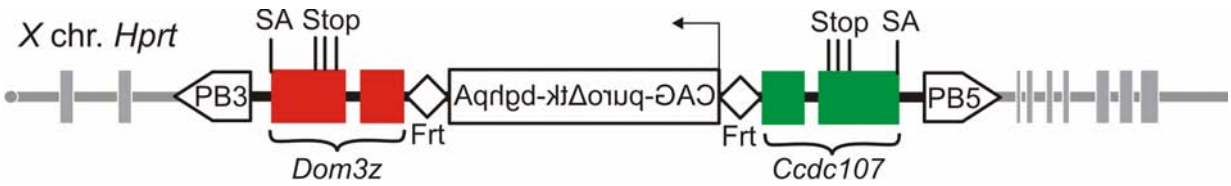
Chapter Three - An efficient mutagenic strategy using the PB transposon

1. Introduction

The aim of this project is to establish a new mutagenic platform, coupled with the *Blm*-deficient ES cell system, to achieve genome-wide mutagenesis using the PB transposon in order to conduct recessive genetic screens. As described in Chapter One, there are several mutagenic agents and modular components available for generating efficient insertional mutagens. Four requirements have been taken into account in my mutagenic strategy: firstly, broad genome-coverage of the insertional mutagen; secondly, the requirement of highly effective gene inactivation; thirdly the ease with which the mutagen can be mapped and finally the establishment of the causal relationship of the genotype and phenotype. The molecular design of my transposon is depicted in Figure 3-1.

Figure 3-1: Design of an inactivating PB transposon targeted into the *Hprt* locus of *Blm*-deficient ES cells.



Dom3z, *ccdc107*, the terminal exon pairs and a portion of the 3' downstream sequence containing the endogenous polyadenylation signal from of the two computationally selected genes *Dom3z* and *ccdc107* respectively; Stop, three stop codons in three different reading frames; SA, endogenous splice acceptor of the penultimate exons of *Dom3z* and *Ccdc107*. The PB transposon was targeted into intron 2 of *Hprt* locus. The *Dom3z* exon pair traps the expression of *Hprt*, thus the resulting ES cells are sensitive to HAT.

1.1. Design principles of the mutagen

Broad genome-coverage of the insertional mutagen

Genome-wide mutagenesis requires the insertional mutagen to be delivered randomly with high efficiency, and two major factors determine this genome coverage. The first and major determinant is the integration characteristic of the insertional mutagen. As described in Chapter 1, the retroviral integration pattern in the mouse genome is significantly non-random, with severe integration "hot spots" (Hansen et al., 2008). The PB transposon integration pattern is significantly more random than retrovirus (Wang et al., 2008b). In addition, PB has a preference to insert into active transcribing units. This bias towards genes is advantageous for insertional mutagenesis. Together with its high transposition efficiency, PB transposon is a good candidate as a mutagen carrier for genome-wide mutagenesis.

There are two ways to deliver the PB transposon, "plasmid-to- genome" integration and intra-genomic mobilisation of a pre-engineered PB transposon within the genome. The "Plasmid-to-genome" delivery method has been successfully used to allow the recovery of homozygous mutants in *Blm*-deficient ES cells (Wang et al., 2008a). However, a careful titration is crucial to restrict the copy number of transposons to one per cell. The second method is intra-genomic mobilisation of the PB transposon. This is achieved by targeted insertion of a PB transposon within the genome, and the transposon is excised from the donor locus and re-integrates in a genome-wide fashion (Wang *et al.*, 2008b; Liang *et al.*, 2009). Re-mobilisation events from the donor locus are relatively inefficient (1 % of total cells electroporated), however they can be enriched using a selection marker that is activated when PB is excised. The outline of this strategy has been introduced in Chapter 1, Figure 1-12. The re-integration rate was approximately 40 % of the total excision events (Wang *et al.*, 2008b; Liang *et al.*, 2009) . When the re-integration events were examined, local hopping was not detected when PB was mobilised from the *Hprt*, although a 10 % local hopping within 2.4 Mb window around the transposon donor site and 18 % residing on the donor chromosome have been observed at *Rosa26* locus (Wang *et al.*, 2008b; Liang *et al.*, 2009). Therefore, PB re-mobilisation from the *Hprt* locus for generating random mutations is an attractive strategy to conduct genome-wide mutagenesis with a tightly controlled copy number to mostly one.

The second determinant is the selection method used for cells with the insertional mutagen integrated. Various methods were discussed to select for loss-of-function mutations in Chapter 1. For instance, the use of the conventional gene trap system relies on the production of a functional reporter protein. This selection strategy eliminates integrations of the mutagen into "wrong" reading frames or "wrong" orientations with respect to the endogenous gene within which the mutagen is inserted. In addition, situations where fusion between the reporter and the endogenous gene have reduced or inactivated the reporter are also eliminated. In addition, integrations into non-expressing loci, genes with weak promoters to drive sufficient expression of the reporter, and non-coding genes will all be eliminated. Therefore, a mutagen in which selection for integration is independent from the requirement for gene inactivation will enhance the coverage of the mutagenesis. In this design (Figure 3-1), the selection marker (puro) for integration is independent from the trapping (mutagenic exon pairs).

Highly effective gene inactivation for loss-of-function screens

Because of the complexity of the mouse genome, which contains 3 billion base pairs and around 30,000 genes, it is essential to achieve efficient gene inactivation to ensure a high portion of integrations in genes result in null mutations. The high mutagenicity is ensured by the gene trapping unit containing a "strong" splice acceptor to "out-compete" the endogenous splice acceptor. The incorporation of stop codons and polyadenylation signals can lead to fusion transcript termination.

In this design (Figure 3-1), the mutagen is bi-directional, with two terminal-exon pairs as gene traps to inactivate gene expression in either orientation. The exon pairs can be selected for containing strong endogenous splice acceptors, polyA signals and stop codons in all reading frames to ensure efficient gene trapping and termination. Furthermore, the incorporation of stop codons before the final mutagen intron-exon splice junction may induce the endogenous non-sense mediated decay (NMD) for fusion transcript degradation. This may further ensure the generation of null mutation even when the mutagen is inserted towards the 3' end of the transcript.

Mutagen mapping and establishment of causal genotype-phenotype relationships

Upon isolation of phenotypic mutants, it is essential to be able to identify the causal gene(s) disrupted by the mutagen. PB transposons provide a unique molecular tag to identify integration sites. However, if there are multiple integrations per cell, it is difficult to assign the causal mutation. Therefore, a single integration of the mutagen per cell is ideal. With the "plasmid-to-genome" integration method, single copy of PB transposon per cell can be achieved in the majority of the cells with careful titration of the amount of the transposon and transposase. However, the balance between efficient genomic integration and the single copy transposon per cell is difficult to strike. Intra-genomic mobilisation of PB transposon can maintain a single copy of mutagen per cell stably and is not restrained with the efficiency of generating sufficient number of mutations.

Precise transposon excision with *PuroΔtk*-mediated selection

PB transposon excision does not generate any footprint. This unique property of PB can be utilised in complete mutant reversion experiments. In mutants with both alleles inactivated (homozygous), two copies of the PB transposon from both alleles of the same locus can be excised. However, the efficiency of obtaining such type of complete reversion events is very low. The incorporation of the *PuroΔtk* cassette (Figure 3-1) allows the efficient isolation of genotype revertants using the drug FIAU to select for PB transposon removal from the genome. The reversion of the disrupted gene can directly address the genotype and phenotype causal relationship, as mutagen removal should lead to the phenotype reversal to wild-type if the mutagen is the cause of the mutant phenotype.

1.2. The DNA mismatch repair pathway as a screening model for proof-of-principle

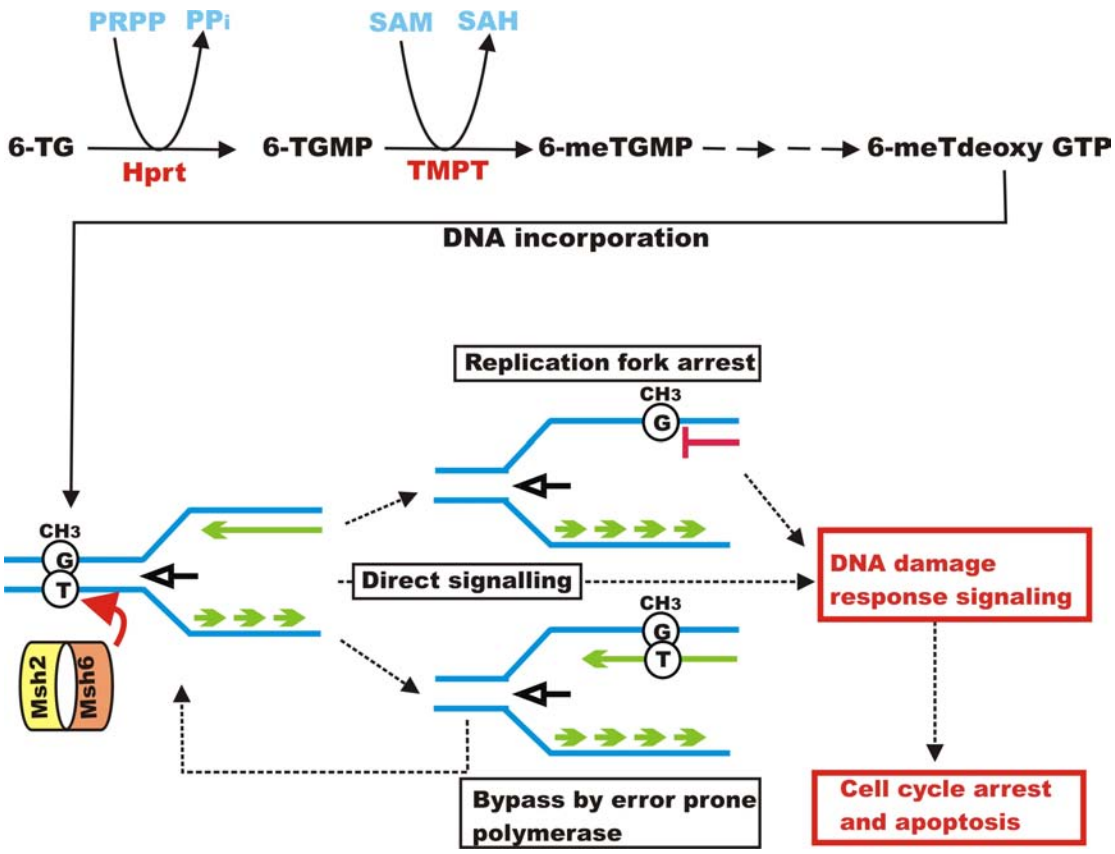
DNA mismatch repair (MMR) is an evolutionarily conserved pathway from bacteria to mammals, which exists to maintain genome integrity during DNA replication and to sense and repair certain types of DNA damage. It has been observed that MMR-deficient cells are more resistant than matched MMR-proficient cells to 6-thioguanine induced cell death (Griffin et al., 1994; Jiricny, 2006). The 6-TG resistant phenotype of MMR-deficient cells has been used

as a phenotypic selection strategy to isolate homozygous mutants in the MMR pathway (Guo, 2004; Wang et al., 2008a).

The mechanism of 6-TG mediated genotoxicity is described below and shown in Figure 3-2. The pro-drug 6-TG can be actively up-taken by cells and enter the purine salvage pathway. Hypoxanthine-guanine phosphoribosyltransferase (Hprt) catalyses the conversion of 6-TG to 6-thio guanosine monophosphate (6-TGMP). 6-TGMP can be methylated by thiopurine *S*-methyltransferase (TPMT) through methyl group transfer from the co-enzyme *S*-adenosylmethionine (SAM) to form 6-methyl thioguanine monophsophate (6-meTGMP). 6-meTGMP can then be subsequently processed to form 6-meTG deoxynucleotide, which can be incorporated into DNA during DNA replication. 6-meTG pairs with T during replication, causing a single base-pair mismatch. In wild type cells, such a mismatch can be recognised and the MMR system is activated for repair. The MMR system may also be directly involved in the cell-cycle checkpoint to trigger cell-cycle arrests and apoptosis. If the modified base is in the template strand, the mismatch will be repeatedly generated by the polymerase and subsequently cause replication fork arrests (Jiricny, 2006). ATR/CHK1-dependent checkpoint signalling cascades are responsible for sensing replication blocks and mediating cell-cycle arrest and apoptosis in unsuccessful mismatch repair attempts (Jiricny, 2006). In addition, the damage could be bypassed by an error-prone DNA polymerase and the DNA damage signalling could be triggered by the repair in the subsequent rounds of replication.

In MMR-deficient cells, the mismatch is not recognised and it does not induce DNA damage signalling; therefore these cells survive in the presence of a genotoxic agent such as 6-TG but the cells bear a large number of mutations in their genome.

Figure 3-2: The mechanistic basis of 6-TG mediated cytotoxicity through MMR system.



