Chapter 6

Isolating large numbers of homozygous mutants in parallel

6.1 Introduction

6.1.1 Aims

In the previous chapter I described the isolation of homozygous mutants by selection for copy number gain in *Blm* deficient cells. Analysis of the selected population revealed that not all the surviving cells are homozygotes, although all have two copies of the selection construct. Thus, using the scheme described, two subcloning steps are necessary to obtain pure clonal homozygous mutants for screening: one initial step to isolate a colony for expansion, and one after double selection to obtain homozygous mutants. This is not practical on a genome wide scale. In this chapter I describe several technical advances to solve this problem by allowing the initial subcloning step to be avoided. Mutants are expanded in a pool, in parallel, and Cre treatment and double selection carried out on this pool. Mutant clones are picked directly from the double selection, resulting in a complex pool with low background. I discuss uses of the libraries generated in this way.

6.1.2 Clonal expansion

When working with a complex pool of cells that will be expanded and selected over a period of time, the problem of clonal expansion arises. This is where cells that will survive the late selection step are present early in the culture, and thus have time to expand to a clone that dominates the selected population (Figure 6.1). The selection steps need to be absolutely stringent, and furthermore the full range of events occurring in the culture needs to be known. The reason for this is best illustrated by considering some numbers. In the homozygote isolation process, there are two selection steps: One to select for a PB integration (G418 for 10 days), and one double G418+Puro step (A further 10 days). A cell that had two integrations will be potentially double resistant from the start, and thus at the time of double selection this clone will have $2^{10} = 1024$ cells. At the opposite end of the scale, a clone undergoing LOH only in the generation immediately prior to selection will have just a single potentially double resistant cell. Thus, if these two clones had been pooled at the start of the experiment, the (useless) cells in the double resistant population with two insertions will vastly outnumber the real homozygote.

6.2 Results

6.2.1 C57BL/6 targeting vector to insert the transposon at the *Hprt* locus.

Choice of cell line

In order to limit the initial copy number of the transposon immediately after mutagenesis, I decided to target it to the X chromosome. Most ES cell lines used for making genetically modified mice are XY, and thus should maintain all loci on the X chromosome outside the pseudoautosomal region as single copy. The obvious choice of locus is *Hprt*, as targeting vectors are available, and the transposon could be inserted such that it disrupts *Hprt* function. Transposition would then be coupled to restoration of *Hprt* function, allowing it to be selected for with HAT.

However, the original *Blm*-deficient cells were from the AB2.2 cell line. This contains a complex mutation at *Hprt* that is not revertible. Therefore the endogenous *Hprt* gene cannot be used for HAT selection in these cells; the usual procedure is to use a human *HPRT* minigene. Because of this mutation, AB2.2-derived cells are not ideally suited for HAT selection in my proposed context. I had originally cloned my transposon into the intron of the *HPRT* minigene, and was planning to insert this at the endogenous (non-functional) *Hprt* locus (see Chapter 3).

Another problem with the original *Blm*-deficient ES cells is that they are a compound heterozygote with respect to *Blm* (Luo *et al.*, 2000). Furthermore, there is evidence that one of these alleles (*m3*) is a hypomorph (McDaniel *et al.*, 2003). This is not necessarily a problem, as the cells do show a Bloom syndrome phenotype, but it is an aspect that could

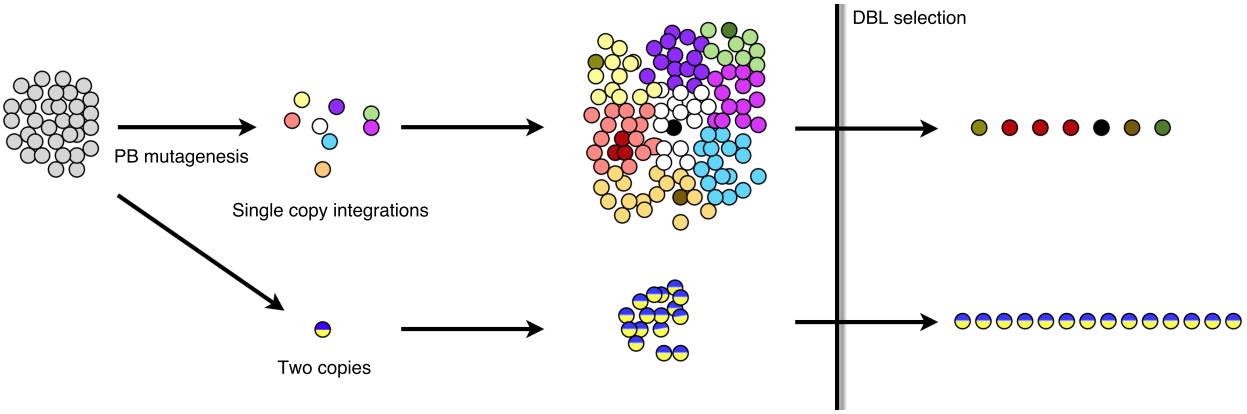

Figure 6.1: Problems caused by clonal expansion in heterogeneous pools of cells. The effect of having two transposon copies present in a minority proportion of cells prior to expansion is illustrated. Different colours represent clones of cells with different transposon insertion sites. All cells in the clone with two copies (bottom) can potentially survive double selection, indicated by a vertical line, while in the single copy clones only the rare homozygotes (represented by dark colours) survive. This leads to the clone with two copies dominating the double resistant population.

be improved on. One potential consequence could be that subclones arise in the culture that have undergone LOH at *Blm*, and therefore display slightly different phenotypes with respect to further LOH or genome instability. Recently, a new *Blm* mutant cell line, $Blm^{e/e}$ was generated in our laboratory by Amy Meng Li (Li, 2010). This cell line has two advantages: (a) it is derived from the JM8.F6 ES cell line (Pettitt *et al.*, 2009), and therefore has a functional *Hprt* gene and (b) the *Blm* locus is homozygous for a genuine null allele. It should be noted that the cells also express GFP and *bsd* constitutively from transgenes at the *Blm* locus; however this is not a problem for my method.

As this cell line is in the C57BL/6 genetic background, I constructed an isogenic targeting vector to insert my transposon (TNN) in such a way as to disrupt *Hprt*.

Retrieval of a *Hprt* fragment from a C57BL/6 BAC

I obtained the tiling path BAC RP23-252K15, which contains the complete *Hprt* coding sequence, and used it to transform EL350 bacteria (see Methods). Using long oligonucleotide primers, I amplified a pBS backbone with 70 bp homology arms. This linear PCR product can be thought of as a circular plasmid, containing two colinear fragments homologous to *Hprt* separated by several kbp with a break between them (Figure 6.2A). When electroporated into recombination competent bacteria containing the *Hprt* BAC, the 'gap' is repaired using the BAC as a homologous template. This resulted in retrieval of the fragment defined by the homology arms into the plasmid backbone (Figure 6.2B,C).

There is a naturally-occurring *Hind*III restriction site in intron two of the mouse *Hprt* gene (NCBI m37 X:50,357,636). I used an adaptor oligonucleotide sequence to introduce a *Nsi*I site into this locus in order to clone TNN as a *Nsi*I fragment from the P2-HPRT-Tn plasmid (Chapter 3). This fragment contains 125 bp of sequence from outside the vector, derived from the human *HPRT* minigene intron. As *Nsi*I cleaves 41 bp inside the PB3 end of the transposon, I had this fragment synthesised with the adaptor oligonucleotide, so the correct PB3 sequence was reconstituted upon ligation (Figure 6.3A). I verified the correct orientation and PB3 reconstitution by restriction digest and sequencing (Figure 6.3B). This targeting vector is named TV28. I also verified the function of the transposon by a transposition assay in ES cells (Figure 6.3C).

Targeting ES cells

I electroporated 10^7 $Blm^{e/e}$ ES cells with 15 μg *Srf*I-linearised TV28 (see Methods). G418-resistant clones were picked after ten days and screened for correct targeting events by PCR directly from colony lysates (Figure 6.4). Ten out of 36 tested were correct, nine of which I tested for targeting at the other junction by Southern blot (Figure 6.5A). All of these confirmed correct targeting, giving a targeting fre-

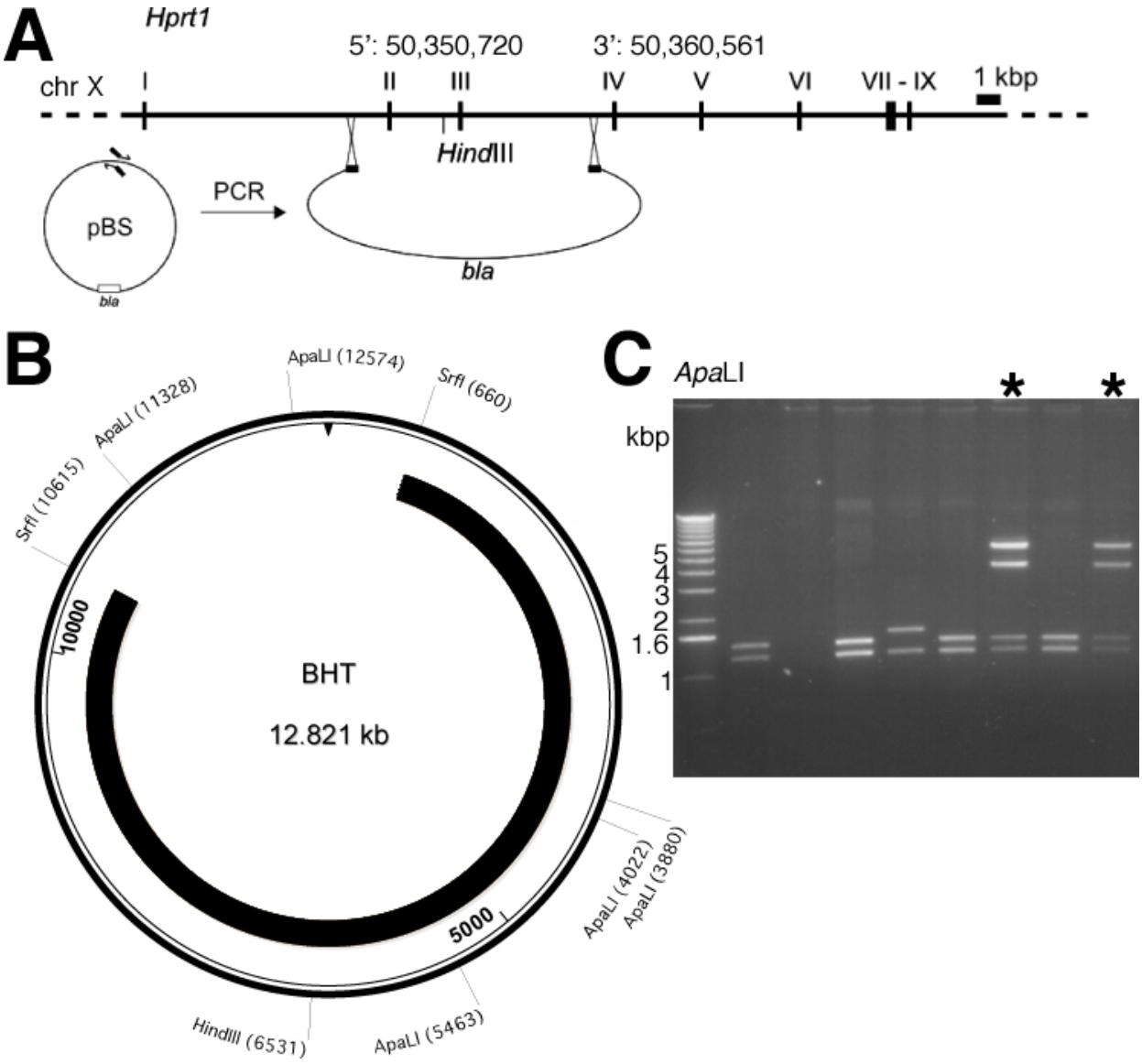

Figure 6.2: Retrieval of a fragment of the C57BL/6 *Hprt* locus. A—Design of the capture vector, produced by amplification of a pBS plasmid backbone with primers tailed with appropriately oriented homologous sequence. B—Map of the plasmid with retrieved fragment. *Srf*I sites were present on the primers used and can be used to linearise the construct for targeting. C—*Apa*LI digest showing correct structure of retrieved fragment. *—correct clone; other clones have only the *Apa*LI bands from the plasmid backbone and may arise from contaminating circular plasmid or recombination-induced recircularisation of the capture vector.

