

# SCIENTIFIC REPORTS



OPEN

## PiggyBac transposon-based polyadenylation-signal trap for genome-wide mutagenesis in mice

Limei Li<sup>1,2,3,\*</sup>, Peng Liu<sup>1,4,\*</sup>, Liangliang Sun<sup>5,\*</sup>, Bin Zhou<sup>2</sup> & Jian Fei<sup>1,6,7,8</sup>

Received: 13 November 2015

Accepted: 23 May 2016

Published: 13 June 2016

We designed a new type of polyadenylation-signal (PAS) trap vector system in living mice, the piggyBac (PB) (PAS-trapping (EGFP)) gene trapping vector, which takes advantage of the efficient transposition ability of PB and efficient gene trap and insertional mutagenesis of PAS-trapping. The reporter gene of PB(PAS-trapping (EGFP)) is an EGFP gene with its own promoter, but lacking a poly(A) signal. Transgenic mouse lines carrying PB(PAS-trapping (EGFP)) and protamine 1 (Prm1) promoter-driven PB transposase transgenes (Prm1-PBase) were generated by microinjection. Male mice doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBase were crossed with WT females, generating offspring with various insertion mutations. We found that 44.8% (26/58) of pups were transposon-positive progenies. New transposon integrations comprised 26.9% (7/26) of the transposon-positive progenies. We found that 100% (5/5) of the EGFP fluorescence-positive mice had new trap insertions mediated by a PB transposon in transcriptional units. The direction of the EGFP gene in the vector was consistent with the direction of the endogenous gene reading frame. Furthermore, mice that were EGFP-PCR positive, but EGFP fluorescent negative, did not show successful gene trapping. Thus, the novel PB(PAS-trapping (EGFP)) system is an efficient genome-wide gene-trap mutagenesis in mice.

The Human Genome Project (HGP) was a large-scale project initiated at the end of the 20<sup>th</sup> century with a focus on gene sequence analysis<sup>1–3</sup>. The HGP had a revolutionary impact on the life sciences and made a significant contribution to the industrialization of modern biomedicine<sup>4–6</sup>. With the sequencing of the genomic DNA of humans and many other organisms, comprehensive studies on gene function, including functional genomics studies, have become an important frontier of life science<sup>7–10</sup>.

Primary techniques in functional genomics include the use of model biological systems with genomic structure similar to that of humans, ontogeny analysis, studies of the development, structure, and functional characteristics of cells and tissues, and studies of complex biological phenomena and related gene functions at the overall biological level<sup>5,8</sup>. Animal models with genetic mutations are commonly used to determine gene function. Numerous methods for generating mutations, including X-rays, chemical mutagenesis, retroviral transfection, and transgenic technology, have been used in gene function research<sup>11–14</sup>; however, these methods do not produce stable expression, frequently affect multiple genes or lead to chromosomal rearrangements, and cannot provide molecular markers to allow collection of mutation information<sup>12</sup>.

Gene targeting technology to produce a specific gene mutation by homologous recombination in embryonic stem (ES) cells, thus generating genetic knock-out or knock-in animals, is currently one of the most important ways of studying the functions of genes<sup>15,16</sup>. Unfortunately, this method is limited by the low probability of homologous recombination and time-consuming animal breeding<sup>17</sup>. Moreover, acquired function-inactivated mutations (null mutations) often produce phenotypes with characteristics different from those of molecular

<sup>1</sup>Research Center for Translational Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China. <sup>2</sup>Department of vascular surgery, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China. <sup>3</sup>Key Laboratory of Arrhythmias of the Ministry of Education of China, East Hospital, Tongji University School of Medicine, Shanghai, China. <sup>4</sup>Department of Cardiology, East Hospital, Tongji University School of Medicine, Shanghai, China. <sup>5</sup>Department of Endocrinology, Changzheng Hospital, Second Military Medical University, Shanghai, 200003, PR China. <sup>6</sup>Metastasis research institute, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China. <sup>7</sup>School of Life Science and Technology, Tongji University, Shanghai, China. <sup>8</sup>Shanghai Research Center for Model Organisms, Shanghai, 201203, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.F. (email: jfei@tongji.edu.cn)

damage associated with diseases. Genome editing technologies that have emerged recently, such as ZFN, Talen, and Crispr-Cas, are widely used to create genetically modified organisms<sup>10,18–20</sup>.

Gene trapping technology allows report vectors to be randomly integrated in the genome or label insertion sites, causing insertion inactivation mutations, mainly by utilizing eukaryotic gene promoters or poly(A) signals to gene trap<sup>21–23</sup>. By pre-determined screening criteria, a large number of cells or individual mutations are obtained at very low cost and used to reveal gene expression patterns and gene functions. Gene trapping combined with random mutations, known as random gene targeting, can produce a large number of mutant species in a short time and allows identification of mutant genetic information easily and precisely<sup>24–26</sup>. Due to its simple principles and high efficiency, random gene targeting is a convenient way to generate large-scale gene mutations and has been widely applied in plants, insects, drosophila, and mice<sup>23,27</sup>. Gene trapping has been used in studies aimed at discovering the functions of a large number of gene sequences. Selectable markers and reporter genes in gene trapping vectors, including enhancer trapping vectors, promoter trapping vectors, and splice acceptor gene trapping vectors, lack their own promoters and are expressed under the control of the endogenous gene's promoter. Therefore, the gene trapping strategies mentioned above do not capture genes that are not expressed or expressed at very low levels in target cells. Therefore, new gene trap vector systems with high efficiency will accelerate the pace of studies aimed at determining the functions of genes. The poly(A) trap vector can select a promoter (which is usually constitutively expressed) and capture genes not expressed in the target cells; therefore, it can capture more genes<sup>26,28,29</sup>.

As part of the mature mRNA, the poly(A) tail is important for exporting the mRNA from the nucleus, inhibiting molecular degradation of mRNA in the cytoplasm, and translation<sup>30–32</sup>. Polyadenylation resulting in a poly(A) tail is essential for production of mature mRNA. The signals required for addition of the poly(A) tail are determined by several cis-elements, including an AAUAAA element and a U-rich upstream sequence element<sup>33</sup>. Mutant poly(A) signals could inhibit mRNA polyadenylation and disrupt transcriptional termination<sup>34</sup>. Therefore, poly(A) signals (PAS) and poly(A) tails are important for proper gene expression<sup>35</sup>. Poly(A) gene trapping is performed by trap vectors without the poly(A) signal of the reporter gene<sup>29,36</sup>.

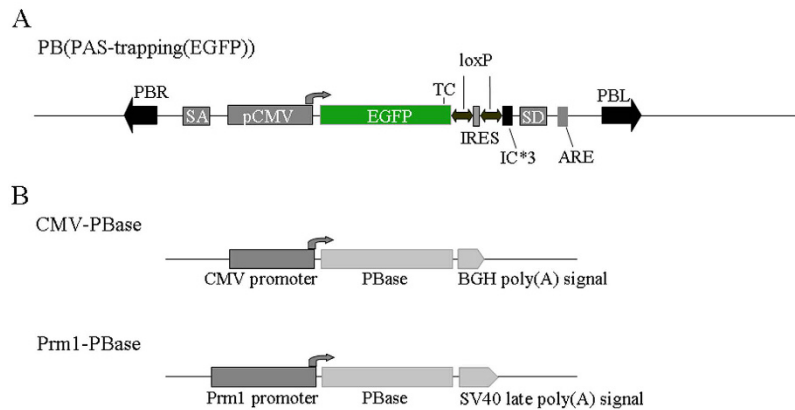
Transposable elements, also known as transposons, are a class of mobile genetic units in the genome. Transposon-mediated mutations in the genomes of plants and animals have been widely reported<sup>37,38</sup>. In lower organisms, transposable elements are used to generate transgenic animals and perform insertion mutagenesis<sup>39–43</sup>. In comparison, the applications of transposons are quite limited in mammals and other vertebrates because of the lack of an efficient and practical transposition system<sup>44</sup>. Significant effort has been expended to establish an efficient mammal transposon system. In 2005, the piggyBac (PB) DNA transposon was applied in genetic analysis of mice and human cells for the first time, demonstrating a practical and efficient system of mammalian transposable elements and providing a new method for conducting large-scale studies of gene function in mammals<sup>45</sup>. The PB transposon has a length of 2472 bp and contains two terminal domains at both ends, as well as a 2.1-kb reading frame encoding transposase in the middle. PB transposons can be removed from the genome and reinserted at the tetranucleotide TTAA site. The minimal terminal domains, containing a 35-bp 5'-end and 63-bp 3'-end, are important for efficient chromosomal integration by the PB transposon<sup>46</sup>. Previous studies have shown that PB transposons have high transposition activity in several species, including insects, drosophila, and mice<sup>47</sup>. However, transposons such as the P-element transposase from *Drosophila* have not been widely in mice because of several limitations, including inefficiency, a limited ability to carry gene fragments, and the close proximity of most new transposon insertions to the original site<sup>48</sup>.

