Chapter Seven – Mobilisation of giant *piggyBac* transposons in the mouse genome

1. Introduction

The development of new technologies that allows the stable delivery of large genomic DNA fragments in mammalian systems is important for genetic research as well as applications in gene therapy. Genomic sequences contain not only protein-coding regions, but also important regulatory elements, which are critical to ensure the appropriate level as well as regulated spatial-temporal gene expression in an organism. Although heterologous-promoter driven cDNA sequences can be readily introduced as transgenic elements, these rarely provide the full repertoire of alternative isoforms, physiological-relevant expression patterns and are prone to silencing. Therefore, the delivery of large contiguous genomic sequences is essential to achieve regulated gene expression.

Episomal vectors based on Epstein-Barr virus (Wade-Martins et al., 2000) and Herpes Simplex Virus type 1 (Hibbitt and Wade-Martins, 2006), have been used to introduce large genomic sequences into mammalian cells. As episomes can be lost without selection pressure, they do not guarantee indefinite expression of the delivered cargo. Long-term expression of a transgene is most reliably achieved by stable integration. Retroviral and lentiviral vectors have been used for this purpose, but their cargo capacity is limited to 10 kb and they are not suited for the delivery of intron-containing cargos. Additionally, these viral systems have immunogenic and tumorigenic potential.

Transfection of naked DNA has been used for large-cargo delivery. Pronuclear injection of bacterial artificial chromosomes (BACs) has been successful for transgenesis of up to 300 kb. However, the integrity, integration site and copy number can not be controlled. BAC vectors have also been used for targeting large cargos to defined genomic positions in ES cells via homologous recombination (Valenzuela et al., 2003), but the efficiency is locus-dependent and can be very low. Recombinases such as Cre have also been used to deliver BACs to a pre-

defined genomic location by recombination-mediated cassette exchange (Wallace *et al.*, 2007; Prosser *et al.*, 2008); however, pre-engineering of target sites in the genome is necessary. While these methods are useful for certain applications, all have limitations and most of them are not able to revert the insertion of large DNA fragments.

DNA transposons have emerged as flexible and efficient molecular vehicles to mediate stable cargo transfer. However, the ability to carry DNA fragments greater than 10 kb is limited in most DNA transposons. The development of a DNA transposon system with the capacity of accommodating large genomic fragment will be an important technology that complements the current methods to facilitate a wide range of genetic and genomic applications. PB has previously been shown to be capable of delivering DNA fragment of up to 10 kb in mice without the significant loss of its transposition efficiency (Ding *et al.*, 2005b). In this chapter, the cargo capacity of PB transposon has been investigated with the aim to develop PB into a giant genomic cargo carrier.