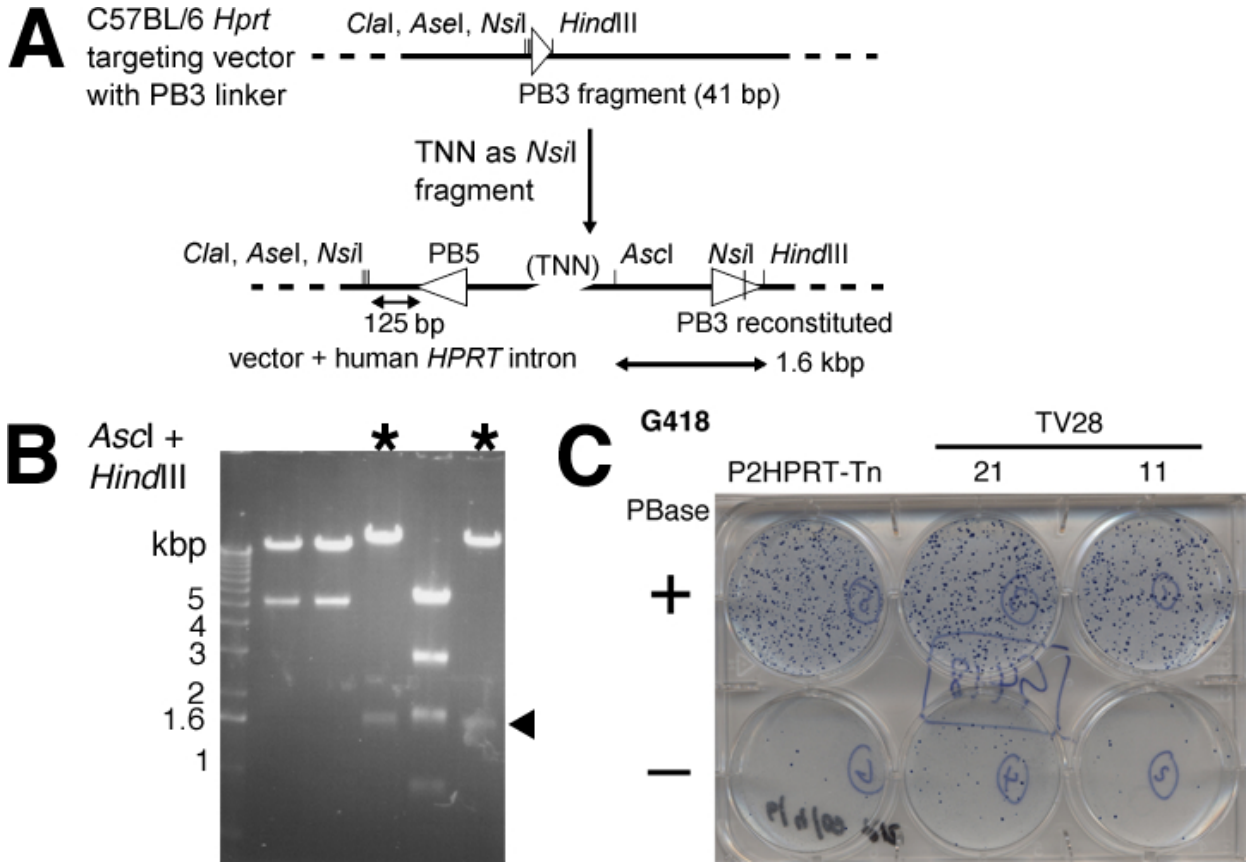

Figure 6.3: Cloning the TV28 targeting vector. A—Cloning scheme. A linker was inserted into the *Hind*III site in the *Hprt* targeting vector, containing part of PB3 distal to the *Nsi*I site. The TNN transposon was cloned as an *Nsi*I fragment from P2-HPRT-Tn. B—Screen of clones from the ligation. Asterisk (*) marks correct clones. Arrowhead indicates the 1.6 kbp band showing that the insert is in the correct orientation to reconstitute PB3. C—The transposon is reconstituted and functional in ES cells. Transient transposition assay with (+) or without (–) pCMV-hyPBase and using the indicated TV28 donor plasmids (two subclones, 11 and 21) or P2-HPRT-Tn.

quency of at least 25%. All clones also showed the expected HAT-sensitive/6TG resistant phenotype, although some contained a small proportion of HAT-resistant cells, presumably wild type (Figure 6.5B). Therefore, to remove the possibility of background HAT resistance from contaminating untargeted cells, I further subcloned the line prior to transposition for library generation. Transfection of subcloned cells with PBase resulted in HAT-resistant subclones as predicted, with no background observed in untransfected cells (Figure 6.5C). This cell line is named B6BTV.

Introduction of an inducible Cre gene into C57BL/6 *Blm* ES cells

In order to use short expansion times in homozygous mutant generation, it is important that the few homozygotes that do segregate in the limited expansion can be efficiently isolated. Transfection and expression of Cre is a limiting factor for the double selection procedure. To improve on this, I introduced an ERT2-Cre gene as for the 129 strain ES cells in the previous chapter. This time I used a vector obtained from Junji Takeda (Osaka University), which introduces an ERT2-iCre-ERT2 gene into the *Rosa26* locus. This gene consists of a mammalian codon-optimised (improved, iCre) Cre coding sequence fused to ERT2 at both termini (Shimshek *et al.*, 2002). Using two ERT2 moieties appears to reduce leakiness (J. Takeda, K. Yusa; personal communication). The targeting involved two steps: One to insert the transgene into the locus, and one to remove the F3-flanked *neo* selection marker using Flp recombinase (FlpO). To make a more generally useful cell line, I targeted this construct to the original $Blm^{e/e}$ cells and then retargeted the resulting cell line (named BRic, **B**L/6 *Blm* **R**osa26-**iC**re) with TV28 as above.

Results of the Cre targeting are summarised in Figure 6.6. I genotyped targeted clones using a PCR assay on colony lysates, which specifically detects the junction on the short arm side of the targeting vector (Figure 6.6A). Rapid PCR genotyping enabled me to directly expand correctly targeted clones, transfect them with a FlpO expression plasmid (PGK-FlpO, Raymond and Soriano (2007)) and screen unselected subclones by PCR for loss of *neo* (Figure 6.6B). As deletion of the *neo* gene was not selected for, it is probable that not all cells in the clone have undergone the deletion, as deletion could have occurred after the first division of the founding cell of the colony. This was evident from G418 sensitivity tests, therefore in order to make a stable line I subcloned some PCR-positive cells. The subclones were a mixture of deleted and undeleted cells as predicted (Figure 6.6C). In parallel to this subcloning, I took one clone that showed the fewest remaining G418-resistant cells for targeting with TV28 as above.

Library generation cell lines

To target the transposon to the BRic cell line, I used a different strategy to simultaneously recover a pure clone that had deleted the *neo* gene from the Cre targeting. I electroporated unsubcloned BRic cells with linearised TV28 as above, and plated the cells in 4-OHT containing medium for 24h. After three days, I replated the cells in puromycin. As the transposon in TV28 was originally in the *neo* orientation, using this scheme provides an internal functional test for inducible Cre activity in the cells and function of the loxP sites and resistance genes in the transposon. No puromycin resistant cells were obtained in a parallel experiment without 4-OHT treatment (Figure 6.7).

I verified correct targeting by PCR screening and functional (HAT sensitivity) test as before (Figure 6.8). Multiple targeted subclones were frozen. These cells were named BRic.TVP or, later, LGP (**L**ibrary **G**eneration **P**uro). In summary, these cells have all the elements desirable for making homozygous mutants easily, i.e.:

- Homozygous null mutation at the *Blm* locus.
- 4-OHT-inducible codon-optimised Cre expressed from the endogenous *Rosa26* promoter (heterozygous).
- TNP transposon integrated at *Hprt* intron two, excision selectable with HAT.
- C57BL/6 genetic background—the same as the reference genome—for easy mapping and comparability.

6.2.2 Generating libraries with the LGP cell line

The scheme for library generation is shown in Figure 6.9. I expanded two subclones, A2 and B4, of the LGP cell line independently to 6-well plates. I transfected adherent cells with 1 μg capped *in vitro* transcribed hyPBase mRNA using the TransMessenger lipofection reagent (Qiagen). The reason for using mRNA was to avoid the possibility of the transposase expression plasmid integrating into the genome of some cells (see Chapter 3). I replated

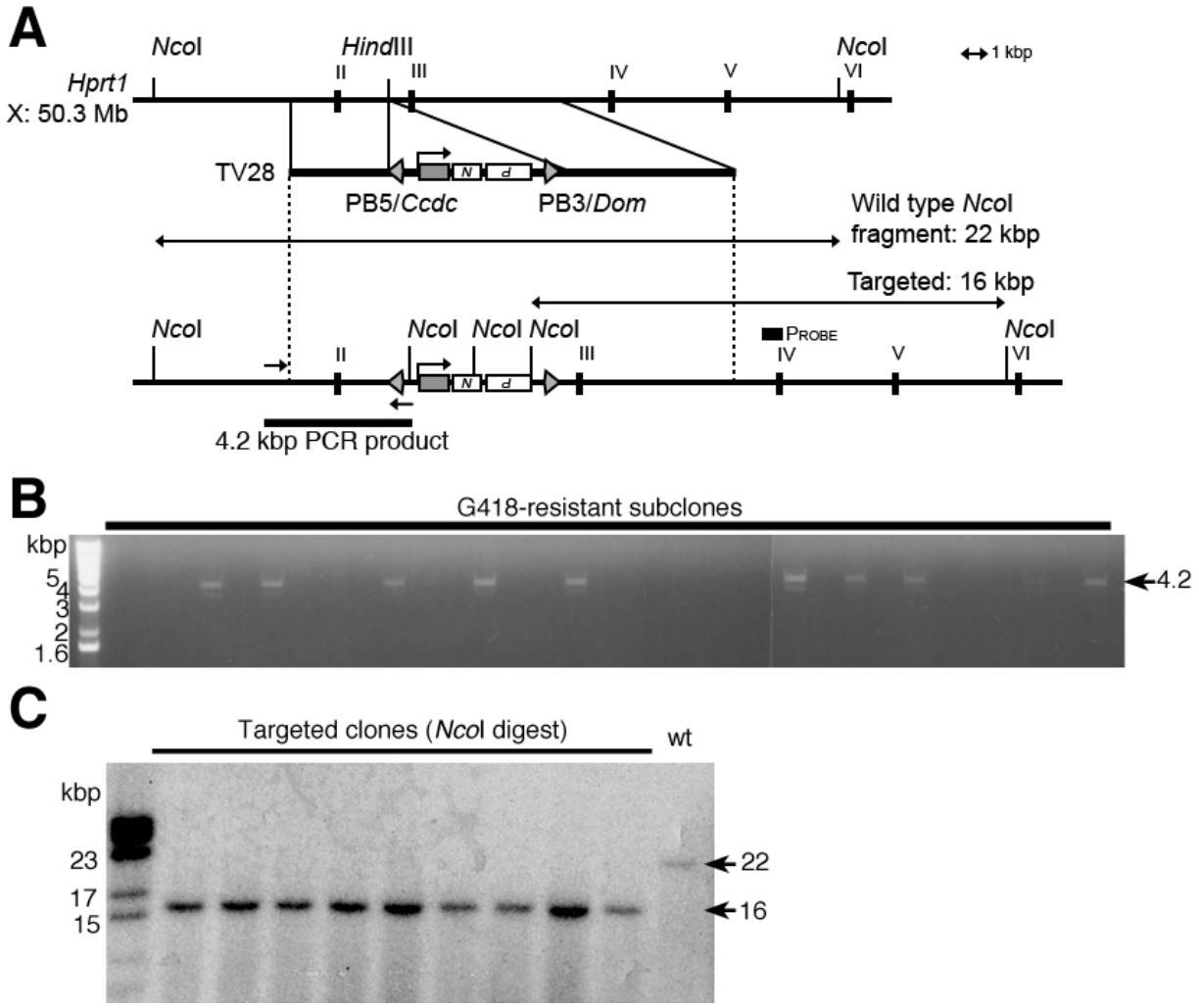

Figure 6.4: Targeting the transposon to *Hprt* in $Blm^{e/e}$ cells using the TV28 vector. A—Targeting scheme. B—PCR screen of lysates from G418-resistant colonies. A 4.2 kbp product is amplified from correctly targeted clones; another product of about 3.5 kbp, still specific to the targeted clones, is also amplified. This may arise from priming elsewhere in the PB repeats. C—Southern blot confirming correct targeting at the 3′ end in nine targeted clones. The probe used is shown in Figure 6.4.

