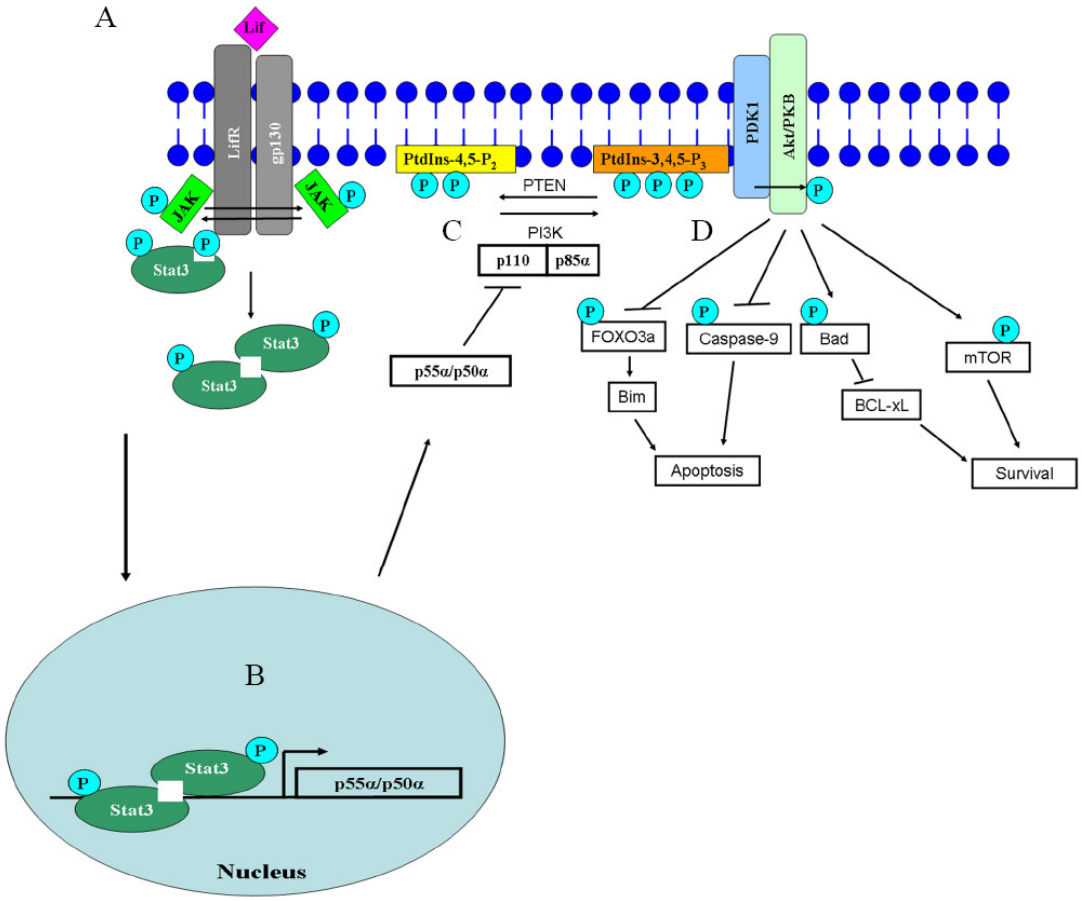


Figure 1.13. Overview of Lif/Stat3-mediated apoptosis. (A) Binding of Lif to its receptor and gp130 results in the activation of JAK which in turn phosphorylates specific tyrosine residues on the LifR, creating docking sites for Stat3. Recruitment of Stat3 to the receptor results in the phosphorylation and activation of Stat3. (B) Phosphorylated Stat3 dimerizes and translocates into the nucleus where it activates the transcription of target genes such as *p55α* and *p50α*. The p55α and p50α proteins are negative regulators of PI3K. (C) PI3K catalyses the phosphorylation of PtdIns-4,5-P_2 to PtdIns-3,4,5-P_3. The reverse reaction is catalysed by PTEN. (D) PtdIns-3,4,5-P_3 creates docking sites for the recruitment of PKB/Akt and PDK1. Akt becomes phosphorylated by PDK1 and is now able to activate/inhibit a plethora of targets such as FOXO3a transcription factor, Caspase-9, pro-apoptotic protein Bad and mTOR, mediating effects of survival. Lif: Leukaemia inhibitory factor; LifR: Lif receptor; JAK: Janus kinase; Stat3: Signal transducer and activator of transcription 3; PtdIns-4,5-P_2: phosphatidylinositol-3,4-biphosphate; PtdIns-3,4,5-P_3: phosphatidylinositol-3,4,5-triphosphate; PDK1: phosphoinositide-dependent kinase 1; P: phosphorylated residue.