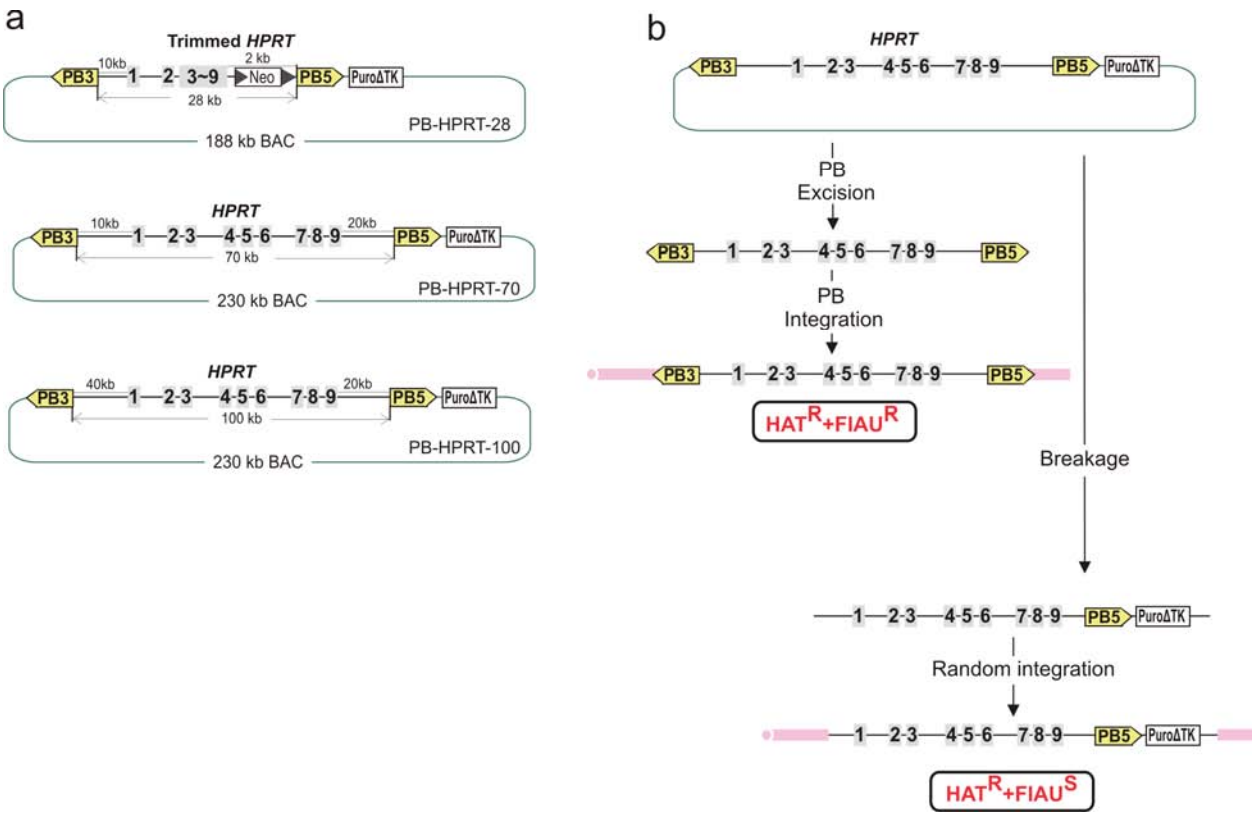
2. Results

2.1. The generation of giant PB transposons

In order to assess the cargo capacity of PB, we have constructed a series of PB transposons with sizes of 28, 70 and 100 kb on a BAC backbone for testing their transposition capability in mouse ES cells (Figure 7-1a). A BAC containing the human *HPRT* gene was used to construct these transposons with the human *HPRT* gene as a positive selection marker. When introduced into the *Hprt*-deficient ES cells, these transposons can complement the *Hprt* deficiency, so that clones in which transposition has occurred could be directly selected in HAT. This BAC was modified by insertion of both PB ITRs by recombineering technology to generate PB transposons with 70 and 100 kb cargos (PB-HPRT-70 and PB-HPRT-100). The 28 kb PB (PB-HPRT-28) was constructed by substituting the genomic regions from exon 3 to 9 with the corresponding part of the *HPRT* cDNA using the PB-HPRT-70 vector and recombineering technology, (Figure 7-1a). The BACs were further modified by insertion of a *PuroΔtk* cassette (Chen and Bradley, 2000) immediately downstream of the PB5'ITR, so that random insertions could be counter-selected. ES cell clones in which the *HPRT* gene has been

inserted by transposition should exclude the *PuroΔtk* cassette and will be resistant to FIAU (Figure 7-1b).

Figure 7-1: Giant PB construction and selection scheme incorporated to detect their transposition in mouse ES cells.

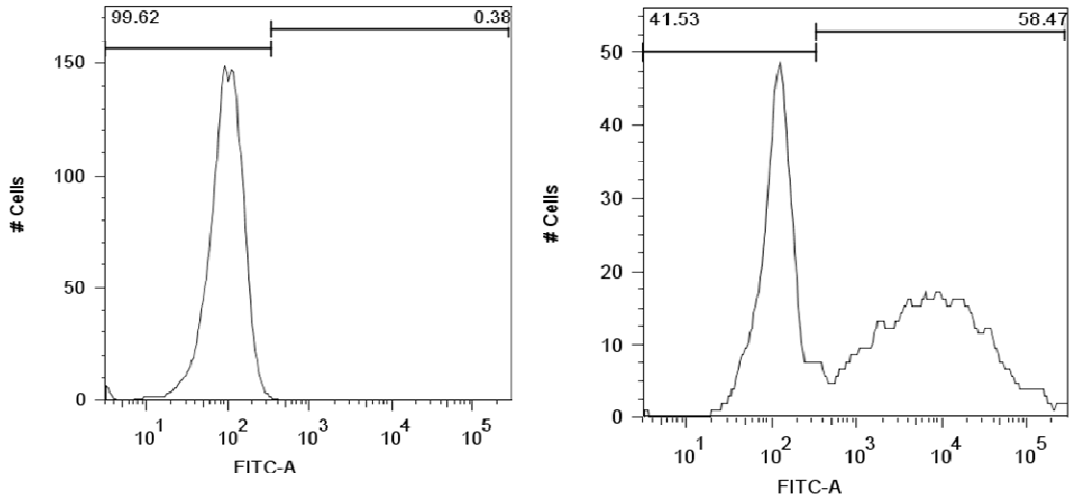


a, Giant PB-HPRT constructs modified from the same BAC. **b**, A scheme of alternative genomic integration pathways. PBase can mediate precise excision of the giant PB from the BAC and insert this into the ES cell genome, generating cells that are resistant both to HAT and FIAU. If physical breakage of the BAC occurs the PB ITRs are separated, random integration of the BAC can occur including the *PuroΔtk* cassette. If the *HPRT* gene on the BAC is intact, the cells will be HAT resistant. If the *puroΔtk* cassette is intact, the cells will be sensitive to FIAU.