In this study, a poly(A) gene trap strategy, known as an expression-independent gene trap, was designed using EGFP as a reporter gene and designated PAS-trapping (EGFP). In order to increase the efficiency of gene trapping and gene targeting, the construct contains an IRES (internal ribosome entry site) cassette to avoid nonsense-mediated mRNA decay (NMD) and an RNA instability element (ARE) to improve splicing efficiency<sup>28,49–51</sup>. Then, taking advantage of the efficient transposition ability of PB, we obtained a PB transposon system carrying the PAS-trapping (EGFP) gene trap vector for gene trapping and insertional mutagenesis, which was designated PB(PAS-trapping (EGFP)). PB(PAS-trapping (EGFP)) can be used to produce transgenic mice capable of producing offspring with a wide range of insertional mutations. A primary feature of PB(PAS-trapping (EGFP)) is that PB transposon-mediated integration events within the genome can be screened conveniently with high efficiency almost independent of the insertion site, because the reporter EGFP gene is driven by a constitutively active promoter. The purpose of the PB(PAS-trapping (EGFP)) system is to allow researchers to generate a variety of species with different genetic mutations, which can be used to study gene function and establish disease models, as well as for research and development in aspects of therapeutic drugs. We hope that adoption of this technology will increase the efficiency of forward transposon-based genetic screening and gene targeting in mice.

## Results

### Construction of a poly(A) gene trap (splice donor gene trap) vector with PB transposable sites and corresponding transgenic mice.

According to the principle of poly(A) gene trapping, we designed a high efficiency gene trap vector system for gene trapping and gene targeting, the PB-mediated poly(A) gene trap vector (Figs 1 and 2 and Supplementary Table S1). Our gene trapping strategy is based on the theory that a poly(A) sequence signaling for the termination of transcription is essential for the proper expression of a gene<sup>29</sup>. Therefore, the reporter gene (EGFP in our case) should be expressed only after transposition upstream of an endogenous PAS. The reporter gene is an EGFP gene with its own promoter and without a poly(A) signal, which is flanked by a pair of PB terminal domains (PBL and PBR) (Fig. 1A). We chose the human CMV promoter to drive expression of EGFP, which is expressed in most murine cell types (Figs 1A and 2). Thus, genes expressed in the target cell will be captured, but genes not expressed in the target cells will also be captured. The RNA splicing acceptor (SA), a 432-bp sequence between the 2nd intron and 3rd exon of the human B-cell lymphoma 2 gene (Bcl-2)<sup>52</sup>, was connected to the 5'-end of the CMV promoter to avoid losing the gene inactivation derived from

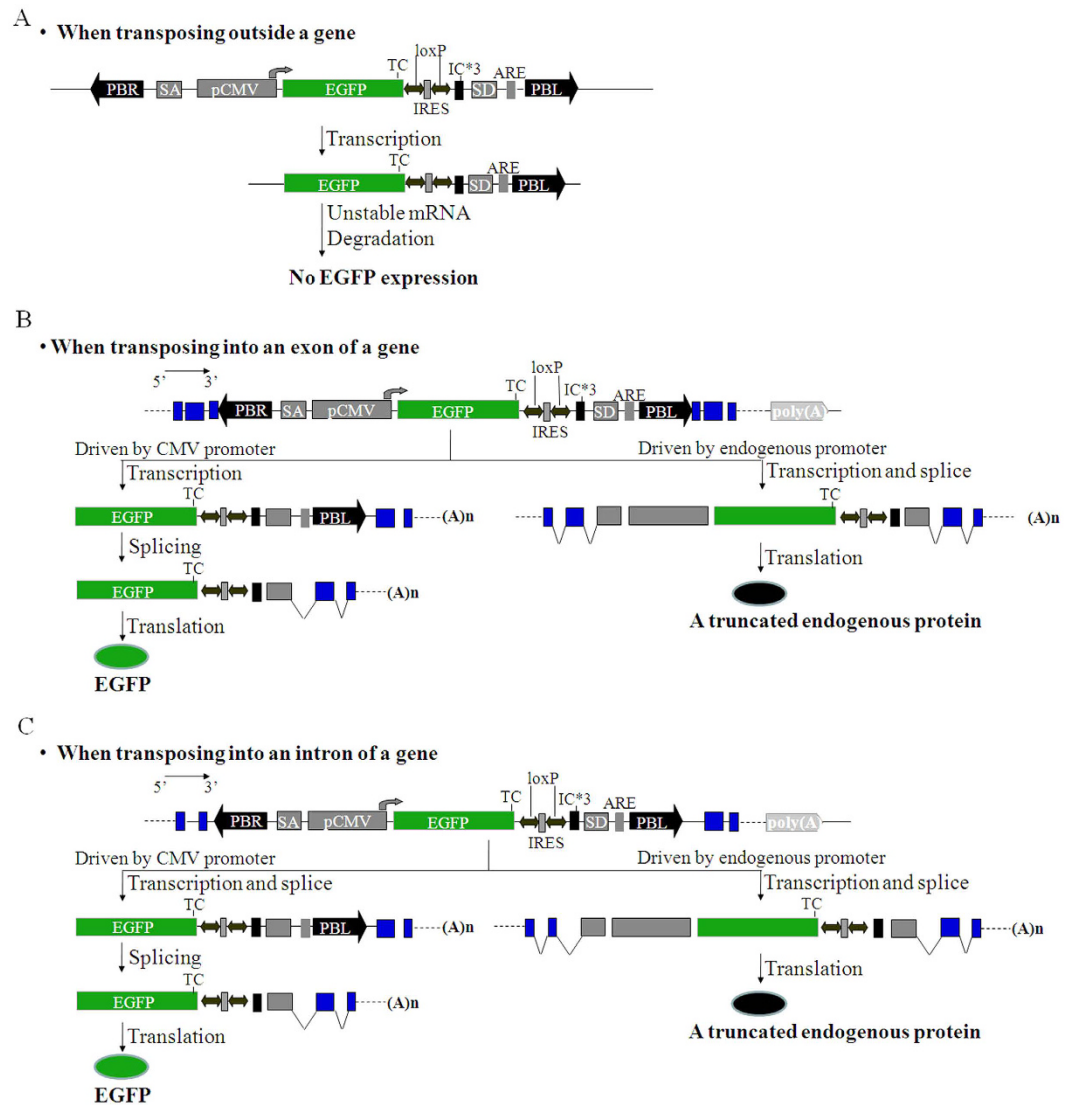


**Figure 1. Schematic diagram of PB(PAS-trapping (EGFP)) gene trapping vector and transposase constructs of the piggyBac Transposition System for Mice.** (A) PB donor constructs, named as PB(PAS-trapping (EGFP)). The reporter gene of PB(PAS-trapping (EGFP)) is an EGFP gene driven by the human CMV promoter. An unstable mRNA signal sequence (ARE) to the 3' end of EGFP was added. The IRES, three ICs and SD were added between EGFP and ARE. There is the option to remove the IRES using Cre, thereby removing the truncated protein (but not GFP). A SA was added in front of EGFP reporter gene. The PAS-trapping (EGFP) cassette was placed between a pair of PB terminal domains (PBL and PBR, black arrows). (B) PB transposase helper constructs. CMV-PBase was constructed by inserting CMV promoters in front of PBase gene, followed by BGH pA. The piggyBac transposase gene (PBase) driven by Prm1 promoters (Prm1-PBase) were followed by SV40 late poly (A) signal. (SA: splicing acceptor; CMV promoter: cytomegalovirus immediate early promoter; TC: termination codon; IRES: internal ribosome entry site; IC: initial codon; SD: splicing donor; ARE: an RNA instability element; PBL: PB repeat left termini; PBR: PB repeat right termini; pA: poly (A); BGH pA: bovine growth hormone poly (A); Prm1: Protamine 1).

vector insertion (Figs 1A and 2C and Figure S1), because mRNA splicing of the endogenous gene spliced it as a part of the intron when PB(PAS-trapping (EGFP)) was inserted into the intron of the gene in mice<sup>26,52–55</sup>. The SA cassette causes insertional mutation when PB(PAS-trapping (EGFP)) is inserted into an intron. An RNA splicing donor (SD) sequence of approximately 174 bp from between the 8th exon and the 8th intron of the murine Hprt gene<sup>49</sup> was connected to the 3'-end of the EGFP reporter gene (Figs 1A and 2). The SD cassette was important for efficient splicing with the 3'-end of the gene in which PB(PAS-trapping (EGFP)) was inserted.