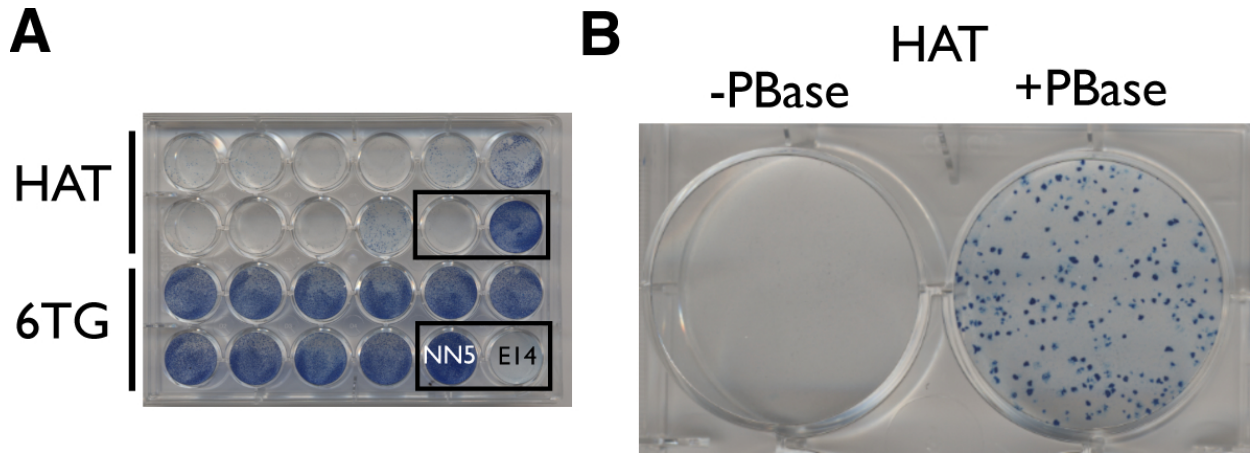

Figure 6.5: Confirmation of targeting and testing transposition from the *Hprt*^{PB} locus. A—Targeted clones are HAT sensitive and 6TG resistant. These have not been subcloned and therefore some HAT-resistant background can be seen in some wells. NN5 and E14 are *Hprt*-negative and positive controls respectively. B—Transfection with PBase results in HAT-resistance from clones that have excised the transposon. The cells used are a subclone from an originally identified targeted clone; no HAT-resistant background is observed without PBase transfection.

the cells 24 hours post transfection and selected in HAT+puromycin to isolate clones in which the transposon had excised and reintegrated elsewhere in the genome. These cells were expanded under HAT and puromycin selection for eight days and in M15 thereafter (with two days in HT medium to allow recovery from HAT selection). The cells were replated during this expansion/selection phase, at which point some were transferred to another plate at low density for counting and analysis of the mobilisation by Southern blot (Figure 6.10).

After expansion for 14 days in total, I changed the medium to 1 μM 4-OHT overnight. The next day I changed the medium to M15 and allowed one day for Cre activity to subside. At this stage I picked a few colonies from a low density plate to assess Cre induction by Southern blot. For the higher density plate I harvested the cells, counted and replated in DBL medium, again at both high and low densities (half and one tenth of the total respectively). I picked colonies from a lower density plating and analysed by Southern blot to determine transposon copy number, orientation and insertion site using the same digest and probe as in Chapter 5. I included clones from the initial mobilisation, background plates and unselected 4-OHT treated plates in the analysis to get a complete picture of each stage of the process.

6.2.3 Sources of background in double-resistant population

Copy number increase on transposition

From Southern blot analysis of double resistant clones generated in this experiment, problems were immediately apparent. First, several clones with two non-allelic insertion sites could be seen even in the HAT+puro resistant clones from immediately after mobilisation (Figure 6.11; lanes 1, 3, 6, 8). Therefore mobilisation from the X chromosome failed to effectively limit the transposon copy number to one. Analysis of genomic DNA from unmobilised LGP cells confirmed that there is only one copy of the transposon (at *Hprt*) in the starting cells (Figure 6.11, far right lanes).

Selection background

Predictably, having clones with two copies present at the start resulted in a background of clones with two non-allelic insertions in the double-resistant population. Moreover, there were some clones that survived double selection but did not have both *neo* and *puro* forms of the transposon as assessed by Southern blot (Figure 6.11, lanes with asterisk). This was surprising, as I had never observed this in the clone-by-clone experiments described in the previous chapter. Therefore it is likely that this background is specific to a comparatively rare set of loci that may have high transcription activity on

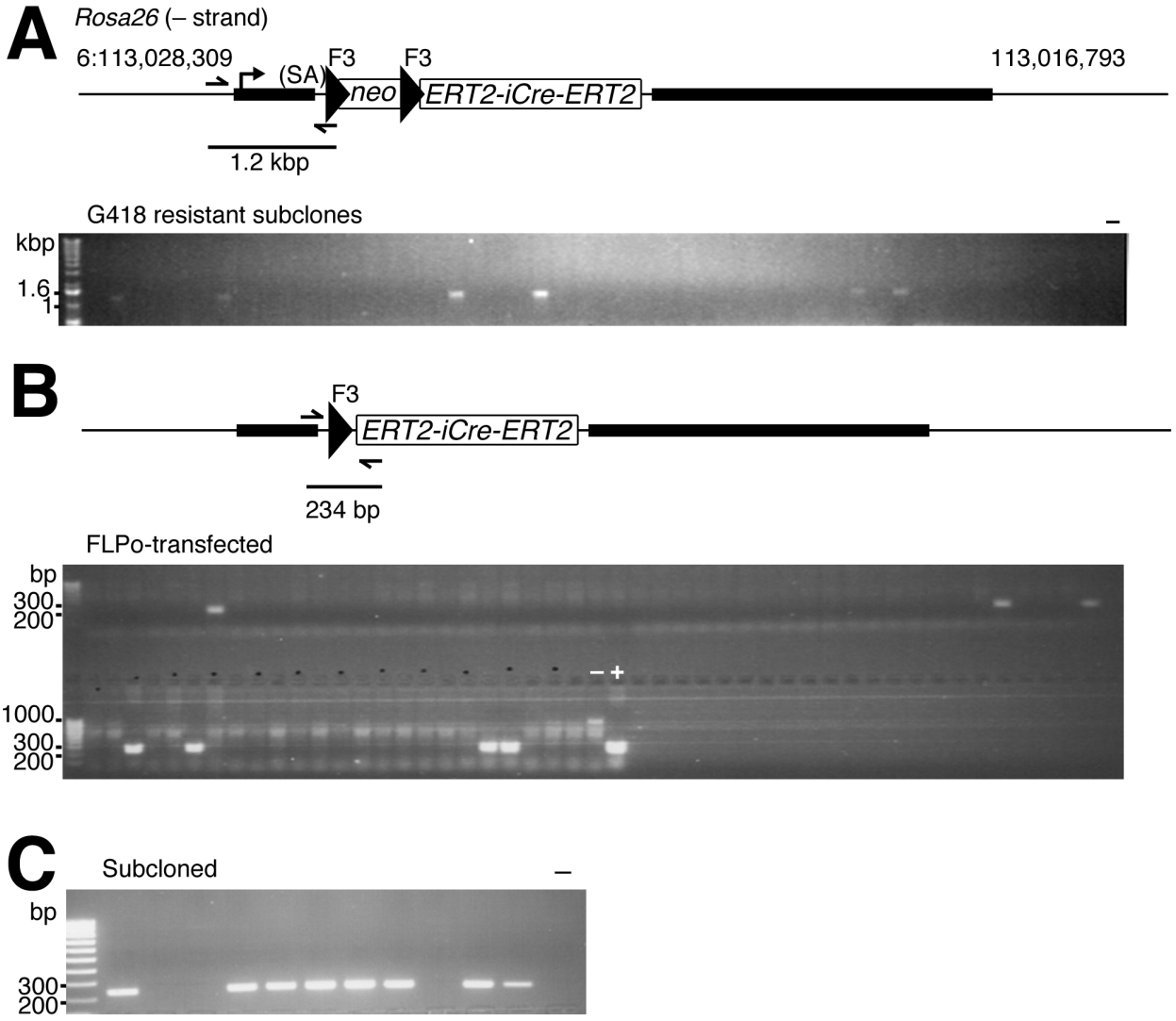

Figure 6.6: Targeting an inducible Cre gene in *Blm*$^{e/e}$ cells. A—Structure of targeted locus. *Rosa26* is shown 5′ to 3′, left to right, although it is on the reverse strand. Homology arms in the targeting vector are shown with a thick line. PCR screen for targeted clones is shown. B—Structure of targeted locus after Flp-mediated removal of *neo*. PCR screen shown. +: DNA from correctly 'popped out' cells (K. Yusa); −: DNA from untransfected cells. A PCR product of around 1 kbp is visible in the negative control and faintly in the clones that have not undergone Flp-mediated deletion. C—Repeated PCR screen on subclones of a correctly 'popped' clone from (B). The *neo* product was visible at long exposure in the negative subclones.

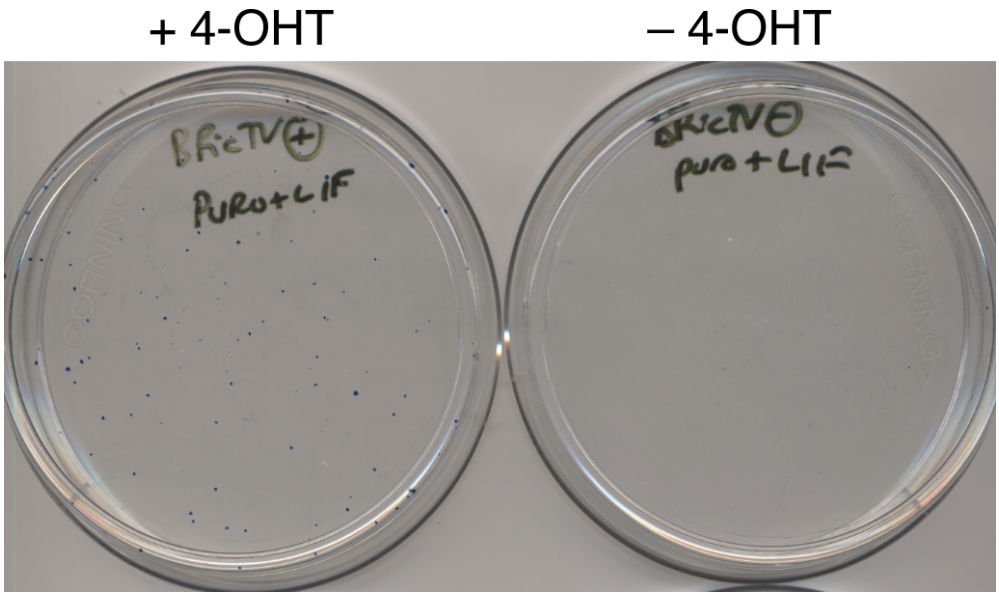

Figure 6.7: Gene targeting combined with 4-OHT treatment to insert the transposon in BRic cells. Puromycin resistant clones were only obtained when cells transfected with the TV28 targeting vector were also treated with 4-OHT.

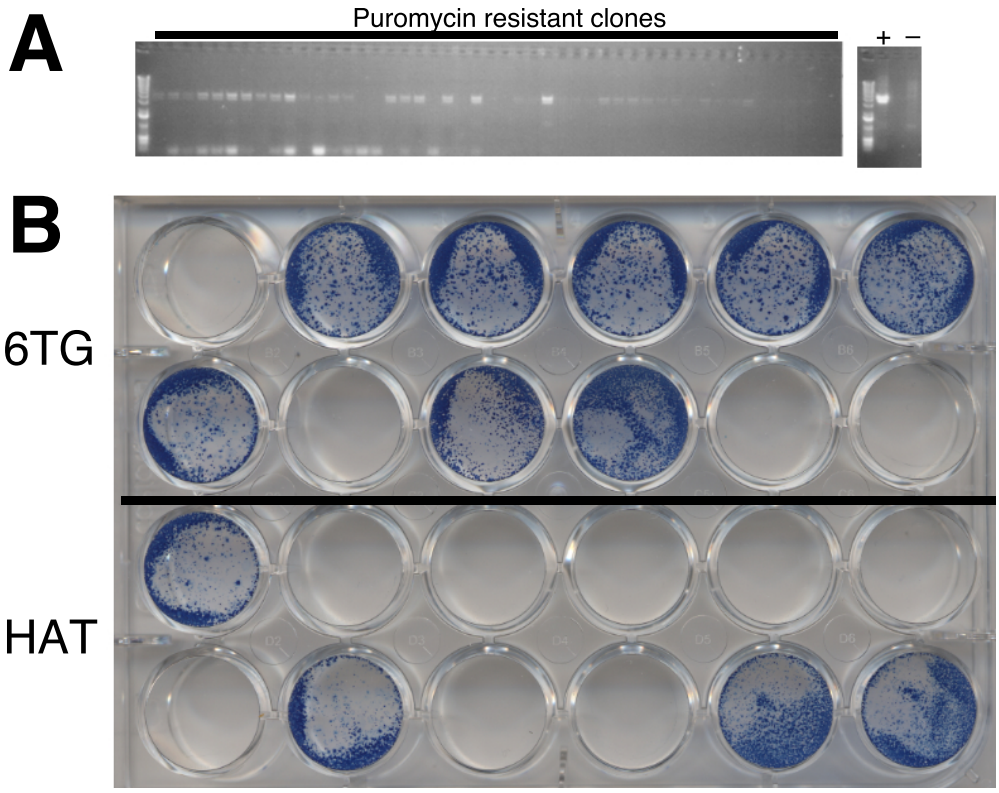

Figure 6.8: Confirmation of targeting in LGP cells picked from plates in Figure 6.7. A—PCR screen at 5′ end. This showed a large number of positives, therefore a functional test was also performed (B). Two HAT-sensitive/6-TG resistant clones were subcloned and frozen in small aliquots.

