

# A *piggyBac* transposon-based genome-wide library of insertionally mutated *Blm*-deficient murine ES cells

Wei Wang, Allan Bradley,<sup>1</sup> and Yue Huang<sup>1</sup>

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom

Cultured mouse or human embryonic stem (ES) cells provide access to all of the genes required to elaborate the fundamental components and physiological systems of a mammalian cell. Chemical or insertional mutagenesis of *Blm*-deficient mouse ES cells can be used to generate genome-wide libraries of homozygous mutant ES cells, which are the substrates for conducting phenotype-driven loss-of-function genetic screens. However, the existing insertional mutation libraries are limited by incomplete genomic coverage. In this study, we have explored the use of *piggyBac* (PB) transposon-mediated mutagenesis to extend the genomic coverage of mutation libraries in *Blm*-deficient ES cells. A library composed of 14,000 individual gene-trap clones was generated and a recessive genetic screen conducted to identify cells with defects in DNA mismatch repair (MMR) genes. Independent mutations in all known genes of the pathway *Msh2*, *Msh6*, *Pms2*, and *Mlh1* were recovered in these screens. The genomic coverage in this library confirms its utility as a new genetic resource for conducting recessive genetic screens in mammalian cells.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

Phenotype-driven genetic screens in diploid mouse and human cell lines have been conducted using RNA interference (Boutros and Ahringer 2008) and chemically or insertionally mutated *Blm*-deficient mouse embryonic stem (ES) cells (Guo et al. 2004; Yusa et al. 2004; Wang and Bradley 2007). RNAi screens depend on the availability and efficient delivery of comprehensive libraries of tens of thousands of expression vectors. For an RNAi screen to be effective, each vector must reduce the level of expression of its specific target gene below a threshold that enables phenotypic selection, which is different for every gene. In contrast, screens in *Blm*-deficient ES cells utilize a single vector or mutagen followed by expansion of the pool of mutants to segregate homozygous mutant cells from a population of cells with heterozygous mutant alleles. A *Blm*-deficient background is required for this strategy because this elevates the rate of mitotic recombination by 20-fold, which is necessary for the generation of homozygous mutants from cells with heterozygous mutant alleles (Luo et al. 2000) at a practical rate. In principle, mutants of any gene expressed in *Blm*-deficient ES cells can be isolated from these libraries with an appropriate screening strategy.

Previously we have described two screens (Guo et al. 2004; Wang and Bradley 2007) conducted in *Blm*-deficient ES cells (Luo et al. 2000). These screens used sectorized libraries that were composed of 10,000 unique gene-trap mutations constructed with a reversible gene-trap retroviral vector (RGTV-1). Although these screens were successful, the complexity of mutations recovered suggested that the coverage of the library was substantially less than 10,000 genes. In the screen for DNA mismatch repair (MMR) genes, seven independent insertions in *Msh6* were recovered, while the screen for retroviral resistant mutants yielded five independent insertions in *Slc7a1*. Given that 12 independent mutations in two genes were isolated from a library of just 10,000

insertions, this implies that this library might contain mutations in 2000–3000 genes only. In addition, several of the expected targets, *Msh2*, *Mlh1*, and *Pms2*, from the DNA MMR screen were not recovered. It is known that retroviral gene trapping is very nonrandom, with some genes being highly represented while others are not found in very large data sets (Hansen et al. 2008). Thus, to conduct a saturating genetic screen in *Blm*-deficient ES cells, an insertional mutagen that provides comprehensive coverage of the genome with relatively low levels of saturation is required.

The resurrection of *Sleeping Beauty*, a Tc1-like transposon from salmonid fish that is active in human and mouse cells (Ivics et al. 1997) provides an alternative to retroviral mutagenesis. Although the *Sleeping Beauty* transposon is particularly effective as a somatic mutagen in the mouse (Collier et al. 2005; Dupuy et al. 2005) it is not active enough for ES cell mutagenesis (Luo et al. 1998). Recently, the *piggyBac* (PB) transposon isolated from the cabbage looper moth *Trichoplusia ni* has been shown to transpose efficiently in mammalian cells (Ding et al. 2005). This element appears to be highly efficient for germ line insertional mutagenesis in mice (Wu et al. 2007). The rates of *piggyBac* transposition in ES cells have also been shown to be very high (Wang et al. 2008). In this study we have explored the use of the *piggyBac* transposon to extend the genomic coverage of insertionally mutated libraries of *Blm*-deficient ES cells. We have generated a family of gene-trap vectors and used these to generate libraries of 14,000 mutant alleles. Screens of these libraries for DNA MMR mutants has resulted in recovery of independent mutations in all known members of the pathway, illustrating that these libraries have excellent genomic coverage.

## Results

### Construction of revertible PB-based gene-trap vectors

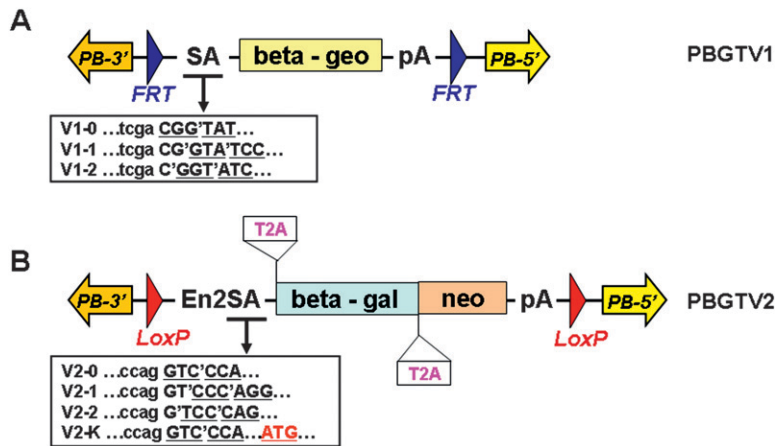
Two series of revertible PB transposon-based gene-trap vectors (PBGTV1 and PBGTV2; Fig. 1) were constructed. These contain the splice acceptor (SA) *Beta-geo* or *SA-T2A-Beta-gal-T2A-Neo* gene-trap cassette flanked by the 5' and 3' PB terminal DNA repeats (5' PBTR and 3' PBTR). *LoxP* or *FRT* sites were inserted at the both ends of

### <sup>1</sup>Corresponding authors.

E-mail [abradley@sanger.ac.uk](mailto:abradley@sanger.ac.uk); fax 01223-494714.

