Effective Expression-Independent Gene Trapping and Mutagenesis Mediated by Sleeping Beauty Transposon

Guili Song a,b, Qing Li a, Yong Long a, Perry B. Hackett c, Zongbin Cui a,⁎

a The Key Laboratory of Aquatic Biodiversity and Conservation, Institute of Hydrobiology, Chinese Academy of Sciences, 7 Donghu Rd., Wuhu, Hubei 430072, China
b Graduate University of the Chinese Academy of Sciences, Beijing 100049, China
c Genetics, Cell Biology, and Development, University of Minnesota, 321 Church St. SE, Minneapolis, MN 55455, USA

Received 3 May 2012; revised 21 May 2012; accepted 28 May 2012
Available online 23 August 2012

ABSTRACT

Expression-independent gene or polyadenylation [poly(A)] trapping is a powerful tool for genome-wide mutagenesis regardless of whether a targeted gene is expressed. Although a number of poly(A)-trap vectors have been developed for the capture and mutation of genes across a vertebrate genome, further efforts are needed to avoid the 3′-terminal insertion bias and the splice donor (SD) read-through, and to improve the mutagenicity. Here, we present a Sleeping Beauty (SB) transposon-based vector that can overcome these limitations through the inclusion of three functional cassettes required for gene-finding, gene-breaking and large-scale mutagenesis, respectively. The functional cassette contained a reporter/selective marker gene driven by a constitutive promoter in front of a strong SD signal and an AU-rich RNA-destabilizing element (ARE), which greatly reduced the SD read-through events, except that the internal ribosomal entry site (IRES) element was introduced in front of the SD signal to overcome the phenomenon of 3′-bias gene trapping. The breaking cassette consisting of an enhanced splicing acceptor (SA), a poly(A) signal coupled with a transcriptional terminator (TT) effectively disrupted the transcription of trapped genes. Moreover, the Hsp70 promoter from tilapia genome was employed to drive the inducible expression of SB11, which allows the conditional remobilization of a trap insert from a non-coding region. The combination of three cassettes led to effective capture and disruption of endogenous genes in HeLa cells. In addition, the Cre/LoxP system was introduced to delete the Hsp70-SB11 cassette for stabilization of trapped gene interruption and biosafety. Thus, this poly(A)-trap vector is an alternative and effective tool for identification and mutation of endogenous genes in cells and animals.

KEYWORDS: Poly(A) trapping; Sleeping Beauty transposon; Insertional mutagenesis; HeLa cells; Zebrafish embryos

1. INTRODUCTION

Gene trapping is an efficient approach for insertional mutagenesis of genes in a target genome. A conventional gene-trap vector consists of a promoterless marker/reporter gene flanked by an upstream splice acceptor and a downstream poly(A) signal (Gossler et al., 1989; Stanford et al., 2001).

When inserted into either an intron or an exon of expressed genes, the marker/reporter gene can be transcribed from the endogenous gene promoter and the transcription is terminated prematurely at the inserted poly(A) site. The nuclear pre-mRNA is then spliced to form a fusion transcript with the trapped exon immediately upstream of the insertion site. Since the processed fusion transcript encodes a truncated, often non-functional version of endogenous protein and the marker/reporter, gene trapping is employed to elucidate gene functions by disrupting expression of trapped genes across a target
For insertional mutagenesis in vertebrates (Medico et al., 2001; Ellingsen et al., 2005), these vectors can be extremely efficient in generating mutations, but have three major disadvantages: limitation in packaging size, mediation of gene expression in undifferentiated cells, and attempts are still undergoing to overcome other problems. In addition, approximately 70% of the protein-encoding genes in the mouse genome have been disrupted by gene trap insertions and some of ESC lines harboring mutations in a single gene have been used for the generation of null mice (De-Zolt et al., 2009). However, the achievement of saturation mutagenesis in a target genome through conventional vectors is difficult because these vectors are not able to capture genes that are not or poorly expressed in undifferentiated cells.