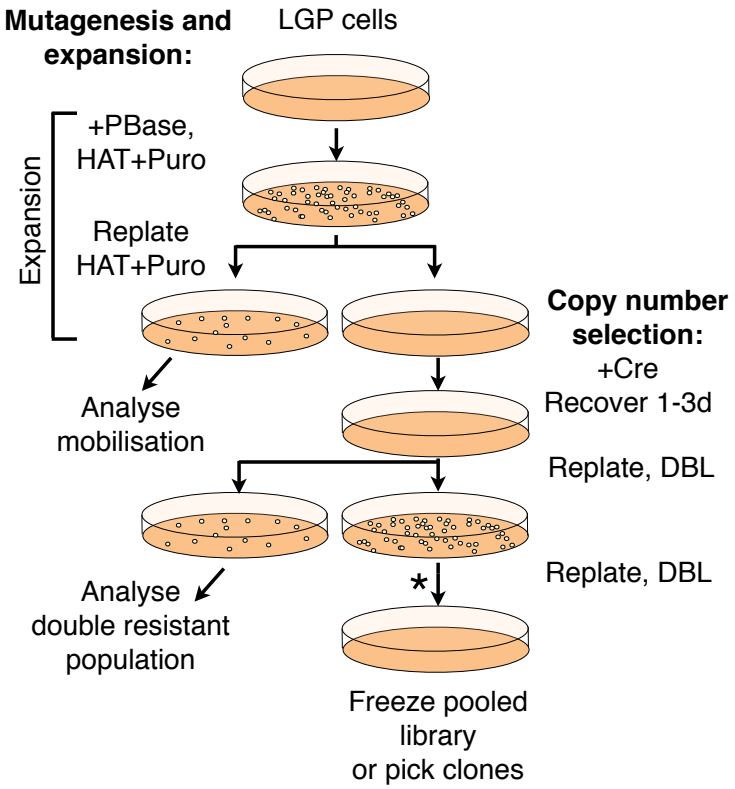

Figure 6.9: General scheme for library generation using the LGP cell line. For LGN cells, HAT+G418+FIAU selection would be used instead of HAT+Puro. The plating scheme shown is for the LGP libraries analysed (Figures 6.11 and 6.13). For the LGNL library analysed in Figure 6.19, clones were replated under DBL selection, at the stage indicated by an asterisk (*).

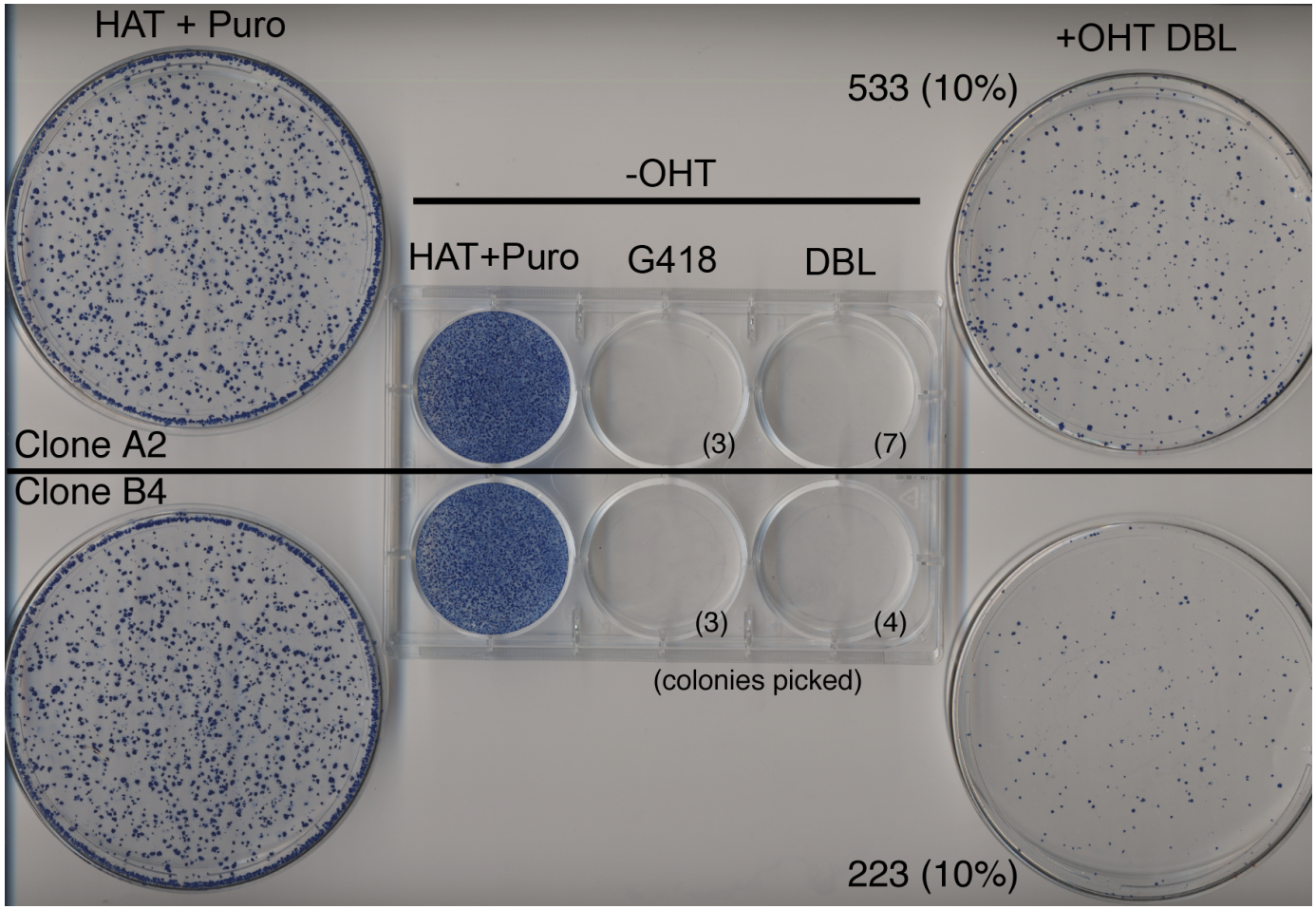

Figure 6.10: Colonies stained at various stages of the library generation process in LGP cells. Two parallel library generations from different starting subclones (A2 and B4) are shown. The plates shown are small proportions of the culture split to a separate plate at each passaging stage and stained 8–10 days later. From left, HAT+puro resistant cells that have mobilised the transposon. These cells are not generally G418- or DBL-resistant without 4-OHT treatment (six-well plate). A few colonies (numbers in brackets) did grow on these plates but were picked for genotyping (Figure 6.11). Finally, a plate containing 10% of the DBL-selected culture is shown with colony numbers.

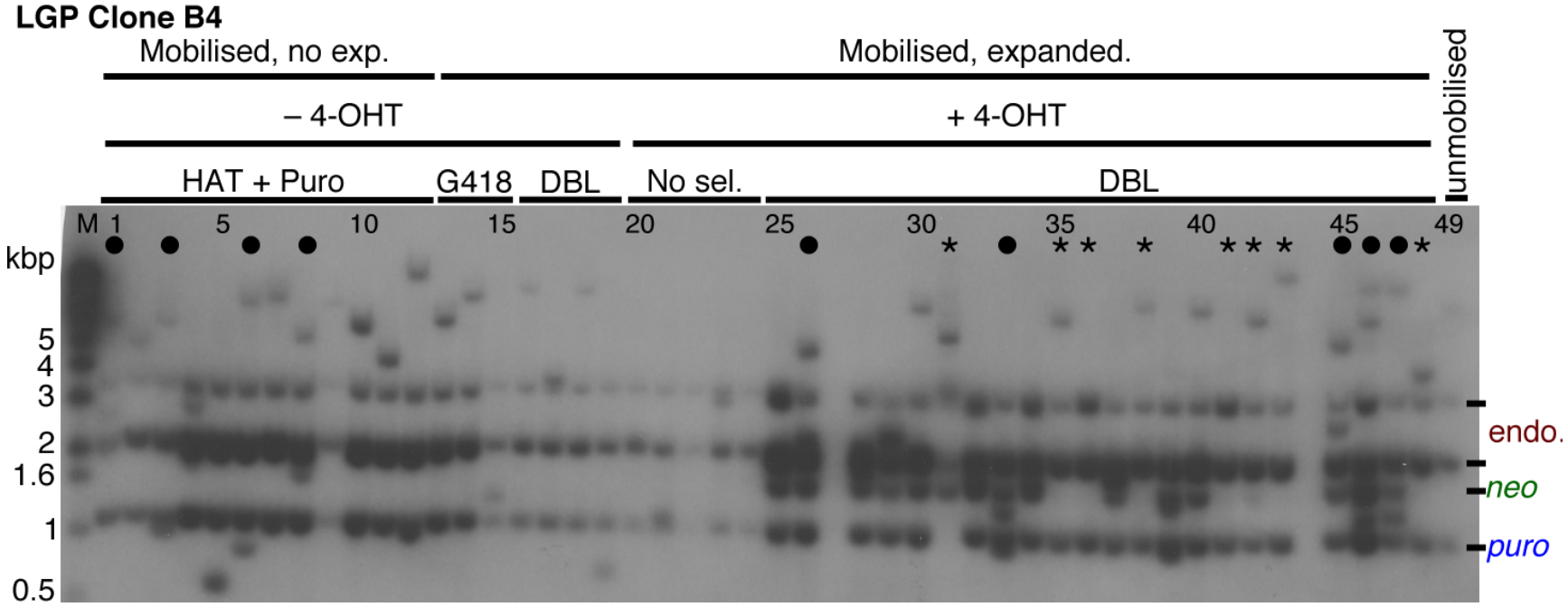

Figure 6.11: Background in double-resistant clones from pooled libraries. Clones from each stage in the generation process from LGP B4 cells were analysed by Southern blot using *Nco*I digested genomic DNA and probe as in Figure 5.6. Several mobilised clones (HAT+Puro) have more than one insertion site (lanes labelled with a dot). Background colonies that are G418- or DBL-resistant prior to 4-OHT treatment have not inverted the resistant construct, as shown by the lack of *puro* band in these clones. Unselected clones treated with 4-OHT show little or no recombination to the *puro* version, suggesting that this was not very efficient. DBL resistant clones that do not show both *neo* and *puro* bands are shown with an asterisk. The far right lane is LGP DNA (unmobilised).

the reverse strand sufficient to express the 'inactive' selectable marker (Figure 6.12A). In these experiments I am interrogating thousands of genomic loci, so it is possible that loci with these properties would be picked up in this experiment but missed in a clone-by-clone context.

If this explanation is correct, there is a simple work-around. If the initial transposition was carried out with a TNN transposon (i.e. in the *neo* expressing orientation rather than *puro*), FIAU could be used to select against loci that can express the $pu\Delta TK$ from a genomic promoter. To convert the TNP transposon in the LGP cell line to TNN, I simply treated cells with 4-OHT and selected in G418. I picked a number of resistant subclones, forming the LGN cell line with the transposon in the *neo* orientation to allow the proposed FIAU counterselection.

Comparison of libraries generated from different clones

I generated libraries starting from two distinct LGP subclones—A2 and B4. The A2 library yielded more double resistant cells, but also had much higher background. This could be due to a problem with the A2 line prior to mobilisation, e.g. aneuploidy, or a stochastic event leading to background that only occurred in the A2 clone. Particularly, 11/24 double resistant clones in this library appeared to have a single insertion site, and arose from selection background (Figure 6.13; lanes 21, 22, 23, 26 etc.).

I expected that clones picked from the 'background' selection plates (i.e. G418 or DBL selection without 4-OHT treatment) would dominate the double resistant population. However, this was not generally the case. In the A2 library, two such clones can be identified (Figure 6.13; lanes 14, 24, 39 and lanes 11 and 36). The B4 library appears relatively complex, despite the background. This indicates that genuine expansion-dependent copy number gain events occur frequently enough to dilute the effect of the background clones to some extent.

Treatment with 4-OHT appeared more effective in the A2 clone compared to the B4 clone (Compare Figure 6.13 lanes 19 and 20 with Figure 6.11 lanes 20–24). In the B4 clone, the intensities of the *neo* and *puro* bands were approximately equal, whereas many subclones of the A2 line displayed bands of unequal intensity. It is possible that some Cre activity remained after plating of the culture in DBL medium, and therefore that some cells continued to switch. An extra replating step or a longer recovery period from 4-OHT induction should be added to ensure that the clones analysed are pure.

6.2.4 A G1-specific transposase to conserve copy number during transposition

PB is known to transpose by a cut-and-paste mechanism, which should be non-replicative. Nonetheless, my data clearly indicate that copy number can increase upon transposition. Considered together with my data from sequencing of transposon excision sites (see Figure 7.9 and discussion in Chapter 7), a possible explanation is that transposition occurs after DNA synthesis, in S or G2 phase. The transposon does increase in copy number after DNA replication, as does every other locus in the genome. While the two copies would normally segregate to different daughter cells, transposition at this stage with reintegration on another chromosome (or the sister chromatid) could result in a daughter cell with two copies (Figure 6.12B). It is not known whether PB transposition is regulated based on cell cycle stage.