E-mail [yue@sanger.ac.uk](mailto:yue@sanger.ac.uk); fax 01223-494714.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.085621.108>. Freely available online through the *Genome Research* Open Access option.



**Figure 1.** *PiggyBac* revertible gene-trap vectors, PBGTV1 and PBGTV2. (A) The PBGTV1 series contain the minimal PB terminal DNA repeats and an adenovirus splice acceptor (SA) *Beta-geo* gene-trap cassette flanked by FRT sites. Three versions were constructed in which the coding sequence of *Beta-geo* has different reading frames. (B) The PBGTV2 series contain the minimal PB terminal DNA repeats and a *loxP* flanked gene-trap cassette consisting of an *En2* splice acceptor (En2SA) *T2A-Beta-gal-T2A-Neo* cassette. Four versions were constructed. In vectors 0, 1, and 2, the coding sequence of *T2A-Beta-gal-T2A-Neo* has different reading frames. In vector K, an ATG start codon was used for the *Beta-gal* coding sequence. pA, bovine growth hormone polyA signal; SA, adenovirus major late gene splice acceptor; En2SA, mouse *engrailed 2* gene splice acceptor; T2A peptide is a self-cleaving short peptide from Picornavirus (Szymczak et al. 2004).

the gene-trap cassette, so that the integrated vectors could be excised with Cre or Flp site-specific recombinases. Excision can also be achieved with PBase, which has the advantage of precisely removing the transposon without leaving a footprint (Thibault et al. 2004; Wang et al. 2008). As the majority of gene-trap insertions are likely to be located in an intron, excision of the transposon would be expected to reverse the mutant phenotype, provided the insertion was responsible for the phenotype. Previous studies (Cary et al. 1989) have reported that the 5' PBTR has potential promoter activity. To avoid its possible influence on the gene-trap cassette, in all vectors of this study the gene-trap cassettes were placed in opposite orientation to the 5' PBTR. To maximize the coverage of genome, each vector series consisted of three or four different versions, in which the coding sequence of *Beta-geo* or *T2A-Beta-gal-T2A-Neo* is designed to be compatible with splicing from exons with different reading frames (Fig. 1).

### Generation of a PB-based genome-wide mutation library

These vectors were individually electroporated into ES cells with or without a PBase expression plasmid, and cells with gene-trap insertions were selected in G418. In the absence of transposase very few G418 resistant colonies were generated, resulting from random insertions of the vector in the genome. In contrast, co-electroporation with PBase generated thousands of G418 resistant colonies (Fig. 2A). The increased number of G418 resistant clones suggests that transposition has occurred. This was confirmed by analysis of the transposon-host junctions by Splinkerette-PCR (SpPCR) followed by sequence analysis (Mikkers et al. 2002), which identified the characteristic features of transposition, namely integration as single units into TTAA sites with no loss of sequence at the host/vector junctions ( $n = 96$ ; data not shown).

A genomic library ideally contains one insertion per cell. To establish conditions to achieve this, the relationship between the quantity of the transposon gene-trap donor plasmid electro-

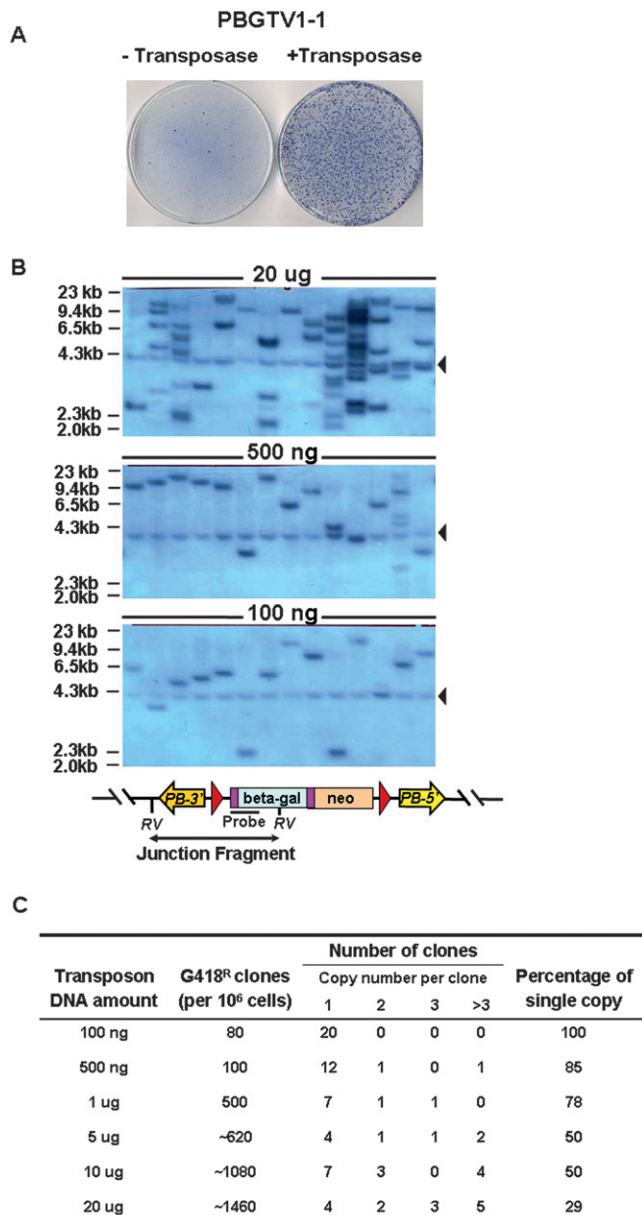
porated and the resulting genomic copy number was titrated. In these transfections, the transposase was maintained at a constant level while the transposon concentration was varied. As expected the transfection efficiency was proportional to the DNA concentration. The transposon copy number was determined by Southern blot analysis on sets of individual clones from each condition, using a restriction enzyme which generates a unique transposon-host junction fragment for each transposon insertion site (Fig. 2B). The copy number analysis is summarized in Figure 2C. At the highest DNA concentration most clones have multiple insertions, while at low DNA concentrations single-copy insertion can be achieved, although the transfection efficiency is also much lower.