To circumvent this problem, a number of expression-independent or poly(A)-trap vectors have been developed and successfully utilized for the identification of differentially expressed genes (Chen et al., 2004; Lin et al., 2006; Tsakiridis et al., 2009). Basic poly(A)-trap vector contains a reporter/selectable marker gene flanked by an upstream constitutive promoter and a downstream splice donor (SD). Integration of a trap cassette upstream of a functional poly(A) sequence from endogenous genes leads to the generation of a stable pre-mRNA and the proper splicing between the trap SD with its SD signal (Shigeoka et al., 2005). However, the majority of poly(A)-trap vectors have a bias toward the last introns of trapped genes because of the activation of nonsense-mediated mRNA decay (NMD) mechanism (Baker and Parker, 2004). NMD promotes the selection of trapping events in the 3'-most intron of target genes as it triggers the degradation of the selectable marker’s transcript based on the presence of a premature termination codon. This disadvantage appears to be resolved in the UPATrap vector through insertion of an internal ribosomal entry site (IRES) and three initiation codons in front of its SD signal (Shigeoka et al., 2005). However, the majority of poly(A)-trap vectors still suffer from an NMD-based handicap and attempts are still undergoing to overcome other problems such as the background SD read-through events. In addition, gene-trap vectors available appear to each have their own insertional “hot spots” (Nord et al., 2007). Thus, the saturation mutagenesis of a target genome can be achieved most economically through the use of a wider range of vectors and approaches (Skarnes, 2005).

Pseudotyped retroviral vectors are now extensively utilized for insertional mutagenesis in vertebrates (Medico et al., 2001; Ellingsen et al., 2005). These vectors can be extremely efficient in generating mutant, but have three major disadvantages: limitation in packaging size, mediation of gene silencing and strong insertional biases (Ellis, 2005; Uren et al., 2005). Recently, several transposon-based vectors have been developed as alternatives for elucidation of gene functions in mouse and zebrafish (Kawakami et al., 2004; Jenkins et al., 2005; Largaespada et al., 2005; Rad et al., 2010; Clark et al., 2011). In comparison with viral vehicles, transposons can carry a large DNA fragment up to 10 kb (Geurts et al., 2003; Zayed et al., 2004; Huang et al., 2010). However, transposon-based vectors have a caveat that less than 10 copies of transposons are usually found in the genome of transgenic animals (Clark et al., 2004; Kawakami et al., 2004). Moreover, over produced transposases appear to be harmful to cellular growth and proliferation (Huang et al., 2010; Gall et al., 2011). Therefore, new strategies were urgently needed to generate novel transposon-based vectors for efficient gene trapping and insertional mutagenesis in vertebrates.

Sleeping Beauty (SB) transposon system is a synthetic member of the Tc1/mariner superfamily, which can transpose in an autonomous and horizontal manner (Plasterk et al., 1999). Sequences related to Tc1/mariner transposons are found in a wide range of host genomes and an active SB transposase gene was resurrected from an extinct DNA transposon-based vectors have a bias that less than 10 copies of transposons are usually found in the genome of transgenic animals (Clark et al., 2004; Kawakami et al., 2004). Moreover, over produced transposases appear to be harmful to cellular growth and proliferation (Huang et al., 2010; Gall et al., 2011). Therefore, new strategies were urgently needed to generate novel transposon-based vectors for efficient gene trapping and insertional mutagenesis in vertebrates.

In this study, we described the generation of an effective poly(A)-trap vector through a combination of three functional cassettes. In the gene-finding cassette, an IRES from the encephalomyocarditis virus (EMCV) was inserted downstream of the selectable marker/reporter gene to suppress the NMD-mediated mRNA degradation of trapped genes and an AU-rich RNA-destabilizing element (ARE) was introduced into an intron in the SD to enrich for authentic trapping events (Xu et al., 1997). In the gene-breaking cassette, an exon containing a strong SA from the carp (Cyprinus carpio) β-actin gene (Liu et al., 1990) and an efficient poly(A) signal followed by a transcriptional terminator element from the human gastrin gene were used to disrupt the expression of trapped genes.