1.5 Genetic control of mammary cell fate

1.5.1 Historical perspectives

The mammary epithelium primarily consists of two main cell types: (1) luminal epithelial cells which line the lumen and (2) myoepithelial/basal epithelial cells which form a surrounding sleeve encompassing the ductal epithelium. Similar to the hematopoietic system, where the various different blood lineages are generated from the HSC, the entire mammary epithelium is also derived from the mammary stem cell (MaSC). Likewise, different signals are required for the specification and cell fate determination of mammary stem and progenitor cells during lineage commitment throughout development. In addition to steroid hormones, transcription factors and cytokines also play important roles in mammary cell fate and lineage commitment (Watson and Khaled, 2008).

The first evidence of the possible existence of adult mammary stem cells was revealed in 1959 when DeOme *et al*. showed that virtually any part of the adult mammary gland could reconstitute the cleared mammary fat pad to generate the entire mammary ductal tree (Deome et al., 1959). Importantly, the cleared fat pad transplantation technique developed in this study formed the basis of an *in vivo* functional test for MaSCs, similar to the transplantation of hematopoietic cells into sub-lethally irradiated host mice. Subsequently, Smith and Medina demonstrated that mammary cells capable of repopulating the mammary gland exist throughout the life span of the adult mice (Smith and Medina, 1988). In addition, using retroviral-tagged mammary epithelial cells, Kordon and Smith revealed the existence of MaSCs and also the presence of three multi-potent but distinct mammary epithelial progenitors (Kordon and Smith, 1998), similar to that described in an earlier study (Smith, 1996). In general, the field of MaSC biology has lagged behind that of HSC research because of a lack of definitive and exclusive stem/progenitor cell markers. However, a breakthrough was achieved in 2006 when two groups demonstrated that a single cell isolated by a combination of cell surface marker was able to reconstitute the entire mammary gland (Shackleton et al., 2006; Stingl et al., 2006). This is the first direct evidence of the existence of the MaSC.

1.5.2 Mammary epithelial hierarchy

1.5.2.1 Characteristics of the mammary stem cell

The isolation of MaSCs was based on expression of the cell surface marker CD24 (heat-stable antigen) and either CD49f (α_6-integrin) (Stingl et al., 2006) or CD29 (β_1-integrin) (Shackleton et al., 2006). Interestingly, CD49f and CD29 are likely to form a functional α_6-β_1-integrin hetero-dimeric complex that mediates the interaction between the epithelial cells and stroma. The subset of lineage-depleted mammary epithelial cells (depletion of hematopoietic and endothelial cells) which expresses medium levels of CD24 and high levels of CD49f/CD29 ($CD24^{med}CD49f^{hi}/CD29^{hi}$) (Figure 1.14A) were found to be enriched in MaSCs (termed mammary repopulation units - MRUs) as determined by transplantation of these cells into cleared fat pads. Stingl *et al.* estimated the frequency of these MaSCs to be 1 per 20 cells of the $CD24^{med}CD49f^{hi}$ subset (Stingl et al., 2006) while Shackleton *et al.* estimated a frequency of 1 per 64 cells of the $CD24^{med}CD29^{hi}$ subset (Shackleton et al., 2006). Consequently, a single cell derived from the MaSC-subset ($CD24^{+}CD49f^{hi}/CD29^{hi}$) was shown to reconstitute the entire mammary gland without the use of supporting cells (Stingl et al., 2006) (Shackleton et al., 2006). A functional role for CD29 in the maintenance of MaSC in the epithelium has been demonstrated recently (Taddei et al., 2008). Deletion of CD29 from the basal compartment of the epithelium abolished the regenerative potential of MaSC, suggesting a critical role of CD29 in the maintenance of MaSC population by mediating basal cell interactions with the extracellular matrix.

The MaSC-enriched subset also expresses low levels of stem cell antigen (Sca1) and appears to occupy a basal position within the mammary epithelium (Stingl et al., 2006) (Shackleton et al., 2006). In addition, the MaSC-enriched subset was also found not to exclude the $Hoechst_{33342}$ dye (Shackleton et al., 2006; Stingl et al., 2006). This is somewhat surprising because the ability to exclude dyes is a characteristic of stem cells. Previous studies have also identified a side population phenotype that contains undifferentiated cells which can give rise to luminal and basal cells (Alvi et al., 2003; Welm et al., 2002). However, this side-population is most likely to represent bi-potent progenitor cells because it is depleted of expression of $CD24^{med}CD49f^{hi}/CD29^{hi}$ and transplantation of these cells at limiting dilutions showed a low repopulating frequency

(Shackleton et al., 2006; Stingl et al., 2006). Interestingly, the MaSC subset appears to be actively cycling as determined by $Hoechst_{33342}$ and pyronin Y staining (Stingl et al., 2006) and this is consistent with another study by Smith *et al.* (Smith, 2005), which showed that *in vivo* long-term label-retaining mammary cells appeared to retain their template DNA strands through asymmetrical cell division. However, the precise relationship between MaSCs and long-term label-retaining mammary cells is unknown. It has been postulated that a minor population of MaSCs that resides in a quiescent state (G0) *in vivo* may be present (Visvader and Lindeman, 2006).