The scheme used to establish a sectorized mutation library is illustrated in Figure 3. The seven vectors from the PBGTV1 and PBGTV2 series were individually transfected into NGG5-3 ES cells at a low DNA concentration so that the majority of clones in each pool

should have single allele mutations. Two pools were established for each vector (14 pools in total), each consisting of ~1000 independent G418 resistant gene-trap clones. Each pool was expanded for a minimum of 14 population doublings to facilitate the segregation of homozygous mutants before further analysis.

To explore the genomic coverage of these libraries we picked 576 clones from two pools established with the PBGTV1 and PBGTV2 transposons and identified the host-transposon junction in 318 of these by Splinkerette-PCR (SpPCR) followed by sequence analysis (Supplemental data set 1). A total of 217 insertions (68%) were uniquely mapped to 212 different Ensembl annotated genes (Mouse Build NCBI m37, Ensembl version 51), which were distributed throughout the genome, consistent with previous observations (Ding et al. 2005; Wang et al. 2008). The coverage of these libraries was compared with the large set of retroviral gene traps in the TIGM OmniBank collection of 276,372 insertion sites available in the International Gene-trap Consortium (IGTC) database (Skarnes et al. 2004; Hansen et al. 2008). Sixteen of these 212 genes (8%) do not appear in the OmniBank collection, and eight appear just once. Thus, even in this small sample set these PB vectors are accessing areas of the genome that are not available even from very large collections of retroviral gene-trap libraries.

We looked for evidence of *piggyBac* insertion hot spots, a known biological limitation in libraries generated with retroviral gene-trap vectors. Of the 212 genes, independent insertions were identified in five (*Smg7*, *Cugbp1*, *Prkcbp1*, *Btd11*, and *Ano6*, shown in Supplemental Table 2). These genes are all large targets, ranging in size from 66 kb to 274 kb, and spanning a total genomic space of 671 kb (0.17% of the "trappable genome" in ES cells). Chi-squared analysis to test the null hypothesis that independent insertions in these five genes had occurred by chance is rejected ( $P < 0.001$ ). Thus, this clustering is caused by the selection imposed for gene-trap events, insertion bias resulting from biological features of the transposase, or the chromatin structure of these particular genes.



**Figure 2.** Copy number analysis of PB insertions. (A) PB transposition in the presence of transposase. Five micrograms of transposon gene-trap vector (PBGTV1-1) were electroporated with or without 10  $\mu$ g of transposase expressing plasmid into  $5 \times 10^6$  AB2.2 cells. A representative picture of plates obtained from G418 selection. (B) Southern blot analysis of gene-trap clones reveals host/transposon junction fragment in each clone. RV, EcoRV. LacZ, probe, an 800-bp Clal fragment from pSAgeo (Soriano et al. 1991). Black triangles indicate the endogenous neomycin locus in NGG5-3 ES cells. (C) Summary of copy numbers obtained using different transfection conditions at a single transposase concentration, 20  $\mu$ g.

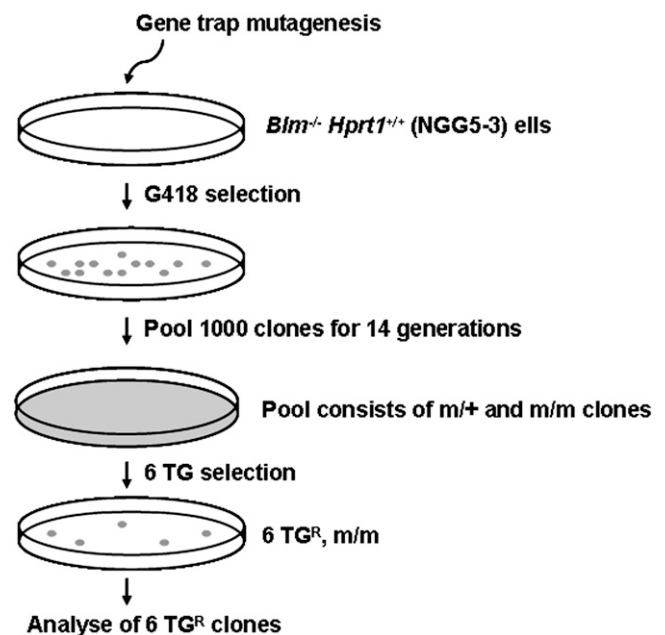
### Screening and analysis of 6-TG-resistant mutants