The tilapia HSP70 promoter (Molina et al., 2001) was employed to drive the expression of SB11 transposase, which allows the remodeling of integrated traps from non-coding sites to new locations and thus increases the opportunity of trapping and mutating endogenous genes. Activities of all components
in this vector were carefully tested in HeLa cells and zebrafish embryos. This novel poly(A)-trap vector is demonstrated to be highly effective in the identification and disruption of endogenous genes in HeLa cells, so it could be a powerful tool for the large-scale genetic screen of functional genes that are essential for embryonic development, organogenesis and the progression of human diseases in cells and model animals.

2. MATERIALS AND METHODS

2.1. Construction of the poly(A)-trap vector

Our trap-vector has been designed to tag as well as to break genes simultaneously (Fig. 1A). These cassettes were sequentially subcloned into the second generation of SB transposon-mediated poly(A)-trap vectors. Images were taken under a Nikon TE2000 fluorescent microscope at 36 h after transfection. Zebrafish embryos at one-cell stage were microinjected with the pSPL3/Finding(intron). Injected embryos at 24 hpf were imaged under a SteReo Lumar V12 microscope from Zeiss.

Fig. 1. Activity of the gene-finding cassette in an intron.

A: a novel poly(A)-trap vector mediated by SB transposon. IR/DR(L) and IR/DR(R), left and right inverted/directed repeats of the SB transposon; SA, splice acceptor from carp β-actin gene; Exon, the exon 2 from the carp β-actin gene; SV40 polyA, SV40 polyadenylation sequence; TT, transcript terminator sequence. TiHsp70, tilapia Hsp70 promoter; SB11, SB11 transposase gene; polyA, a poly(A) signal from carp β-actin gene; SV40, SV40 promoter; Reporter, kanamycin (Neo) or EGFP gene; IRES, internal ribosome entry site; SD, splice donor; ARE, AU-rich RNA-destabilizing element. B: the finding cassette was subcloned at the Sac II/Xho I site of a modified pSPL3 vector to generate the pSPL3/Finding(intron), which was used for transient transfection of HeLa cells at 80% confluence. Images were taken under a Nikon TE2000 fluorescent microscope at 36 h after transfection. Zebrafish embryos at one-cell stage were microinjected with the pSPL3/Finding(intron). Injected embryos at 24 hpf were imaged under a SteReo Lumar V12 microscope from Zeiss.
transposon pT2/HB (Cui et al., 2002). An exon-trapping plasmid pSPL3 (Wang et al., 2005) was utilized for testing the transcriptional and splicing activities of gene-finding and breaking cassettes. The TiHsp70 promoter was obtained from the tilapia genome using primers 5'-ctgctgagagcctcgagactctcgtg-3' and 5'-gaccgggtctttgacttcg-3', then inserted at the Nhe I/ Age I site upstream of SB11 transposase gene. The SA and SD sequences from carp β-actin gene were amplified using primers 5'-cttgctagccatttcgacagcactcgc-3' and 5'-agagctgtatctgactacgttcg-3', then inserted at the Nhe I/ Age I site of plasmid pSPL3 (Wang et al., 2005) to generate a transposon pT2/HB (Cui et al., 2002). An exon-trapping plasmid pSPL3 was utilized for testing the activities of gene-finding and breaking cassettes in the poly(A)-trap vector to determine the relative expression levels of reporter gene and exons in the components of pSPL3-derived vectors by qRT-PCR assays. The nested PCR products were separated on the 1.5% agarose gel and specific DNA bands were purified and cloned into pZero2/TA vector for sequencing. DNA sequences of PCR products were used for blast the human genome in the ENSEMBL and NCBI database.