Estrogen plays an important role in ductal morphogenesis and also in alveologenesis as discussed above (Chapter 1.4.2.1). ERα is an important prognostic marker for breast cancer and ERα-negative breast cancers are those that are resistant to therapy and usually have a bad prognosis. Expression of ERα appears to be restricted to 10-30% of luminal cells in the ducts and alveoli (Asselin-Labat et al., 2006; Mallepell et al., 2006) and ERα-positive cells appear to be non-dividing cells that instruct adjacent epithelial cells to proliferate via paracrine regulatory effects (Clarke et al., 1997; Mallepell et al., 2006; Russo et al., 1999; Seagroves et al., 1998). The basal MaSC-enriched subset was found to be ERα-negative and PR-negative (Asselin-Labat et al., 2006). It has been hypothesized that these ERα-negative and PR-negative MaSCs undergo asymmetric division to generate ERα-positive cells, which subsequently proliferate in response to estrogen and secrete paracrine factors that regulate adjacent ERα-negative cells (Dontu et al., 2004).

Using the same cell surface markers (CD24/CD49f), the luminal and basal cells of the mammary epithelium can be clustered into two distinct populations: $CD24^{hi}CD49f^{+}$ (luminal) and $CD24^{+}CD49f^{hi}$ (basal) (Figure 1.14B). The ability to separate these two populations is critical as it allows the systematic characterization of the progenitor and differentiated cell populations within the mammary epithelium.

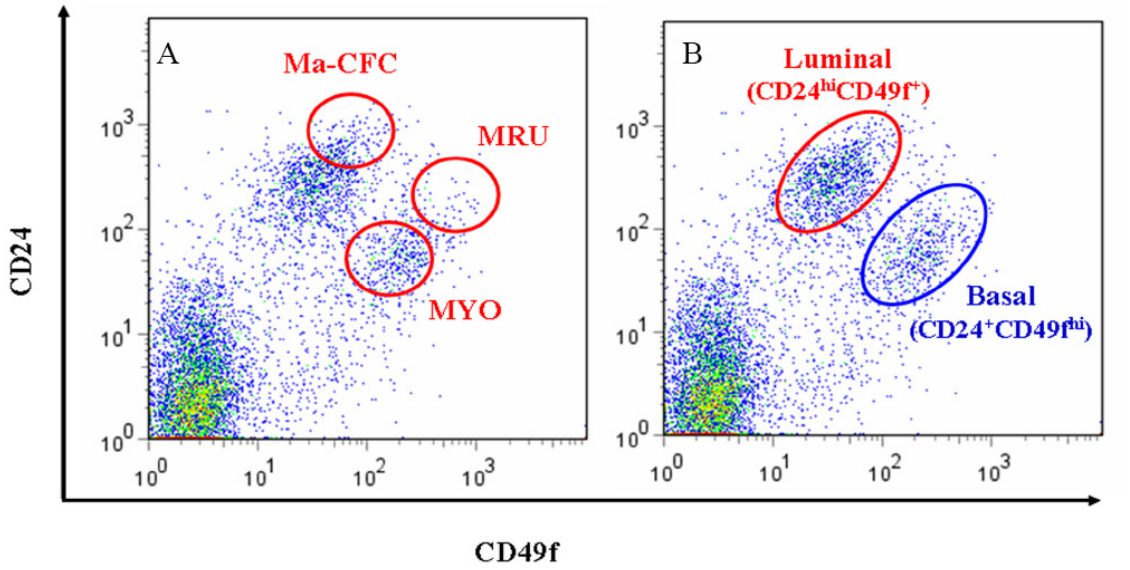


Figure 1.14. Characterization of murine epithelial cells using cell surface markers. (A) Distribution of mammary repopulating units (MRU)-CD24^{med}CD49f^{hi}; mammary-colony forming cells (Ma-CFC) C24^{hi}CD49f^{lo} and myoepithelial cells (MYO) CD24^{lo}CD49f^{lo}. **(B)** Distribution of CD45^{-}Ter119^{-}CD31^{-} epithelial cells according to their CD24 and CD49f expression. Luminal cells are characterized by CD24^{hi}CD49f^{+} whereas basal cells are characterized by C24^{+}CD49f^{hi}.