The coverage of the pools and the efficiency of the mutagenesis was functionally assessed by screening for mutations in DNA mismatch repair (MMR) genes, since the selection is direct and several known MMR genes should be recovered from this screen. The 14 pools were selected in 6-thioguanine (6-TG) and resistant

clones were observed in nine pools. To investigate the relationship among clones in the same pool, the transposon-host junction fragments were cloned and sequenced. This analysis identified pool specific junction fragments verifying that clones from the same pool were daughters derived from a small number (1–4) of founders. A total of 14 independent clones were recovered (Supplemental Table 3), five of which had PB insertions in known MMR genes: *Msh2*, *Msh6*, *Pms2*, and *Mlh1* (Fig. 4A). Each clone was examined with gene and host-transposon-specific PCR to determine if the mutation was homozygous or heterozygous. Four of five insertions in the MMR genes (*Msh2*-m1, *Pms2*-m1, *Mlh1*-m1, and *Msh6*-m1) were homozygous (Fig. 4B), while *Mlh1*-m2 and others were heterozygous and were not examined further.

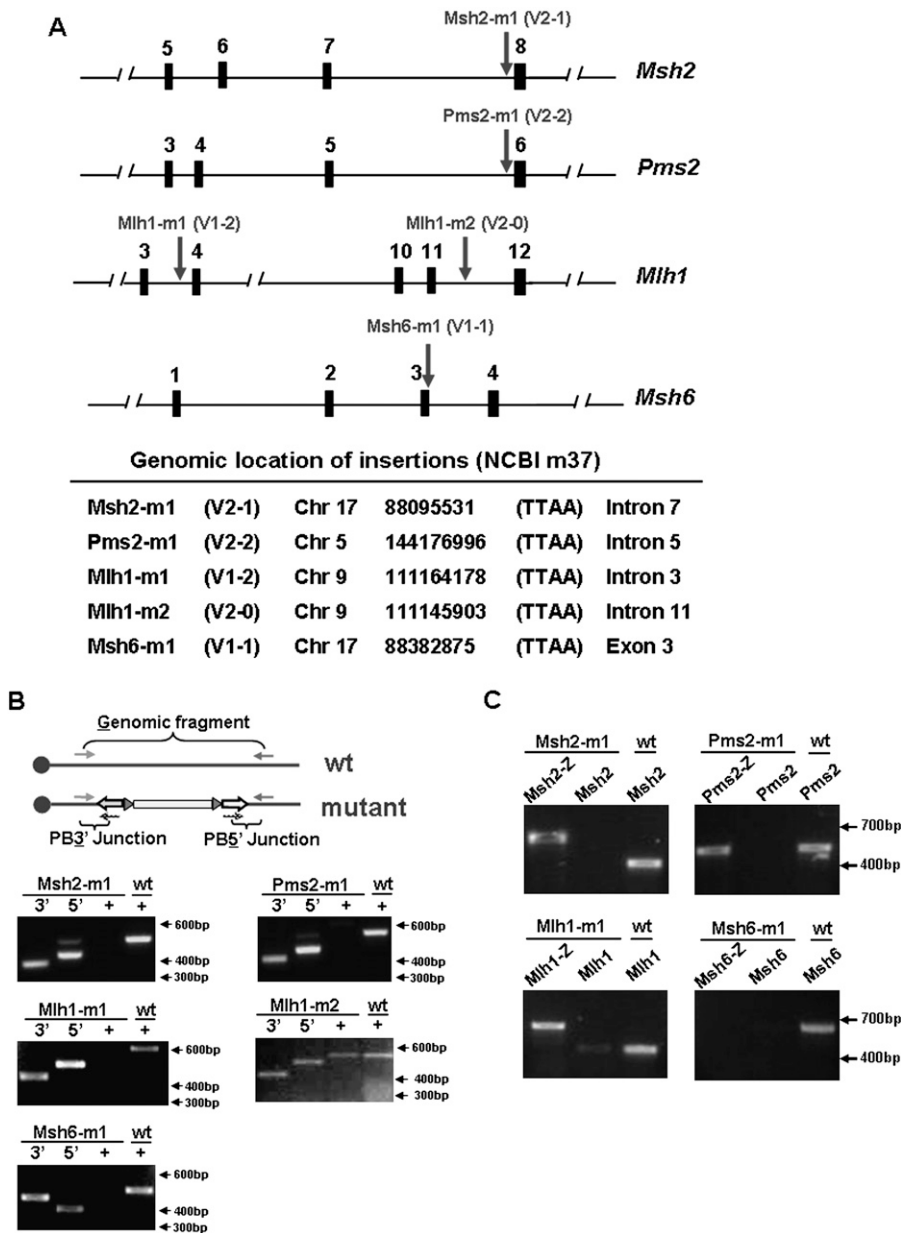
In all but one mutant, the gene-trap cassette had inserted into an intron of the trapped gene, in the appropriate transcriptional orientation and into the transposase TTAA target site. However, the insertion in *Msh6*, was in a TTAA site in exon 2 opposite to the orientation of transcription which will disrupt the normal *Msh6* transcript. To understand how this clone survived G418 selection for gene-trap events, 5' RACE analysis was performed which identified a *Beta-gal* transcript fused to exon 22 of the *Fbxo11* gene, which lies adjacent to *Msh6* but is transcribed in the opposite orientation (Supplemental Fig. 1).