If this hypothesis is correct, the copy number increase could be avoided by limiting transposition to G1 phase of the cell cycle. An interesting study describing a fluorescent based cell cycle indicator suggested a possible way to achieve this. Sakaue-Sawano *et al.* made two complementary fluorescent protein fusions fused to degradation signal sequences from two cell cycle regulated proteins. Cells expressing the GFP derivative-Geminin fusion that they describe are green, except in G1 and early S phase, where Geminin is ubiquitylated and degraded. Similarly, cells that express a RFP derivative fused to a CDT1 fragment fluoresce red, but only in G1 and early S phase, after which CDT1, a replication origin licensing factor, is degraded. The fusion fragments were from human genes, but worked effectively in mice too. Therefore, fusing the PB transposase to the CDT1 fragment in a similar way might also limit its expression to G1 and early S.

I prepared cDNA from human iPS cell RNA and PCR amplified the human *CDT1* fragment encoding amino acids 30–120 as described in Sakaue-Sawano *et al.* (2008), using primers tailed with 50 bp arms with homology to the hyPBase expression plasmid. My strategy was to fuse the CDT1 to the PBase C-terminus, and introduce a linker of three amino acids between the two (Figure 6.14A). At the DNA level, this linker included an *Asc*I site to allow cloning of longer linkers if required, as previous attempts to make a PBase-ERT2 fusion had shown that PBase activity was affected by the fusion—in this case only a C terminal fusion with a positively-

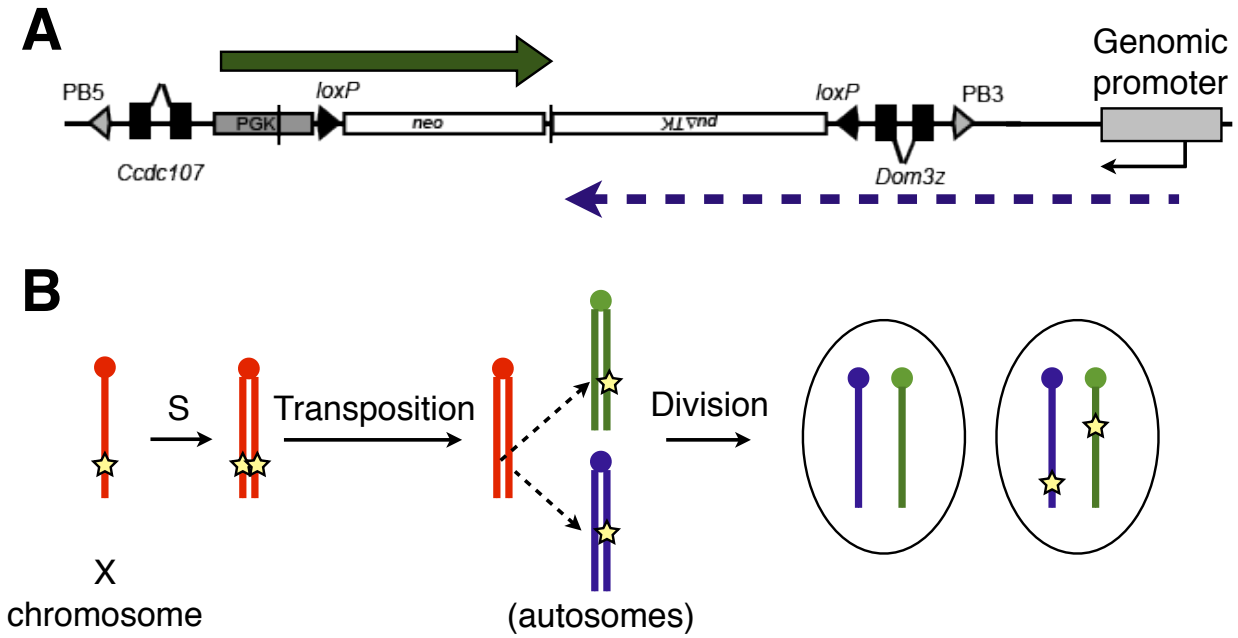

Figure 6.12: Possible sources of background in pooled libraries. A—Transcription of the reverse strand at certain genomic loci leading to expression of the 'off' resistance gene. B—Possible mechanism for the isolation of cells with two non-allelic copies, based on transposition occurring after DNA synthesis (S phase).

charged linker showed activity (Cadiñanos and Bradley, 2007). However, this could have been associated with the ERT2 domain rather than the PBase.

I linearised pCMV-hyPBase with *Not*I and co-electroporated heat-induced SW106 bacteria (see Methods) with the gel-purified fragment with the CDT1 PCR product with homology arms. Recombination in the bacteria reconstitutes a circular plasmid with the correct fusion gene, selected for by growth on ampicillin. Colonies were screened by PCR and checked by restriction digest after retransformation.

The fusion protein was active as a transposase in ES cells, displaying slightly lower activity than hyPBase without the fusion partner (Figure 6.14B). There are no commercial antibodies available that recognise the PB transposase. I tried using a commercially available antiserum raised against amino acids 7–106 of human CDT1 (Abcam ab52731). However this reacted with many different proteins on a Western blot of ES cell extracts, and did not detect any extra proteins in cells that had been transfected with a PB-CDT1 expression plasmid (Figure 6.15). Furthermore, I could not detect a band of the predicted size for endogenous CDT1 in protein extracts from a human cell line. Therefore I was unable to show cell cycle regulation of the fusion protein by Western blot. In an attempt to detect the fusion protein I made a plasmid designed to express a N-terminal FLAG epitope-tagged version of PB-CDT1. However, this transposase was not active in a transposition assay in ES cells (not shown). Further work needs to be done to verify the cell cycle regulation at the protein level.

6.2.5 Mobilisation using G1-specific transposase and FIAU counterselection

I went ahead with a mobilisation experiment using LGN cells and PB-CDT1 *in vitro* transcribed mRNA. This time I used medium containing HAT, G418 and FIAU (HGF) to select clones with excised and reintegrated transposons at loci that will not result in selection background. In this case, mobilisation efficiency was very low—only 142 clones were obtained. However, all clones had single copy integrations when analysed by Southern blot (Figure 6.16). Whereas seven of 20 clones analysed from mobilisations using hyPBase had two non-allelic insertions, all 18 analysed using PB-CDT1 had single copy integrations. This shows that the PB-CDT1 fusion protein does effectively limit the copy number of the transposon during transposition. I pooled

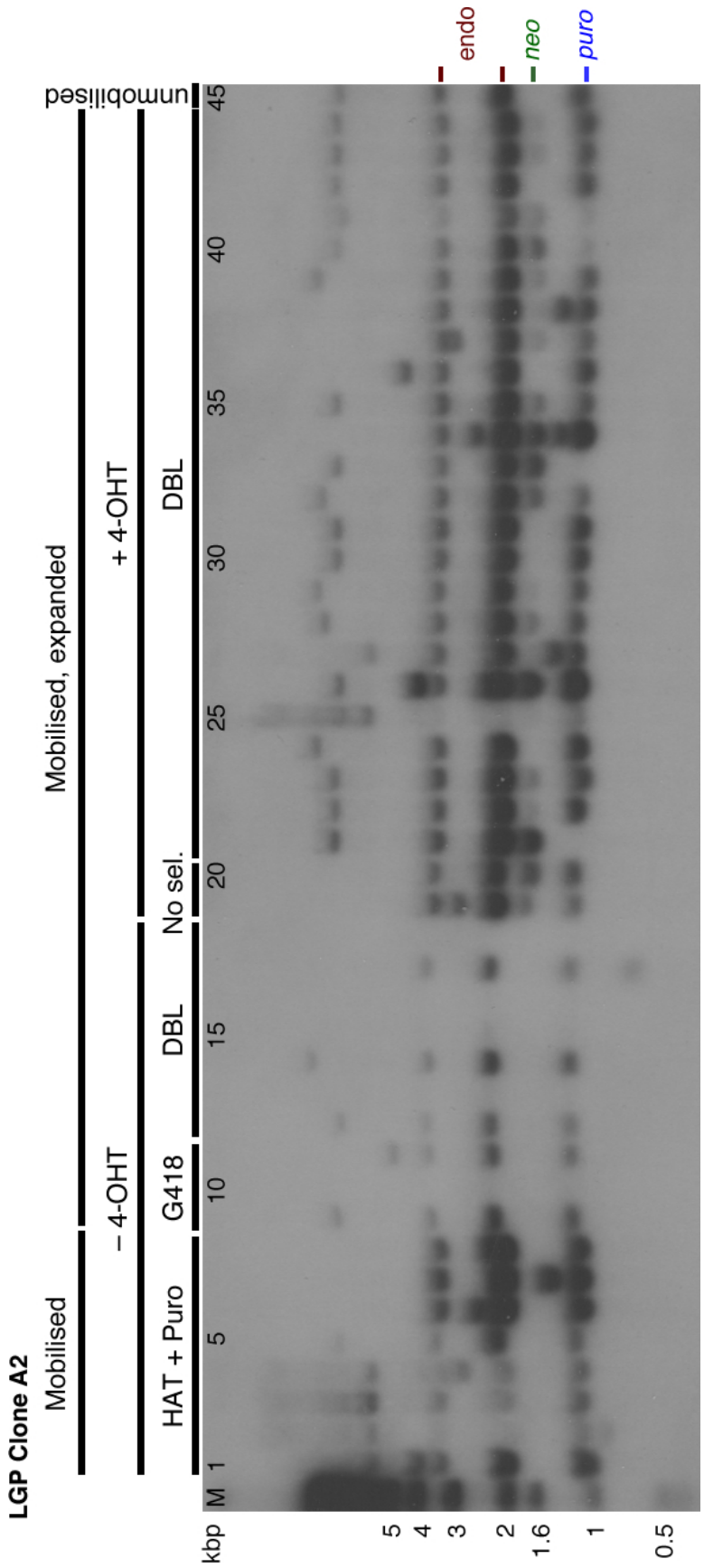

Figure 6.13: Analysis of clones from the LGP A2 library. Clones from various stages of library generation were analysed by Southern blot as in Figure 6.11. This library showed much higher background in the double resistant population compared to the LGP B4 library. Cre induction, however, appeared more efficient (lanes 19 and 20).

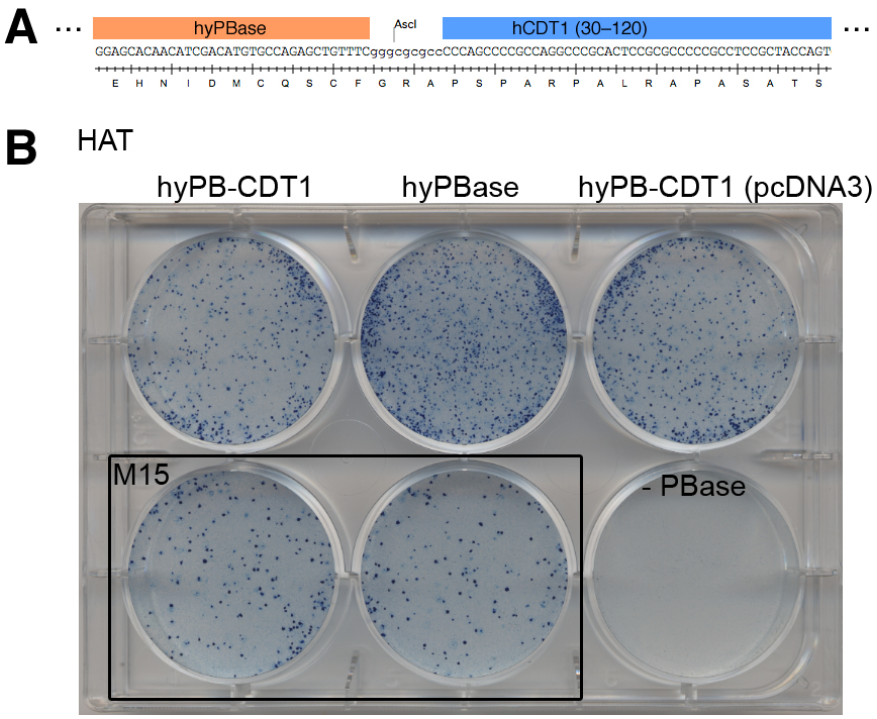

Figure 6.14: Structure and function of PBase-CDT1 fusion protein. A—DNA and protein sequence at the junction between the hyPBase coding sequence and the fused human CDT1 fragment. B—The fusion does not impair function of the transposase. Results of mobilisation in LGP cells are shown. ES cells were electroporated with 10 μg of the plasmid indicated. Expression of the transposase is driven by the CMV promoter in all cases. Top row—HAT selection, indicating transposition. Bottom row—cloning efficiency plates for hyPB-CDT1 and hyPBase transfections (unselected), and negative control (HAT selected). The top right well differs only in that PB-CDT1 is expressed from pcDNA3, which also uses the CMV promoter but contains an independently expressed *neo* gene.