2.2. Transposition detection of giant PB transposons in ES cells

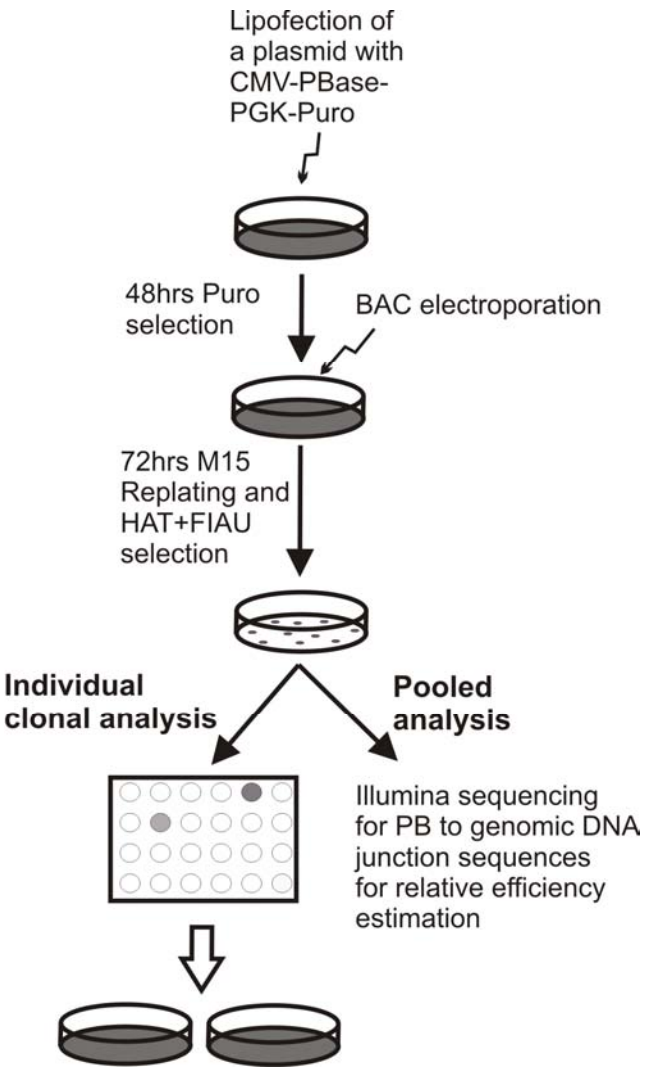
The *Hprt*-deficient AB2.2 mouse ES cell line was transiently transfected using the lipofection method with one of the two types of the *piggyBac* transposase (PBase); the mammalian codon optimized version, mPBase (Cadinanos and Bradley, 2007) or a hyperactive form HyPBase (Yusa, *et al* Submitted). These PBase-expressing plasmids also contain a puromycin selection cassette so that ES cells expressing PBase could be enriched by a pulse puromycin selection (Taniguchi et al., 1998). As a negative control, a plasmid co-expressing enhanced green fluorescent protein (eGFP) and the puromycin resistant cassette was used. With this enrichment method, greater than 50 % of the ES cells were expected to be expressing PBase given that they were eGFP positive (Figure 7-2). Three days after PBase transfection, the BACs harbouring different-sized PB transposons were introduced by electroporation and the cells were replated in HAT and FIAU containing medium to select for ES cells with a stable integration of PB. The HAT and FIAU resistant colonies were either picked and analysed individually or pooled together for high throughput PB-transposition identification and mapping using the Illumina sequencing platform. Figure 7-3 shows the entire experimental scheme.

Figure 7-2: Transfection efficiency of the AB2.2 ES cells determined by the eGFP expression after a pulse of puromycin selection.



Negative control eGFP transfected

Figure 7-3: Outline of the experimental scheme.



All three PB-HPRT transposons gave rise to HAT and FIAU double-resistant colonies and the colony numbers exceeded those in the non-PBase control, Table 7-1. Unexpectedly, the number of double-resistant colonies did not vary greatly with the size of the PB transposon. However, the number of double-resistant colonies increased significantly when HyPBase was supplied compared to mPBase suggesting that transpositions occurred. HAT and FIAU resistant colonies can be generated by two competing mechanisms: transposition or random integration of the BAC with the loss of the *PuroΔtk* cassette. The proportion of random

integration events was higher with the 70 and 100 kb PB transposons judged by the number of HAT and FIAU double-resistant colonies in the non-PBase control. The larger transposons are expected to have a higher background of HAT and FIAU double- resistant colonies because the *PuroΔtk* cassette is located 20 kb from the 3' end of the *HPRT* gene whereas the *PuroΔtk* cassette in the 28 kb PB is only separated from the *HPRT* stop codon by 2 kb (Figure7-1a).

Genuine transposition can be distinguished from the random integration by analyzing sequences adjacent to the PB inverted repeats. If PBase-mediated integration occurred, both ends of the PB ITRs should be flanked by mouse genomic sequences together with PB's signature recognition site TTAA (Ding *et al.*, 2005a; Wang *et al.*, 2008b; Liang *et al.*, 2009). If random integration occurred, the original BAC vector sequences adjacent to PB ITRs will be present. Illumina sequencing technology was used to identify a large number of genuine PB transposition events from the random integrations. HAT and FIAU resistant colonies were pooled from each experimental condition; genomic DNA was extracted and subjected to paralleled paired-end sequencing to identify the PB5' ITR – genomic junctions, Figure 7-4.