The effect of the homozygous transposon insertions on the expression of MMR genes was assessed by reverse transcription (RT)-PCR using primer pairs specific for the exons, which spanned



**Figure 3.** Strategy to generate and isolate 6-TG resistant clones. *Blm*<sup>-/-</sup>, *Hprt1*<sup>+/+</sup> (NGG5-3) cells were used to generate gene-trap mutations by transposon gene-trap mutagenesis. The gene-trapped clones were selected with G418, pooled and expanded for 14 generations. After expansion, each pool, which consists of heterozygous and homozygous clones, was screened using 6-TG, and resistant clones were picked for further analysis. 6-Thioguanine (6TG) is utilized by the purine nucleotide synthesis pathway and is incorporated into replicating DNA leading to mismatches within the DNA duplex. The MMR complex recognizes this type of mismatch and initiates multiple rounds of repair, which generates single- and double-strand DNA breaks. These activate signaling pathways and lead to cell cycle arrest and apoptosis. Thus, 6-TG is a sensitive selective agent for the isolation of DNA MMR mutants.





**Figure 4.** Analysis of 6-TG<sup>R</sup> clones. (A) Insertion sites of transposons in four DNA mismatch repair genes shown by arrows. The insertion–host junctions were identified by SpPCR and sequence analysis. The vector and reading frames are indicated. (B) Homozygosity analysis of five mutant clones; Msh2-m1, Pms2-m1, Mlh1-m1, Mlh1-m2, and Msh6-m1. Genomic PCR analysis was conducted with primers for the PB3' and PB5' transposon/host junctions (mutant alleles, marked as "3'" and "5'", respectively) or with genomic primers which span the insertion site (wild-type alleles). The wild-type genomic fragments "+" are absent in gene-trap mutant clones but were amplified from Blm-deficient ES cells (marked as "wt"), which indicated that transposon insertions are homozygous in all cases except Mlh1-m2. (C) Wild-type and fusion transcripts in gene-trap mutants. RT-PCR was performed with primers for the upstream exon of the relevant mutated gene to *lacZ* gene (marked as "Msh2-Z, Pms2-Z, Mlh1-Z, and Msh6-Z") as well as for the wild-type alleles. The wild-type transcripts are detected in nonmutated Blm-deficient control samples (marked as "wt") but were absent in all four mutant clones. Fusion transcripts of the expected size were detected in three mutant clones, Msh2-m1, Pms2-m1, and Mlh1-m1.

the insertion site. In all four clones, the endogenous transcript could not be detected (Fig. 4C). The expected *lacZ* fusion transcript could be detected in the three clones with intronic insertions; however, a *Msh6-lacZ* fusion transcript was not detected in the

clone with the insertion in exon 2, consistent with the 5' RACE result mentioned above.

#### Excision of PB by PBase re-expression

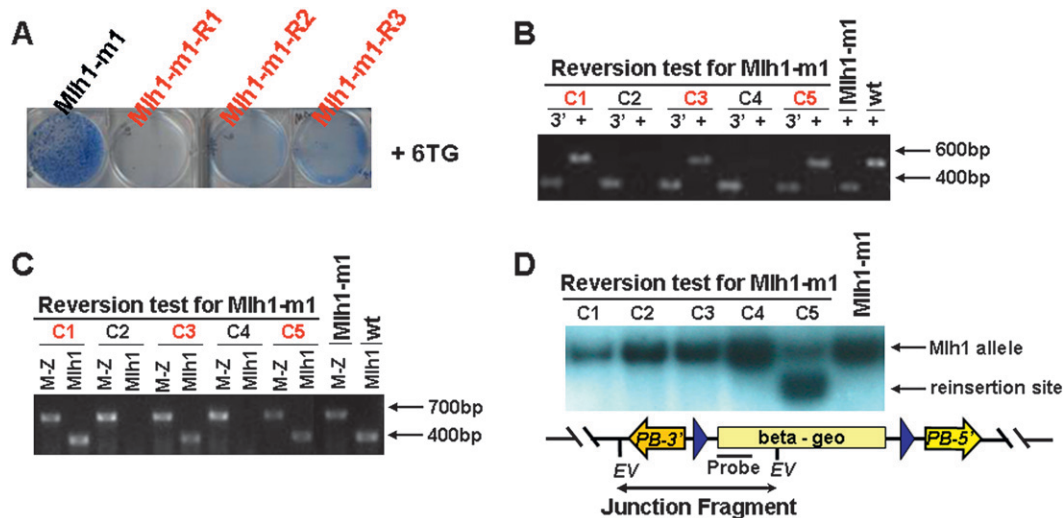
To establish causality between the transposon insertion and a mutant phenotype these transposons were designed to be reversible. The gene-trap cassette can be removed with a site-specific recombinase, *Flp* (PBGTV1), *Cre* (PBGTV2), or the entire transposon can be removed with PBase. Recombinase-mediated excision of the gene-trap cassette would leave the PB inverted repeats in the locus. In the majority of cases this residual element will be located in an intron and is not expected to interfere with normal splicing. For insertions in exons, such as the one in Msh6-m1, reversion can be accomplished with PBase because of precise excision of PB.

To confirm the reversibility of these alleles with PBase, the Mlh1-m1 mutant clone was transiently transfected with a PBase expression plasmid. Single-cell clones were isolated and sib-selected to test their sensitivity to 6TG. Three out of 96 tested clones were 6TG sensitive and their genotype was confirmed to have reverted to heterozygosity by genomic PCR analysis (Fig. 5A,B). RT-PCR analysis for *Mlh1* revealed that these clones were positive for *Mlh1* expression, as expected (Fig. 5C). To confirm that these were true revertants we looked for evidence of mobilization of the transposon. One of the revertant clones had acquired a new allele (Fig. 5D), which is consistent with the 40% reinsertion rate for PB (Wang et al. 2008).