See main text for the detailed description. PRPP: 5-phospho-ribosyl-1α-pyrophosphate; PPi: pyrophosphate ion; SAM: *S*-adenosyl methionine; SAH: *S* –adenosylhomocysteine; TMPT: thiopurine *S*-methyltransferase; 6-TGMP: 6-thio guanosine monophosphate; 6-meTGMP: 6-methylthioguanine monophosphate; 6-methyl-thio deoxy GTP: 6-methylthio deoxyl guanine triphosphate.

In this chapter, I describe how the mutagenic units used in the transposon were identified and how these cells were assessed functionally. The MMR pathway was used as a proof-of-principle screening system to validate this newly established mutagenic strategy.

2. Results

2.1. Mutagenic exon-pair selection

Selection of mutagenic exon-pair was achieved by scanning the annotated mouse genome computationally for terminal exon-intron structures which met defined criteria for efficient mutagenesis. This part of the work was done in conjunction with Steve Pettitt, another graduate student, in the Bradley group.

The computational selection criteria for the terminal exon pairs are as follows:

- Terminal exon pairs together with the intron in between the exons are defined as a mutagen unit.
- The mutagen unit must be relatively small (less than 2000 bp) to ensure highly efficient PB transposition.
- The mutagen unit must be present in all transcript variants of the endogenous gene, i.e. the mutagen exons are constitutively spliced, to ensure that the splice acceptors are relatively strong.
- Stop codons are present in the penultimate exon of the other two non-natural reading frames around 50 to 55 bp from the exon-intron junction.
- The mutagen unit must not contain any domain structure or structure/signal for cellular organelle localisation.

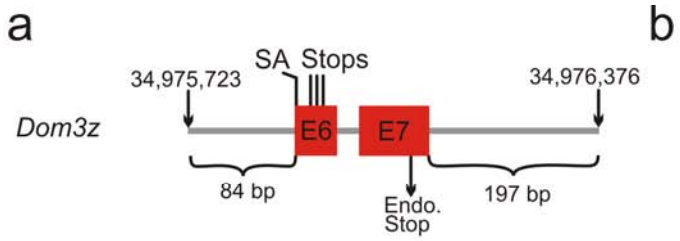
A short list of genes which contained mutagen units fitting the top four criteria were then manually curated at the protein sequence level and through literature searching to confirm the candidates that satisfy the final criterion. This was to avoid any possible fusion protein product having dominant-negative functionality which would complicate the mutant phenotype. Out of eight finalists, the two smallest mutagen units from the genes *Dom3z* and *Ccdc107* were selected.

Dom3z is located in the gene dense MHC class III regions on Chromosome 17, coding for a protein which is homologue to *C. elegans* DOM-3 and human *DOM3Z*. It binds to a nuclear

exoribonuclease, involved in the processing of 5.8S rRNA (Xue et al., 2000). *Ccdc107*, short for Coiled-coil domain-containing protein 107 precursors, resides on Chromosome 4 and this protein's function is currently unknown.

Sequence modifications were also conducted to introduce additional stop codons into the natural coding frames of the two mutagen units by site-directed mutagenesis. This was done by Steve Pettitt. Thus both mutagen units contain stop codons in all reading frames before the final exon-intron junction. This is an observed structure, which can induce a natural surveillance mechanism, nonsense mediated decay (NMD) pathway, to degrade mRNAs with premature stop codons (PTCs) (Chang et al., 2007). Thus these stop codons in the penultimate exon may be recognised as PTCs to induce the NMD pathway and degrade the trapped fusion transcripts. This feature may be particularly useful when the mutagen is inserted in the 3' end of a gene, and the fusion product may have sufficient endogenous protein structures to function normally. Figure 3-3 shows the structures and the locations of the stop codons of the mutagen pairs.

Figure 3-3: Mutagen *Dom3z* and *Ccdc107* exon-pair structures.

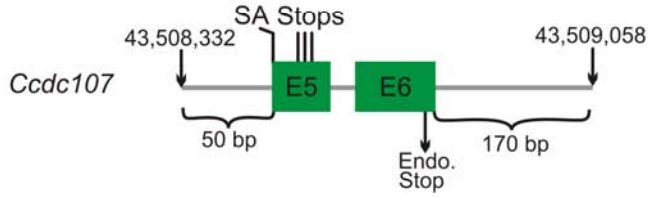


b

Stops in *Dom3z* E6:

Frame	Codon	Position*	Status
+1	TGA	16	Engineered
+2	TGA	35	Endogenous
+2	TAG	77	Endogenous
+3	TAG	57	Endogenous
+3	TAG	3	Endegenous

+1 is the natural reading frame for *Dom3z*.



d

Stops in *Ccdc107* E5:

Frame	Codon	Position*	Status
+1	TGA	3	Endogenous
+1	TAA	48	Endogenous
+1	TAG	36	Endogenous
+2	TGA	16	Engineered
+3	TAG	11	Engineered

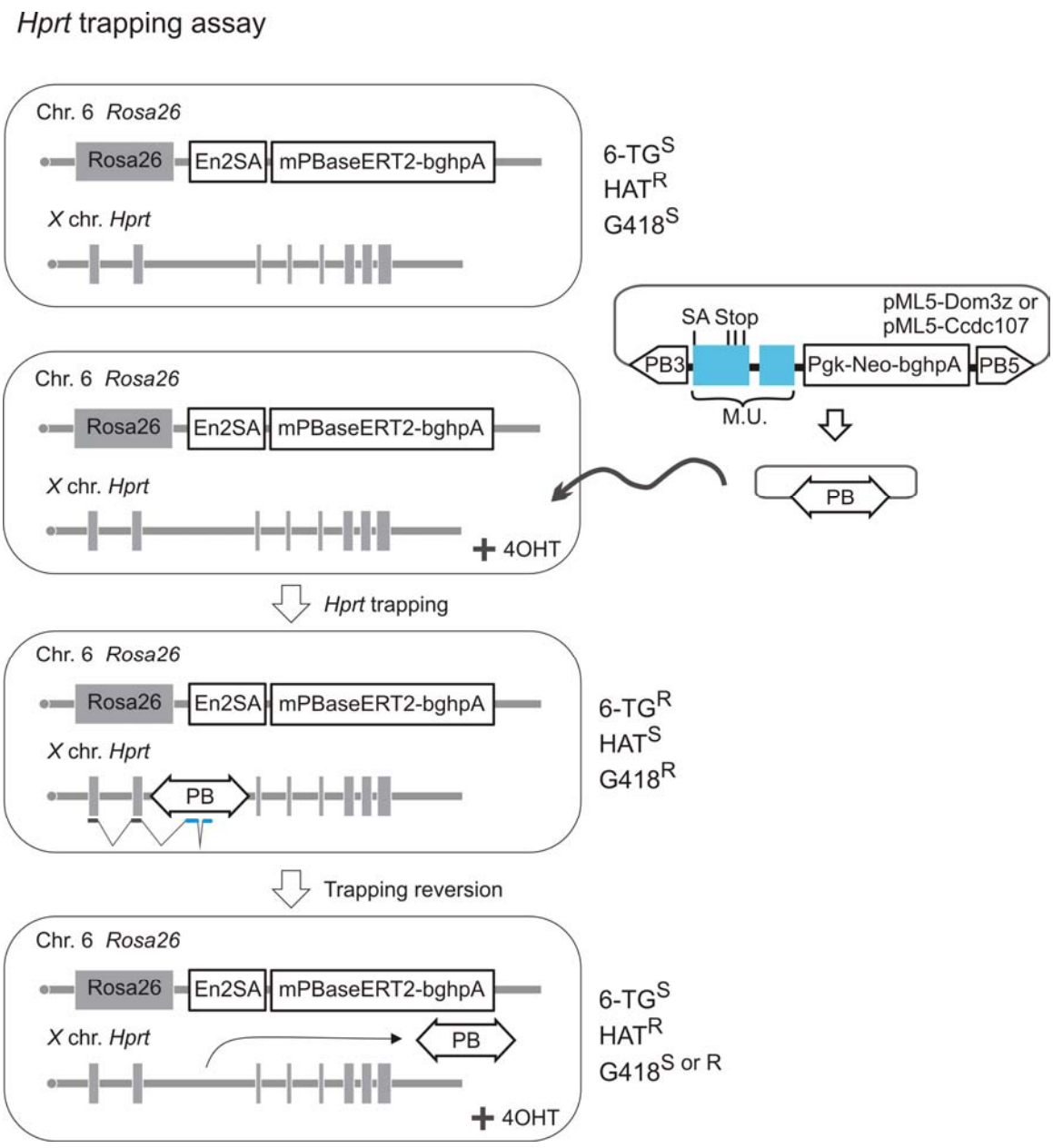
+3 is the natural reading frame for *Ccdc107*.

a and c, Schematic diagrams showing the mutagen structures. The coordinates indicate the genomic positions of the fragments based on NCBI37. SA, splice acceptor; Stops, stop codons; Endo. stop, endogenous stop codon. b,d, Endogenous and engineered stop codons within the penultimate exon of the mutagens to cover all reading frames. Position* indicates the nucleotide position of the "T" in the stop codons. Position 1 defines as the first nucleotide of the penultimate exon.

2.2. Validations of mutagen units using an *Hprt* trapping assay

Since the mutagen units were computationally selected, it was critical to assess their gene-inactivation efficiencies experimentally. The first assay, an *Hprt* trapping assay was designed to test the strength of the splice acceptor in competition with the endogenous splice acceptor of the trapped gene for generating null mutations, Figure 3-4.

Figure 3-4: Schematic representations of the *Hprt* trapping assay.



A saturating amount of PB transposon plasmids were introduced into an Hprt-proficient ES cell line (AB1) previously targeted with an inducible mPBase (mPBaseERT2) in the *Rosa26* locus, 3' to the Rosa26 endogenous promoter. mPBase can be induced with 4-OHT which results in its nuclear localisation. *Hprt* trapping events can be selected with drugs G418 and 6-TG. If the 6-TG resistant phenotype is caused by PB transposon inserting within *Hprt*, phenotypic rescue to HAT resistance is possible by reactivating mPBase by 4-OHT to re-mobilise the PB transposon out of the *Hprt* gene.

An ES cell line expressing an inducible form of the mPBase (AB1-R26$^{mPBaseERT2/+}$) was used for this assay, which provides a temporal control of transposase activation. In this cell line, mPBase was fused with a modified version of the human estrogen receptor ligand binding domain to its C-terminus (Cadinanos and Bradley, 2007). The mPBaseERT2 is driven by the constitutively active endogenous *Rosa26* promoter (Cadinanos, unpublished).

2.2.1. *Hprt* trapping assay with 6-TG and G418 dual selection

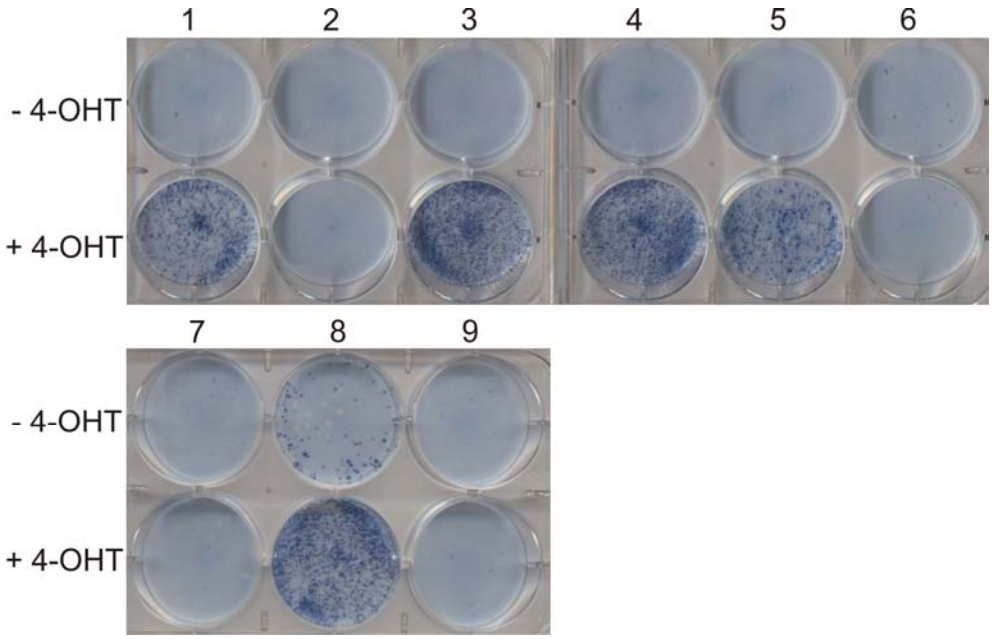
Two constructs (pML5-Dom3z and pML5-Ccdc107) were made with PB transposon harbouring the individual mutagen exon-pairs together with a NGK-Neo expression cassette. Integration of these mutagen-containing PB transposons into the *Hprt* locus with the correct orientation can inactive the *Hprt* expression, thus these cells can be selected with 6-TG to access the trapping ability.