Transposition events were identified for all three transposons when either mPBase or HyBase was used, Table 7-1. The absolute number of transposition events dropped significantly as the cargo size increased from 28 kb to 70 kb, however, the 70 kb and 100 kb PB transposons showed a similar number of integration events. The proportion of transposition events among the HAT and FIAU double-resistant colonies was lower with the 70 and 100 kb PB transposons than the 28 kb PB transposon, reflecting the higher rate of random integrations of large transposons as seen in the HAT and FIAU resistant colony number in the non-PBase control. HyPBase-mediated transposition was approximately four times that of mPBase for the larger transposons and seven times for the 28 kb transposon. Albeit at lower efficiency, wild type PBase can mediate transposition with large cargos, suggesting that the large-cargo capacity is an intrinsic property to the PB system, not acquired as a result of modifications to PBase.

Table 7-1: Transposition efficiency of different sized PB transposons and versions of the PBase.

Transposons	mPBase		HyPBase		eGFP
	HAT+ FIAU	Transpositon Events (%)*	HAT+ FIAU	Transpositon Events (%)*	HAT+ FIAU
PB-HPRT-28	39	18 (46 %)	183	131 (72 %)	9
PB-HPRT-70	56	9 (16 %)	104	26 (25 %)	37
PB-HPRT-100	77	5 (7 %)	103	30 (29 %)	47

The number of transposition events was determined using massive-parallel sequencing from pooled HAT and FIAU resistant clones. *: Percentage of transposition events as a fraction of HAT and FIAU double-resistant colonies. The transposition events are assumed to be one per cell.

Figure 7-4: Schematic representation of the experimental platform (a) and bioinformatics (b) analysis to identify transposition events using the Illumina sequencing.