#### Discussion

One factor which determines the success of a dominantly selected phenotypic screen conducted in mammalian cells is the sensitivity and specificity of the selection system deployed to isolate the mutant clones. Other important variables are the degree of saturation of the genome achieved with the mutagen, a low level or absence of spontaneous background mutations, and the ability to causally relate the mutation with the selected phenotype. The malleability of the ES cell enables configuration of the genome to facilitate selection of mutants in

diverse pathways. However, engineering an ES cell genome to enable selective isolation of mutants is usually conducted before the construction of a library. Thus, reliable and robust technologies to repetitively generate libraries with good coverage of the



**Figure 5.** Reversibility of Mlh1-m1 mutation with PBase. (A) Low-density survival assay demonstrating 6-TG sensitivity of three revertant clones. (B) Genomic PCR analysis of 96 reversal candidates with primers for the PB3' IR and sequences upstream of insertion site of *Mlh1* (marked as "3'") detected the transposon/host junction of the expected size. Representative candidate clones including three reverted ones are shown. The genomic PCR product fragments (marked as "+") are absent in two mutant clones (C2 and C4) but were amplified from three revertants (C1, C3, and C5). (C) Expression of wild-type *Mlh1* and fusion *Mlh1-lacZ* transcripts in gene-trap mutants. RT-PCR with primers for exon2 of *Mlh1* and *lacZ* (marked as "M-Z") detected fusion transcripts of the expected size from all 96 clones. The wild-type transcripts (marked as "Mlh1") are absent in clones C2 and C4 but are found in the three revertants (C1, C3, and C5). (D) Southern blot analysis of Mlh1-m1 revertant clones identifies a new integration of the transposon in revertant clone 5. EV, EcoRV. Probe, LacZ.

genome are required. RNAi libraries offer one solution, although there remain significant logistical issues with handling complex mixtures of tens of thousands of different expression vectors. Mutagenesis in Blm-deficient ES cells (Guo et al. 2004; Yusa et al. 2004; Wang and Bradley 2007) provides an alternative, which takes advantage of established and reliable mutagenesis methods.

The most direct way to identify the causal genetic lesion in Blm-deficient ES cell screens has been to use an insertional mutagen delivered by a retroviral vector such as RGTV-1 (Guo et al. 2004). Although this has led to the successful identification of known and previously unknown genes required for DNA MMR and retroviral infection cells (Guo et al. 2004; Wang and Bradley 2007), it has become apparent that there was a substantial bias in the genome coverage in these libraries. Analysis of the very large retroviral gene-trap insertion site data sets illustrates the degree of this bias. The largest resource termed TIGM OmniBank II, contains more than 350,000 retroviral ES cell gene-trap clones (Hansen et al. 2008). Although mutations in 10,433 unique genes can be found in this library, the density of insertions in each gene is very uneven. For instance, 27% of the genes have only a single insertion, while hundreds of genes each have hundreds of independent insertions. Although the total coverage of the expressed genome in ES cells is represented in this library, it is not practical to generate 350,000 gene-trap clones for a screen.

In this study, we built a series of PB transposon-based gene-trap vectors and used these to generate sectored libraries of mutants in Blm-deficient ES cells. The total library was composed of 14,000 individual gene-trap clones, the genomic coverage of which was assessed by cloning the insertion sites in 217 clones from two vectors. Analysis of these insertion sites revealed that 8% of these genes were not in the OmniBank II data set and another 4% occur just once. Thus, these PB vectors are accessing genes that have not been touched by 20-fold saturating retroviral gene-trap mutagenesis.

The analysis of the 217 *piggyBac* insertions revealed a pattern that was predominantly random, but five genes contained two insertions each, indicating that *piggyBac* gene-trap mutagenesis also has hot spots. We compared the five genes with two insertions in our study with a previous set of 467 *piggyBac* gene traps (Wang et al. 2008). The five genes with two PBGTV hits were not found in this data set, although four different genes (*Ckap5*, *Dnajc5b*, *Sertad2*, and *Cog5*) were trapped more than once (Supplemental Table 2). The nine genes with two independent *piggyBac* insertions have highly variable retroviral insertions in the TIGM OmniBank II library, ranging from 0 to 252 independent insertions. To further explore the relationship between the *piggyBac* and retroviral gene-trap insertions hot spots we examined the representation of the 207 genes with one PBGTV insertion with the retroviral insertions in OmniBank II. Excluding the 16 PB insertions which are not found in OmniBank II, the average number of retroviral gene-trap clones is 26, which is broadly consistent with the saturation of this resource. However, several of these genes are retroviral hotspots, for instance, *Pig1* has 485 insertions and *Nmnat2* has 248 insertions. Although there is some evidence of overlapping hot spots, we can conclude that the coverage is different.

We have functionally assessed the coverage of these libraries by screening for DNA MMR mutants. Mutations in four known MMR genes were identified in these screens, which indicated that there was greater genome coverage compared with the library generated with RGTV-1 which had a similar number of unique insertion sites. In our previous screen using RGTV-1, seven independent homozygous mutations in *Msh6* and two independent mutations in *Dnmt1* were recovered but mutations in three known MMR related genes, *Msh2*, *Pms2*, and *Mlh1*, were not identified. Our failure to recover insertions in these genes is consistent with the fact they appear to be cold spots for retroviral gene-trap mutagenesis. Within the IGTC database of 350,000 retroviral gene-trap ES cell clones *Msh6* and *Dnmt1* have 55 and 43 independent

insertions, respectively, while *Mlh1* and *Pms2* are represented by just three and three clones each. Insertions in *Msh2* are found 31 times; thus, these mutations are occurring approximately once in every 10,000 clones, which would explain their absence from our original library of this size. In contrast, screens of the PB libraries identified mutations in four different genes among the five unique mutants analyzed.