2.4. Splinkerette-PCR assays

Genomic DNAs were extracted from individual G418-resistant single-cell colonies. Splinkerette-PCR assays were performed as described in previous studies (Cui et al., 2002; Uren et al., 2009) to obtain the genomic DNA sequences adjacent to inserted SB transposons. The purified genomic DNA was digested with BfaI and the digested genomic DNAs were ligated with a linker obtained from the annealing of long-stand and short-stand oligos. The primary PCR was performed under the following conditions: 1 cycle at 95°C for 1 min; 10 cycles at 95°C for 10 s and 65°C for 2 min, decrease 0.5°C per cycle; 20 cycles at 95°C for 10 s and 65°C for 2 min; 70°C for 10 min. The first-round PCR product was diluted for the nest-PCR assays under the conditions: 1 cycle at 95°C for 2 min; 30 cycles at 95°C for 20 s, 61°C for 30 s and 72°C for 2 min; 72°C for 10 min. Oligos and primers are listed in Table S1. The nested PCR products were separated on the 1.5% agarose gel and specific DNA bands were purified and cloned into pZero2/TA vector for sequencing. DNA sequences of PCR products were used for blast the human genome in the ENSEMBL and NCBI database.

2.5. Transcriptional expression analysis

Total RNA samples were prepared from transfected cells, embryos or individual cell colonies using the TRIZOL reagent from Invitrogen, and treated with RNase-free DNase at 37°C for 30 min and then at 85°C for 10 min. The first-strand cDNAs were synthesized using oligo(dT) primers in the RevertAid™ First Strand cDNA Synthesis Kit from Fermentas (USA) according to the manufacturer’s instructions. Various fusion transcripts of genes from cells transfected with pSPL3-derived vectors and G418-resistant cell colonies were examined under the conditions: 1 cycle at 95°C for 5 min; 30 cycles at 95°C for 30 s, 60°C for 30 s and 72°C for 3 min; 72°C for 10 min. Primers used for transcriptional analysis are listed in Table S1. The PCR products were subjected to 1% agarose gel electrophoresis and sequencing.

Quantitative real-time PCR (qRT-PCR) assays were conducted to determine the relative expression levels of reporter gene and exons in the components of pSPL3-derived vectors by using the SYBR® Green Real-time PCR Master Mix from TOYOBO (Japan) on the Bio-Rad iQ5 2.0 machine. An absolute quantification method was used to measure the copy numbers of transcripts for exons in pSPL3/Breaking(intron) and pSPL3-E3/Breaking(exon). A 10-fold dilution series containing 10^2–10^6 copies of molecules was prepared from a template sample of known concentration. The serial 10-fold dilutions and samples were then assayed in the same run. A standard curve was obtained by plotting cycle threshold (Ct) values against log-transformed concentrations of serial 10-fold dilutions (Fig. S1). The copy numbers of transcripts in each sample were calculated through a comparison of Ct values from the standard curve.

To quantitatively determine the relative expression level of trapped genes in single-cell colonies, gene-specific primers (18–22-mer, Table S1) that span at least one intron were

2.2. Cells and zebrafish embryos

HeLa cells (ATCC® CCL-2TM) were cultured under atmospheric condition (95% air and 5% CO2) at 32°C in Dulbecco’s modified Eagle medium from GIBCOTM-Invitrogen Corporation (USA), containing fetal bovine serum (10%, w/v), penicillin (100 U/mL), streptomycin (100 mg/mL) and amphotericin B (2 μg/mL). The culture medium was replaced two to three times a week.

AB inbred strain of zebrafish (Danio rerio) was reared in a re-circulating water system and maintained at standard conditions. Naturally fertilized zebrafish embryos were staged by hours post-fertilization (hpf) and embryos at one-cell stage were microinjected with plasmids containing one of functional cassettes in the poly(A)-trap vector to determine the activities of gene-finding and breaking cassettes in vivo.