1.5.2.2 Mammary luminal progenitors

Different types of mammary progenitor cells can be detected under various *in vitro* culture conditions (Dontu et al., 2003; Smalley et al., 1998; Stingl et al., 2001). Mammary colony forming cells (Ma-CFCs) refer to progenitors that can form discrete mammary colonies *in vitro* (Stingl et al., 2001). In the mouse mammary epithelium, most Ma-CFCs are localized within the $CD24^{hi}CD49f^{lo}$ subset (Figure 1.14A) (Stingl et al., 2006). This subset lacks any repopulating capability and is found within the luminal profile and has been shown to express high levels of cytokeratin 6, a putative progenitor cell marker (Hu et al., 2006; Stingl et al., 2006). Thus the $CD24^{hi}CD49f^{lo}$ subset most likely represents luminal progenitors which are responsible for generating the luminal lineages.

Transplantation of mammary epithelial cells into a cleared mammary fat pad generates outgrowths which consist of (1) lobular (secretory) cells only, (2) ducts only and (3) both lobular cells and ducts (Smith, 1996). This demonstrates that three distinct multi-potent mammary cell types exist in the murine mammary gland (Smith, 1996). All three cell types are most likely derived from a common ancestor, but the precise progenitor cells that generate the lobule-restricted or the ductal-restricted outgrowths have not been identified. Recently, CD61 has been identified as a luminal epithelial progenitor marker (Asselin-Labat et al., 2007). Lineage-negative $CD29^{lo}CD24^{+}CD61^{+}$ subset is highly enriched for Ma-CFCs (Asselin-Labat et al., 2007). Another luminal progenitor population which is enriched for Ma-CFCs has also been identified (Sleeman et al., 2007). This subset of progenitor cells exhibits low levels of CD133/prominin expression, suggesting that luminal differentiation is accompanied by reduced CD61 but increased CD133 expression, resulting in a mature luminal ($CD61^{-}CD133^{+}$) luminal cell (Vaillant et al., 2007). Interestingly, a subset of luminal progenitors was found to express ERα (Asselin-Labat et al., 2007). Consistent with this study, an ERα-positive luminal progenitor subset has been identified and characterized by John Stingl (Stingl and Watson; manuscript in preparation). Using cell surface markers CD49b (α2-integrin) and Sca1 (Ly-6A/E), the luminal subset can be classified into progenitor ($CD49b^{+}$) or differentiated ($CD49b^{-}$) cells based on the presence/absence of CD49b (Figure 1.15) (Stingl and Watson; manuscript in preparation). In addition, the luminal progenitor

population can be separated into Sca1$^-$ and Sca1$^+$ luminal progenitors. The CD49b$^+$Sca1$^+$ subset shows expression of ERα while almost all the CD49b$^+$Sca1$^-$ cells are ERα-negative (Stingl and Watson; manuscript in preparation). Further analysis of the Sca1$^-$ and Sca1$^+$ luminal progenitors revealed that the Sca1$^-$ fraction showed relatively high levels of expression of milk proteins. Hence, these results revealed that two functionally distinct luminal progenitor populations exist in the mammary epithelium and it has been postulated that the CD49b$^+$Sca1$^-$ (ERα-negative) subset represents alveolar progenitors while the CD49b$^+$Sca1$^+$ (ERα-positive) subset represents ductal progenitors (Stingl and Watson; manuscript in preparation). Definitive proof of this hypothesis is needed by transplanting these two different luminal progenitor populations into mammary fat pad.