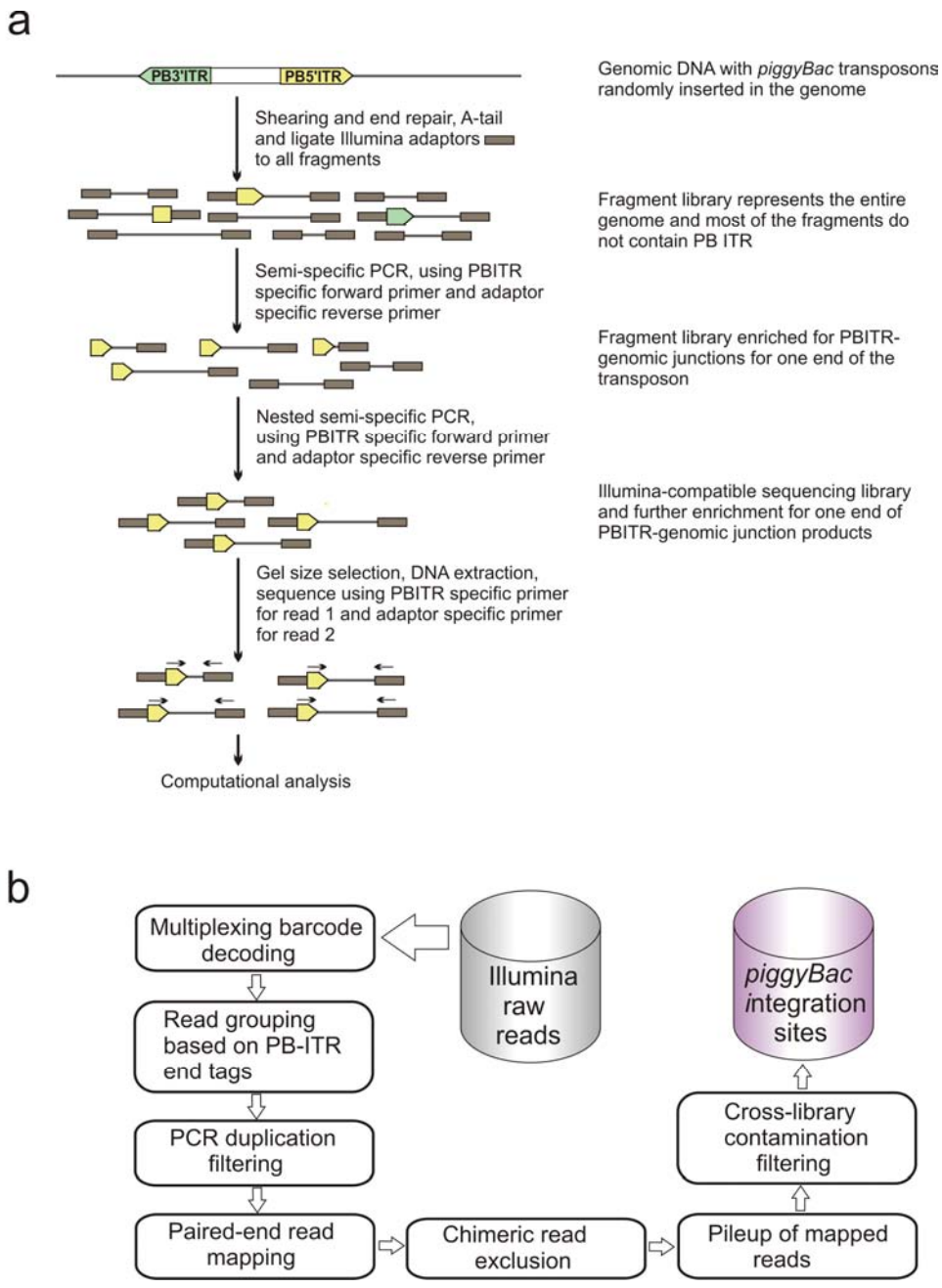


Figure **a** was adapted from Daniel J. Turner's original figure and **b** was adapted from Zemin Ning's original figure.

2.3. Individual clone analysis of giant PB genomic integrations

In a separate experiment, double-resistant colonies were generated and analyzed individually to identify integration sites using Splinkerette PCR, Figure 7-5a. For each transposon insertion analyzed, both the PB5' and PB3' ITR – host genome junction sequences were contiguous in the mouse genome (Figure 7-5b). Analysis of the transposon copy number in these clones by Southern blotting also revealed that almost all the PB-mediated integrations were single-copy (Figure 7-5c).

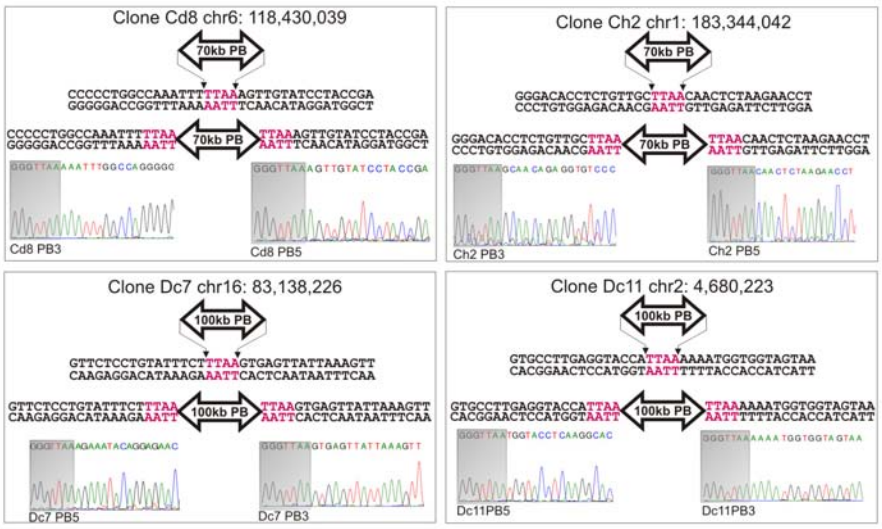
One of the major advantages of using transposition to deliver large genomic DNA fragments is that cargo integrity is expected to be maintained. To examine this, we used a custom high-resolution (average probe spacing of 130 bp) comparative genomic hybridization (CGH) array, covering the entire human *HPRT*-containing BAC. The samples and control were mixed with equal amount of human male DNA to provide a baseline for array normalization, Figure 7-6a. Four independent clones with 70 kb (Cd8 and Ch2) and 100 kb (Dc7 and Dc11) PB insertions were assessed. The regions of copy number gain in all the clones precisely matched the regions flanked by the PB ITRs (Figure 7-6b,c). Within these regions, the human DNA sequences were continuous and did not contain any detectable change (Figure 7-6b). Thus, the large cargos mobilized as PB transposons remained intact in all cases.

Figure 7-5: Giant PB transposition with single copy integration per cell.

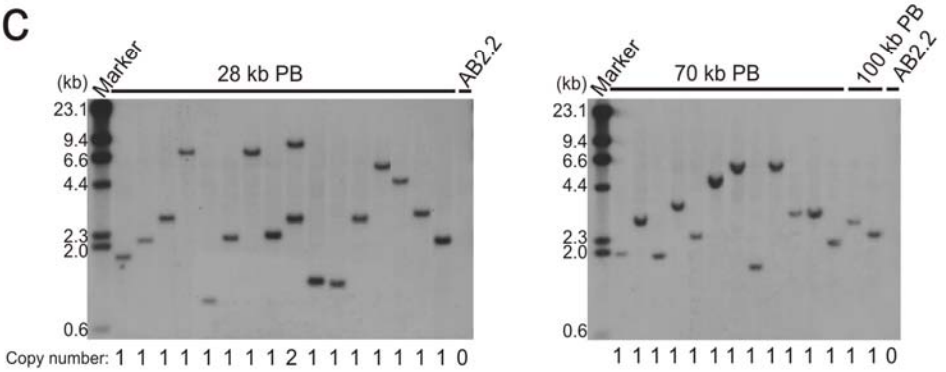
a

Transposons	mPBase HAT+FIAU^R analyzed	mPBase Transposition event	HyPBase HAT+FIAU^R analyzed	HyPBase Transposition event
PB-HPRT-28	32	4	55	23
PB-HPRT-70	43	0	69	13
PB-HPRT-100	70	0	64	2