In this report, we have described two series of gene-trap vectors, each consisting of three or four different versions, in which the coding sequence of *Beta-geo* or *T2A-Beta-gal-T2A-Neo* has different reading frames. Genes, which have one or a few introns, can generate a functional gene trap by a vector with the correct reading frame or by the use of an internal ribosome reentry site (IRES). Internal reinitiation is not always efficient enough to provide adequate levels of G418 resistance for selection. Thus, the combination of these vectors will increase the coverage of genome.

Using Cre-*loxP* or Flp-*FRT*-mediated recombination, the integrated gene-trap cassette can be removed, leaving a single *loxP* or *FRT* site and part of mutagen fragment at the insertion site (Araki et al. 1997; Schaft et al. 2001). Although the residual LTR or PB repeats are believed not to affect gene function in the majority of cases, if the vector was inserted into a 5' UTR region, a promoter region or an exon, the remaining fragment may interfere with the expression of the host gene, resulting in a nonreversible phenotype. The *piggyBac* gene trap vectors are, however, reversible following re-expression of the transposase, which mobilizes the element out of the locus leaving the locus in an intact wild-type sequence configuration (Cary et al. 1989; Wang et al. 2008). Mobilization can be further demonstrated by confirming a new transposon insertion site in the genome of the reverted cell.

The molecular tag provided by the inserted mutagenesis in the gene-trap mutations allows high throughput determination of genomic integration sites, which are most efficiently determined by splinkerette PCR (Hom et al. 2007). SpPCR from retroviral repeats is complicated by the many endogenous retroviral LTRs in the mouse genome. Since there is no analog of the PB transposon in the mouse genome, the SpPCR method is highly specific and reliable, which supports high throughout applications of this technique.

## Methods

### Cell culture

ES cells were cultured on a layer of mitotically inactive SNL76/7 feeder cells and transfected by electroporation as described previously (McMahon and Bradley 1990).

### *Hprt1*<sup>+/+</sup>, Blm-deficient ES cells

The NGG5-3 Blm-deficient ES cell line been described previously (Guo et al. 2004). This Blm-deficient ES cell line is derived from the AB2.2 ES cell line, which carries a loss-of-function mutation of the X-linked *Hprt1* gene. However, HPRT1 is required to convert non-toxic 6-TG into an analog which can be incorporated into DNA, which then serves as a substrate for DNA mismatch repair. To generate HPRT1-positive Blm-deficient cells, two copies of *Hprt1* were targeted into both alleles of the *Gdf9* locus on chromosome 11 in Blm-deficient (m3/m4) cells (Luo et al. 2000).

### Gene-trap vector construction

The *piggyBac* transposon vector containing 313 bp of the 5' inverted terminal DNA repeat (TR) and 235 bp of 3' inverted TR was pro-

vided by Dr. Xiaozhong Wang (Northwestern University). The *piggyBac* 5' and 3' inverted repeats were PCR amplified and cloned upstream and downstream from the *FRT* flanked *SA-Beta-geo* or the *loxP* flanked *En2SA-T2A-Beta-gal-T2A-Neo* gene-trap cassettes (Skarnes et al. 1992; Szymczak et al. 2004), and named PBGTV1 and PBGTV2, respectively. The sequence of these vectors is available from GenBank under accession numbers FJ495177 (PBGTV1-0), FJ495178 (PBGTV1-1), FJ495179 (PBGTV1-2), FJ493252 (PBGTV2-0), FJ493253 (PBGTV2-1), FJ493254 (PBGTV2-2), and FJ493255 (PBGTV2-K).

### Generation of gene-trap mutations

A single-cell clone of NGG5-3 was isolated and expanded to minimize background mutations. Five million cells were coelectroporated with 1  $\mu$ g of each gene-trap vector and 20  $\mu$ g of a PBase expression plasmid pCAGGS-PBase, in which a native *piggyBac* transposase is under the control of the CAG promoter. After electroporation the cells were plated onto 90-mm feeder plates and G418 selection (180 mg/L active ingredient) was initiated 24 h later and continued for 8 d until individual ES cell clones were visible. The G418 resistant clones within each plate were pooled and expanded for 4 d before initiating 6-TG selection.

### Screening for 6-TG-resistant ES cells

Fifteen million cells were seeded per 150-mm feeder plate and 2  $\mu$ M, 2-amino-6-mercaptapurine (6-TG, Sigma) selection was initiated 24 h after plating and continued for 8 d. 6-TG resistant clones were allowed to grow for 4 more days without selection before colonies were picked and expanded for molecular analysis.

### Isolation of the transposon–chromosome junction

Isolation of the transposon–chromosome junction was performed using the Splinkerette-PCR method as described (Mikker et al. 2002). The protocol was modified for use with ES cells on 96- or 24-well plates. Briefly, genomic DNA was isolated from ES cell colonies on 96- or 24-well plates; 2–4  $\mu$ g DNA was digested with Sau3AI and ligated with the corresponding Splinkerette adaptors HMSp–Sau3AI (generated by annealing Splinkerette oligos HMSpBb–Sau3AI with HMSpAa). A first round of PCR was carried out with Splinkerette primer HMSp1 and PB transposon primers PB5'-1 or PB3'-1. Then 1% volume of the PCR product was directly used for second-round nested PCR, which was carried out with Splinkerette primer HMSp2 and the PB transposon primers PB5'-2 or PB3'-2. The nested PCR products were purified by High Pure 96 UF cleanup Plate (Roche) and were used for sequencing with primer PB5'-seq or PB3'-seq separately. Primers for Sp-PCR and sequencing are shown in Supplementary Table 4.