Conventional poly(A)-trap systems have two primary shortcomings<sup>49,51</sup>. One limitation of conventional poly(A)-trap systems is the biased gene trapping that occurs during random integration<sup>49</sup>. Several research groups reported that most screened mutations mediated by conventional poly(A)-trap vectors occurred near the 3'-end of the affected gene<sup>49,56</sup>. When the translation termination signal is too far away from the poly(A) signal, mRNA surveillance mechanisms in eukaryotic cells consider the mRNA to be abnormal and degrade it<sup>49,55</sup>. Therefore, the reporter gene is usually not translated because of the long distance between the translation termination signal and the captured poly(A) in the case of the insertion of the trap vector into the 5'-region of the gene. Fortunately, an IRES (internal ribosome entry site) sequence and three ICs (initiation codons) in all three reading frames were proven to avoid the mRNA-surveillance mechanism, as described previously<sup>28,49</sup>. This IRES sequence, with a LoxP site on both sides and three ICs in all three reading frames, was added between the EGFP report gene and SD sequence in our system (Fig. 1A). Abnormal proteins encoded by IRES mediated by the PB-mediated poly(A) gene trap vector may be harmful in insertion-mutant mice. Therefore, the inserted IRES sequence was deleted using a Cre-mediated LoxP-deletion system. The other primary limitation of conventional poly(A)-trap systems is the production of false-positive clones caused by SD read-through events<sup>51</sup>. Poly(A)-signal trapping vectors obtain false-positive clones by screening with selectable markers or reporter genes if such selectable markers and reporter genes are expressed as a result of transcriptional read-through when they are inserted into a non-gene region or endogenous non-coding 3'-region (Figure S2)<sup>57,58</sup>. When vectors with and without an RNA instability element (ARE) were compared, the vector with an ARE showed a reduction of approximately 1.7-fold in the number of trapped G418-resistant colonies in comparison with the vector lacking an ARE<sup>50,51</sup>. Therefore, the ARE has the ability to reduce false-positive results<sup>50</sup>. In order to solve this limitation of poly(A) trap vectors, an ARE sequence from human GM-CSF precursor cDNA was added behind the SD sequence (Fig. 1A)<sup>59</sup>. This unstable mRNA signal sequence can lead to instability in mRNA transcribed from EGFP after the trap vector is inserted into a non-gene region or endogenous non-coding 3'-region (Fig. 2A and Figure S2), preventing EGFP protein synthesis<sup>50,51,57,58</sup>.

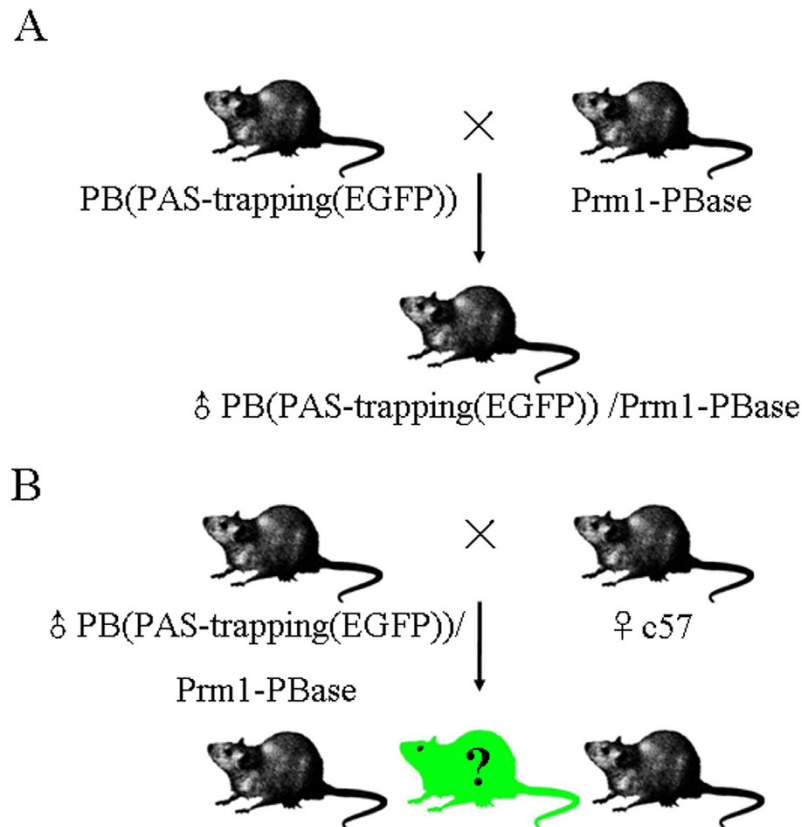
We constructed transgenic mice with the PB(PAS-trapping (EGFP)) vector. We also produced a PB transposase driven by the CMV promoter (CMV-PBase, Fig. 1B). PB transposons efficiently transpose between vectors and chromosomes via a “cut and paste” mechanism<sup>45,60–62</sup>. PB transposons can be removed from vectors and reinserted at tetranucleotide TTAA sites in the genome, preventing concatemer formation during delivery of the vector into chromosomes<sup>61,62</sup>. In order to avoid integration of multiple copies of the PB(PAS-trapping (EGFP)) vector in a head-to-tail array at donor sites<sup>63</sup>, we utilized PB transposons for cut-and-paste gene delivery. Before it was microinjected into fertilized mouse oocytes to generate PB(PAS-trapping (EGFP))-positive transgenic mice,



**Figure 2. Illustration depicting PB(PAS-trapping (EGFP)) gene trapping strategies and its mechanism of action.** (A) The lack of a dedicated poly(A) signal for the EGFP leads to no EGFP expression after transposition outside a gene. (B) Schematic diagram of gene trap and insertional mutation by PB(PAS-trapping (EGFP)) gene trap vector when transposition into an exon of a gene with the same direction of the endogenous gene reading frame. (C) Schematic diagram of gene trap and insertional mutation by PB(PAS-trapping (EGFP)) gene trap vector when transposition into an intron of a gene with the same direction of the endogenous gene reading frame.

the PB(PAS-trapping (EGFP)) vector was mixed with the CMV-PB vector at a mass ratio of 1:2 in the form of a circular plasmid. Genomic DNA was isolated from pup tail samples and used for genotype identification by PCR. We obtained 10 PB(PAS-trapping (EGFP)) single-positive transgenic mice.

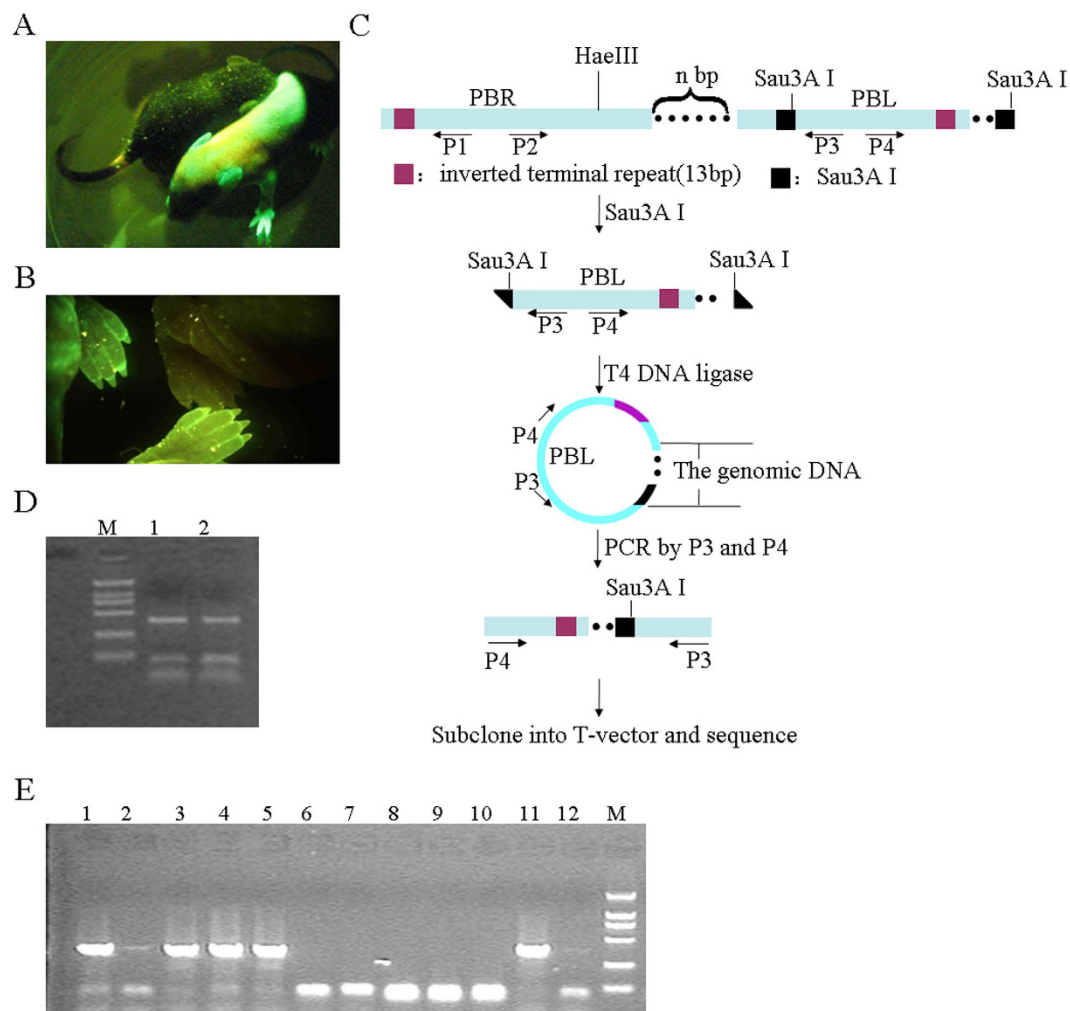
**Transposition strategy of PB(PAS-trapping (EGFP)) in male mice.** Spermatogonia with various mutations produce sperm with the corresponding mutations via the expression of PB transposase in spermatogonia, allowing corresponding individual mutants to be obtained by subsequent reproduction (Fig. 3). In order to obtain transgenic mice expressing PB transposase in mouse spermatogonia, we selected the mouse *prm1* promoter to drive the expression of PB transposase in spermatogonia (Fig. 1B)<sup>64,65</sup> and generated Prm1-PBase mice through transgenic technology. Six founder mice of Prm1-PBase transgenic mice were obtained and crossed with C57 mice. The progenies of Prm1-PBase transgenic founders were screened for PBase expression. Founder mouse no. 5 had the highest PBase expression and was selected for further study. Male mice doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBase were obtained by mating the PB(PAS-trapping (EGFP)) and Prm1-PBase transgenic mouse lines. With the expression of PB transposase in the male reproductive system, spermatogonia obtain various mutations and produce sperm containing these mutations, so that corresponding mutants are obtained by reproduction (Fig. 3).