In order to maximise the probability of transposon integration into the *Hprt* locus, a saturating amount (50 µg) of pML5-Dom3z and pML5-Ccdc107 plasmids were independently electroporated into 1x10^7 AB1-ROSA26$^{mPBaseERT2/+}$ cells. Electroporated cells were plated in six individual 90 mm plates. 1 µM 4-OHT was added 16 hours post-electroporation and sustained for two days. G418 selection was initiated 24 hours post-electroporation and was applied for five days to select cells with PB transposons integrated. G418 and 6-TG double selection was applied from day six onwards until colonies formed. Five clones were obtained from pML5-Dom3z transfected cells and four clones from pML5-Ccdc107 transfected cells.

The double resistant clones were further tested for the reversion of *Hprt* activity to establish the causality of PB transposon integration and the *Hprt* inactivation. The double resistant cells were plated in duplicate and 1 µM 4-OHT was applied o/n to one plate of the pair. HAT selection was added to all cells and was initiated 24 hours after the first application of 4-OHT. Five out of nine clones could be reverted to HAT resistance, suggesting the rest had inactivated *hprt* as a result of negative background mutations, Figure 3-6.

Splinkerette PCR was used to try to identify PB integration sites in the *Hprt* locus for the five Hprt-deficient clones which yielded revertants. In these experiments, a large amount of PB plasmid was used, thus there will be many PB integrations per cell. The Splinkerette PCR reaction does not amplify all insertion sites evenly, thus making it difficult to identify specific integrations in *Hprt*.

Figure 3-5: Reversion analysis of the 6-TG resistant colonies.



1-4 were clones isolated from pML5-Dom3z electroporation; 5-9 were clones isolated from pMl5-Ccdc107 electroporation. The colonies were plated in duplicate and incubated overnight with or without 4-OHT. The next day, the cells were selected using HAT.

2.2.2. 4-OHT induction optimisation

It has been previously estimated that 47 % of the PB integrations reside in genes and 80 % of the insertions in genes land in actively transcribed genes in ES cells (Liang et al., 2009). Based on this information, the number of cells with PB insertions required to obtain cells with the *Hprt* locus inactivated can be estimated, Table 3-1. The calculation assumes that the average copy number of the independent PB integrations is five when using the PB transposon saturation transfection method, the trapping efficiency of the mutagen is 100 %, and only one

Hprt trapping event per cell can occur. Using this calculation, the total number of cells with PB inserted required to generate five *Hprt*-trapped clones is $1.5x10^5$ cells. With the protocol used with the saturating amount of PB transposon, the number of cells harbouring integrated PB transposon within their genome can be over 10 % of the survival cells after electroporation (Wang et al., 2008b), i.e. at least $5x10^5$ cells should contain PB integrations with the 50 % survival rate post-electroporation. Therefore, per mutagen tested, at least 15 *Hprt* trapped cells should be generated per $1x10^7$ cells electroporated with the protocol used. This estimation indicates that the actual efficiency of obtaining *Hprt* trapped clones was low in this experiment.

Table 3-1: Estimate of number of cells required to obtain *Hprt* trapping events.

Hprt trapping efficiency estimate	
Total No. of cells with PB inserted	N
No. of cells with PB integrated, assuming 5 insertions/cell	5N
No. of insertions in genes (47 %)	2.5N
No. of insertions in actively transcribed genes in ES cells (80 %)	2N
Hprt size proportional to the actively transcribed genome size (34 kb/ $1x10^6$ kb*)	$6.8 x10^{-5}N$
Correct orientation to trap *Hprt*, assuming 100% trapping efficiency (50%)	$3.4 x10^{-5}N$
Value of N, if Hprt trapping in one cell is obtained	$3 x10^4$ cells

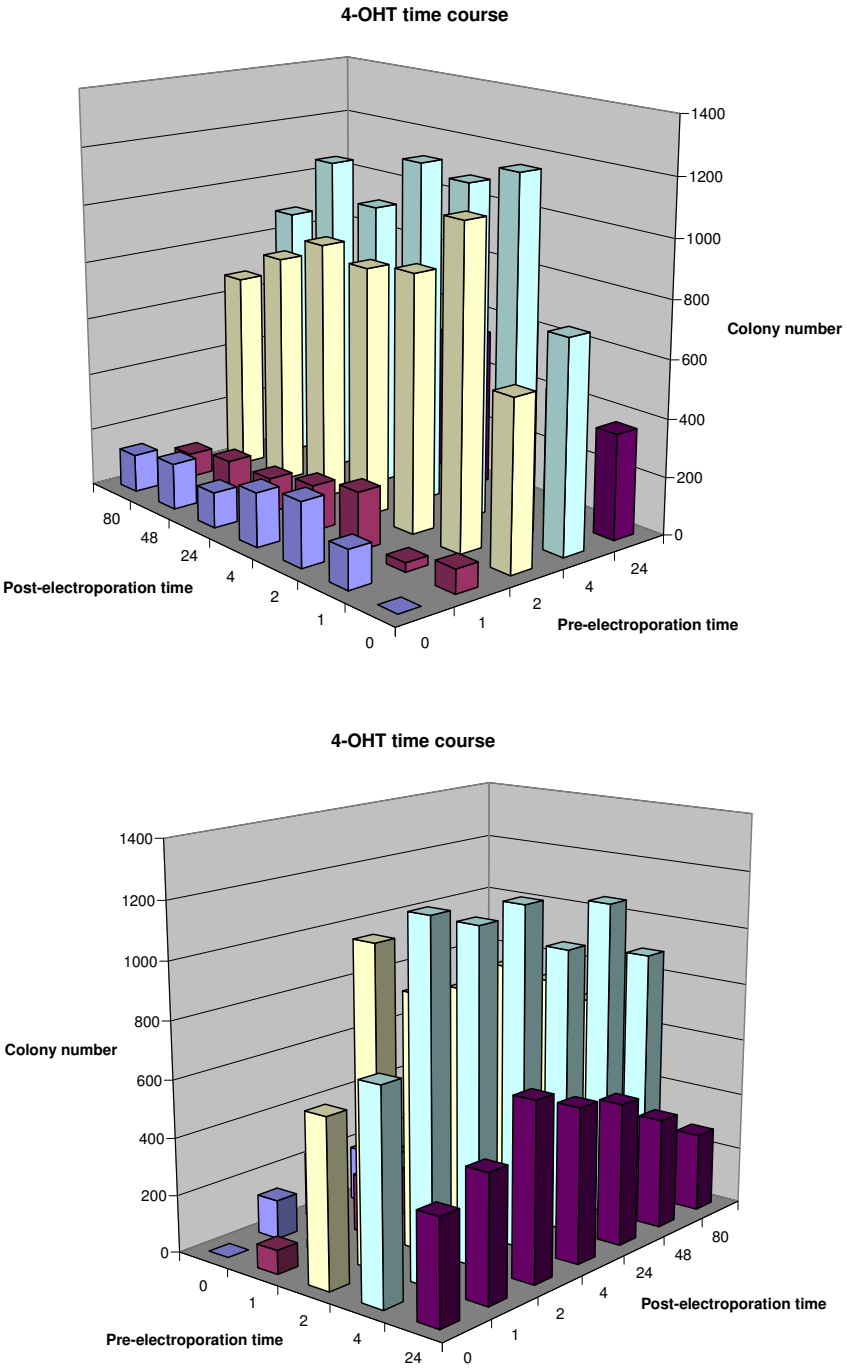
This table represents the *Hprt* trapping efficiency calculation. * $1x10^6$ bp is the total size of the genes that are actively transcribed in ES cells, and the size of each gene is defined as the absolute value of the differences between start and stop codon positions. The actively transcribed gene list and their start and stop codon coordinates were obtained from the gene trap consortium (http://www.sanger.ac.uk/Postgenomics/genetrap/).

The low efficiency with which the 6-TG resistant colonies were obtained could be due to a suboptimal 4-OHT induction protocol. Additionally, the high cell number plated may also give false-negative results due to the cross-killing Hprt-deficient cell by adjacent Hprt-proficient cells.

In order to optimise the induction condition, a pre- and post- electroporation time-course matrix of 4-OHT was set up to obtain the optimal 4-OHT induction protocol in order to achieve high transposition efficiency. For each time point, $1x10^7$ AB1-ROSA26$^{mPBaseERT2/+}$ cells were incubated with 4-OHT for the length of time required 0, 1, 4, and 24 hours before electroporation. Subsequently 1 µg of a plasmid with PB transposon harbouring PGK-Neo expression cassette (pML5) was electroporated. One twentieth of the electroporated cells were plated in six-well plates with 4OHT present for post-electroporation time points 0, 1, 2, 24, 48, and 80 hours. G418 selection was initiated 24 hours post electroporation. The results are summarised in Figure 3-6.

Short term pre-electroporation incubation with 4-OHT between two to four hours showed a significant elevation of transposition events overall. The dramatic effect of pre-electroporation incubation with 4-OHT provides an indication of the fast kinetics of PB transposition. The effect of post-electroporation incubation with 4-OHT was less overt, but prolonged incubation with 4-OHT showed a trend of reducing number of G418 resistant colonies. This may be due to continuous PB transposition, with each excision event having 40 - 50% probability of transposon lost. Overall, the optimal 4-OHT induction condition was a four-hour pre-electroporation incubation coupled with a 24 hour 4-OHT post-electroporation incubation.

Figure 3-6: 4-OHT induction protocol optimisation.



The top panel is viewed from the front while the bottom panel is viewed from the back.

2.2.3. *Hprt* trapping assay with sequential G418 and 6-TG selections

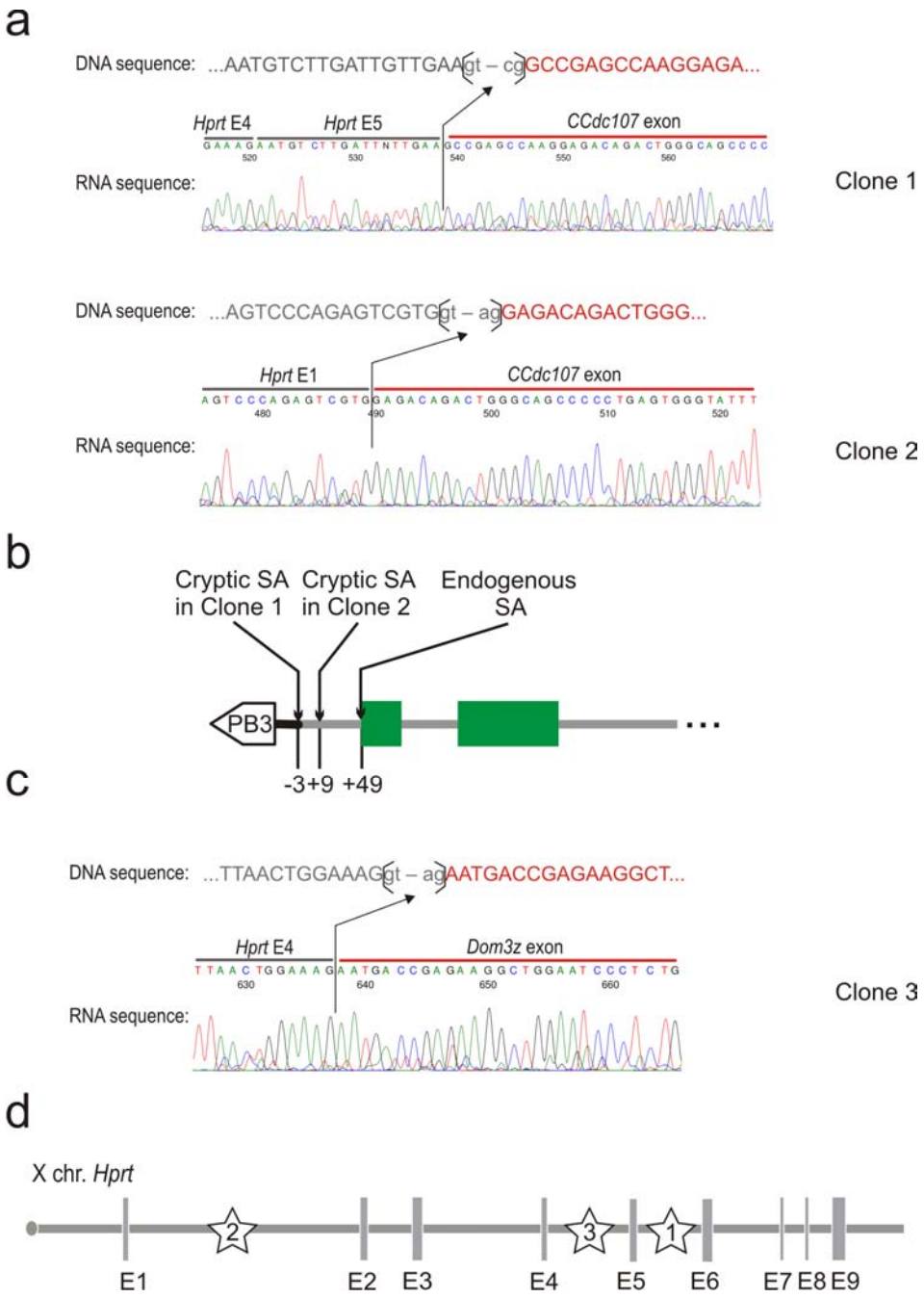
Sequential selection was used with the optimised 4-OHT induction in order to maximise the efficiency of obtaining the *Hprt* trapped clones. Selection using G418 a few days prior to G418 and 6-TG dual selection has several advantages. Firstly, initial G418 selection can eliminate a large proportion of non-transfected cells, thus enriching the cells with PB transposon integrated. Secondly, the replating of G418-selected cells at low density can reduce the cross killing of Hprt-deficient cells by the neighbouring Hprt-proficient cells. In addition, delayed selection of 6-TG allows the decay of *Hprt* mRNA and protein to avoid false negative results. Finally, PB transposon may have a delayed integration into the *Hprt* locus to give rise to a sub-population of the sibling cells within a clone containing Hprt-proficient cells. Without replating, these *Hprt*-trapped daughter cells can be crossed killed under 6-TG selection by sharing metabolites with their Hprt-proficient sibling cells.