### Rapid isolation of the 5' end of gene-trap transcripts by 5'-RACE

5'-RACE reactions were performed with the 5'-RACE system (Roche). In brief, RNA was isolated from mutant clone Msh6-m1 cultured in one well of a 24-well plate, and 5  $\mu$ g of total RNA was used for generating the first-strand cDNA. Then the single-strand cDNA was tailed by PolyC tailing at 37°C. By using the tailed cDNA as a template, first and second rounds of PCR were performed. The lacZ primers and 5' RACE universal primers for the 5'-RACE reaction are provided in Supplemental Table 3. The PCR product of the mutant Msh6-m1 fusion transcript from *lacZ* to 5' end was TA cloned for sequencing.

## PBase reversal

Five million cells were electroporated with 20 µg *piggyBac* transposase expression plasmid, plated on 90-mm feeder plates, and 2 d later the cells were replated at low density (1000 cells/10-cm plate) to allow the growth of individual ES cell colonies. Clones were picked into 96-well plates and the deletion of the PB transposon was confirmed by PCR. 6-TG resistance was assessed by low-density plating ( $10^4$  cells/well) in 24-well plates. After selection for 8 d, clones were visualized by staining with 2% methylene blue in 70% ethanol.

## Acknowledgments

We thank Dr. Xiaozhong Wang of Northwestern University for providing the *piggyBac* transposon vector and PB transposase expressing plasmid; Dr. Ge Guo and Dr. Pentao Liu for their comments on this manuscript; Dr. Zikai Xiong, Frances Law, and Alistair Beasley for help with tissue culture; and Wendy Mustill and Patrick Biggs for technical assistance in 5' RACE. Y.H. acknowledges support from the Royal Society UK and National Basic Research Program of China (2005CB522507). This work was supported by the Wellcome Trust and Sanger Institute Grant number 79643.

## References

- Araki, K., Imaizumi, T., Okuyama, K., Oike, Y., and Yamamura, K. 1997. Efficiency of recombination by Cre transient expression in embryonic stem cells: Comparison of various promoters. *J. Biochem.* **122**: 977–982.
- Boutros, M. and Ahringer, J. 2008. The art and design of genetic screens: RNA interference. *Nat. Rev. Genet.* **9**: 554–566.
- Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., and Fraser, M.J. 1989. Transposon mutagenesis of baculoviruses: Analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**: 156–169.
- Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**: 272–276.
- Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T. 2005. Efficient transposition of the *piggyBac* (PB) transposon in mammalian cells and mice. *Cell* **122**: 473–483.
- Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. 2005. Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* **436**: 221–226.
- Guo, G., Wang, W., and Bradley, A. 2004. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* **429**: 891–895.
- Hansen, G., Markesich, D., Burnett, M., Zhu, Q., Dionne, K., Richter, L., Finnell, R., Sands, A., Zambrowicz, B., and Abuin, A. 2008. Large-scale gene trapping in C57BL/6N mouse embryonic stem cells. *Genome Res.* **18**: 1670–1679.
- Hom, C., Hansen, J., Schnütgen, F., Seisenberger, C., Floss, T., Irgang, M., De-Zolt, S., Wurst, W., von Melchner, H., and Noppinger, P.R. 2007. Splinkerette PCR for more efficient characterization of gene trap events. *Nat. Genet.* **39**: 933–934.
- Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**: 501–510.
- Luo, G., Ivics, Z., Izsvak, Z., and Bradley, A. 1998. Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **95**: 10769–10773.
- Luo, G., Santoro, I.M., McDaniel, L.D., Nishijima, I., Mills, M., Youssoufian, H., Vogel, H., Schultz, R.A., and Bradley, A. 2000. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet.* **26**: 424–429.
- McMahon, A.P. and Bradley, A. 1990. The *Wnt-1 (int-1)* proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**: 1073–1085.
- Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., and Berns, A. 2002. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nat. Genet.* **32**: 153–159.
- Schaft, J., Ashery-Padan, R., van der Hoeven, F., Gruss, P., and Stewart, A.F. 2001. Efficient FLP recombination in mouse ES cells and oocytes. *Genesis* **31**: 6–10.
- Skarnes, W.C., Auerbach, B.A., and Joyner, A.L. 1992. A gene trap approach in mouse embryonic stem cells: The lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes & Dev.* **6**: 903–918.
- Skarnes, W., von Melchner, H., Wurst, W., Hicks, G., Nord, A., Cox, T., Young, S., Ruiz, P., Soriano, P., Tessier-Lavigne, M., et al. 2004. A public gene trap resource for mouse functional genomics. *Nat. Genet.* **36**: 543–544.
- Soriano, P., Friedrich, G., and Lawinger, P. 1991. Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J. Virol.* **65**: 2314–2319.
- Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A. 2004. Correction of multi-gene deficiency in vivo using a single “self-cleaving” 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22**: 589–594.
- Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M., Buchholz, R., Demsky, M., Fawcett, R., Francis-Lang, H.L., et al. 2004. A complementary transposon tool kit for *Drosophila melanogaster* using *P* and *piggyBac*. *Nat. Genet.* **36**: 283–287.
- Wang, W. and Bradley, A. 2007. A recessive genetic screen for host factors required for retroviral infection in a library of insertionally mutated *Blm*-deficient embryonic stem cells. *Genome Biol.* **8**: R48. doi: 10.1186/gb-2007-8-4-r48.
- Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. 2008. Chromosomal transposition of *PiggyBac* in mouse embryonic stem cells. *Proc. Natl. Acad. Sci.* **105**: 9290–9295.
- Wu, S., Ying, G., Wu, Q., and Capecchi, M.R. 2007. Toward simpler and faster genome-wide mutagenesis in mice. *Nat. Genet.* **39**: 922–930.
- Yusa, K., Horie, K., Kondoh, G., Kouno, M., Maeda, Y., Kinoshita, T., and Takeda, J. 2004. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**: 896–899.

Received August 28, 2008; accepted in revised form January 22, 2009.