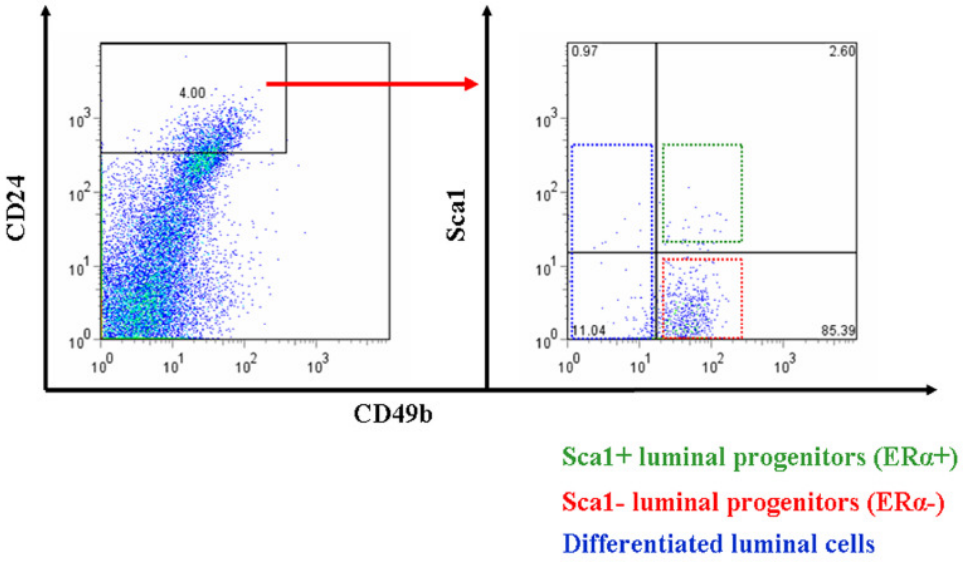


Figure 1.15. FACS profile of murine luminal epithelial cells. Luminal mammary epithelial cells (CD24^{hi}) can be separated into luminal progenitors (CD49b-positive) and differentiated luminal (CD49b-negative) populations. The luminal progenitors can be further classified by their expression of Sca1. The Sca1$^+$ and Sca1$^-$ mammary epithelial cells form colonies *in vitro*, demonstrating their progenitor property.

1.5.2.3 Myoepithelial/basal cells

The myoepithelial/basal cells occupy a $CD24^{+}CD49f^{hi}/CD29f^{hi}$ position in the mammary flow cytometric profile which partially overlaps with the MaSC-enriched subset (Figure 1.14B). Hence, the MaSC-enriched subset also contains myoepithelial/basal cells (Shackleton et al., 2006; Stingl et al., 2006). The Ma-SC-enriched and myoepithelial/basal subsets showed elevated expression in cytokeratin 5 and 14, smooth muscle actin, smooth muscle myosin, vimentin and laminin transcripts (Stingl et al., 2006). Thus the myoepithelial/basal subset contains a mixture of MaSC and committed myoepithelial/basal cells but may also contain a putative common luminal-myoepithelial progenitor and/or myoepithelial progenitor.

1.5.2.4 Current model of mammary epithelial hierarchy

Based on the recent data, a model of the hierarchy of mammary epithelial cells has been proposed (Figure 1.16). The basal ERα/PR-negative MaSC may divide asymmetrically to generate a putative common bi-potent progenitor. This common progenitor would then give rise to the luminal and myoepithelial progenitors. The luminal progenitor consists of both ERα-positive and ERα-negative subsets which may represent ductal and alveolar progenitors respectively. Differentiation of the ERα-positive ductal progenitor produces ERα-positive ductal cells. In addition, the ERα-positive ductal progenitor may secrete paracrine factors that may induce the proliferation and/or differentiation of adjacent ERα-positive and ERα-negative luminal progenitors. The ERα-negative alveolar progenitors proliferate during pregnancy and differentiate to form secretory alveolar cells. The myoepithelial progenitors would give rise to myoepithelial cells.