AB1-ROSA26$^{mPBaseERT2/+}$ cells were treated with 4-OHT following the optimised protocol to induce PBase subsequently. 50 µg of pML5-Dom3z and pML5-Ccdc107 plasmids were electroporated into 1x10^7 4-OHT induced cells, independently. The electroporated cells were plated initially on a 90 mm plate and G418 was initiated 24 hours post-electroporation. G418 selection was sustained for four days, and colonies started to emerge. Cells were then trypsinised and these cells were plated in eight 90 mm plates at 2.5x10^5 cells per plate. 6-TG and G418 selection was applied directly after replating, and the selection was sustained for 12 days. In total, 20 double resistant colonies were obtained from the pML5-Dom3z transfected cells and 18 from pML5-Ccdc107 cells. The double resistant colonies were subjected to the reversion analysis as described before. Following 4-OHT induction, five clones yielded *Hprt* revertants in pML5-Dom3z transfected cells while six clones were obtained from pML5-Dom3z transfected cells. The non-revertants could due to the spontaneous null mutations of the *Hprt* locus.

Splinkerette PCR was performed to identify the PB integration sites within the *Hprt* locus. Three clones (one from pML5-Dom3z and two from pM5-Ccdc107) out of the 11 yielded the PB-*Hprt* genomic junction fragments. RT-PCR was performed using a mutagen exon-specific

primer (either Dom3z or Ccdc107 depending on the orientation of the integration in relation to the *Hprt* transcription) and an *Hprt*-exon-specific primer on these three clones to confirm the trapping of *Hprt* and this provided an opportunity to access the splicing structures of the fusion transcripts, Figure 3-7. Sequencing of the RT-PCR products of the fusion transcripts confirmed that the splice junction between the *Hprt* and Dom3z mutagenic exons were as predicted with the endogenous splice acceptor mediating the trapping (Clone 3, Figure 3-7b). However, in neither case of the Ccdc107 mutagen-mediated trapping, was the endogenous splice acceptor from the penultimate Ccdc107 exon used. Instead, two independent cryptic splice acceptors were identified (Clone 2 and Clone 3, Figure 3-7a). The cryptic splice acceptor indentified from Clone 1 was present at the cloning junction between the endogenous DNA fragment containing *Ccdc107* and the vector sequence; whereas the cryptic splice acceptor observed in Clone 2 was within the endogenous intronic sequence of *Ccdc107* (Figure3-7a,b). Both these splice acceptors were located 5' of the endogenous splice acceptor of the exon 5 of the *Ccdc107* gene.

Figure 3-7: PB transposon-mediated *Hprt* trapping.

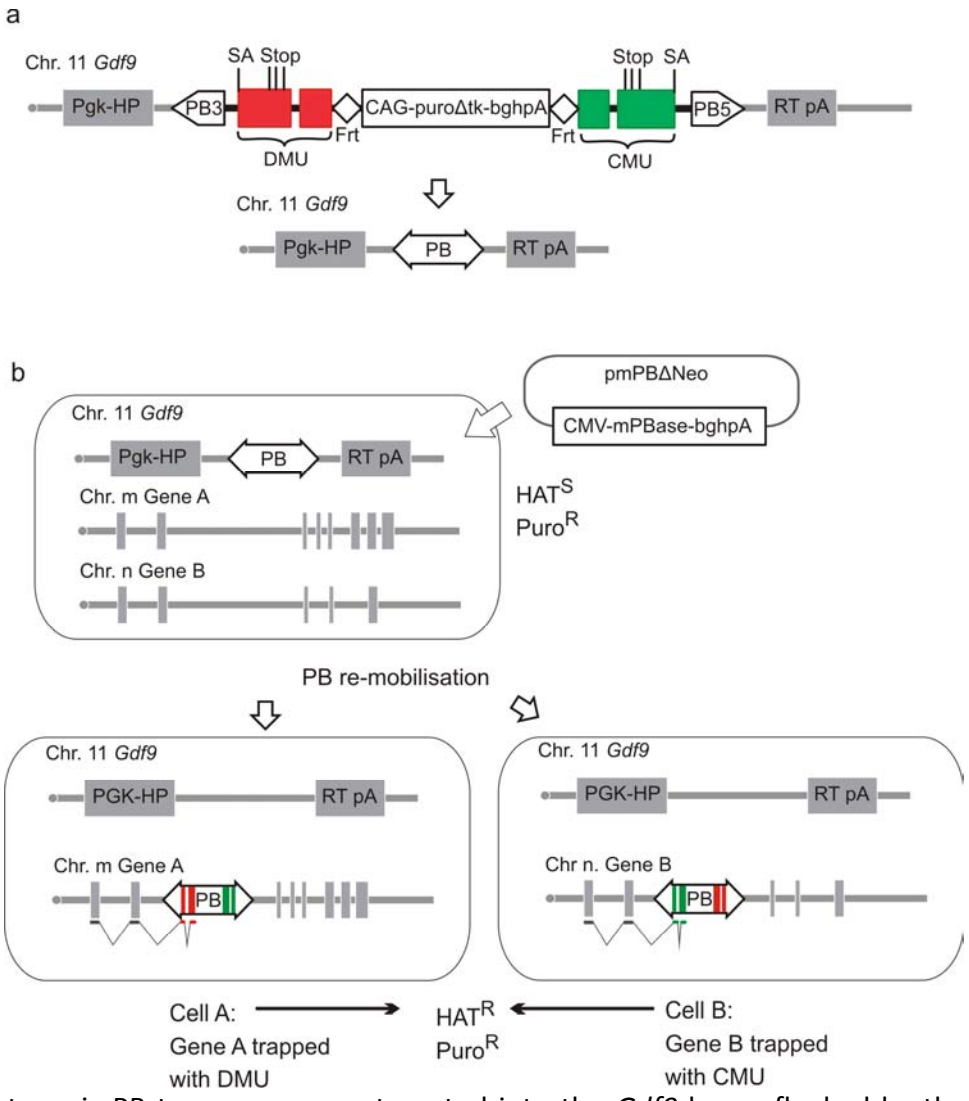


a,c, RT-PCR sequence traces of the three revertible 6-TG resistant clones. b, schematic summary of the two cryptic splice acceptors identified with the Ccdc107 mutagen. The number indicates the location of the splice acceptors with the first base from the cloned endogenous Ccdc107 sequence to be "0". The base position is referred to the first "G" in the exonic sequence after the splice junction. d. Summary of the integration sites of the three revertible 6-TG resistant clones in the *Hprt* locus, with the stars representing the integration sites and the number within the star denoting the clone I.D.

2.3. Validation of mutagen units using a random trapping assay

The second validation assay, a random trapping assay was designed to test whether the endogenous splice acceptors of the penultimate exons from *Dom3z* and *Ccdc107* are capable of trapping endogenous genes in different genomic contexts. The schematic representation of the assay is shown in Figure 3-8.

Figure 3-8: Schematic representation of the random trapping assay.

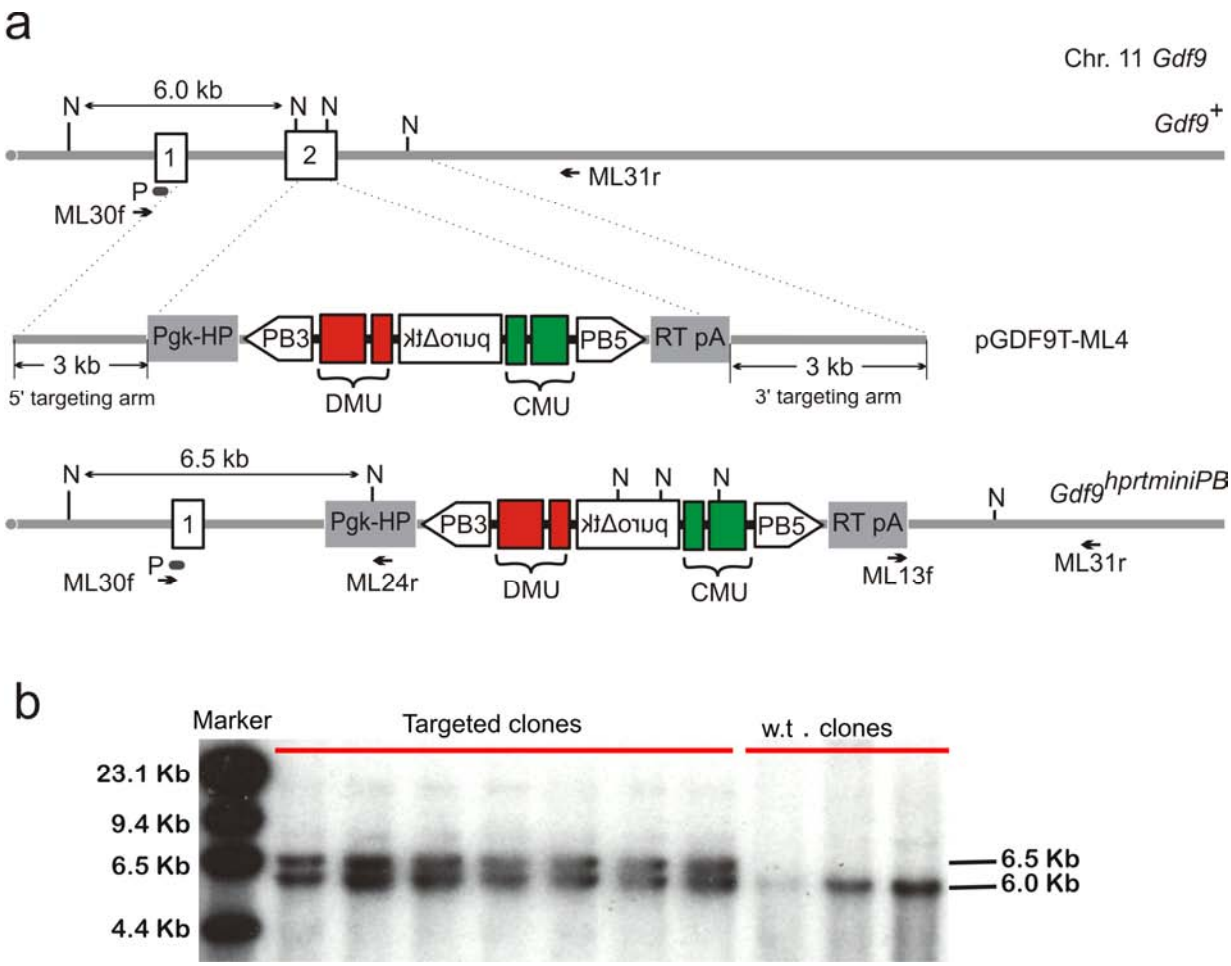


a, The mutagenic PB transposon was targeted into the *Gdf9* locus, flanked by the 5' and 3' part of the *HPRT* minigene cassette. HP, 5'part of the human HPRT cDNA; RT-pA, 3' part of the human *HPRT* cDNA with polyA signal. b. Post mPBase introduction, PB re-integration events in introns can be used to characterise the trapping capability of the mutagen units (DMU, Dom3z mutagen unit and CMU, Ccdc107 mutagen unit) in different genomic contexts. Both Dom3z and Ccdc107 mutagen units can be assessed independently.

2.3.1. Generation of a cell line with an autosomal single-copy PB transposon

A cell line was made in order to conduct the random trapping assay. A single copy of the mutagenic PB transposon was targeted into the autosomal locus *Gdf9*, flanked by the two portions of the *HPRT* minigene, Figure 3-9a. A targeting vector pGDF9T-ML4 was constructed by inserting the mutagenic PB transposon into an existing Gdf9 targeting vector by recombineering (Chapter Two). The targeting vector was linearised by *Pme*I, purified by ethanol precipitation and electroporated into $1x10^7$ NN5 ES cells. The NN5 cell line was derived from AB2.2, in which the endogenous *Hprt* locus was inactive (Kuehn et al., 1987). Puromycin selection was applied 24 hours post-electroporation and 48 colonies were analysed by long-range PCR for both the 5' and 3' junctions of the targeting arms to the contiguous genomic regions. Seven clones with positive results for long-range PCR for both targeting arms were further confirmed by Southern blotting to be positive, thus the targeting efficiency is 15 % at this locus using this construct, Figure 3-9. The resulting cell line was termed NN5-Gdf9$^{hprtminiPB/+}$.