b

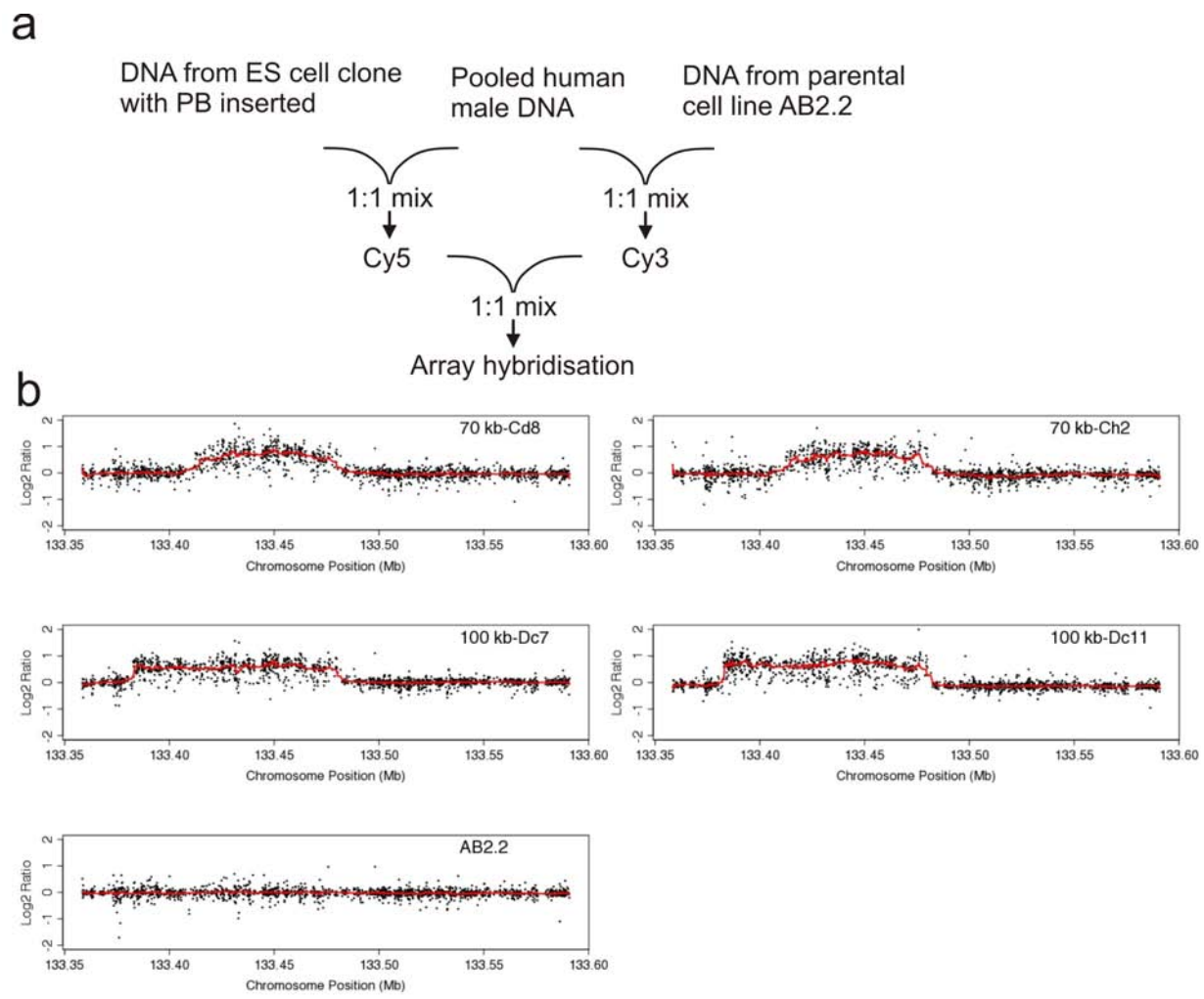


c



a, Transposition events of different sized PB transposons and versions of the PBase from single-colony analysis. **b**, Precise integration of giant PB transposons at the expected TTAA site. The chromosomal coordinates of the first T corresponds to the PB recognition site TTAA are shown (NCBI m37). **c,** Southern blot illustrating the copy number of the PB mediated largo-cargo integrations using the PB5'ITR as the detection probe. The genomic DNA was digested with *Spe*I and *Xba*I.