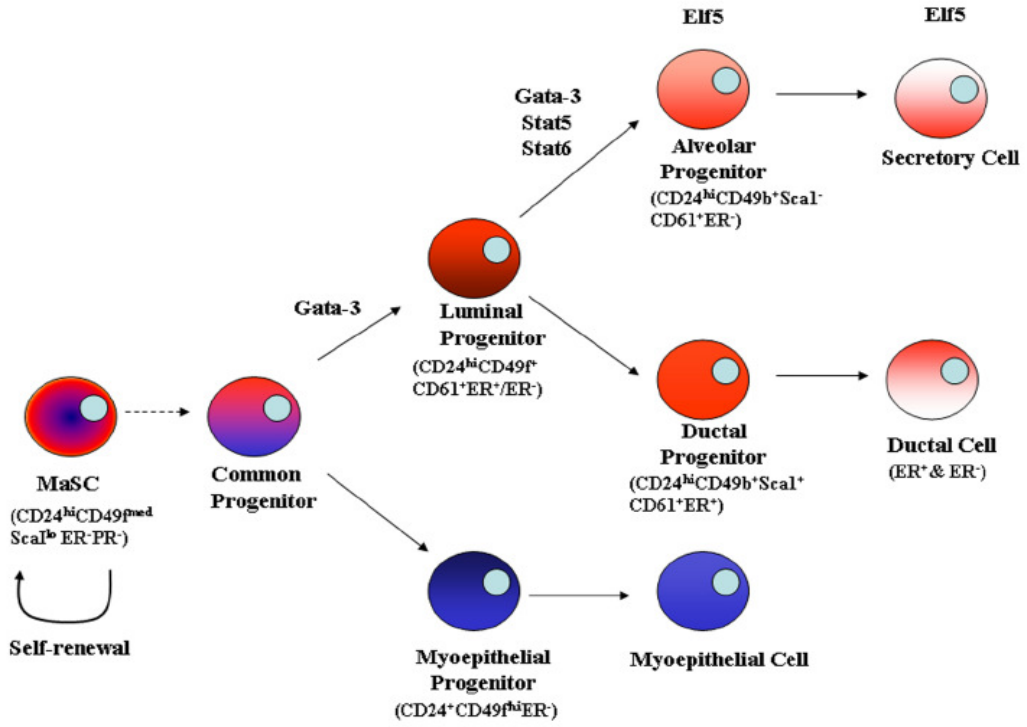


Figure 1.16. Hierarchy of murine epithelial cells. Model of epithelial hierarchy and transcription factors implicated in lineage commitment in the murine mammary gland. The mammary stem cell (MaSC) exhibits high levels of integrins and is negative for the estrogen receptor and the progesterone receptor. Proliferation and differentiation of the MaSC gives rise to a common progenitor which would generate the lineage-restricted luminal and myoepithelial progenitors. Two phenotypically distinct populations of luminal progenitors exist in the epithelium: $CD49b^{+}Sca1^{-}(ER\alpha^{-})$-putative alveolar progenitors and $CD49b^{+}Sca1^{+}$ $(ER\alpha^{+})$-putative ductal progenitors. Proliferation and differentiation of these progenitors generates the differentiated luminal cells. MaSC: Mammary stem cells; ER: estrogen receptor α; PR: progesterone receptor; Stat: Signal transducer and activator of transcription.

1.5.3 Lineage commitment during puberty and pregnancy

1.5.3.1 Transcription factors

Although the mammary epithelial hierarchy has been proposed, the precise molecular mechanisms that regulate mammary stem/progenitor cell proliferation and differentiation into the various epithelial lineages remains poorly understood. However, recent publications have identified several transcription factors that are important for lineage commitment during mammary gland development (Asselin-Labat et al., 2007; Khaled et al., 2007; Kouros-Mehr et al., 2006; Oakes et al., 2008), adding a different perspective to the existing paradigm that steroid hormones and prolactin are the principal temporal regulators of mammary gland lineage commitment and differentiation.

The Gata family of transcription factors plays an important role in cell fate specification in many different tissues such as the skin and the immune system. In the immune system, *Gata-3* has been shown to be critical for T cell specification and in T_H2 cell differentiation (Rothenberg, 2007a). Recently, *Gata-3* has been shown to be essential for mammary gland development and the maintenance of differentiated luminal cell fate (Asselin-Labat et al., 2007; Kouros-Mehr et al., 2006). In the first study, *Gata-3* was identified to be the most highly enriched transcription factor in the developing and mature mammary luminal cells (Kouros-Mehr et al., 2006). Conditional deletion of *Gata-3* using MMTV-Cre resulted in the failure of TEB formation and profound structural defects (Kouros-Mehr et al., 2006). In addition, there was a selection pressure to retain a functional *Gata-3* allele in the surviving mammary outgrowths, suggesting that *Gata-3* is essential for mammary development. Subsequently, conditional deletion of *Gata-3* using WAP-rtTA-Cre (where Cre expression is induced by doxycycline administration) revealed that *Gata-3* is necessary for maintenance of the luminal epithelium in the adult mammary gland (Kouros-Mehr et al., 2006). Acute loss of *Gata-3* (5 days after doxycycline treatment) led to the expansion of undifferentiated luminal cells as demonstrated by a dramatic reduction of the levels of differentiated luminal markers (Kouros-Mehr et al., 2006). In contrast, chronic loss of *Gata-3* (14 days after doxycycline treatment) led to luminal cell death, leading to severe lactational defects. Importantly, these defects occurred because the lobulo-alveoli did not develop properly, rather than because of any apoptosis or abnormalities in cell proliferation. Similar results were

obtained in a separate study where *Gata-3* was deleted using the WAP-Cre mice where Cre recombinase was expressed in the epithelium during pregnancy (Asselin-Labat et al., 2007). Deletion of *Gata-3* resulted in failed alveoli development and lactogenesis. Additionally, consistent with the first study which demonstrated that loss of *Gata-3* resulted in an increase in undifferentiated luminal cells, Asselin-Labat *et al.* revealed that loss of *Gata-3* caused a dramatic increase in the pool of $CD29^{lo}CD24^{+}CD61^{+}$ luminal progenitors, suggesting a block in luminal differentiation (Asselin-Labat et al., 2007). Over-expression of *Gata-3* in the $CD29^{lo}CD24^{+}CD61^{+}$ luminal progenitors was sufficient to induce terminal differentiation of these cells, eventually cumulating in the expression of milk protein transcripts such as *β-casein* and *Wap* (Asselin-Labat et al., 2007). Collectively, these results indicated that *Gata-3* is required for the maintenance and differentiation of the luminal lineage (Asselin-Labat et al., 2007). The effects of loss of *Gata-3* in the mammary gland resembled those of *ERα* deficiency, suggesting that these two genes may function in the same pathway (Mallepell et al., 2006). Subsequently, Gata-3 was shown to bind directly to the promoter of forkhead transcription factor (*FOXA1*), which has been suggested to be essential for estrogen signalling in the mammary gland and required for binding of ERα to chromatin (Carroll et al., 2005). It has also been shown that *Gata-3* is required for estrogen-mediated breast cancer cell growth and ERα was shown to directly stimulate *Gata-3* expression (Eeckhoute et al., 2007); however whether this feedback loop applies to the normal mammary gland remains to be established.