Figure 3-9: Gene targeting of the *Gdf9* locus to generate the NN5-*Gdf9*$^{hprtminiPB/+}$ cell line.



a, Cartoon representation of the DNA structures of the wild type *Gdf9* allele (*Gdf9*$^{+}$) and the targeted allele (*Gdf9*^{hprtminiPB}). N, *Nco*I recognition site; P, southern probe; DMU, Dom3z mutagenic unit; CMU, Ccdc107 mutagenic unit; ML30f and ML24r are primers used for 5' long-range PCR; ML13f and ML31r are primers used for 3' long-range PCR. b, Southern blot confirmation of the targeted clones. Marker, Lamda *Hind*III ladder.

2.3.2. The random trapping assay

1x10^{7} NN5-*Gdf9*$^{hprtminiPB/+}$ cells were electroporated with 25 µg of mPBase containing plasmid (pmPBΔNeo), and the electroporated cells were plated in a 90 mm plate and selected with HAT and Puro the next day for the clones with PB transposon has been excised from the Gdf9-HPRT locus and re-integrated in the genome. In total, 1,300 colonies were obtained from the electroporation. Ninety six Double-resistant colonies were picked and were subjected to

Splinkerette PCR to identify the locations of re-integrated PB transposons. All clones with uniquely identifiable genomic locations (84 clones) were further analysed for insertions within introns in genes which were transcriptionally active in ES cells based on the genome-wide ES cell gene trap resource (with a total of 165,778 trapped events are present in the database). Clones with PB transposons present in introns were examined and five selected in which the Dom3z mutagenic unit is in the same orientation as the gene, while another five clones were selected on the basis of the orientation of the Ccdc107 mutagenic unit, Table 3-2.

The ten clones were subjected to RT-PCR analysis to determine the efficiency of gene trapping in these randomly-selected genes using a gene-specific upstream primer with a mutagen specific terminal-exon primer. The chromosomal locations and the position of the PB transposons with the selected genes were described in Table 3-2, and Figure 3-10a shows the RT-PCR results.

Table 3-2: Ten PB reintegration clones selected from the random trapping.

Gene	Chr	Location	Start	End	Mutagen	Clone
Tmem131	1	Intron 4	36919002	36919531	*Ccdc107*	D7
Dut	2	Intron 4	125082204	125082534	*Ccdc107*	G10
Fstl4	11	2nd last intron	52993736	52994119	*Dom3z*	E8
A24Rik	15	Intron 17-18	72649435	72649901	*Dom3z*	C4
Sfrs3	17	Intron 2	29173721	29174157	*Ccdc107*	E4
Sema6a	18	Intron 1	47487960	47488080	*Dom3z*	F8
Undcd3	11	Intron 1	6097207	6097470	*Ccdc107*	H8
stk22s1	7	Intron 6-7	52208693	52208790	*Dom3z*	A8
Ran	5	Intron 4	129527403	129527558	*Dom3z*	F2
Tmem50a	4	Intron 1	134469873	134470347	*Ccdc107*	B3

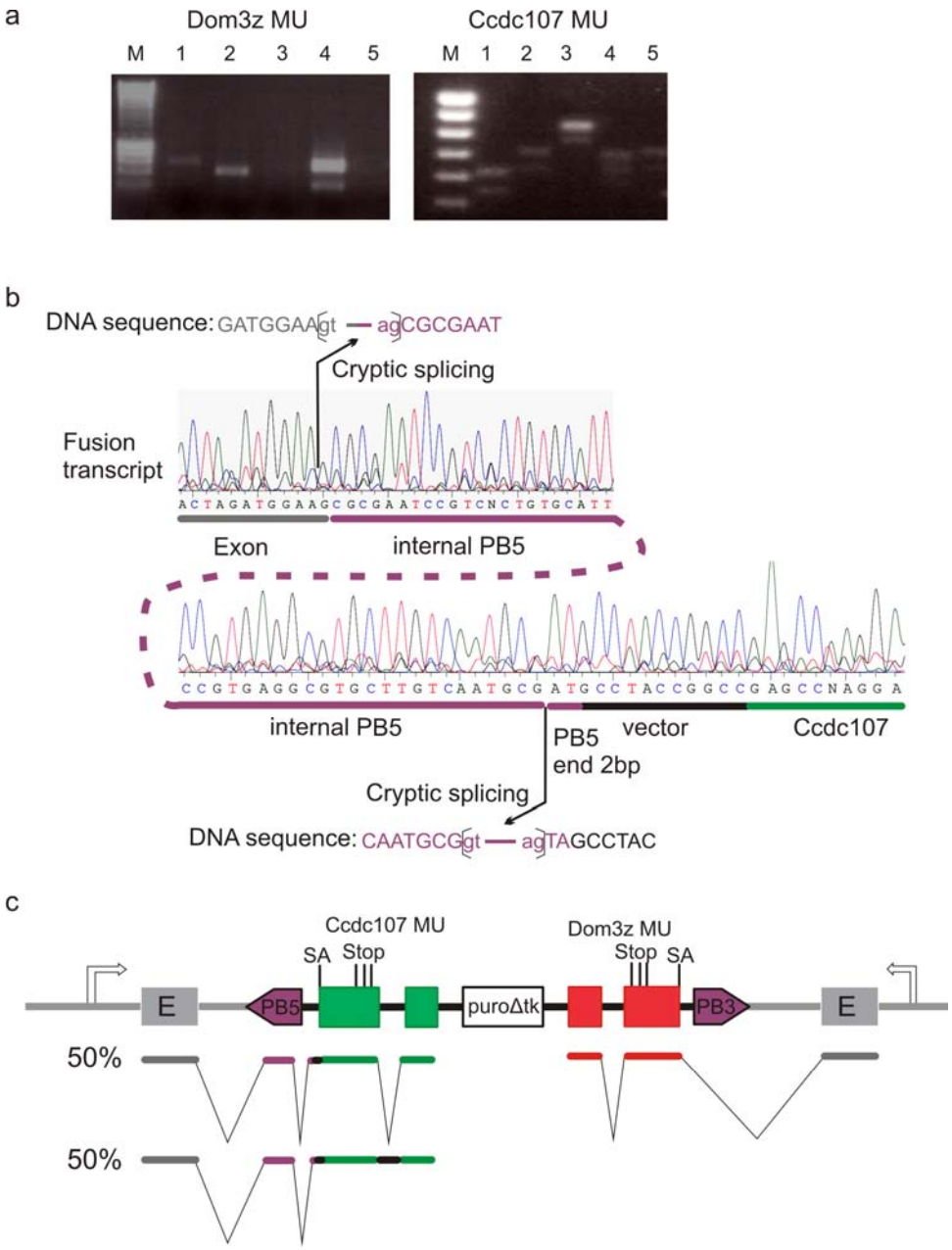
The locus-specific primers for these clones can also be found in Table 2-4 Chapter Two. NCBIm36 was used for the mapping of the PB integration sites.

In clones where Dom3z was the trap, three out of the five clones analysed showed PCR products representing the fusion products. Sequencing of the RT-PCR fusion products confirmed that the endogenous splice acceptor from the Dom3z penultimate exon was mediating the trapping as predicted. Two clones did not yield any PT-PCR products. The minor product in Clone 4 was non-specific PCR amplification. There are two possibilities to explain this, firstly, the failure for Dom3z to mediate trapping and secondly highly efficient NMD effect. Further experiments will be required to distinguish the two possibilities.

In clones where the Ccdc107 mutagen unit mediated trapping, all clones generated fusion products (Figure 3-10a). However, unexpectedly, every clone produced doublet PCR bands with similar intensity. Sequence analysis of these PCR products revealed further complexity of these fusion transcripts. The two PCR products were produced due to the alternative splicing of the intron between the two Ccdc107 terminal exons. More interestingly, a cryptic splice acceptor and donor were present within the PB5 sequence, which trapped transcripts coming from the upstream exon in all five cases, Figure 3-10b,c.

Taken together, the trapping mediated from the Ccdc107 orientation did not occur as predicted due to the presence of cryptic splice acceptors within the PB5 sequence and the alternative splicing of the intron between the mutagenic terminal exons. However, trapping is observed in all random loci tested. The trapping mediated from the Dom3z orientation occurs as predicted, however, two clones did not show fusion product amplification. Figure 3-10c summarises the trapping structures for both mutagenic units. In conclusion from this analysis, the mutagenic PB transposon should yield the gene-trap mutations 80 % of the total integration events where PB transposon land in genes.

Figure 3-10: Characterisation of the trapping events mediated by the mutagenic units.



a, RT-PCR results from ten clones studied. Genes trapped with the Dom3z mutagenic unit (left): 1, *Fstl4*, 2, *A24Rik*, 3, *Sema6a* 4, *stk22s1*, 5, *Ran*; Genes trapped for Ccc107 mutagenic unit (right): 1, *Tmem131*, 2, *Dut 3*, *Sfrs3*, 4, *Undcd3*, 5, *Tmem50a*. b, Sequence trace from no.3 of Ccdc107 mediated trapping, when PB transposon inserted within intron 2 of Sfrs3 on chromosome 17. One cryptic splice donor and two cryptic splice acceptors were observed within the PB5'ITR. c, Schematic summary of the trapping characteristics of the mutagenic PB transposon. The colour schemes in b and c are coordinated so that they correspond to the origins of the DNA/RNA sequences.

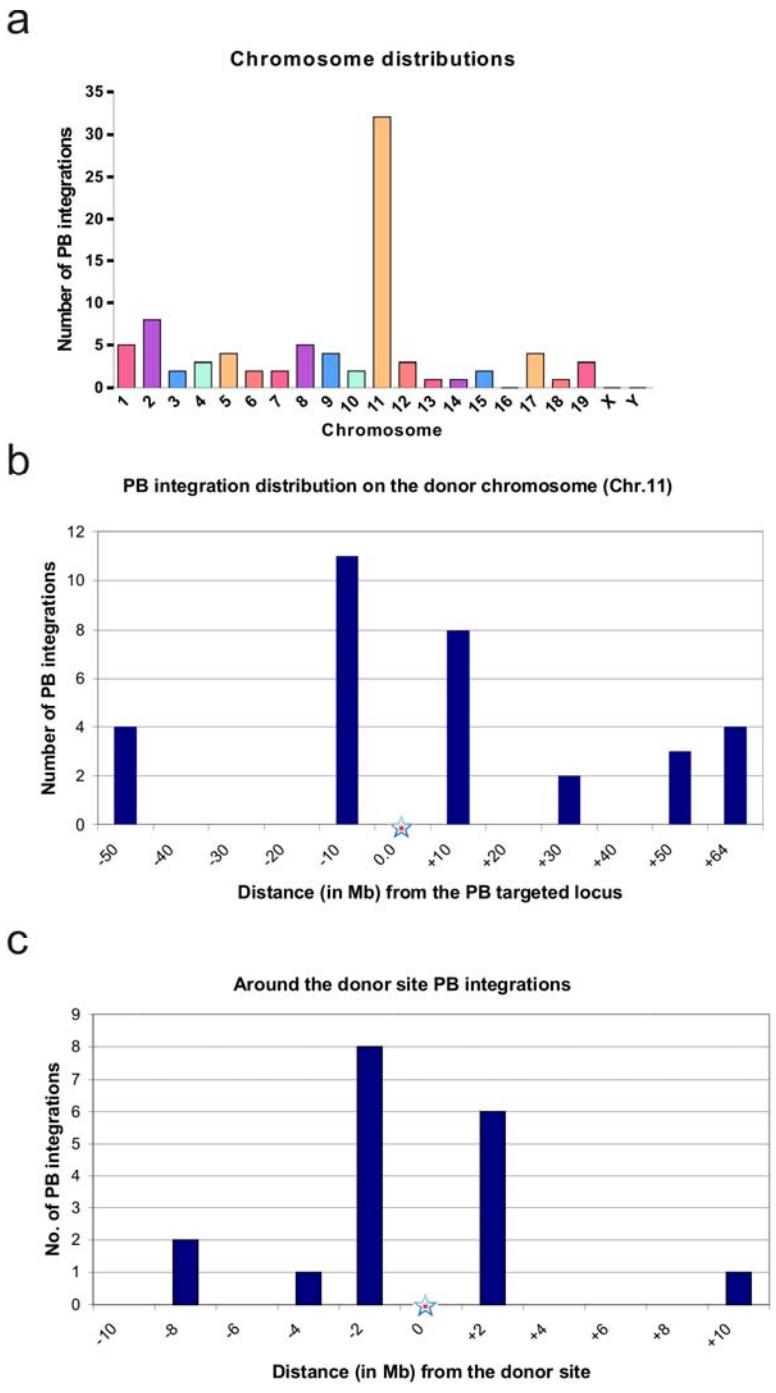
2.4. Local hopping of PB around the *Gdf9* locus

From the 96 puromycin and HAT double resistant clones generated in the random trapping assay, 86 clones were mapped to unique locations in the mouse genome with high alignment quality. Genome integration sites were analysed in order to gain insight into the PB transposon re-mobilisation characteristics following excision from the *Gdf9* locus.