**Figure 3. Overview of the breeding scheme used to generate mutant mice in the mouse germline.** GFP was used as a marker to monitor transposition events. **(A)** Obtaining the male mice doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBase. Transgenic mouse lines carrying PB(PAS-trapping (EGFP)) and protamine 1 (Prm1) promoter-driven PB transposase transgenes (Prm1-PBase) were generated by microinjection. The male mice doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBase were obtained by mating the PB(PAS-trapping (EGFP)) and Prm1-PBase transgenic lines. **(B)** Generating mutant mice in the mouse germline. The mice doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBase were crossed with WT females to generating offspring with various insertion mutations.

**Generation of mice carrying insertion mutations and identification of insertion sites.** LM-PCR was performed to determine the transposition sites of these 10 PB(PAS-trapping (EGFP)) single-positive transgenic mice (Fig. 4). Briefly, tail DNA digested with *Sau3AI* or *HaeIII* was ligated, after which transposition sites were amplified by PCR with transposon-specific primers<sup>45</sup>. We found that the PB(PAS-trapping (EGFP)) transposition cassette in PB(PAS-trapping (EGFP)) transgenic founder mouse no. 106 was inserted into the 2nd intron of spastic paraplegia 20 (*Spg20*) on the 3rd chromosome. However, the EGFP gene had a reading direction opposite that of *Spg20*; therefore, EGFP protein was not translated (Figure S3A). In order to facilitate subsequent analysis of insertion sites, we selected founder mouse no. 106, which expressed a single copy of the PB(PAS-trapping (EGFP)) transposition cassette at a single insertion site and showed no EGFP fluorescence on the surface, for further study.

PB(PAS-trapping (EGFP)) transgenic founder mouse no. 106 was mated with Prm1-PBase transgenic mice founder mouse no. 5. Two male PB(PAS-trapping (EGFP))/Prm1-PBase transgenic mice were obtained and mated with wild-type C57 mice. Fifty-eight pups from seven litters were screened under a fluorescence microscope and subjected to PCR analysis. One mouse from a litter of eight mice resulting from the pairing of PB(PAS-trapping (EGFP))/Prm1-PBase transgenic mice and wild-type C57 mice emitted significant fluorescence (Fig. 4A,B). Among the eight pups, six PB(PAS-trapping (EGFP)) transgene-positive mice, including the fluorescent subject, were identified by conducting PCR on genomic DNA (Table 1). Transgene insertion sites were identified in the six obtained PB(PAS-trapping (EGFP)) transgene-positive mice using LM-PCR (Fig. 4C,D,F). The PB(PAS-trapping (EGFP)) transposition cassette of the fluorescence-positive mouse was inserted into an intron of the DEDD (death effector domain-containing) gene (Table 2 and Fig. 5A), while the PB(PAS-trapping (EGFP)) transposition cassettes of the remaining five mice were all located in the 2nd intron of *Spg20* on the 3rd chromosome (Table 2 and Figure S3A). The reading direction of EGFP in the fluorescence-positive mouse was consistent with that of the DEDD gene. The EGFP reporter gene was expressed only when trapping the poly(A) signal of a gene (Table 2).



**Figure 4. Screening for mutant mice.** (A,B) Screening for mutant mice performed by GFP expression. Newborn mice were examined by fluorescence microscopy. (C,D,E) Molecular analysis of flanking genomic DNA sequences at sites of new transposon insertion. (C) Schematic diagram of reverse PCR for identification of insertion sites, after digested with restriction enzyme *Sau3A I*. (D) The PCR products were amplified from ligation products of the *Sau3A I*-digested genomic DNA with T4 ligase. (E) Clone selection of the positive clones, who had the PCR product (obtained in Fig. 4D) by PCR. After electrophoresis of the PCR products in 1.2% agarose gels, the desired nucleic acid fragments were purified and imported into pMD18-T Simple vector.

No	EGFP-PCR positive frequency <sup>1</sup>	New insertions frequency <sup>2</sup>
190–197	4/8 = 0.5	1/4 = 0.25
407–413	3/7 = 0.429	1/3 = 0.333
400–406	2/7 = 0.286	1/2 = 0.5
418–426	5/9 = 0.556	1/5 = 0.2
491–506	3/11 = 0.273	1/3 = 0.333
483–490	3/8 = 0.375	1/3 = 0.333
427–434	6/8 = 0.75	1/6 = 0.167
Total	26/58 = 0.448	7/26 = 0.269

**Table 1. Summary of PB transposition in 106# line.** <sup>1</sup>EGFP-PCR positive per litter. <sup>2</sup>New insertion of the EGFP-PCR positive.

We found 26 pups that were trapping vector-positive, of which five pups showed green fluorescence on their body surface. The EGFP reporter gene was expressed only when trapping the poly(A) signal of a gene. Genomic walking analysis confirmed that the EGFP gene was inserted inside an endogenous gene with the

NO.	Flanking sequences <sup>1</sup>	Chr <sup>2</sup>	Gene Name/GenBank accession number <sup>3</sup>	Intron or Exon <sup>4</sup>	Direction <sup>5</sup>	EGFP fluorescence
106	TTAATTTTAAAA	3	Spg20	2 <sup>th</sup> intron	Opposite <sup>6</sup>	Negative <sup>8</sup>
191	TTAAAGGAAAGAA	3	Not inside known gene	/	/	Negative
503	TTAAGTAAATCATT	10	Not inside known gene	/	/	Negative
418	TTAAAGATTTATTT	12	Dgkb	19 <sup>th</sup> intron	Consistent <sup>7</sup>	Positive <sup>9</sup>
432	TTAACAGTCACTCT	1	Dedd	1 <sup>th</sup> intron	Consistent	Positive
490	TTAATTTCTTCTT	3	Gm9733	1 <sup>th</sup> intron	Consistent	Positive
506	TTAAGGCGAGAAAT	3	Ift80	11 <sup>th</sup> intron	Consistent	Positive
695	TTAAAGACAGGGTT	3	Fam96a	2 <sup>th</sup> intron	Consistent	Positive

**Table 2. Distribution of transposon insertion sites in the PB(PAS-trapping (EGFP))/Prm1-PBase system.** Spg20: spastic paraplegia 20, Dgkb: Diacylglycerol kinase beta, Dedd: death effector domain-containing, Ift80: intraflagellar transport 80 homolog (Chlamydomonas), Gm9733 and Fam96a: a putative gene. <sup>1</sup>Nucleotide composition of flanking sequences from PB(PAS-trapping (EGFP)) insertion site. <sup>2</sup>Distribution of PB(PAS-trapping (EGFP)) insertions in the chromosome. <sup>3</sup>Gene/GenBank accession number, in which PB(PAS-trapping (EGFP)) was inserted. <sup>4</sup>Distribution of PB(PAS-trapping (EGFP)) insertions in genes. <sup>5</sup>The direction of the EGFP gene in the PB(PAS-trapping (EGFP)) vector within the endogenous gene reading frame. <sup>6</sup>The direction of the EGFP gene in the PB(PAS-trapping (EGFP)) vector was opposite with the direction of the endogenous gene reading frame. <sup>7</sup>The direction of the EGFP gene in the PB(PAS-trapping (EGFP)) vector was consistent with the direction of the endogenous gene reading frame. <sup>8</sup>The EGFP fluorescence negative mice. <sup>9</sup>The EGFP fluorescence positive mice.