The prolactin-regulated Ets transcription factor *Elf5* has recently been shown to specify alveolar cell fate (Oakes et al., 2008). Loss of one allele of *Elf5* in the mouse resulted in females that displayed defective lobulo-alveoli development and reduced milk secretion (Zhou et al., 2005) and retroviral over-expression of *Elf5* in *PrlR*-deficient mammary cells rescued alveologenesis (Harris et al., 2006). Loss of both alleles of *Elf5* caused embryonic lethality; therefore Oakes *et al.* transplanted $Elf5^{-/-}$ mammary cells into cleared fat pads of $Rag1^{-/-}$ host mice and analyzed the effects of *Elf5* deletion on mammary development (Oakes et al., 2008). While ductal morphogenesis was unaffected, alveologenesis during pregnancy was severely affected (Oakes et al., 2008). Analysis of these mammary glands by flow cytometry revealed an accumulation of

$CD29^{lo}CD24^{+}CD61^{+}$ luminal progenitors, similar to that observed in the *Gata-3*-deficient mammary glands, suggesting that loss of *Elf5* resulted in the blockade of luminal differentiation. Quantitative PCR analysis of the $CD29^{lo}CD24^{+}CD61^{+}$ luminal progenitors in both the virgin and 12.5 dpc gestation mammary glands demonstrated expression of both *Gata-3* and *Elf5*. However, immuno-staining showed that they were expressed in different cell populations (Oakes et al., 2008). The differential expression patterns of both genes were also observed in the $CD29^{lo}CD24^{+}CD61^{-}$ mature luminal cells. Whereas expression of *Gata-3* was high in the $CD29^{lo}CD24^{+}CD61^{-}$ mature luminal cells in the virgin but decreased at late gestation, *Elf5* exhibited the exact opposite expression patterns – low expression in the virgin glands but high expression in the late gestation mammary glands. In addition, over-expression of *Elf5* in the virgin glands disrupted ductal morphogenesis and caused precocious alveologenesis and milk production (Oakes et al., 2008). This prompted the authors to suggest that different populations of Gata-3-responsive and Elf5-responsive cells exist within the $CD29^{lo}CD24^{+}CD61^{+}$ luminal progenitor population and conclude that *Elf5* is important for specifying alveolar cell fate. Whether Gata-3 and Elf-5 co-operate in luminal progenitor cells to regulate alveoli differentiation is still unknown.

Several other transcription factors such as the helix-loop-helix inhibitor (*Id2*), *C/EBPβ*, *Stat5* and *Stat6* have also been implicated in alveolar differentiation. Khaled *et al.* demonstrated that Stat6 and its upstream cytokines IL4 and IL13 are required for the expansion of luminal lineage during early gestation (Khaled et al., 2007). Loss of *Stat6* or double knockout of *Il4* and *Il13* resulted in a dramatic reduction in the number of alveoli during gestation, with up to 70% reduction at day five of gestation compared to the wild-type mice (Khaled et al., 2007). Conversely, loss of *Socs5*, a negative regulator of Stat6 caused precocious alveolar development. Hence, Stat6 signalling is important for the regulation of mammary luminal progenitors during lactation.

An interesting perspective to mammary lineage commitment has been added with the recent studies of Gata-3 and Stat6 in mammary development (Watson and Khaled, 2008). As discussed above, both Gata-3 and Stat6 are required for development of the luminal lineage in the mammary gland. In the hematopoietic system, naïve T cells can be polarized into T_H1 or T_H2 cells by cytokines. Both Gata-3 and Stat6 have been shown to

be required for T_H2 cell polarization (Ansel et al., 2006). Interestingly, mammary epithelial cells (MEC) secrete type-1 cytokines IL12a, interferon-γ (IFNγ) and tumour necrosis factor (TNF) in culture. However upon lactogenic stimuli, the MEC switched to secreting type-2 cytokines such as IL4, IL13 and IL5 (Khaled et al., 2007). This demonstrates a role for these immune cell cytokines in epithelial cell fate decision and suggests that the genetic control of mammary epithelial cell fate and hematopoietic cell fate might be conserved. Collectively, these studies demonstrate that temporal regulation of transcription factors in addition to steroid hormones and prolactin signalling, are also required for mammary cell specification and lineage commitment (Figure 1.16).