Surprisingly, 32 out of 84 insertions were found on chromosome 11 (38 %), Figure 3-11a. In the vicinity of the donor site, 14 out of 32 reintegrations on chromosome 11 (44%) were within a 2.4 Mb of the donor site, Figure 3-11b,c. Although the dataset analysed here is small, the local hopping bias towards the donor chromosome and the donor site is clear. The number of centromeric and telomeric integrations also seems to be higher than other regions on chromosome 11 apart from the donor site.

Since the re-mobilisation selection strategy in this experiment is independent from the gene trapping status, the gene preference of PB integration can be analysed in an unbiased fashion. Of the 84 integrations, 42 landed in genes (50 %), 15 of which are on chromosome 11, and six were within the 2.4 Mb regions surrounding the donor site. It is known that PB-mediated integration has a bias towards genes compared with a simulated prediction for random integrations (Liang et al., 2009). For my small set of 36 insertions in genes which are not at the donor region, 30 insertions were within genes which have been trapped previously in mouse ES cells using a selection-based trapping strategy (a total of 165,778 trapping events in the database), which indicates that these genes are actively transcribed in mouse ES cells (http://www.sanger.ac.uk/PostGenomics/genetrap/). Only two out of six integrations within the 2.4 Mb region have not been previously trapped. The PB transposon donor site is within a gene-dense region that is actively transcribed. The characteristics of this donor region may bias for local PB integrations within this region as the PB transposon integration is favoured to chromosomal regions that are actively transcribed, possibility due to the easy access to open chromatin structures.

Figure 3-11: Local hopping observed with PB transposon mobilised from the *Gdf9* locus.



a, Genome-wide distribution of PB re-integration, with a total of 84 events analysed. b,c, Local clustering of re-integrations surrounding the donor site. The star represents the donor locus where PB was inserted by gene targeting. “0” is the targeted locus of PB transposon. “- “ is 5’ away from the donor site and “+” is 3’ downstream of the donor site.

2.5. Proof-of-principle of the mutagenic strategy in a DNA mismatch repair screen

In order to experimentally validate that the mutagenic strategy using my PB transposon could be coupled with the *Blm*-deficient ES cell system for recessive genetic screens, a proof-of-principle screen was conducted using the NN5-Gdf9$^{hprtminiPB/+}$ cell line using the *Gdf9* locus as the donor site, to identifying components of the DNA mismatch repair pathway with 6-TG selection as the phenotypic readout.

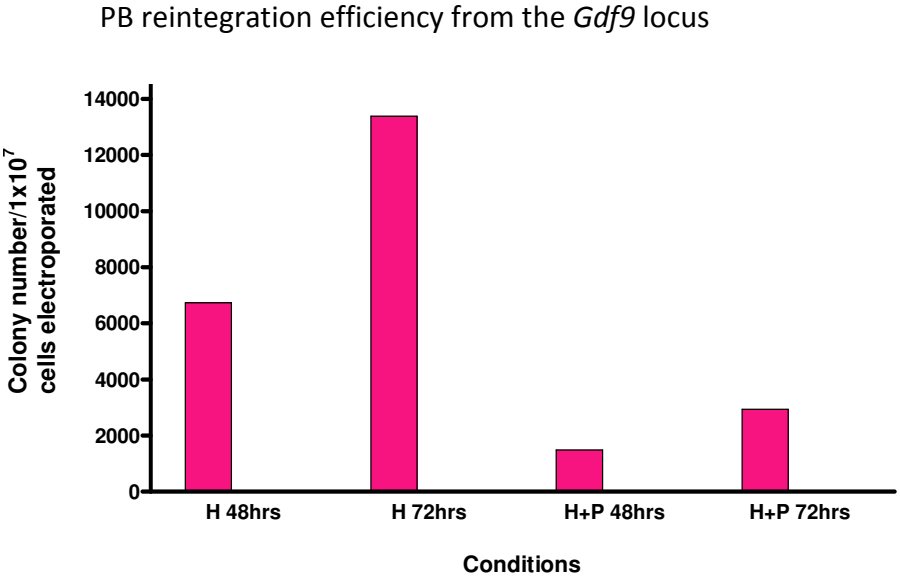
2.5.1. PB re-integration efficiency estimation in NN5- Gdf9$^{hprtminiPB/+}$ cells

In order to provide an estimate of the efficiency of mutant generation by PB re-integration using NN5- Gdf9$^{hprtminiPB/+}$ cells, the number of clones with PB transposon excised from the donor *Gdf9* locus and reintegrated in the genome could be measured. The mutagenic transposon was mobilised by electroporating 1x10^7 NN5- Gdf9$^{hprtminiPB/+}$ cells with 25 µg of mPBase expression plasmid (mPBΔNeo). The electroporated cells were plated in ten 90 mm plates equally and the cells were selected with either HAT containing medium alone or HAT with puromycin. The selection was initiated either 48 or 72 hours post-electroporation, allowing sufficient time for PB re-integration to occur, Figure 3-12. A negative control experiment was carried out in parallel, without the electroporation of PBase expression plasmid.

In the control electroporation without the PBase, no colony was formed under any selection condition. The numbers of HAT or HAT and Puro double resistant colonies were two fold higher when selection was initiated 72 hours rather than 48 hours post-electroporation, but Hprt-deficient ES cells (i.e. cells without PB excision from the donor site) can be cross-rescued by Hprt-proficient cells (cells with PB excised from the donor site), giving rise to mixed colonies. The number of HAT and Puro double-resistant colonies was a quarter of HAT resistant colonies. The reintegration efficiency estimated here was lower than previously estimated without the positive selection for re-integration (Liang et al., 2009). The discrepancy between the two measurements is due to the timing of the puromycin selection in this experiment relative to the PB reintegration kinetics. The previous measurement was based on HAT-based excision selection (Liang et al., 2009); therefore, delayed PB

reintegration events may represent a small proportion of cells within a colony and could be detected by the Splinkerette PCR method. However, such kind of colonies can not survive under direct puromycin selection. A 48-hour post-electroporation selection scheme provided 1,500 HAT and puro double-resistant colonies per $1x10^7$ cells electroporated. This condition provides good mutant complexity per pool for the mutant library construction.

Figure 3-12: PB re-integration efficiency excising from the *Gdf9* locus.



H, HAT; P, puromycin.

2.5.2. Library construction and screening

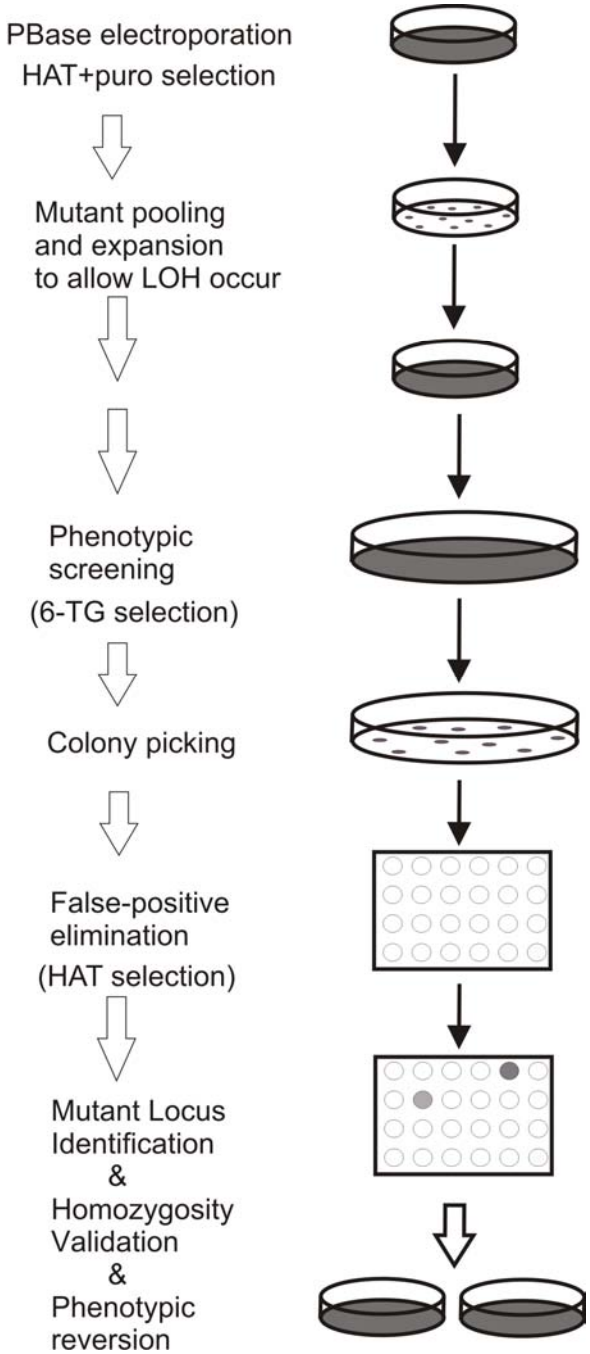
Ten electroporations were conducted with the electroporation condition described above, and the cells resulting from each electroporation was plated in a 90 mm plate were selected in HAT and puromycin containing medium until colonies were formed, which yielded approximately 1,300 heterozygous mutants per 90 mm plate. The colonies were pooled in each 90 mm plate and passaged into a fresh 90 mm plate until the plate was confluent ($3x10^7$ cells/90 mm plate). The cells were then passaged 1:1 from a 90mm plate to a 150 mm plate, and 2 μM 6-TG selection was commenced 24 hours after plating to select for DNA MMR mutants. The experimental scheme is shown in Figure 3-13. MMR mutant isolation using 6-

TG selection can be conducted at high cell density, as 6-TG-induced cytotoxicity is only effective in cells with proficient MMR mechanism. MMR mutant cells are insensitive to the accumulation of mismatch repairs in the genome, therefore they survive even surrounding cells are continuously supplying them with genome-toxic metabolites derived from 6-TG.

After eight days of 6-TG selection, around 100 - 200 colonies were formed on each 150 mm plate. The number of 6-TG resistant colonies generated was too high to be real, as it is unlikely that 1,000 to 2,000 genes are involved in the MMR system. In addition, previous screens using libraries with a similar heterozygous mutant complexity to this experiment yielded less than twenty clones were produced (Guo, 2004; Wang et al., 2008a). There may be two possibilities for the generation of such large amount of 6-TG resistant clones. Firstly, LOH events within MMR relevant genes occurred very early on during the expansion, thus the homozygous daughter cells expanded several generations before 6-TG selection was initiated. Secondly, false positive clones may present due to loss of the *HPRT* minigene, as Hprt-deficient ES cells are resistant to 6-TG selection. The method to distinguish the two possibilities is to select these clones individually with HAT. If the clones were generated due to loss of the *HPRT* minigene, they should be sensitive to HAT selection. As the *HPRT* minigene is on one of the homologous chromosomes of an autosomal locus, the rate of losing the *HPRT* minigene in *Blm*-deficient ES cells is high, approximately one per 2,000 cells post PB transposon excision. Cells deficient in *Hprt* are 6-TG resistant irrespective to the MMR status, as 6-TG can not be converted and incorporated into DNA synthesis, thus contributing to the majority of the 6-TG resistant colonies formed.

600 Colonies were picked in total from all ten 150 mm plates. The cells in 96-well plates were selected under HAT for four days. Seventeen HAT resistant colonies were obtained from these 600 colonies. Therefore, the large proportion of false positive clones was highly likely due to loss of the *HPRT* minigene on the *Gdf9* locus by LOH.

Figure 3-13: Schematic representation of the experimental procedures for isolating homozygote MMR mutants.



The seventeen 6-TG and HAT double resistant clones were expanded and the PB integration sites were identified using the Splinkerette PCR. The results are summarised in Table 3-3. Seven independent insertions were found within the seventeen clones, and four were mapped within 600 kb region in four different genes surrounding the PB donor locus. These are likely to reflect the local hopping effect observed previously. In these clones, the PB integration sites are unlikely to relate the 6-TG resistant phenotype. One insertion was mapped to an intergenic region. One PB integration site was mapped to a gene *Rrp9* (ribosomal RNA processing 9), which is a component of a nucleolar small nuclear ribonucleoprotein particle, snoRNP, thought to participate in the processing and modification of the pre-ribosomal RNA (UniProtKB/Swiss-Prot). *Rrp9* therefore is unlikely to be a candidate of the MMR system. It is likely that spontaneous mutations in other MMR components have occurred in these clones, as *Blm* deficiency promotes mutagenesis and the conversion to homozygosity. The MMR screen using 6-TG promotes the accumulation of mutations in the genome; therefore background spontaneous mutations in this screen can be elevated compared to other screens. A final PB integration sites for three daughter clones was in intron 1 of *Msh6*, a known gene essential for the DNA mismatch recognition. The insertion sites are summarised in Table 3-3.