2.3. Transfection and selection of genetin-resistant cell colonies

One day prior to transfection, approximate 3 × 10^5 HeLa cells in 2 mL of culture medium were seeded on 35 mm culture dishes and cultured overnight. Cells at 70%–80% confluence were transfected with the FuGENE 6 reagent from Roche (USA). Two days after transfection, the cells were trypsinized and 10% of these cells in selective medium containing 800 μg/mL of genetin (G418) were evenly seeded onto 10 cm dishes. The selective medium was replaced twice a week until the formation of single-cell colonies. Different single-cell colonies were separately harvested and expanded in medium containing 300 μg/mL of G418 for further analysis.
designated to amplify 150–200 bp fragments from the genomic DNA adjacent to the inserted SB transposons. PCR reactions were run in triplicates on 96-well plates and each reaction contains 5 μL of diluted (1:10) cDNA template from 2 μg of total RNA, 100 mmol/L of each primer and 10 μL of 2 × SYBR Green I Master Mix in a volume of 20 μL. The reaction conditions are as follows: 1 cycle at 95°C for 3 min; 40 cycles at 95°C for 10 s and 60°C for 30 s; 26 cycles at 70°C for 30 s. The expression of GAPDH gene was used as the reference to calculate the relative expression of trapped genes in single-cell colonies using the 2^-ΔΔCt method (Livak and Schmittgen, 2001). All of specific primers are listed in Table S1.

2.6. Southern blotting

HeLa cells (3 × 10^5) were harvested by centrifugation at 1200 r/min for 5 min. The cell pellets were re-suspended in 300 μL 1 × PBS buffer (0.8% NaCl, 0.02% KCl, 0.144% Na2HPO4, 0.024% KH2PO4) and then lysed at 65°C for 6 h by addition of 300 μL DNA extraction buffer containing 10 mmol/L Tris (pH 8.0), 100 mmol/L EDTA (pH 8.0), 0.5% SDS, and 400 μg/mL protease K. The total DNA was purified by using the E.Z.N.A.™ Tissue/DNA Kit from OMEGA/ Bio-Tek (USA). Total genomic DNA (~20 μg) was digested with the EcoRI I at 37°C overnight, separated on a 0.7% agarose gel and transferred onto a positively charged nylon membrane from Roche. The probes were obtained from Neo or SB11 coding sequences by PCR amplification with primer pairs of NEO-F/NEO-R (Table S1), and labeled according to the DIG High Prime DNA Labeling and Detection Starter Kit II from Roche. Hybridization and immunological detection were processed according to the manufacturer’s procedures.

2.7. Remobilization of integrated transposons

The remobilization of SB transposons integrated in the genome of HeLa cells was examined by placing the G418-resistant HeLa cells at 37°C for 5 days in order to induce the expression of SB11 transposase gene. Since transposition events occur in only a small portion of cells and G418-resistant cells in the same colony may contain trap inserts at different loci, we designed two reverse primers containing 7 base-pair sequences (TACAGTA or TACTGTA) at their 3' ends to amplify a DNA fragment around the footprints after SB transposon remobilization using a PCR-based method described previously (Liu et al., 2004). PCR assays were performed under the conditions: 95°C for 5 min; 34 cycles at 95°C for 30 s, 58°C for 30 s and 72°C for 30 s; 72°C for 10 min. PCR products were purified and sequenced.

2.8. Statistical analysis

Data were expressed as means ± standard deviation and student’s t-test was performed using the SPSS version 15.0 for windows (Illinois, USA) to determine the significant difference (P < 0.05 or P < 0.01) between two groups.

3. RESULTS

3.1. Generation of a novel poly(A)-trap vector

The design of our novel poly(A)-trap vector is based on three principles: that it should be sensitive in finding genes across a target genome, effective in disrupting gene expression and suitable for large-scale mutagenesis. To reach these goals, we constructed a SB transposon-based poly(A)-trap vector pT2/Poly(A)-Trap (Fig. 1A), which contains three functional cassettes that are required for gene-finding, gene-breaking and large-scale mutagenesis, respectively. In the gene-finding cassette, a modified EMCV/IRES sequence from the pIRE2-EGFP (Clontech, USA) was placed downstream of the reporter gene encoding neomycin or EGFP. In addition, a modified SD from the exon/intron 1 of carp β-actin gene (Liu et al., 1990) and an ARE from the human GM-CSF gene (Xu et al., 1997) were inserted downstream of the IRES to reduce the background SD read-through events. The gene-breaking cassette consists of an enhanced SA from carp β-actin gene (Liu et al., 1990), an SV40 poly(A) signal and a transcriptional terminator sequence from human gastrin gene (Sato et al., 1986). The combination of these elements assured the disruption of trapped gene expression. In the efficiency cassette, the activation of the tilapia Hsp70 promoter (TiHsp70) at 37°C allows the inducible expression of SB11 transposase, which is necessary for the remobilization of trapping inserts from non-coding sites. Moreover, the Cre/LoxP system was utilized to mediate the deletion of TiHsp70-SB11 cassette for stabilization of trapped gene interruption and biosafety. Activities of major components in pT2/Poly(A)-Trap were carefully examined in HeLa cells and developing zebrafish embryos.