Figure 7-6: Regional high density CGH array analysis to determine the PB cargo integrity.



c

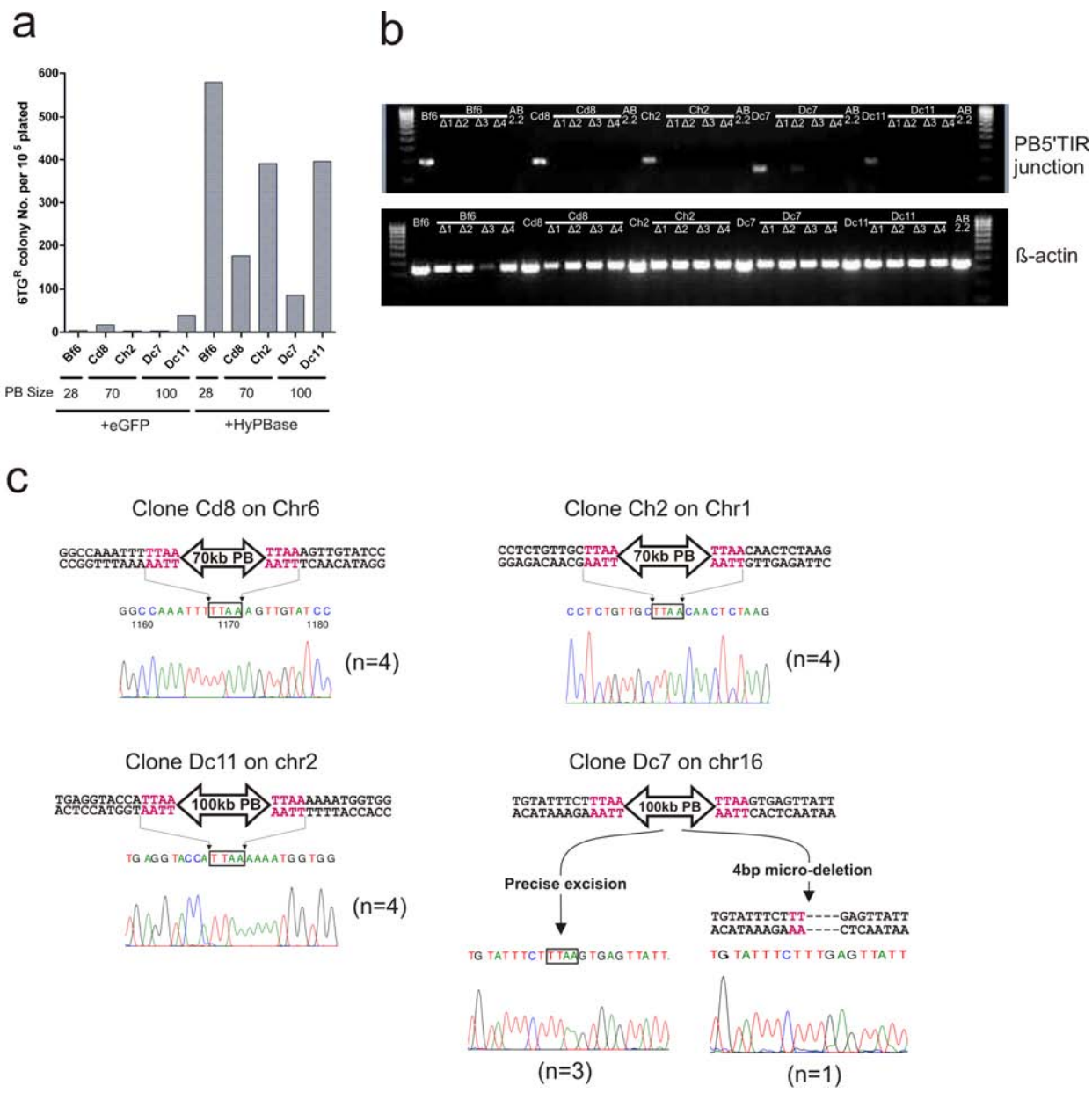
Coordinates on X (NCBI36)	70 kb PB		100 kb PB	
	Actual	Array	Actual	Array
Start	133,412,009	133,412,083	133,382,256	133,382,623
End	133,481,773	133,481,626	133,481,773	133,481,626

a, The pooled human male DNA was used as a reference to detect the gain of copy number for the *X*-linked genomic DNA in the ES cell clones with PB integrated. b, Regional CGH analysis showing the gain of an extra copy of the human *HPRT* gene delivered by PB transposition. The red line was calculated as the running median of the Log_2 value of each CGH probe to aid the visualization. **c**, Comparison of CGH array read-outs for the positions of copy number gain and the actual positions.