Table 3-3: Details of seven independent integration sites identified from the MMR screen.

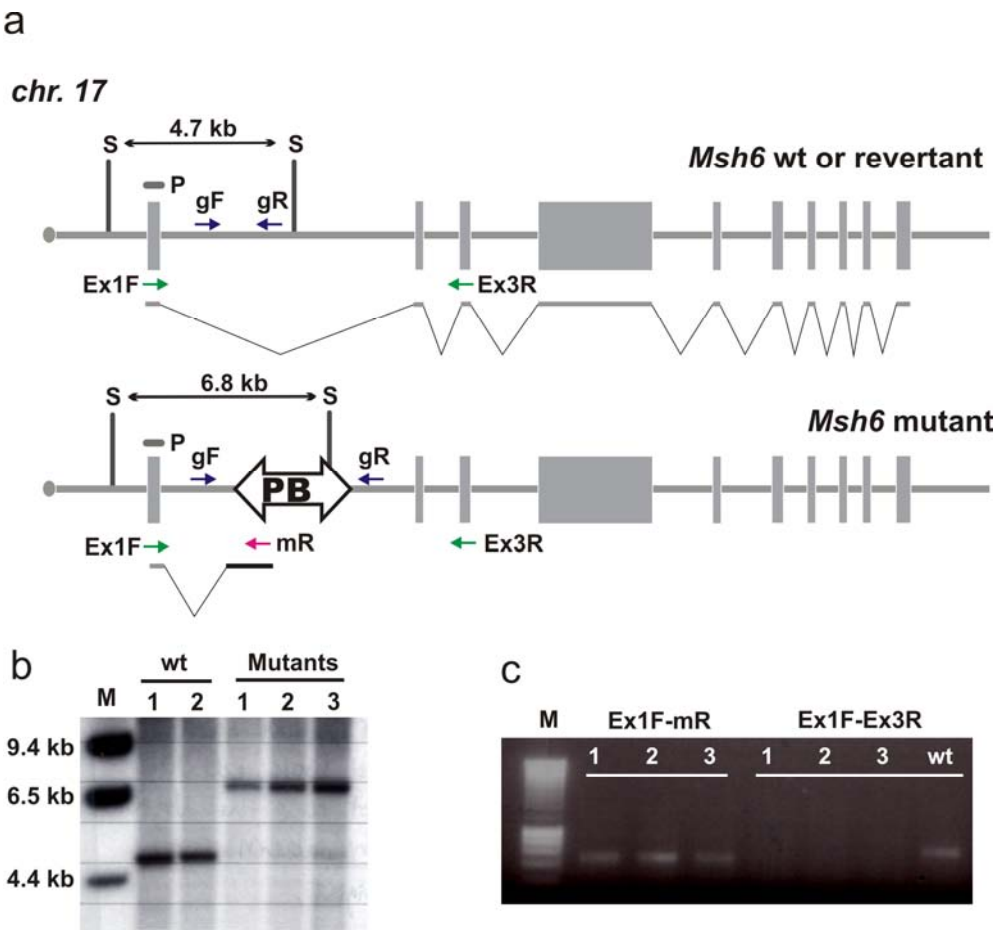
Pool	Chr.	In gene	Samples	Start	End	PBITR	Strand	Length
1	11	Aff4	1	53166717	53166863	5	C	147
3	11	Fst14	3	52847204	52847285	5	F	82
5	11	Il13	4	53447234	53447397	5	F	164
7	17	Msh6	5,6,7	88376055	88376181	5	F	127
6	9	Rrp9	8,10,14,16,17	106380210	106380327	3	C	118
6	11	Sept8	9,11,12,13	53355879	53356054	5	F	176
6	5	No	15	127578422	127578557	3	C	136

Note that the integrations on chromosome 11 are within 600 kb from the donor site. NCBIm36 was used for the integration-site mapping.

2.5.3. Mutant analysis and validation

This *Msh6* mutant was analysed for its homozygosity status. An external DNA probe was used so that both wild type and mutant alleles could be detected by Southern analysis. This clone was confirmed to be homozygous by Southern blotting, Figure 3-14a,b. Further analysis of the transcript by RT-PCR confirmed that the mutant was null for *Msh6*, Figure 3-14c.

Figure 3-14: *Msh6* mutant validation.



a, Schematic representation of the wild type (revertant) and mutant allele, with Southern blotting detection strategy (b) and the primers used for genomic PCR and RT-PCR (c). S, *Spe*I recognition site; P, Southern blotting probe. b, Southern blotting to confirm the homozygosity status of the mutant. c, RT-PCR analysis. 1, 2, 3, are three sister clones from the mutant. The primers used for the PCR reaction are indicated above the samples. wt, wild type control; M, marker. mR, reverse primer from the terminal exon of *Dom3z*.

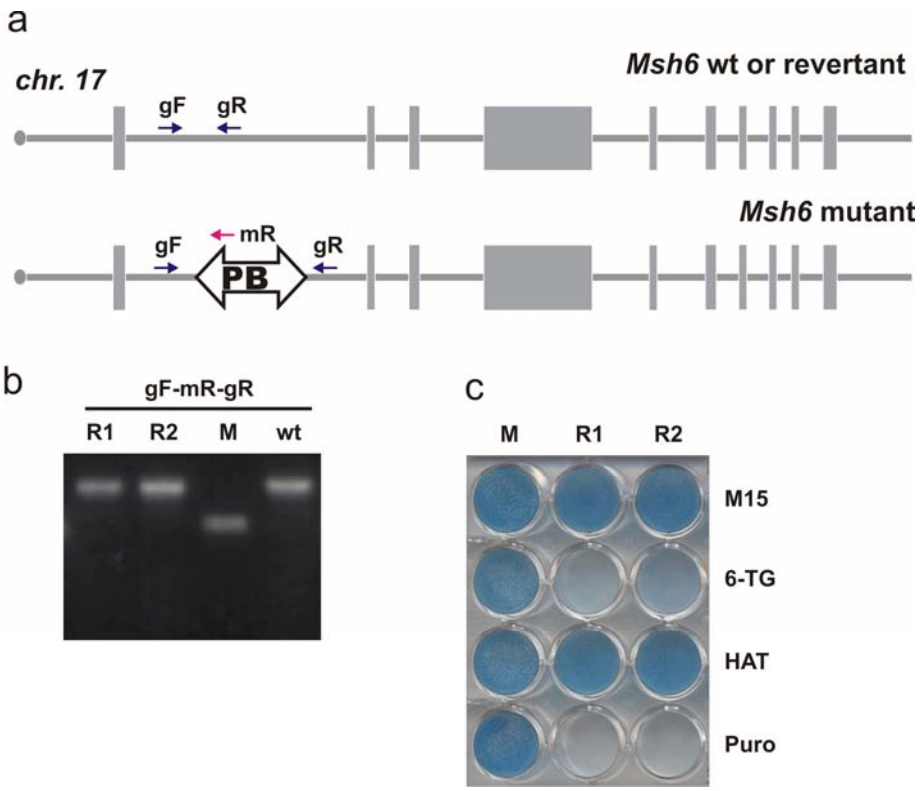
Msh6 is a known gene for the MMR pathway and homozygous mutant of this locus can be used to functionally validate the use of PBase to revert the genotype of this mutant, further confirming the genotype-phenotype causality. Removal of the mutagenic transposon from the intron of the *Msh6* gene allows the transcription of the wild type mRNA, thus the revertant cells become sensitive to 6-TG selection. A single copy of PB transposon removal allows the wild-type transcription to resume at one allele, however, in certain cases, heterozygous mutants may show a haploid insufficient phenotype, making the causality difficult to be confirmed. Therefore, complete geno-type reversion is ideal. The efficiency of the dual-transposon removal could be very low. Therefore, by coupling the PB transposon removal with a negative selection scheme, the revertant clones can be easily isolated by direct selection of cells without any PB transposons (containing *puroΔtk* expression cassette) residing in the genome.

The $3x10^6$ *Msh6* mutant ES cells were electroporated with 20 µg of mPBase expression plasmid (mPBaseΔNeo) and the cells were plated into a 90 mm plate and grown without any selection for three days to allow the transcripts for puroΔtk to decay prior to conducting the negative selection. To select for complete genotype revertants, $1x10^5$ cells were plated in a 90 mm plate and FIAU selection was initiated the following day. As a background control, pBlueScript plasmid was electroporated instead of PBase expression plasmid. Colonies from the mPBase-electroporated cells were picked for further genotype validation and phenotypic profiling.

In total, 148 colonies were formed in the cells electroporated with PBase whereas 30 colonies were present in the control plate. This slight background observed in the control was likely due to the presence of a small fraction of ES cells without the PB transposon mixed within the mutant clone or spontaneous mutations generated within the *tk* coding region of the puroΔk cassette. Triple-primer PCR was conducted using locus-specific genomic primers up- and downstream of the integration site, together with transposon-specific primer from Dom3z sequence, Figure 3-15a,b. FIAU-resistant revertant cells had lost both copies of the transposons (from both homologous chromosomes) as predicted. A panel of drug selections

was also conducted to confirm the phenotype of the revertants, Figure 3-15c. Genotype-reverted cells also showed a wild-type phenotype, i.e. sensitive to 6-TG selection. The loss of both copies of the PB transposon was also reflected by the loss of puromycin resistance. Both mutant and revertants were HAT resistant, suggesting that the *HPRT* minigene was not lost in all cases.

Figure 3-15: *Msh6* mutant rescue analysis.



a, Schematic representation of the wild type/revertant and the mutant allele. b, triple-primer competition PCR to genotype the revertants. c, Phenotypic profile of the mutant and the revertant cells. R1 and R2, are two independent revertant clones; M, mutant, wt, wild type.

3. Discussion

This chapter has described the establishment and experimental validations of a new mutagen and a strategy to deploy intra-genomic re-mobilisation of a *piggyBac* transposon to generate genome-wide heterozygous mutants. This strategy incorporates the aims of unbiased genome-wide coverage, efficient mutagenesis, easy identification of the mutation and reversion to establish genotype-phenotype causality.

In this strategy, I have generated a novel mutagenic PB transposon and inserted it into the ES cell genome by gene targeting, providing a stable single copy per cell transposon. Upon the supply of PB transposase, cells with PB excised from the donor locus and re-integrated elsewhere can be enriched. This enrichment for the re-integration events is dependent on a dual positive-selection strategy for the reactivation of *Hprt* transcription upon PB excision from the donor site and a positive selection marker within the PB transposon. Because of the high transposition efficiency of PB, in theory, intra-genomic mobilisation is sufficient in this design to provide enough heterozygous mutants to cover the whole ES cell-expressed genome in merely twenty 90 mm culture plates.

3.1. Molecular design of the gene-inactivating PB transposon

The molecular design within the PB transposon is aimed at maximal gene inactivation, with a non-selective trapping in both orientations mediated by terminal-exon pairs selected from the mouse genome. It should be mutagenic in either orientation in intronic and exonic positions in all genes, irrespective of their protein-coding potential, gene expression levels or reading frames. The use of endogenous exons as efficient gene traps has genetic basis, and it was explored and experimentally validated in this chapter. Most of the widely used gene-trap cassettes contain a splice acceptor, which has been directly cloned from naturally occurring viral or mammalian sequences. For example, *βgeo* relies on the adenoviral splice acceptor from the viral major late transcript to disrupt gene expression (Friedrich and Soriano, 1991), and the mammalian *Engrailed-2* splice acceptor is also frequently used (Collier et al., 2005). The popularity of these splice acceptors is historical and other endogenous exons within the mouse genome are likely to contain suitable splice acceptors for the purpose of efficient gene

trapping in many genomic contexts. Therefore, we computationally scanned the mouse genome to identify good mutagenic terminal-exon structures with criteria that allow the selection of "strong" mutagenic units other than the conventional splice acceptors.

The final two candidates were experimentally validated for their mutagenicity in two complementary assays, an *Hprt* trapping assay and a random trapping assay. The *Hprt* trapping assay is a stringent measurement for the strength of the trapping, as only cells with a near or complete null mutation of *Hprt* can give rise to 6-TG resistance. In these experiments, several colonies were isolated with a 6-TG resistant phenotype and the phenotype was revertible upon PBase re-introduction. In many cases, the exact insertion sites could not be mapped due to the technical challenges of mapping PB integration sites when there are many in each cell. However, several independent clones with PB transposons were mapped in independent integration sites within the *Hprt* were eventually detected, and the *Hprt* inactivation in all cases was mediated by the mutagenic units. Considered together, the data suggest that the mutagenic units efficiently cause gene inactivation.