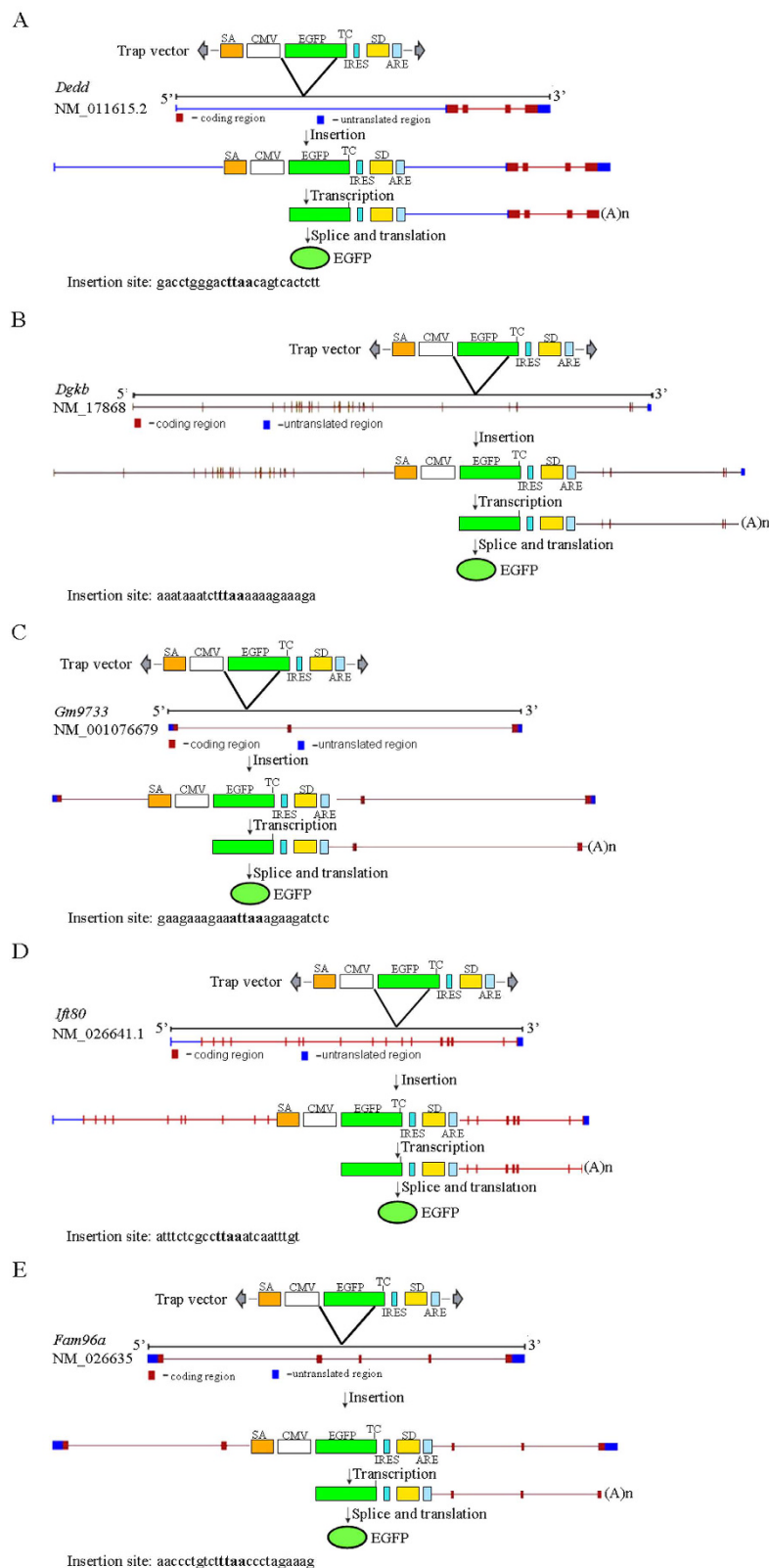
same transcription direction in each fluorescent mouse. The five independent insertion sites were mapped in the EGFP-expressing mice: DEDD (inserted in the 1st intron) (Table 2 and Fig. 5A), DGKB (diacylglycerol kinase beta, inserted in the 19st intron) (Table 2 and Fig. 5B), Gm9733 (a putative gene, inserted in the 1st intron) (Table 2 and Fig. 5C), Ift80 (intraflagellar transport protein 80 homolog, inserted in the 11th intron) (Table 2 and Fig. 5D), Fam96a (inserted in the 2nd intron) (Table 2 and Fig. 5E). However, no successful trapping events were identified in the 21 pups that lacked an EGFP signal, of which 19 pups showed the insertion identified in founder mouse no. 106, whereas 2 pups showed a new insertion that was not inserted into any known gene (Figure S3B,C). Our results indicate that EGFP fluorescence on the body surface of newborn mice is a reliable marker for screening endogenous gene-trapped mice.

Among the 58 pups from seven litters mentioned above, we found that 44.8% (26/58) of pups were transposon-positive progenies by conducting PCR on genomic DNA. Identification of insertion sites in transposon-positive mice showed that there were 7 transposon-positive offspring with new transposon integrations and 19 transposon-positive offspring with transposon integrations identical to that of founder mouse no. 106. The transposition capacity (rate of generating a new site or transposition) of the PB(PAS-trapping (EGFP)) system was 26.9% (7/26) (Table 1). About 71.43% (5/7) of the new trap insertions mediated by the PB transposon were in transcriptional units (Table 2). In addition, 100% (5/5) of the EGFP fluorescence-positive mice had new trap insertions mediated by the PB transposon in transcriptional units. Moreover, the direction of the EGFP gene was consistent with the direction of the endogenous gene reading frame.

After comparing the genomic sequences of the seven mice with new insertion sites (Table 1) with sequences from several public databases, we found that the PB transposons of these mice were all inserted into TTAA sites. Four PB transposons were located in chromosome 3, one was located in chromosome 1, one was located in chromosome 10, and one was located in chromosome 12. These results, as well as the fluorescence microscopy observations, show that fluorescence on the body surface of newborn transgenic mice can be observed only when the transposon has been inserted into a gene and the reading frames of EGFP and the endogenous gene are in the same direction (Table 1).

## Discussion

With the rapid accumulation of sequence data from humans, animals, and plants, determination of the biological functions represented by the sequence data has become urgent<sup>3,66</sup>. In the field of gene function studies, gene knockout is currently the most widely used research method. However, due to the low throughput of knockout mouse production, the number of types of knockout mice is quite limited<sup>15,67,68</sup>. Therefore, there is an urgent need for efficient and high-throughput screening methods for discovering unknown genes and analyzing their functions. In the present study, we utilized PB transposable elements and a PAS-trapping (EGFP) gene trapping vector to achieve gene trapping and insertional mutagenesis in mice. Our results demonstrate that PB transposon-mediated poly(A) gene trapping occurred in the reproductive system after mating male PB(PAS-trapping (EGFP))/Prm1-PBase transgenic mice with wild-type female mice (Figs 3 and 4). Using our PB(PAS-trapping (EGFP)) system, numerous pups carrying mutations in different sites were generated, after which captured genes were detected easily under a fluorescence microscope (Fig. 4 and Table 2). There are several advantages to using our method to generate mutant mice: (1) mutant mice can be identified quickly and non-invasively by EGFP fluorescence, (2) mutants can be obtained easily by mating transposon-positive mice with wild-type mice, (3) it requires less labor than producing knockout mice, and (4) it is easy to detect insertion sites by PCR.



**Figure 5. Five independent insertion sites were mapped in EGFP-expressing mice. (A)** *Dedd* (inserted in 1th intron) was inserted by PB (PAS-trapping (EGFP)) trap vector. **(B)** *Dgkb* (inserted in 19th intron) was inserted by PB (PAS-trapping (EGFP)) trap vector. **(C)** *Gm9733* (inserted in 1th intron) was inserted by PB (PAS-trapping (EGFP)) trap vector. **(D)** *Ift80* (inserted in 11th intron) was inserted by PB (PAS-trapping (EGFP)) trap vector. **(E)** *Fam96a* (inserted in 2th intron) was inserted by PB (PAS-trapping (EGFP)) trap vector. *Dgkb*: Diacylglycerol kinase beta, *Dedd*: death effector domain-containing, *Ift80*: intraflagellar transport 80 homolog (*Chlamydomonas*), *Gm9733* and *Fam96a* are putative genes.