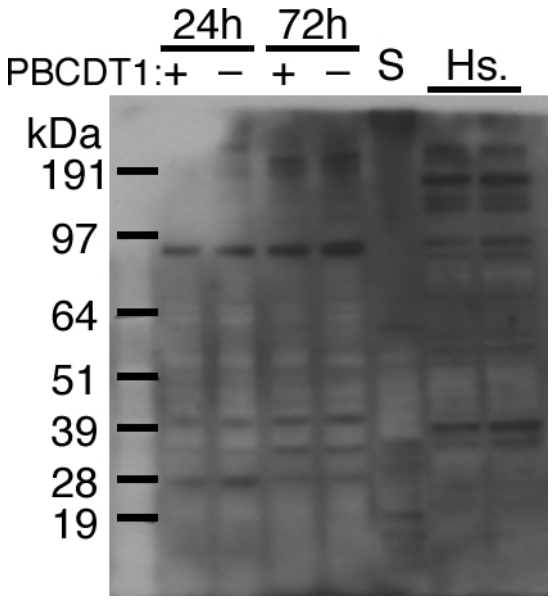

Figure 6.15: Western blot using anti-CDT1 antibody. Protein extracts tested (10 μg loading each) are JM8A3 cells at 24 or 72h post transfection with a PB-CDT1 (+) or GFP (–) expression plasmid. S—a stable G418-resistant cell line following pcDNA3-PB-CDT1 transfection [protein appears degraded]. Hs—control protein extracts from human cells. Predicted sizes—Human/mouse endogenous CDT1/Cdt1: 63 kDa; PB-CDT1 fusion: 87 kDa. The secondary antibody was HRP-conjugated rabbit anti-mouse IgG (Abcam ab6728).

all clones from this mobilisation experiment and expanded them for double selection.

The low efficiency of mobilisation in this case is likely to have been a technical problem with this particular experiment, as reintegration was readily observed in an experiment where LGP cells were transfected with PB-CDT1 (Figure 6.17A). Moreover, including FIAU counterselection did not adversely affect the number of recovered reintegration events in a pilot experiment conducted in B6BTV cells (Figure 6.17B).

6.2.6 Results of double selection with PB-CDT1 and FIAU counterselection

I expanded the cells for 20 days, as the initial cell number was low, then replated the cells in medium containing 1 μM 4-OHT. As the efficiency of Cre induction was low in the LGP B4 clone previously, where adherent cells were induced, I treated cells in suspension this time. After incubation overnight, during which cells attached to the plate, I replaced the medium with M15. After two further days I replated cells in DBL. Counting colonies from a replica plating at low density indicated that the culture contained 7,200 DBL-resistant colony-forming units (cfu) at this point. I checked the efficiency of Cre induction, and also verified that no double resistant colonies grew without 4-OHT treatment (Figure 6.18). In this case the Cre induction was very efficient, as opposed to the previous experiment with LGP cells where OHT treatment was carried out on adherent cells (see Figure 6.11, lanes 20–24).

To ensure that I picked pure clones, after four days I replated the cells under DBL selection. I picked double-resistant colonies and analysed them by Southern blot as before. This time, all clones analysed displayed fragments from both *neo* and *puro* forms of the transposon. This shows that FIAU counterselection effectively removed loci with expression of the resistance gene that is not oriented with the PGK promoter (Figure 6.19).

Despite starting from an apparently all single copy population, 19 of 45 double resistant subclones analysed from this library still have two non-allelic copies of the transposon (Figures 6.19 and 6.20). However, these are distinguished from the clones with two copies in the first library by the fact that all but one (lane 16) share one band, the size of which is consistent with the $Hprt^{PB}$ donor locus. Such clones were not detected in the initial mobilisation, suggesting they were only present at a low

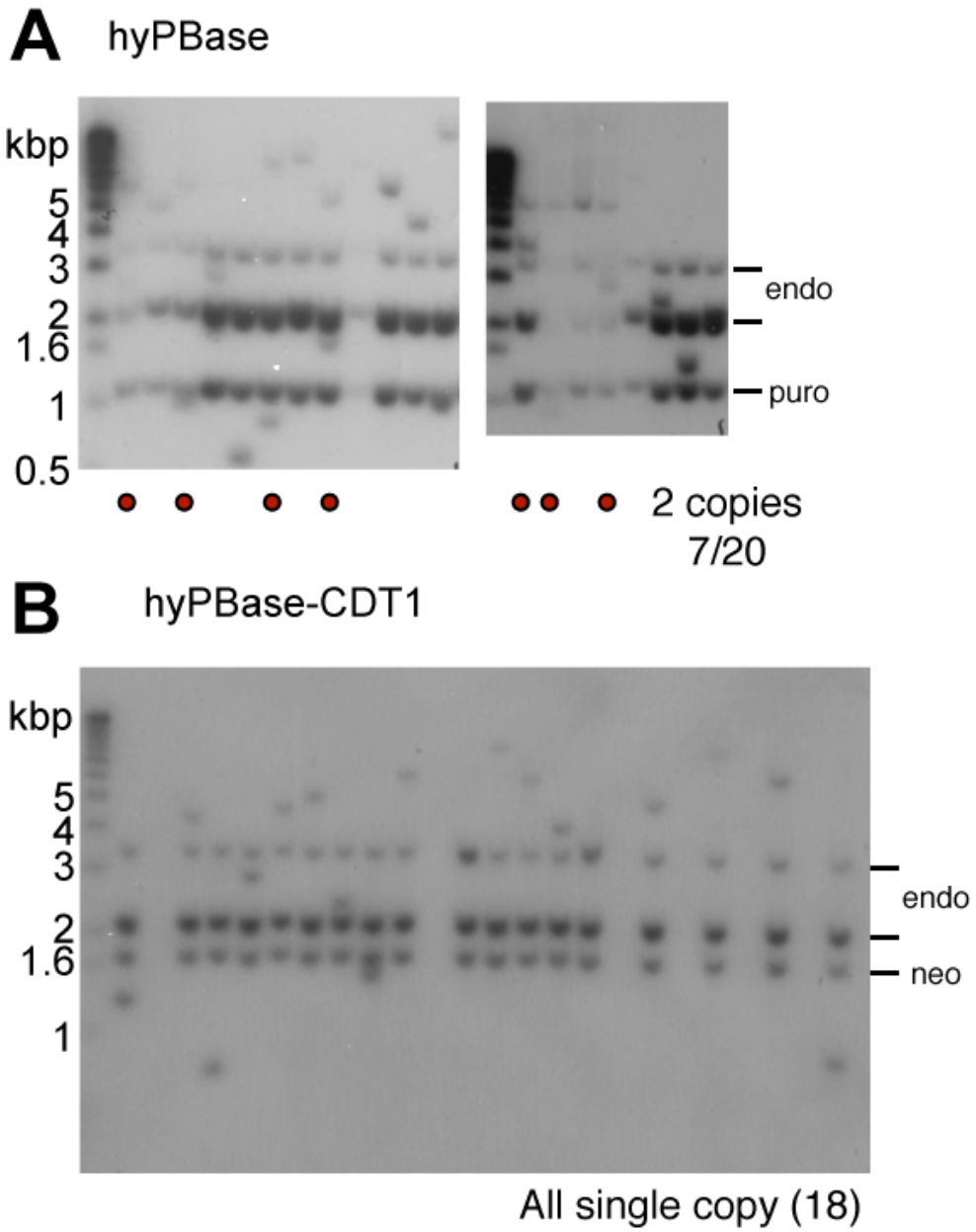

Figure 6.16: Mobilisation using the PB-CDT1 fusion protein preserves copy number. *Nco*I digests probed with the PB-CCdc probe as in Figure 6.11. A—HAT+Puro resistant clones from two mobilisation experiments in LGP cells using hyPBase mRNA. Several clones with two insertion sites are visible (lanes marked with a red dot). B—HAT+G418+FIAU resistant clones from mobilisation in LGN cells using PB-CDT1 mRNA. All clones have a single insertion site.

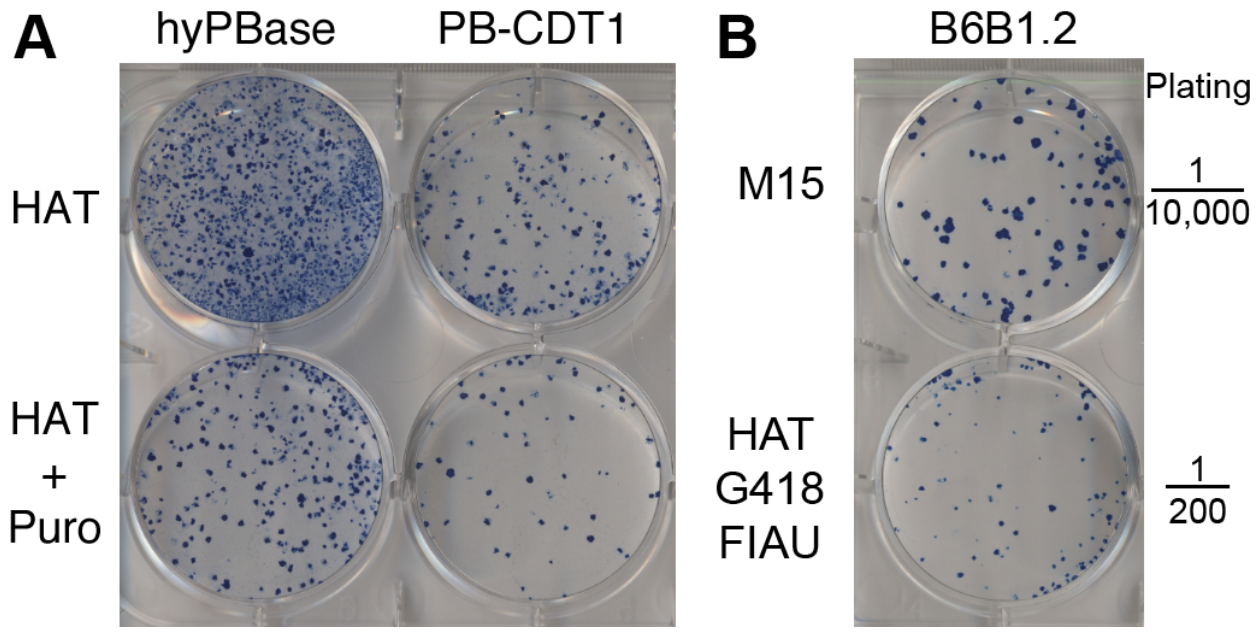

Figure 6.17: A—The fusion with the CDT fragment does not compromise transposon reintegration. LGP cells were transfected with hyPBase or PB-CDT1 expression plasmid as indicated and selected in HAT (excision) or HAT+Puro (excision and reintegration). B—An experiment using B6BTV cells, identical to LGN except for the lack of an inducible Cre gene, showing that reintegration events can be efficiently recovered using FIAU counterselection. The proportion of the mobilised culture plated is shown on the right

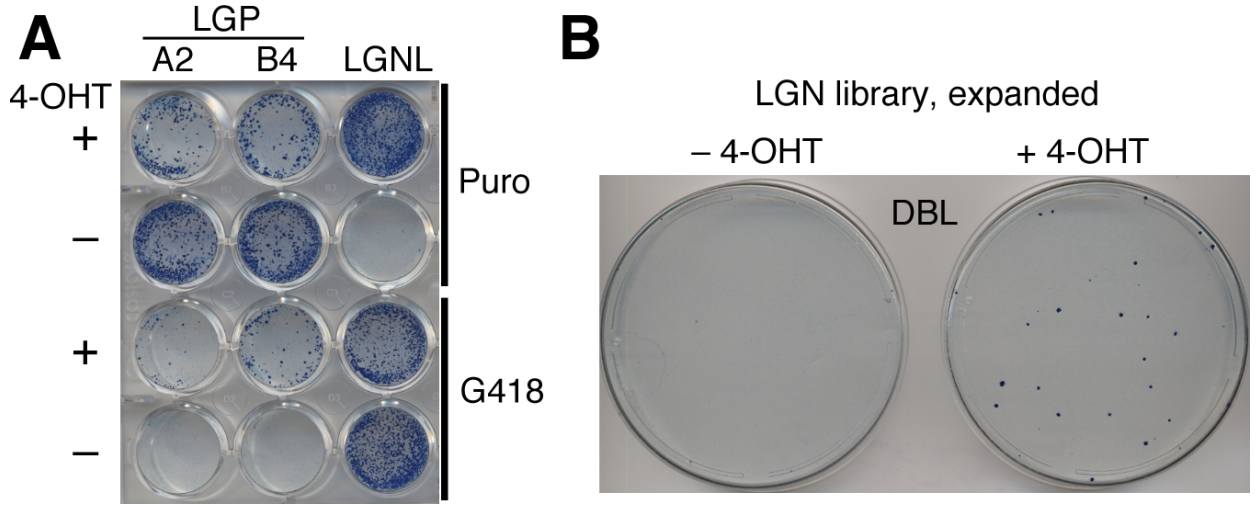

Figure 6.18: Verification of 4-OHT sensitivity and specificity in the LGNL1 library. A—4-OHT treatment efficiently induces switching to G418 resistance. Identically treated LGP cells are plated for comparison; in this case the switch is from puro to G418 resistance. B—4-OHT treatment is required for the isolation of double resistant cells.