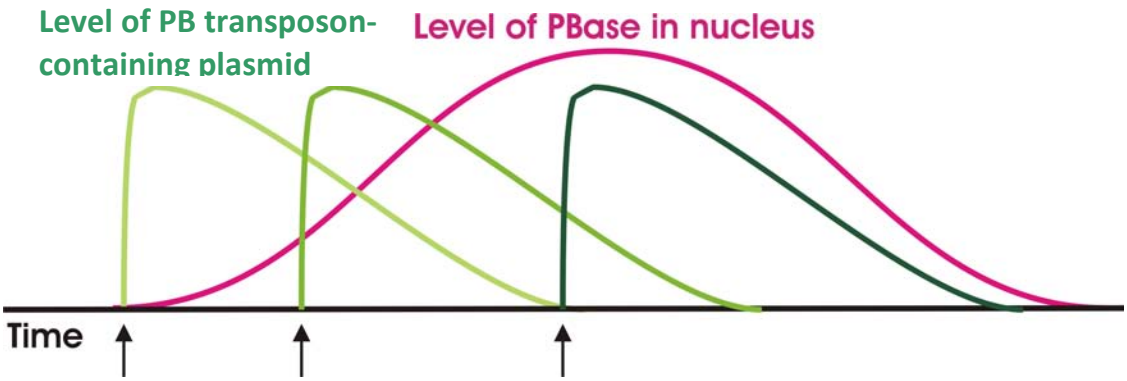
One limitation of this assay is that it does not test mutagen trapping capability in different genome contexts. Therefore, I conducted a random trapping assay to address this issue, using a cell line in which the mutagenic PB transposon was inserted into an HPRT minigene which was itself targeted into the *Gdf9* locus. PB was excised from the donor locus *Gdf9* following PBase expression, and reintegrated in a genome-wide fashion. A set of ES cell clones with PB re-integrated within genes expressed in ES cells were selected and assessed for the efficiency of the two mutagenic units within the PB transposon. Overall based on the data from two independent trapping assays, the mutagenic transposon are predicted to be able to mediate trapping 80 % of all intragenic integrations.

3.2. *piggyBac* possesses a fast transposition kinetics

A PBase inducible cell line, AB1-ROSA26$^{mPBaseERT2/+}$, has been established with the advantage of a tight temporal control over the PBase activity in the nucleus by 4-OHT addition. A pre- and post- PB transposon introduction time course was analysed with the aim to obtain high

efficiency of "plasmid-to-genome" transposition using the inducible PBase system. A short period of PBase induction (2-4 hours) prior to the electroporation of the PB transposon-containing plasmid dramatically elevated the transposition efficiency by five fold compared to without the pre-electroporation treatment. This suggests that PBase-catalysed transposition reaction occurs very fast when PBase and PB transposon first encounter and the longer the overlapping period between PBase and the PB transposon, the higher efficiency of obtaining transposon integrations, Figure 3-16. The post-electroporation incubation with 4-OHT was less obvert. However, sustained 4-OHT incubation post-electroporation consistently decreased the overall number of G418 resistant colonies. This may reflect the continuous intra-chromosomal transpositions occurring in each cell after the initial PB integration. The rate of transposon loss during intra-chromosomal transposition is approximately 60 %, measured from the single PB transposon remobilisation from a genomic donor locus (Liang et al., 2009). This means that with the sustained supply of PBase, the copy number of the integrated transposons are gradually lost overtime *via* repeated cycles of intra-chromosomal mobilisation.

Figure 3-16: A model to account for the PB transposition kinetics.



The green curves represent the level of PB transposon-containing plasmid in the nucleus with an initial maximal level at the point of electroporation and decay overtime due to cell division and DNA degradation. The pink curve represents the level of PBase present in the nucleus induced by 4-OHT, with the initial accumulation upon 4-OHT addition and decay overtime due to protein turnover after 4-OHT withdrawal. The arrows below the timeline represent the different time points when the PB transposon containing plasmids were introduced by electroporation.

3.3. Local hopping characteristics of *piggyBac*

The PB intra-genomic mobilisation was also investigated using a single copy PB mobilised from the *Gdf9* locus on chromosome 11. Clear local hopping was observed on the donor chromosome and predominantly surrounding the donor site. The possible local hopping effect of PB intra-chromosomal transposition has been noted previously for intra-genomic mobilisation from the *Rosa26* locus on Chromosome 6 (Wang et al., 2008b) but not observed on the *X*-linked *Hprt* locus (Wang *et al.*, 2008b; Liang *et al.*, 2009), with assay set-ups similar to the one described here, i.e. without any trapping-based integration selection. The results from these different loci are directly compared and summarised in Table 3-4.

Table 3-4: Local-hopping comparison among different genomic loci.

	Gdf9 locus		*Rosa26* locus		*Hprt* locus[a]		*Hprt* locus[b]	
	Insertions	% Total	Insertions	% Total	Insertions	% Total	Insertions	% Total
Total	84	100 %	264	84 %	93	100 %	79	100 %
PB Donor Chr.	32	38 %	47	32 %	8	9 %	5	6 %
PB Donor site*	14	17 %	25	14 %	3	3 %	0	0 %

*, 2.4 Mb region surrounding the donor site was analysed; for *Hprt* locus, 2.4 Mb window was set outside the 33.6 kb *Hprt* gene, as the integration site was not mentioned in the paper. Datasets for Rosa26 locus and *Hprt* locus were obtained from supplementary materials of Wang and co-workers (Wang et al., 2008b). a, data from (Wang et al., 2008b) b, data from (Liang et al., 2009), and the integration events in this data were combined with two versions of the PBase, i.e. the wild-type insect PBase and the mPBase.

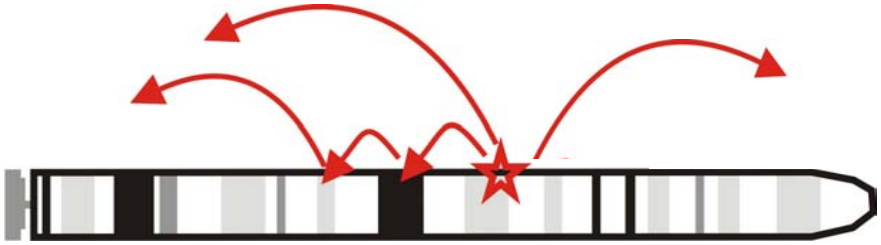
Out of the three PB transposon donor loci, the *X*-linked *Hprt* locus does not show a donor chromosome local hopping pattern, although there may be a slight tendency towards local re-integrations around the donor locus observed in one of the *Hprt* excision data sets (Wang et al., 2008b). Both the *Gdf9* locus and *Rosa26* locus show a local hopping effect, however at the *Gdf9* locus, this appears to be more extensive.

There are several possibilities contributing to the local hopping of PB transposon. Firstly, the experimental condition is slightly different with different versions and amounts of PBase introduced. The amount of nuclear PBase protein level can affect the pattern of transposition. The more PBase-containing plasmid transfected or the more stable PBase protein is (i.e.

mammalian codon-optimised PBase for stable expression *v.s.* wild-type PBase with insect-derived sequence), the longer the PBase protein can exist to catalyse the transposition events continuously. Secondly, PB mobilisation has been described to be sensitive to methylation. Methylation reduces the transposition efficiency of PB (Wang et al., 2008b). Therefore, if there is extensive methylation in an area around the donor locus, the PB transposon may "get stuck" within this region due to inefficient transposition in methylated genomic sequences. Additionally, if the transposon itself is methylated after targeted insertion into the genome, its re-mobilization efficiency can be significantly reduced. Thirdly, the chromatin structure may affect re-mobilisation pattern and efficiency. PB has a strong preference for integrating in actively transcribed genes (Liang et al., 2009), suggesting that PB has a preference of integrating into open chromatin regions. If the donor site is surrounded by "inaccessible" closed chromatin, PB transposition may also be retarded to mobilise out of the donor region. Chromosome 11 has a very high gene density and the region surrounding the *Gdf9* locus is also rich in actively transcribed genes in ES cells. The donor site/chromosome integration preference may be partially due to the open chromatin bias. A final possibility is that *Blm*-deficiency may influence PB transposition. The ES cell line used for the PB transposon re-mobilisation from the *Gdf9* locus was derived from a *Blm*-deficient ES cell line (Guo, 2004), whereas other remobilisation work was based on a Blm-proficient ES cell background. The influence of Blm deficiency and PB transposition is not yet known. However, experimental evidence shown in Chapter 4 seems to suggest that Blm-deficiency negatively affect the PB-mediated transposition, as identical PB transposon targeted into the same position within the *Hprt* locus in Blm-proficient or deficient ES cells showed differential re-mobilisation efficiency (Chapter 4, Figure 4-4). It is possible that *Blm*-deficient ES cells may not be able to repair the DNA double strand breaks (DSBs) generated upon PB excision. Continuous cycles of PB excisions and re-integrations within the same cell cause cell death if these genomic damages can not be fixed efficiently and precisely. Genome-wide remobilisation may be achieved through continuous transpositions, thus cells with less cycles of continuous transposition may be selected for in this scenario. The precise mechanism is yet to be discovered.

For the purpose of genome-wide mutagenesis, the use of the *Hprt* locus is a better choice to maximise genome-wide coverage. However, the observation of local hopping is an interesting phenomenon to provide us with an insight into the kinetics of intra-genomic transposition of *piggyBac*, Figure 3-17. Using a temporal controllable PBase expression (PBaseERT2), this kinetic aspect of the *piggyBac* transposition can be dissected. Such investigations not only can provide insight into the fundamental characteristics of PB transposition, but may also highlight potential risks in using PB in genetic studies and clinical medicine for its possible "non-tagged" mutagenesis of the genome.

Figure 3-17: Possible intra-genomic mobilisation kinetics of the PB transposon.



Genome-wide reintegration of the PB transposon can be achieved in two ways. One possibility is that PB can directly re-integrate into a locus randomly after being excised from the donor site. Another possible mechanism is through several rounds of local hopping to "escape" the donor site. This local-hopping escape model can be extended to the reintegration into a proximal chromosome, which is spatially local to the donor site.

3.4. A recessive genetic screen using the established mutagenesis strategy

A DNA mismatch repair screen was conducted using a library generated with the established mutagenesis strategy and a known gene was identified from this screen. The inability to uncover other MMR genes is mainly due to the bias re-integrations of the PB transposon surrounding the PB transposon donor site. The local hopping of PB transposon surrounding the 2.4-Mb region of the donor site was measured to be 17 % of the total number of re-integrations. This means that in this mutant library with 10,000 mutants, only 8,300 clones contain integrations away from the donor site. Thus, the number of heterozygous mutations was limited in this library. However, a known MMR gene, *Msh6*, was isolated from the screen

suggesting that the established mutagenesis strategy was sufficient to couple with the *Blm*-deficient background for recessive genetic screens. It has been previously observed that intra-chromosomal mobilisation from the *Hprt* locus did not give rise to local hopping, Table 3-4. Therefore, using the *Hprt* locus as the donor site will be better choice to mediate genome-wide mutagenesis.

In this proof-of-concept MMR screen, other HAT and 6-TG double-resistant clones were identified, which were not obvious candidates for MMR pathway and most of these clones contain PB integration sites mapped to regions locally surrounding the *Gdf9* locus. This suggests that these clones may possess background mutations in the MMR pathway genes, and the PB integrations are not the casual mutations. These PB integrations are unlikely to be causal to the 6-TG resistant phenotype and background mutations may have occurred in these clones. *Blm* deficiency promotes conversion of heterozygous to homozygosity, thus spontaneous mutations generation during ES-cell culturing may be converted to homozygosity with 20-fold enhancement in *Blm*-deficient background compared to wild-type cells. In addition, the MMR screen using 6-TG selection may enhance the random background mutation rate, as 6-TG is genotoxic pro-drug and its metabolites can be incorporated into the genomic DNA of cells to generate point mutations. Therefore, it is important to conduct genetic rescue experiment to validate the causal link between the PB integration sites and the phenotype. If the integration is irrelevant to the phenotype, excision of the PB transposon will not revert the mutant phenotype to wild type.

3.5. Complete genotype reversion using PBase with FIAU selection

For the purpose of genetic screens, it is important to be able to establish a causal relationship between the gene mutated and the phenotype. The best way to establish this connection is through genetic rescue experiments. Upon mutagen removal or complementation of a wild-type copy of the mutated gene, the mutant phenotype should also revert to wild type.

Mutagenesis mediated with DNA transposons has the significant advantage of their simple removal. PB transposition has the unique property of excision without footprint. Thus,

excision of the mutagenic PB transposon from a gene can completely revert the mutant phenotype to wild type. Even insertions in exons can be fully reverted. Because the reversion efficiency per transposon is approximately 1 %, identifying complete reversion events of homozygous mutants is not efficient enough without any selection strategy, especially in "difficult-to-excise" genomic contexts. In my transposon design, a *puroΔtk* cassette was introduced to facilitate the selection for revertants using FIAU. This strategy was demonstrated in this chapter using a homozygous *Msh6* mutant obtained from a DNA MMR genetic screen. Many genotype-reverted colonies could be readily obtained and their drug-selection profile proved that these were phenotypic reversions.

The selective excision of PB transposon is also widely useful for other purposes requiring PB transposon removal without reintegration elsewhere in the genome. For example, integration-free induced pluripotent stem (iPS) cells can be generated with PB transposon carrying the Yamanaka factors to reprogram somatic cell types. The transgenes can be subsequently removed from the genome by PBase supply and FIAU selection (Yusa et al., 2009).