The piggyBac transposon system has been successfully used to create random insertion mutant mice<sup>45,69–72</sup>. However, methods of identifying mice with mutant insertions are not well established. In the present study, we designed a poly(A) gene trapping system, in which a target gene is captured using a reporter gene as bait. Insertion of a DNA vector containing a reporter gene into the genome results in inactivation or mutation of the endogenous gene and simultaneously activation or expression of the reporter gene, allowing for identification of insertion mutagenesis<sup>50,73</sup>. The reporter gene used in our study was an EGFP gene with a ubiquitously expressed CMV promoter (Fig. 1), which facilitated detection of offspring with insertion mutations by *in vivo* fluorescence microscopy (Figs 3 and 4). Gene trapping strategies can be divided into three main types: enhancer trap, promoter trap, and splice acceptor gene trap. Enhancer trap vectors have not been widely applied in mice because the mutagenesis rate of enhancer traps has not been reported<sup>74</sup>. In contrast, the mutagenesis rate of promoter traps is high, while their exon insertion frequency is very low<sup>75</sup>. Because the insertion occurs in an intron, the gene trap efficiency of splice acceptor gene trap vectors is 50-fold greater than that of promoter traps<sup>73</sup>. However, alternative splicing may cause low-level generation of wild-type transcripts and less efficient allele mutations<sup>76</sup>. The poly(A) trap vector can capture a wider range of genes by using a constitutively expressed promoter to capture genes not expressed in target cells<sup>28,29</sup>. In our study, 100% (5/5) of the EGFP fluorescence-positive mice had new trap insertions in transcriptional units (Fig. 5 and Table 2). No gene trap events were identified in mice that were EGFP-PCR-positive, but EGFP fluorescence-negative (Figure S3 and Table 2). Our results demonstrate that the poly(A) signal and poly(A) tail are essential for proper gene expression.

A highly efficient SA was joined to the 5' end of the CMV promoter to enhance the efficiency of insertion mutation, as described previously<sup>49,50,52</sup>. This design disrupts endogenous gene function when trapping vectors are inserted into the intron region, as shown in Fig. 2 and Figure S1. In mammalian cells, if the distance between a stop codon and the last exon-exon junction is more than 60 nt, mRNAs with premature stop codons are degraded by nonsense-mediated mRNA decay (NMD)<sup>49,52</sup>. In our PB(PAS-trapping (EGFP)) system, an IRES sequence and three ICs were added to avoid the mRNA-surveillance mechanism, as described previously<sup>49,50,52</sup>. As shown in Fig. 5 and Table 2, we screened five EGFP fluorescent-positive mice, in which the PB(PAS-trapping (EGFP)) trapping vectors were all more than 60 nt from the last exon-exon junction. These results demonstrate that our system allows the use of EGFP to screen mutant mice in which insertion occurred in the first few introns of an endogenous gene (Table 2 and Fig. 5). In order to reduce the number of false-positive results obtained using the fluorescent microscope, an unstable mRNA signal sequence was used to make mRNA transcribed from EGFP unstable when the trapping vector was inserted into a non-gene region<sup>51,59</sup> (Fig. 2 and Figure S2). For example, EGFP fluorescence was not detected in mouse #191 or #503 (Figure S3B,C), in which the PB(PAS-trapping (EGFP)) vectors were not inserted into known genes.

The introduction of the PB(PAS-trapping (EGFP)) system into the reproductive system allows mice to produce offspring with insertional mutations, which can be screened using a high-throughput method. The transposition capacity (rate of generating a new site or transposition) of the progeny of the male PB(PAS-trapping (EGFP))/Prm1-PBase transgenic mice was 26.9% (7/26) (Table 1), which was higher than that of the sleeping beauty (SB) transposition. The SB transposition capacity was about 20%<sup>77</sup>. EGFP expression was not easily detected in some progeny of the SB transposon-transgenic mice<sup>63,77,78</sup>. EGFP expression may be influenced by DNA methylation or NMD<sup>78,79</sup>. However, in our system, five of the seven new trap insertions showed EGFP fluorescence on the body surface.

Our PB(PAS-trapping (EGFP)) system is functional, but has some limitations. For example, the PB(PAS-trapping (EGFP)) system may produce EGFP fluorescence-positive mice in which the PB(PAS-trapping (EGFP)) trapping vector is inserted outside a gene, because of the presence of *cis*-elements for mRNA polyadenylation in the insertion site. However, an unstable mRNA signal sequence was used to reduce the rate of false-positive results<sup>52,63</sup>. In addition, the system PB(PAS-trapping (EGFP)) may produce insertion mutagenesis in EGFP fluorescence-negative mice, because the direction of the EGFP gene in the vector may be opposite the direction of the endogenous gene reading frame. Finally, embryonic lethality precluded screening of some mice.

In our system, PB transposon-mediated poly(A) gene trapping occurred in the reproductive system after mating male PB(PAS-trapping (EGFP))/Prm1-PBase transgenic mice with wild-type female mice (Figs 3 and 4). Numerous pups having different PB-insertion sites were generated, after which captured genes were detected easily under a fluorescence microscope (Fig. 4 and Table 2). Moreover, all transposition events occurred exclusively at TTAA sites (Table 2). Consequently, mice with various mutations can be continuously produced, while an embryo bank containing diverse mutants can be established by mating transgenic mice with wild-type mice. This mutant mouse resource will facilitate studies of gene function and allow the establishment of new disease models to support research aimed at producing new therapeutic drugs. The PB(PAS-trapping (EGFP)) transposon system provides a new method for conducting large-scale studies of gene function in mammals. In comparison with traditional methods such as gene knockout and chemical mutagenesis, PB transposon-based poly(A) gene trapping is convenient, fast, and practical. Therefore, the PB(PAS-trapping (EGFP)) system should enable functional genomics research in mice and other higher animals to be conducted more efficiently. In addition, the adoption of this technology will increase the efficiency of forward transposon-based genetic screening in mice.

## Materials and Methods

**Animals.** C57BL/6J mice, PB(PAS-trapping (EGFP)) transgenic mice, and PB transposase transgenic mice were obtained from Shanghai Research Center for Model Organisms (Shanghai, China). All animals were housed in a specific-pathogen-free facility with food and water provided *ad libitum*. All animal experiments were approved by the Institutional Animal Care and Use Committee of Shanghai Research Center for Model Organisms and conducted in accordance with government guide-lines for animal care.

### Construction of a poly(A) gene trap (splice donor gene trap) vector with PB transposable sites.

Two LoxP sequences were synthesized. The restriction sites for the first LoxP sequence were BglII and SacI, whereas those of the second LoxP sequence were EcoRI and SalI. The two LoxP sequences were inserted into pEGFP-C1 (Clontech Laboratories, Mountain View, CA, USA) to generate pEGFP-C1-LoxP2. IRES was amplified from a pIRES vector (Clontech Laboratories, Mountain View, CA, USA) and cloned into the SacI and EcoRI sites of pEGFP-C1-LoxP2 to generate pEGFP-C1-LoxP2-IRES. The SD (splicing donor), approximately 174 bp of the sequence between the 8th exon and 8th intron of the hypoxanthine phosphoribosyltransferase 1 (*Hprt*) gene, was amplified from mouse genomic DNA and inserted into the KpnI and ApaI sites of pEGFP-C1-LoxP2-IRES to generate pEGFP-C1-LoxP2-IRES-SD.

The unstable mRNA signal, a 54-bp fragment from the cDNA of human granulocyte-macrophage colony-stimulating factor precursor (GM-CSF precursor), was synthesized and cloned into the XbaI and MluI sites of pEGFP-C1-LoxP2-IRES-SD to generate pEGFP-C1-LoxP2-IRES-SD-in. PB L was amplified from PB3XP3GFPaf and cloned into the XbaI and MluI sites of pEGFP-C1-LoxP2-IRES-SD-in to generate pEGFP-C1-LoxP2-IRES-SD-in-L, as shown in Fig. 1A. PB R was amplified from PB3XP3GFPaf and cloned into the NotI and KpnI sites of pEGFP-C1-LoxP2-IRES-SD-in-L to generate the gene trap vector, PB(PAS-trapping (EGFP)).