1.5.3.2 Notch signalling

Notch signalling plays important roles in cell fate decision and lineage specification in several tissues such as the hematopoietic system where temporal regulation of Notch signalling is critical for lymphocyte development (Radtke et al., 2004). Not surprisingly, Notch signalling is also required for T_H1/T_H2 cell fate determination (Ansel et al., 2006; Radtke et al., 2004). In mammals, there are four Notch receptors (Notch1-4) and five ligands (Jagged1 and 2, Delta-like-ligand1, 3 and 4). Notch receptors are synthesized as precursor proteins that are cleaved upon transport to the cell membrane and are expressed as hetero-dimers. Ligand binding to the extracellular portion of the Notch receptor triggers a cascade of proteolytic reaction that cumulates in the liberation of the cytoplasmic domain of Notch receptor (ICN). The ICN translocates to the nucleus and binds to the transcription factor recombination-binding protein-Jκ (RBP-Jκ), converting it from a transcription repressor to a transcription activator. ICN also recruits several co-activators such as mastermind-like protein (MAML) and p300, eventually resulting in the activation of Notch target genes, including the hairy enhancer of split (*Hes*) genes. Mammary fat pads transplanted with *Rbpj*κ-deficient mammary cells showed normal ductal morphogenesis, suggesting that RBP-Jκ is not required for the establishment of luminal and basal cells (Buono et al., 2006). Instead, it was revealed that lobulo-alveolar development during pregnancy was affected. Importantly, there was an accumulation of epithelial cells with expression of both luminal (cytokeratin 18) and basal (cytokeratin 14) markers and there was also an excess of luminal cells expressing

the p63 basal marker. In addition, there was a transient amplification of cytokeratin 6-positive luminal cells that could reflect either a block in differentiation or a slower progression through the luminal differentiation pathway (Buono et al., 2006). This study suggests that loss of *Rbpj*κ during pregnancy results in luminal cells that acquire basal-like features and that the Notch signalling pathway through RBP-Jκ is important for the maintenance, but not the establishment of luminal cells. Other studies have also implicated the importance of Notch signalling during lobulo-alveolar development during pregnancy. Over-expression of Notch1 and Notch3 ICN using the mouse mammary tumour virus (MMTV) provirus resulted in impaired ductal and lobulo-alveoli development during pregnancy and the transgenic mice developed mammary tumours, confirming the oncogeneic potential of Notch1 ICN *in vivo* (Hu et al., 2006). In addition, Notch3 has been shown to be critical for the commitment of bi-potent progenitors to the luminal lineage *in vitro* (Raouf et al., 2008). Collectively, these studies indicate the importance of Notch signalling in mammary development and suggest that temporal regulation of the levels of canonical Notch signalling pathway is crucial to proper lobulo-alveolar development.

1.6 Thesis project

The work of this thesis is based on a simple question: "Do Bcl11 proteins play important roles in development and cell fate decisions in the mammary gland?" The purpose of this project was to identify unique patterns of *Bcl11* expression in the mouse and to characterize their roles in development and lineage commitment.

Current expression data is derived primarily from reverse transcriptase PCR and *in situ* hybridization; whilst being informative, the resolution and sensitivity of these techniques at a single cell level is limited. Hence, the first aim of this thesis was to generate *lacZ*-tagged reporter alleles of both *Bcl11* genes using a new recombineering based approach (Chan et al., 2007). These mice were then used to characterize the spatial and cellular expression of *Bcl11* genes in the embryo and the adult tissues.

The second aim of this study was to characterize the roles of *Bcl11a* and *Bcl11b* in development and lineage commitment in specific tissues. The mammary gland provides a great model system to study various aspects of cellular proliferation,

differentiation and apoptosis. With the recent purification of mammary stem cells and the availabilities of different cell surface markers to isolate different subsets of mammary epithelial cells, the mammary gland has become an ideal system to study lineage commitment and cell fate decisions. These, together with the identification and validation of two *BCL11A* mutations in human breast cancer samples (Wood et al., 2007) prompted me to investigate the roles of *Bcl11* genes in mammary development.

In this thesis, I first characterized the expression of the *Bcl11* genes in various mammary lineages using *lacZ*-tagged reporter mice in combination with flow cytometric analysis. Subsequently, loss-of-function and gain-of-function approaches were used to dissect the roles of the *Bcl11* genes in embryonic and postnatal mammary development and in lineage commitment. Finally, I investigated the underlying molecular mechanisms implicated in the phenotypes.