2.4. Analysis of chromosomal excision capability of giant PB

ES cell lines with 28 kb (Bf6), 70 kb (Cd8 and Ch2) or 100 kb (Dc7 and Dc11) PB transposons were transiently transfected with HyPBase and enriched for expression with a pulse puromycin selection. Following a period of culture to allow for the decay of *HPRT* mRNA and protein products, the cells were plated at low density and selected for the loss of *HPRT* in 6-TG. PB excision was observed for all clones tested with efficiencies ranging from 0.1 % to 0.6 % of the total number of cells plated (Figure 7-7a). 6-TG resistant colonies derived from each of the four parental PB-containing clones were examined by genomic PCR using transposon-host specific primer sets for each integration site, and none exhibited the PB ITR-host genomic junctions (Figure 7-7b). PB does not normally leave a footprint upon excision. We therefore sequenced the transposon excision sites in all of the 6-TG resistant clones to check for their intactness. Precise excision was observed in all clones derived from the three donor sites (Cd8, Ch2, and Dc11). However, one out of four 6-TG resistant clones derived from the Dc7 clone showed a micro-deletion (Figure 7-7c). Micro-deletions upon PB excision have been reported previously (Wang et al., 2008b), suggesting that this low frequency of imprecise excision is not due to the size of the cargo.

Figure 7-7: PBase mediated excision of giant PB transposons from the ES cell genome.



a, Genomic excision efficiency of five ES cell clones containing giant PB transposons with different cargo sizes following transfections with either HyPBase or eGFP (control) **b**, Molecular analysis of individual 6-TG resistant colonies to evaluate fidelity of excision events. Excision of PB eliminates the PB-host junction fragment amplified in the parental lines. **c,** Analysis of PB excision identified one clone with a micro-deletion (clone Dc7) following the excision of a 100 kb PB transposon. "n" represents the number of 6-TG resistant colonies with the shown sequence traces of the excision site.

3. Discussion

In this chapter, the mobilisation of giant *piggyBac* transposons of up to 100 kb from exogenous BAC vectors and endogenous genomic loci in mouse ES cells was demonstrated. Mobilisation of giant PB transposons achieves stable but precisely revertable genomic insertions. Importantly, large DNA cargos remain intact during transposition and the copy number of the delivery is predominantly one.

In the vector-to-chromosome transposition assay, the efficiency of transposition dropped as the cargo size increased, Table 7-1. This could be caused by the lack of integrity of the BAC vector during preparation and electroporation. Breakage between PB ITRs prevents transposition and stimulates random integration. The chance of a break occurring between the PB ITRs increases with the distance between them. In the chromosomal transposition assay, the frequency of excision appeared to be less dependent on the size of the transposon than the integration site. This supports the view that one of the major factors influencing vector-to-chromosome transposition is the continuity of the BAC DNA between the PB ITRs, rather than inherent limits in the transposition reaction per se.

In the system described here, transposition events can be enriched by negative selection using FIAU, because transposition uncouples the PB transposon from a negatively selectable *puroΔtk* cassette on the BAC backbone. It follows that the tighter the linkage between the positive selection marker in the transposon and the negative selection cassette, the greater the degree of enrichment for transposition events. The tight linkage in the 28 kb transposon allowed us to achieve 70 % transposition efficiency after positive-negative selection, Table 7-1a. This selection scheme is very useful to enrich for transposition by selecting against random integrations of the BACs.

We have demonstrated that giant PB transposons effectively deliver intact large genomic DNA fragments with a controllable copy number. This is useful in many genetic applications such as BAC transgenesis and genetic complementation. Another DNA transposon, *Tol2*, has

also recently been shown to deliver a 70 kb genomic DNA for transgenesis (Suster et al., 2009). The additional ability of *piggyBac* to cleanly excise large genomic DNA fragments provides a valuable genome engineering technology for creating *in vitro* and *in vivo* gains and losses of large genomic regions.

PB-mediated integration of large genomic fragments can provide "permanent complementation" with prolonged and physiologically-regulated gene expression. It also avoids the complications of viral vectors, which can induce host-immune responses and tumorigenesis. Viral vectors contain the polII promoter activity within the 5' viral long terminal repeat (LTR), therefore insertions of the viral vectors may ecpotically drive the expression of surrounding genes or part of the gene if insertions is present within a gene. If the surrounding gene is a proto-oncogene, such as *C-myc* or *Ras*, over-expression of such genes can initiate tumorigenesis. On the contrary, PB ITRs do not contain promoter signals, thus they have a low risk of inducing tumour formation. Although PB integration is random, specific integration sites of PB can be screened to identify permissive locations that are not likely to affect normal function for clinical applications. The development of giant PB transposons will be valuable for therapeutic gene delivery of large genomic sequences in patient-specific iPS cell lines to combat a range of human genetic diseases.

Giant PB transposons are comparatively simple to construct. In principle, a genome-wide resource of PB-BACs could be generated using recombineering technology (Chang et al., 2007). Such a resource can be used in genetic screens and in complementation studies. Transient expression of PBase to mediate giant PB transposition does not require prior genome modification, thus giant PB libraries can be used in most cell types and organisms.

Taken together, the work presented here provides a framework for using *piggyBac* to mobilise large genomic DNA fragments. This will open the door to a wide range of future applications in genetics and genomic research as well as clinical medicine, which have been difficult to conduct previously with other tools.