The primers used in the construction of the gene trap vector were as follows:

FLoxP1: 5' gatcctaataactcgtatagcatacattatacgaagttatgagct 3'  
 RLoxP1: 5' cataactcgtataatgtatgctatacgaagtatttaa 3'  
 FLoxP2: 5' aattcataactcgtatagcatacattatacgaagttatg 3'  
 RLoxP2: 5' tcgacataactcgtataatgtatgctatacgaagttatg 3'  
 FIRES: 5' attgagctccctcctcccccccccta 3'  
 RIRES: 5' gctgaattccatccatcgggtgtgggcaagcttatc 3'  
 FSD: 5' agtggtagcactcagggattgaaatgtaag 3'  
 RSD: 5' aatggcccgagcgtctgactgactacag 3'  
 FIn: 5' gatccatcagtaataattatattatatttttaaaatattattattattatttaagt 3'  
 RIn: 5' ctagactaaataaataaataaataatatttttaaaatataaataatattactgatg 3'  
 FpBac L: 5' ggatctagagtcgacctcagcagcatgcaag 3'  
 RpBac L: 5' aagacgctgacaatgttcagtcagagac 3'  
 FSA: 5' cgggagctcctgtatcttaagatggctgg 3'  
 RSA: 5' gccacgctgcatattatttctactgc 3'  
 FpBac R: 5' aatggtagcagatgtttgtttgacggac 3'  
 RpBac R: 5' atgtagcagcagatcaaaacgaaatcgacg 3'

**Construction of the PB transposase expression vector.** *Prm1-PB transposase.* In order to obtain transgenic mice expressing PB transposase in the spermatogonia, the Prm1 promoter was amplified from C57 mouse genomic DNA<sup>64,65</sup> and cloned into multiple BglII and HindIII cloning sites of pGL3-Basic (Promega, Madison, WI, USA), after which the cDNA of the PB transposase amplified from pMA-Bac<sup>45</sup> was inserted into the HindIII and XhoI sites to generate pGL3-Prm1-PBBase. The correct reading frame was confirmed by sequencing. The mouse Prm1 upstream promoter-driven PB transposase expression vector is shown in Fig. 1B.

The primers used in the construction of the Prm1 upstream promoter-driven PB transposase expression vector were as follows:

FPrm1: 5' ctaagatctgttttactagagccacc 3'  
 RPrm1: 5' tcgaagcttcatggtgctggctggccg 3'  
 FPBase: 5' agcaagcttgatgttcttagacgatg 3'  
 RPBase: 5' atgctcagtcagaacaactttggcac 3'

*CMV-PB transposase.* The PB transposase cDNA was inserted into the HindIII and XhoI sites of pcDNA3.1 to generate CMV promoter-driven PB transposase expression vector CMV-PB.

**Construction of PB(PAS-trapping (EGFP)) and PB transposase transgenic mice.** *Construction of PB(PAS-trapping (EGFP)) transgenic mice.* CMV-PB was mixed with PB(PAS-trapping (EGFP)) at a mass ratio of 1:2 in the form of a circular plasmid. The mixed vectors were microinjected into fertilized mouse oocytes, after which the injected eggs were transplanted into the reproductive systems of pseudopregnant mice to produce pregnancy. Genomic DNA was isolated from the tails of newborn pups and used for genotype identification by PCR to identify PB(PAS-trapping (EGFP))-positive transgenic mice.

*Construction of PB transposase transgenic mice.* The Prm1-PBBase expression cassette was digested from pGL3-Prm1-PBBase with NotI and AlwNI. Positive transgenic mice containing the vector were generated by the microinjection technique described above.

**Construction of a transgenic mouse line doubly positive for PB(PAS-trapping (EGFP)) and Prm1-PBBase and generation of mice carrying insertion mutations.** PB(PAS-trapping (EGFP)) and Prm1-PBBase double-positive transgenic male mice (PB(PAS-trapping (EGFP))/Prm1-PBBase) were screened from the offspring of the PB(PAS-trapping (EGFP)) and Prm1-PBBase single-positive transgenic lines (Fig. 3). The PB(PAS-trapping (EGFP))/Prm1-PBBase mice were mated with wild-type female mice to produce offspring with various insertion mutations.

**PCR and sequence analysis.** Genomic DNA from the tail of each EGFP-positive mouse was extracted by phenol-chloroform. The concentration of DNA was measured by assessing the absorbance at 260 nm. The DNA were digested with restriction enzyme HaeIII or Sau3AI and stored overnight at 37 °C, followed by inactivation at 68 °C for 45 min and overnight ligation of the digested genomic DNA with T4 ligase at 16 °C (Fig. 4C). The resulting ligation products were amplified by PCR as follows: 94 °C for 5 min 30 s; 35 cycles of 94 °C for 30 s, 58 °C for 30 s, and 70 °C for 3 min; 72 °C for 10 min. After electrophoresis of the PCR products in 1.2% agarose gels, nucleic acid fragments were purified and imported into pMD18-T Simple vector (TaKaRa Bio Inc., Otsu, Shiga, Japan). Finally, after transformation of *Escherichia coli* cells with the pMD18-T Simple vector and clone selection, positive clones were sequenced.

The primers used in identification of insertion sites were as follows:

P1: 5' cctcgatatacagaccgataaaacacatgc 3'

P2: 5' agtcagtcagaacaacttggcacatc 3'

P3: 5' cttgacctgccacagaggactattagagg 3'

P4: 5' cagtgcactaccgcattgacaagcagc 3'

**Observation in mice under a fluorescence microscope.** Fluorescence was visualized using an inverted microscope (Nikon Eclipse TE300, Nikon, Tokyo, Japan). After selecting a 510/20-nm bandpass filter for EGFP detection, EGFP-positive mice were identified.

**Bioinformatics analysis.** The nucleotide sequences of captured genes were compared with those available from public genome databases, including the Mouse Genome Database at the University of California (UCSC) and Ensembl mouse genome database. In order to determine the specific position of the endogenous gene in which the trapped gene was located, sequence alignment was conducted using the UCSC database. BLAST homology searches were conducted using the non-redundant expressed sequence tag (EST) database of the US National Center of Biotechnology Information (NCBI). EST clone information, including raw data, was obtained from the NCBI Unigene database.

## References

- Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
- Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
- Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
- Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187–197 (2011).
- Snyder, M., Du, J. & Gerstein, M. Personal genome sequencing: current approaches and challenges. *Genes Dev* **24**, 423–431 (2010).
- Gonzaga-Jauregui, C., Lupski, J. R. & Gibbs, R. A. Human genome sequencing in health and disease. *Annu Rev Med* **63**, 35–61 (2012).
- Lappalainen, T. Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res* **25**, 1427–1431 (2015).
- Waldron, D. Functional genomics: Large-scale mouse phenotyping. *Nat Rev Genet* **16**, 499 (2008).
- Marx, V. A blooming genomic desert. *Nature methods* **11**, 135–138 (2014).
- Hilton, I. B. & Gersbach, C. A. Enabling functional genomics with genome engineering. *Genome Res* **25**, 1442–1455 (2015).
- Oliver, P. L. & Davies, K. E. New insights into behaviour using mouse ENU mutagenesis. *Hum Mol Genet* **21**, R72–81 (2012).
- Carlson, C. M. & Largaespada, D. A. Insertional mutagenesis in mice: new perspectives and tools. *Nat Rev Genet* **6**, 568–580 (2005).
- Lund, A. H. *et al.* Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. *Nat Genet* **32**, 160–165 (2002).
- Johnson, R. M. *et al.* MuLV-related endogenous retroviral elements and Flt3 participate in aberrant end-joining events that promote B-cell leukemogenesis. *Genes Dev* **28**, 1179–1190 (2014).
- Austin, C. P. *et al.* The knockout mouse project. *Nat Genet* **36**, 921–924 (2004).
- White, J. K. *et al.* Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**, 452–464 (2013).
- Rosen, B., Schick, J. & Wurst, W. Beyond knockouts: the International Knockout Mouse Consortium delivers modular and evolving tools for investigating mammalian genes. *Mamm Genome* **26**, 456–66 (2015).
- Ul Ain, Q., Chung, J. Y. & Kim, Y. H. Current and future delivery systems for engineered nucleases: ZFN, TALEN and RGEN. *J Control Release* **205**, 120–127 (2015).
- Gaj, T., Gersbach, C. A. & Barbas, C. F. 3<sup>rd</sup> ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol* **31**, 397–405 (2013).
- Hisano, Y. *et al.* Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep* **5**, 8841 (2015).
- Zambrowicz, B. P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
- Friedel, R. H. *et al.* Gene targeting using a promoterless gene trap vector (“targeted trapping”) is an efficient method to mutate a large fraction of genes. *Proc Natl Acad Sci USA* **102**, 13188–13193 (2005).
- Kontarakis, Z. *et al.* A versatile strategy for gene trapping and trap conversion in emerging model organisms. *Development* **138**, 2625–2630 (2011).
- Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
- De-Zolt, S. *et al.* High-throughput trapping of secretory pathway genes in mouse embryonic stem cells. *Nucleic Acids Res* **34**, e25 (2006).
- Mayasari, N. I. *et al.* Mixture of differentially tagged Tol2 transposons accelerates conditional disruption of a broad spectrum of genes in mouse embryonic stem cells. *Nucleic Acids Res* **40**, e97 (2012).
- Alvarado, M. C. *et al.* Gene trapping with firefly luciferase in Arabidopsis. Tagging of stress-responsive genes. *Plant Physiol* **134**, 18–27 (2004).
- Song, G., Li, Q., Long, Y., Hackett, P. B. & Cui, Z. Effective expression-independent gene trapping and mutagenesis mediated by Sleeping Beauty transposon. *J Genet Genomics* **39**, 503–520 (2012).
- Niwa, H. *et al.* An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J Biochem* **113**, 343–349 (1993).
- Wilusz, C. J., Wormington, M. & Peltz, S. W. The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* **2**, 237–246 (2001).