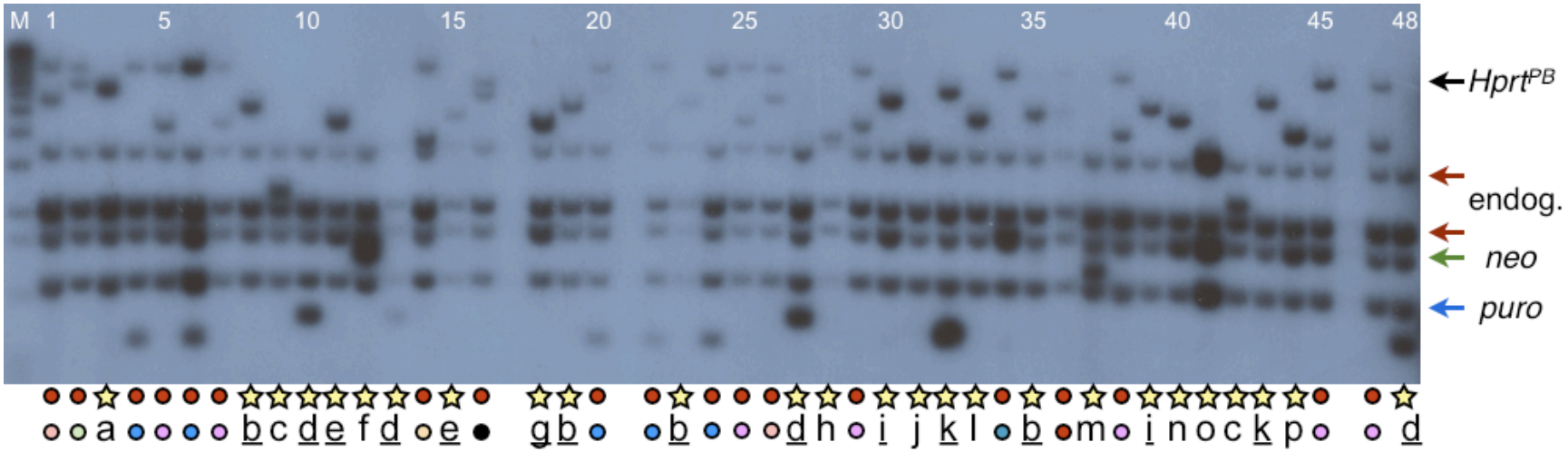

Figure 6.19: Analysis of double-resistant clones generated with G1-specific mobilisation and FIAU counterselection. Southern blot analysis with *Nco*I and PB-Ccdc probe as before. All clones show *neo* and *puro*-expressing forms of the transposon; therefore there is no selection background in this library. Most clones that show two non-allelic insertions (lanes marked below with dots) still have a band consistent with the $Hprt^{PB}$ donor locus (10 kbp). Starred lanes have allelic insertions and are potential homozygous mutants. Clonal relationships are indicated by letters (allelic insertions) or colours (non-allelic). Letters are underlined where the clonal relationship is supported by mapping data rather than band size. Further analysis in Figure 6.20.

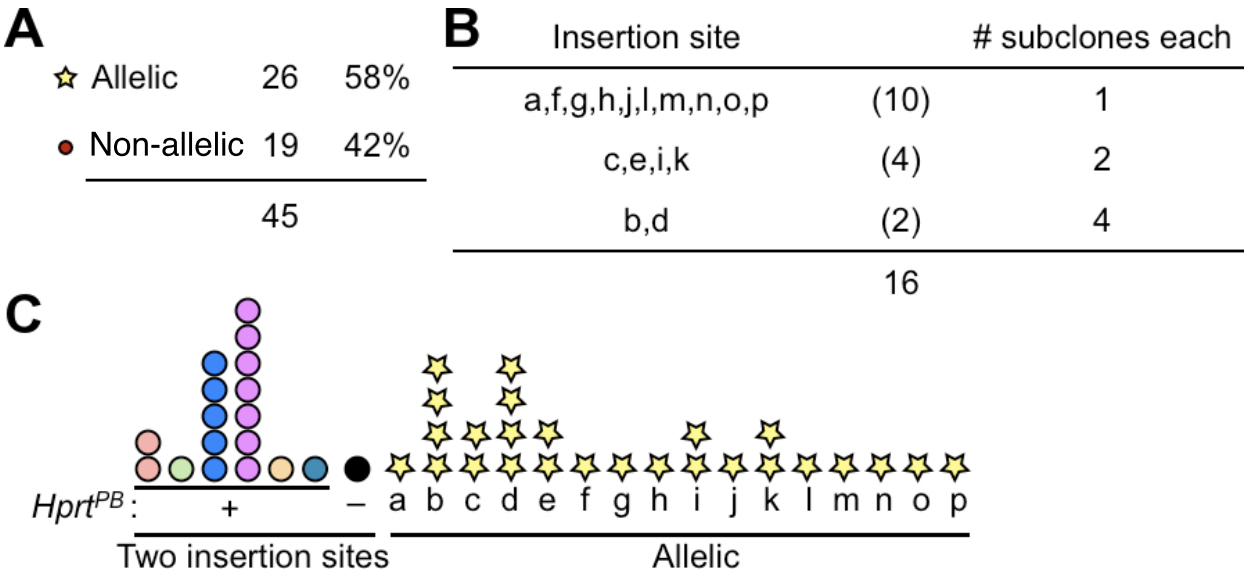

Figure 6.20: Analysis of clonal relationships between LGNL1 double-resistant clones. A—Table showing proportion of clones allelic vs. non-allelic insertions. B—Breakdown of allelic clones by insertion site. The number of subclones representing each insertion site is given in the right column. C—graphical representation of clonality data for both non-allelic (dots; colours correspond to Figure 6.19) and allelic insertions (stars). Each symbol represents a single subclone, each column a different insertion site (or site combination for non-allelic clones).

level prior to double selection.

The remaining 26 clones (58%) displayed the expected band pattern for cells with two allelic transposon insertions. There were at least 16 different insertion sites among the clones analysed on the blot, although this is a lower limit as there is some ambiguity for clones with similar size insertion-specific bands (Figure 6.20). Analysis of the clonal relationships was supported in some cases by mapping the insertion sites for clones with allelic insertions (Table 6.1). This shows that the method can generate complex libraries that are not dominated by clonal expansion from early LOH events. Some insertion sites mapped to the X chromosome. These clones presumably arise from aneuploidy in the culture. The double resistant allelic subclones need to be genotyped individually in order to determine how many are genuine homozygotes and how many arise from aneuploidy and retain the wild type locus.

6.2.7 Some allelic mutants retain the wild type locus

I genotyped the allelic mutants that I was able to map to see if they were genuine homozygous mutants. By PCR I was able to show that five of the double resistant subclones retained the wild type locus, and five were genuine homozygous mutants (Figure 6.21A). Of the five homozygous subclones identified, three were from one parental clone (clone b in Figure 6.19, Figure 6.21B).

6.3 Discussion

6.3.1 Sources of background in mutants isolated from complex pools

The experiments described in Chapter 5, in which mutants were expanded in isolation rather than as a pool, had not revealed any selection background. However, when I applied the method on a large scale, I isolated some clones that had not undergone Cre-mediated inversion of the selection construct. As the selection conditions are similar (with respect to cell density, drug concentration etc.), I interpret this as a consequence of sampling hundreds or thousands of loci simultaneously. The effect was not specific to either of the selectable markers or drugs, further suggesting a problem with the regulatory elements of the construct rather than the G418+puro selection itself.

A potential explanation is that there is a small fraction of potential transposon insertion sites that can express the resistance gene in the opposite di-

Lane	Gene	Chr	Position	Ori	Mapped ends	Clone
3	u/s of *Fubp3*	2	31,426,112	+	5+3	a
8	No gene	18	56,838,595	-	5+3	b
9	*Herc1*	9	66,347,379	+	3 only	c
10	No gene	17	34,010,713	+	3 only	d
11	*Trim47*	11	115,977,569	-	5 only	e
12	No gene	X	34,351,776	-	3 only	f
13	No gene	17	34,010,713	+	3 only	d
15	*Trim47*	11	115,977,572	-	3 only	e
18	No gene	X	12,924,110	-	5+3	g
19	No gene	18	56,838,595	-	5+3	b
23	No gene	18	56,838,595	-	5+3	b
27	No gene	17	34,010,713	+	3 only	d
28	*Gabbr1*	17	37,200,698	-	3 only	h
30	No gene	7	38,812,480	+	3 only	i
31	Refseq transcript	14	22,289,968	+	5+3	j
32	u/s of *Jarid2*	13	44,814,764	+	5+3	k
33	No gene	9	58,117,021	+	3 only	l
35	No gene	18	56,838,595	-	5+3	b
37	*Zmym3*	X	98,615,188	-	5+3	m
39	No gene	7	38,812,480	+	5+3	i
40	No gene	18	38,539,945	+	5 only	n
41	No gene	18	35,183,268	+	5+3	o
43	u/s of *Jarid2*	13	44,814,764	+	5+3	k
44	No gene	18	56,838,595	-	5+3	p
48	No gene	17	34,010,713	+	3 only	d

Table 6.1: Mapping data for LGNL1 clones with allelic insertions. Lane and clone columns refer to Figure 6.19. Ori, orientation: + indicates PB5 centromeric, PB3 telomeric.

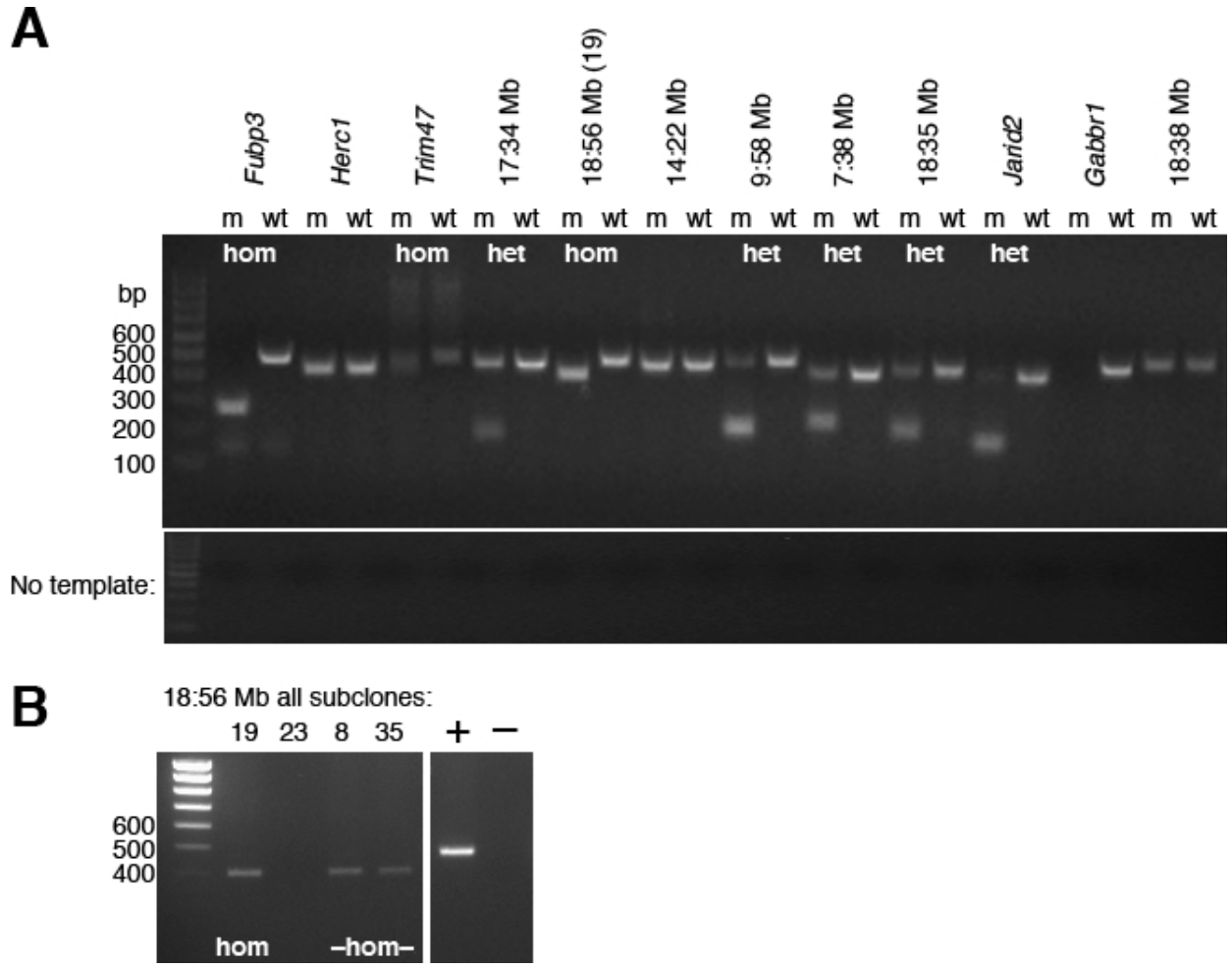

Figure 6.21: PCR genotyping of allelic LGNL1 subclones. A—PCR assays using three primers to detect mutant and wild type alleles for mapped allelic LGNL1 clones. m: DNA from isolated mutant. +/wt: wild type LGN DNA. The mutant product is smaller than the wild type in each case. Some PCRs do not amplify a product from the mutant, this could be due to incorrect mapping or primer incompatibility. Different primer sets and mutants are indicated by gene for insertions in or near a gene, and chromosome : position (Mb) for intergenic insertions. B—PCR results for additional subclones of clone b (18:56 from part A).

rection to the PGK promoter. This may be due to high levels of transcription or translation initiation on the appropriate strand, despite the presence of stop codons in the mutagen, which immediately precedes the 'unexpressed' resistance gene. There is accumulating evidence that a large proportion of the genome is transcribed to some degree (Cheng *et al.*, 2005; Gustincich *et al.*, 2006), and the presence of polyadenylation signals in the transcripts from the resistance genes would stabilise such transcripts. I added a counter-selection step with FIAU as an interim solution to this problem. However, in the long term the construct could be designed more intelligently—one such design would be to use a single polyadenylation site as well as a single promoter for both resistance genes, which would mean that any inappropriate transcripts of the resistance gene are unlikely to be stable. Such a construct has been developed, although not yet tested in a pooled format to my knowledge (K. Horie, J. Takeda *et al.*, unpublished).