31. An, J. J. *et al.* Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* **134**, 175–187 (2008).
32. Bava, F. A. *et al.* CPEB1 coordinates alternative 3'-UTR formation with translational regulation. *Nature* **495**, 121–125 (2013).
33. Martin, G., Gruber, A. R., Keller, W. & Zavolan, M. Genome-wide analysis of pre-mRNA 3' end processing reveals a decisive role of human cleavage factor I in the regulation of 3' UTR length. *Cell Rep* **1**, 753–763 (2012).
34. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108**, 501–512 (2002).
35. Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev* **25**, 1770–1782 (2011).
36. Salminen, M., Meyer, B. I. & Gruss, P. Efficient poly A trap approach allows the capture of genes specifically active in differentiated embryonic stem cells and in mouse embryos. *Dev Dyn* **212**, 326–333 (1998).
37. Nagayoshi, S. *et al.* Insertional mutagenesis by the Tol2 transposon-mediated enhancer trap approach generated mutations in two developmental genes: *tcf7* and *synembryn*-like. *Development* **135**, 159–169 (2008).
38. Zhang, Z. & Saier, M. H., Jr. Transposon-mediated adaptive and directed mutations and their potential evolutionary benefits. *J Mol Microbiol Biotechnol* **21**, 59–70 (2011).
39. Mosimann, C. & Zon, L. I. Advanced zebrafish transgenesis with Tol2 and application for Cre/lox recombination experiments. *Methods Cell Biol* **104**, 173–194 (2011).
40. Yergeau, D. A., Kelley, C. M., Zhu, H., Kuliye, E. & Mead, P. E. Forward genetic screens in *Xenopus* using transposon-mediated insertional mutagenesis. *Methods Mol Biol* **917**, 111–127 (2012).
41. Yergeau, D. A. & Mead, P. E. Manipulating the *Xenopus* genome with transposable elements. *Genome Biol* **8** Suppl 1, S11 (2007).
42. Scali, C. *et al.* Post-integration behavior of a Minos transposon in the malaria mosquito *Anopheles stephensi*. *Mol Genet Genomics* **278**, 575–584 (2007).
43. Venken, K. J. *et al.* MiMIC: a highly versatile transposon insertion resource for engineering *Drosophila melanogaster* genes. *Nature methods* **8**, 737–743 (2011).
44. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221–226 (2005).
45. Ding, S. *et al.* Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**, 473–483 (2005).
46. Solodushko, V., Bitko, V. & Fouty, B. Minimal piggyBac vectors for chromatin integration. *Gene Ther* **21**, 1–9 (2014).
47. Grabundzija, I., Izsvak, Z. & Ivics, Z. Insertional engineering of chromosomes with Sleeping Beauty transposition: an overview. *Methods Mol Biol* **738**, 69–85 (2011).
48. Rio, D. C., Barnes, G., Laski, F. A., Rine, J. & Rubin, G. M. Evidence for *Drosophila* P element transposase activity in mammalian cells and yeast. *J Mol Biol* **200**, 411–415 (1988).
49. Shigeoka, T., Kawaichi, M. & Ishida, Y. Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells. *Nucleic Acids Res* **33**, e20 (2005).
50. Tsakiridis, A. *et al.* Expression-independent gene trap vectors for random and targeted mutagenesis in embryonic stem cells. *Nucleic Acids Res* **37**, e129 (2009).
51. Ishida, Y. & Leder, P. RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucleic Acids Res* **27**, e35 (1999).
52. Baker, K. E. & Parker, R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol* **16**, 293–299 (2004).
53. Keng, V. W. *et al.* A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol* **27**, 264–274 (2009).
54. Carlson, C. M. *et al.* Transposon mutagenesis of the mouse germline. *Genetics* **165**, 243–256 (2003).
55. Seto, M. *et al.* Alternative promoters and exons, somatic mutation and deregulation of the Bcl-2-Ig fusion gene in lymphoma. *Embo J* **7**, 123–131 (1988).
56. To, C. *et al.* The Centre for Modeling Human Disease Gene Trap resource. *Nucleic Acids Res* **32**, D557–559 (2004).
57. Matsuda, E. *et al.* Expression profiling with arrays of randomly disrupted genes in mouse embryonic stem cells leads to *in vivo* functional analysis. *Proc Natl Acad Sci USA* **101**, 4170–4174 (2004).
58. Osipovich, A. B., Singh, A. & Ruley, H. E. Post-entrapment genome engineering: first exon size does not affect the expression of fusion transcripts generated by gene entrapment. *Genome Res* **15**, 428–435 (2005).
59. Shaw, G. & Kamen, R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**, 659–667 (1986).
60. Mitra, R., Fain-Thornton, J. & Craig, N. L. piggyBac can bypass DNA synthesis during cut and paste transposition. *Embo J* **27**, 1097–1109 (2008).
61. Ivics, Z. *et al.* Transposon-mediated genome manipulation in vertebrates. *Nature methods* **6**, 415–422 (2009).
62. Claeys Bouuaert, C. & Chalmers, R. M. Gene therapy vectors: the prospects and potentials of the cut-and-paste transposons. *Genetica* **138**, 473–484 (2010).
63. Horie, K. *et al.* Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. *Mol Cell Biol* **23**, 9189–9207 (2003).
64. Zambrowicz, B. P., Harenda, C. J., Zimmermann, J. W., Brinster, R. L. & Palmiter, R. D. Analysis of the mouse protamine 1 promoter in transgenic mice. *Proc Natl Acad Sci USA* **90**, 5071–5075 (1993).
65. Peschon, J. J., Behringer, R. R., Brinster, R. L. & Palmiter, R. D. Spermatid-specific expression of protamine 1 in transgenic mice. *Proc Natl Acad Sci USA* **84**, 5316–5319 (1987).
66. Wallace, H. A. *et al.* Manipulating the mouse genome to engineer precise functional syntenic replacements with human sequence. *Cell* **128**, 197–209 (2007).
67. Georgi, B., Voight, B. F. & Bucan, M. From mouse to human: evolutionary genomics analysis of human orthologs of essential genes. *PLoS Genet* **9**, e1003484 (2013).
68. Schofield, P. N., Hoehndorf, R. & Gkoutos, G. V. Mouse genetic and phenotypic resources for human genetics. *Hum Mutat* **33**, 826–836 (2012).
69. Wang, W., Bradley, A. & Huang, Y. A piggyBac transposon-based genome-wide library of insertionally mutated Blm-deficient murine ES cells. *Genome Res* **19**, 667–673 (2009).
70. Rad, R. *et al.* PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**, 1104–1107 (2010).
71. Wang, W. *et al.* Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proc Natl Acad Sci USA* **105**, 9290–9295 (2008).
72. Chakraborty, S., Ji, H., Chen, J., Gersbach, C. A. & Leong, K. W. Vector modifications to eliminate transposase expression following piggyBac-mediated transgenesis. *Scientific reports* **4**, 7403 (2014).
73. Stanford, W. L., Cohn, J. B. & Cordes, S. P. Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* **2**, 756–768 (2001).
74. Korzh, V. Transposons as tools for enhancer trap screens in vertebrates. *Genome Biol* **8** Suppl 1, S8 (2007).
75. Schebelle, L. *et al.* Efficient conditional and promoter-specific *in vivo* expression of cDNAs of choice by taking advantage of recombinase-mediated cassette exchange using FLEX gene traps. *Nucleic Acids Res* **38**, e106 (2010).
76. Voss, A. K., Thomas, T. & Gruss, P. Efficiency assessment of the gene trap approach. *Dev Dyn* **212**, 171–180 (1998).
77. Fischer, S. E., Wienholds, E. & Plasterk, R. H. Regulated transposition of a fish transposon in the mouse germ line. *Proc Natl Acad Sci USA* **98**, 6759–6764 (2001).

78. Takeda, J., Keng, V. W. & Horie, K. Germline mutagenesis mediated by Sleeping Beauty transposon system in mice. *Genome Biol* **8** Suppl 1, S14 (2007).
79. Horie, K. *et al.* Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. *Proc Natl Acad Sci USA* **98**, 9191–9196 (2001).

### Acknowledgements

Supports from Tongji University Outstanding HR Plan (Grant no. 2014KJ081) and the National Natural Science Foundation of China (Grant no. 81100586) are acknowledged. This study also supported by China Postdoctoral Science Foundation (Project No. 2012M510905) and the Shanghai Postdoctoral Sustentation Fund (Project No. 12R21417200).

### Author Contributions

J.F. and L.L. designed the experiments; L.L. performed the research; L.L., L.S. and B.Z. analyzed the data; J.F., L.L. and P.L. wrote the manuscript. All authors reviewed the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Li, L. *et al.* PiggyBac transposon-based polyadenylation-signal trap for genome-wide mutagenesis in mice. *Sci. Rep.* **6**, 27788; doi: 10.1038/srep27788 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>