These problems illustrate the difficulties arising from strong selection for a trait (double resistance) from a complex pool. In the clone-by-clone method, such 'rare' events as two-copy mutagenesis or locus-specific background are not serious, as even if they do occur, the other mutants are being cultured separately and are protected from contamination by the resulting mass of double resistant cells (see Figures 5.5 and 6.1). In the pooled format, these events adversely affect the complexity and usefulness of the double-selected libraries, as they result in potential double resistant cells present at the start. Therefore, the method needs to be refined to precisely target the event of interest (transposon copy number increase) and avoid the possibility of background completely—i.e. making these 'rare' events 'impossible'. My approach was to use FIAU counterselection mentioned above along with a novel cell-specific transposase to limit copy number in the mutagenesis step.

6.3.2 PB transposition and the cell cycle

Little is known about the cell cycle dependence of PB transposition, if any. My hypothesis for the transposon copy number increase post mobilisation is that transposition can occur after DNA is replicated. With this in mind, I modified the PB transposase by adding a degradation signal from the G1 and early S phase specific CDT1 protein. Although I was unable to characterise the expression pattern of the fusion protein directly using the reagents available, the fusion protein appeared to have the desired effect of limiting the transposon to one copy per cell after transposition from *Hprt* (Figure 6.16). This supports the idea that transposition to different loci after replication can result in cells with two non-allelic copies. Once again, this was not apparent in the clone-by-clone experiments, or rather clones with two non-allelic copies in this case were assumed to arise from repeated plasmid-to-genome transposition.

A side-by-side comparison of the PB-CDT1 fusion protein and the hyPBase protein that it is derived from indicates that the activity of the fusion, measured in terms of excision to give HAT resistance in LGP cells, is about half that of hyPBase (Figure 6.14B). A change of this magnitude is consistent with loss of expression in mid-S to G2, but is not formally separable from another effect of the CDT1 moiety (e.g. steric hindrance of the PBase).

6.3.3 Generation and uses of pooled mutant libraries

Library generation

The various library generation experiments that I have done have resulted in constant refinement of the protocol. A step-by-step experimental protocol, representing my current methods, is provided in the appendix to this thesis. This protocol incorporates the following elements that are required to obtain useful libraries:

- Use of *in vitro* transcribed mRNA to prevent integration of the transposase plasmid.
- Use of the PB-CDT1 fusion protein to limit initial copy number.
- FIAU counterselection to remove locus-specific selection background.
- Measurement of key parameters: starting clone number, average clonal expansion, Cre efficiency, double-resistant clone number.
- Characterisation of complexity and usefulness of libraries by Southern blot using the PB-Ccdc probe.

Some parameters still need to be fully optimised, particularly the expansion time to obtain sufficient complexity and a high proportion of homozygotes. This is currently based on theoretical consideration of the LOH rate, rather than empirical evidence.

The final experiments described in this chapter demonstrate a mutant library where 58% of double-resistant subclones are allelic mutants. These are

potential homozygotes, although some will still retain a wild type locus due to aneuploidy as described in Chapter 5. The results of PCR genotyping indicate that some are genuine homozygotes and some still retain the wild type allele, as for the clone-by-clone experiments. It is encouraging that all subclones were genuine mutants for the insertion site that was identified in several double resistant subclones (Figure 6.21). The library was of reasonable clonal complexity, with 16 different insertion sites identified from 26 double resistant subclones. This indicates that many clones are undergoing LOH. An expansion of 20 days was used in this experiment; a shorter expansion could reduce the redundancy still further, as discussed in Chapter 5. Of the 42% of the subclones that had two non-allelic insertions, almost all appeared to retain the $Hprt^{PB}$ donor locus. There may be further improvements to the protocol that could circumvent this. If these clones do indeed arise from aneuploidy *prior* to mobilisation, the cells could be sorted by DNA content prior to PBase transfection. Alternatively, if an arrayed library is to be made, a simple PCR screen that detects the unjumped locus could identify these clones. It may be possible to design a quantitative PCR (or Southern blot) assay to determine the relative amount of unjumped *Hprt* in a number of pooled libraries, and thus choose the best to use in screens.

Uses of pooled mutant libraries enriched for homozygotes

Libraries generated using this method are enriched for homozygous mutants by several orders of magnitude. In the expanded population prior to double selection, it is likely that there are of the order of 1,000 heterozygous cells for each homozygote. This is the population previously used for dominant screens in *Blm*-deficient cells. After my double selection procedure, clones with two allelic copies are readily visible in clones picked and analysed by Southern blot. Fifty-eight per cent were potential homozygotes. Therefore, depending on the actual percentage of real homozygotes compared to aneuploid cells, the proportion of useful cells is likely to be between half and one quarter of the library. This represents an enrichment of 250–500×.

I have documented a number of problems above, which mean that the enrichment is not complete. Is the level of enrichment obtained sufficient to use these libraries for genetic screens?

One obvious improvement is that the number of cells required for screening is vastly reduced. In an unenriched library, the order of 10^8 cells are typically screened (10,000 clones × 10,000 cells per clone; Guo (2004)). Using enriched libraries, screening the order of ten cells per clone should be sufficient (Figure 6.22). Thus, good genome coverage could be obtained by screening around 100,000 cells, which are easily accommodated in one well of a six-well plate, requiring only a few ml of medium. This may be important for some applications—for example, where resistance to dangerous or hard to obtain substances is being studied.

A second application of these libraries that I am interested in investigating is annotation of weak resistance or sensitivity phenotypes. If the library is grown for a period of time under stress or weak selection, mutants with a fitness advantage will increase their relative representation in the pool, and *vice versa*. Such screens have been successfully carried out in yeast and bacteria, and their extension to a mammalian system would greatly assist in functional annotation of the genome. As detailed in Chapter 3, Illumina sequencing can be used to determine the composition of mutants in a pool. Thus, sequencing pools of mutants expanded with or without stress or selection should allow the appropriate measurements to be made (Figure 6.22). This is only possible with enriched libraries, as the presence of a mutant is determined by insertion site sequencing, and therefore heterozygous cells will also contribute to the signal.

This method is likely to work best on libraries with short expansion times, as it depends on a high proportion of cells with each measured insertion site being homozygous mutants. If the library contains wild type retaining cells with the same insertion site as genuine homozygous mutants (see Chapter 5), these may mask an effect and lead to a loss in sensitivity. Although the mutant would not be recovered in this case, sequencing is very high throughput and should allow large libraries to be investigated. Multiple insertions per gene showing the same change in abundance would give confidence in the results. As many genes are now knocked out or have targeting vectors available (International Mouse Knockout Consortium *et al.*, 2007), methods that provide functional information are vital in prioritising further studies.

Finally, is this method suitable for the construction of an arrayed homozygous library, in a manner similar to the yeast deletion collection? Clearly in any such library there will be some 'junk', arising either from clones with non-allelic insertions, aneuploid clones that retain a wild type locus, or severe redundancy in the library (Figure 6.22). This needs to be taken into account when picking clones for an

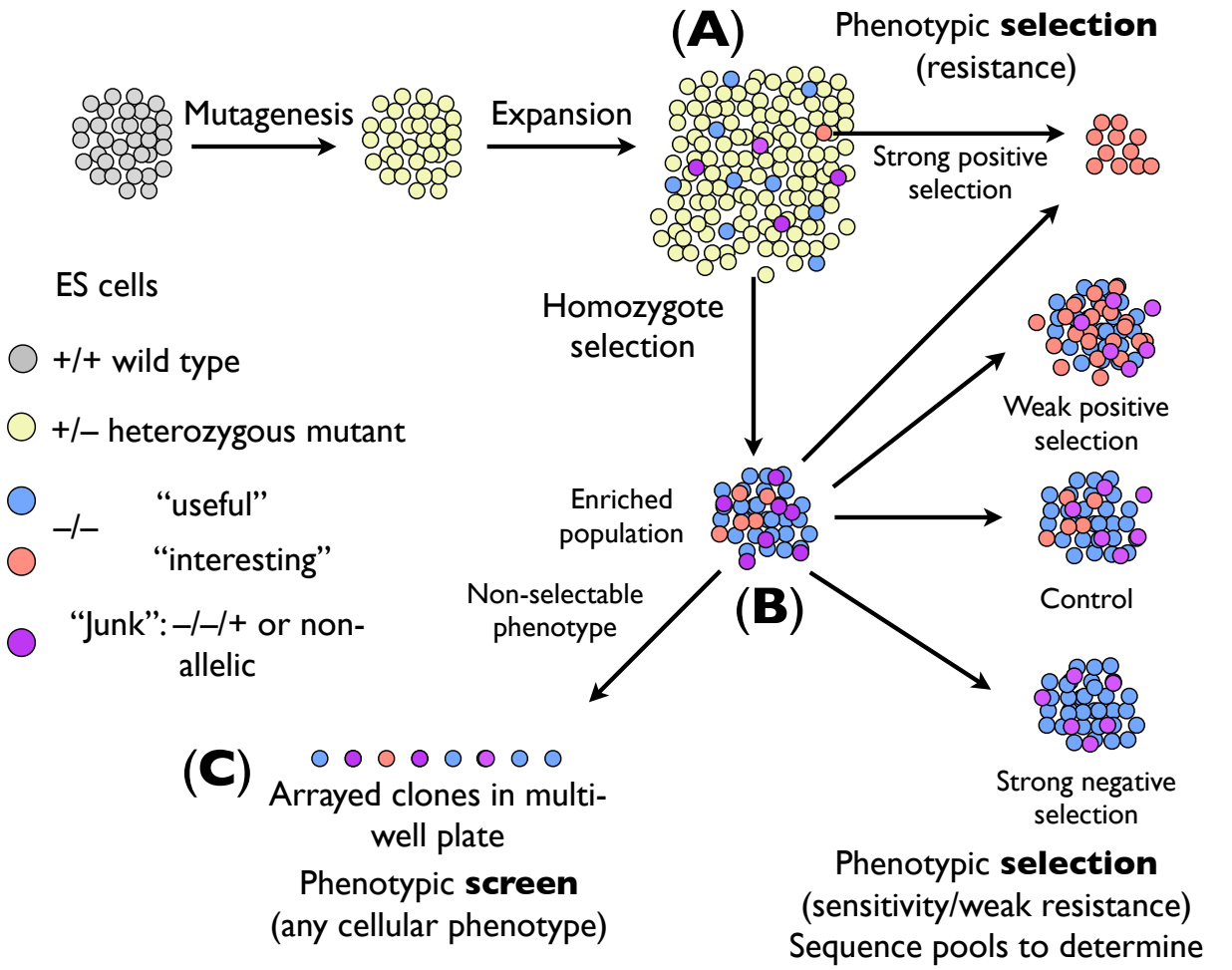

Figure 6.22: Uses of enriched libraries. A—Unenriched library, screening only by strong positive selection. B—Homozygote-enriched population isolated by copy number selection, screen by positive selection (fewer cells) or by sequencing more weakly or negatively selected pools (see text). C—Arrayed library made by subcloning homozygote-enriched population.

arrayed library, and the decision on how much junk to tolerate will be a logistical one. The method presented here to characterise the complexity and usefulness should be a guide in this regard. It appears that this may vary between experiments due to the stochastic nature of events that give rise to double resistant cells. Therefore I have taken care to isolate multiple subclones of the library generation cell lines, which can be used to generate many different libraries in parallel. These can then be characterised, and the best ones chosen to take further for screens.

6.3.4 Conclusions

There are several sources of background encountered when selecting for copy number increase on a genome-wide scale. These include copy number gain during transposition, and inappropriate expression of the second resistance gene in the construct. Copy number during transposition appears to be conserved when transposition is limited to the expression period defined by a CDT1 fragment, i.e. G1 and early S phase. This represents evidence that PB transposition can occur throughout the cell cycle. Using the $puro\Delta TK$ gene to select against inappropriate expression of the second gene eliminates this source of background. Several other adjustments to the library generation method can be used to reduce the background. The method described results in incomplete, but significant, enrichment of allelic two-copy mutants in the culture.