3.2. Activities of gene-finding cassette components in an intron

Although there are several versions of poly(A)-trapping vectors that are used in various vertebrate systems (Chen et al., 2004; Lin et al., 2006; Tsakiridis et al., 2009), there is room for improvement to avoid potential 3'-bias of integration and SD read-through. Accordingly, we generated a gene-finding cassette by insertion of a modified IRES at the eleventh ATG site downstream of the reporter gene and an ARE element into the intron of SD signal. The activity of gene-finding cassette was examined by amplification of a Sac II/Hho I fragment in pT2/PolyA-Trap into the multiple cloning sites (MCS) of pSPL3 to generate a pSPL3/Finding (intron) vector (Fig. 1B). The pSPL3 vector is originally designed to search potential exon sequences in a genomic DNA fragment. As shown in Fig. 1B, strong EGFP expression was detected under a fluorescence microscope in 88% of HeLa cells transfected with pSPL3/Finding(intron) plasmids. Similarly, EGFP expression was found in 78% of embryos microinjected with pSPL3/Finding(intron) plasmids.

To examine whether the poly(A) signal and ARE element are essential for the EGFP expression, three pSPL3/
Finding(intron)-derived constructs including pSPL3/Finding(intron)ΔpolyA, pSPL3/Finding(intron)ΔARE and pSPL3/Finding(intron)ΔAREΔpolyA, were generated (Fig. 2A). As shown in Fig. 2B, strong EGFP expression was found in 82% of HeLa cells transfected with pSPL3/Finding(intron), but weak EGFP expression was detected in 5% of cells transfected with pSPL3/Finding(intron)ΔpolyA. Similarly, deletion of the poly(A) signal significantly reduced the ratio of EGFP-expressing cells transfectected with pSPL3/Finding(intron)ΔpolyA. It is worth noting that the ratio of EGFP-expressing cells in two poly(A)-containing constructs (from 82% to 80%) was not significantly affected by deletion of the ARE, but markedly increased in two non-poly(A)-containing constructs (from 5% to 44%) by deletion of the ARE. These data indicate that the ARE is necessary for the reduction of the SD read-through background events. Numbers of EGFP-expressing cells transfected with these vectors are summarized in Fig. 2C.

To characterize further the effects of the poly(A) signal and the ARE element on the stability of EGFP mRNA, total RNA was extracted from transfected cells and qRT-PCR was performed to determine the relative levels of EGFP mRNA (Fig. 2D). The results showed that the relative levels of EGFP mRNA were significantly higher in cells transfected with pSPL3/Finding(intron)ΔpolyA than in cells transfected with pSPL3/Finding(intron). These data suggest that the ARE element is crucial for the reduction of SD read-through events and that the poly(A) signal is important for the stability of EGFP mRNA.
mRNA was isolated for qRT-PCR analysis 24 h after transfection with pSPL3/Finding(intron), pSPL3/Finding(intron)ΔpolyA, pSPL3/Finding(intron)ΔARE and pSPL3/Finding(intron)ΔARE/polyA. As shown in Fig. 2D, the relative EGFP mRNA level in pSPL3/Finding(intron) ΔpolyA-transfected HeLa cells was significantly lower than that in pSPL3/Finding(intron)-transfected HeLa cells (P < 0.01) and deletion of the ARE in pSPL3/Finding(intron) ΔpolyA led to a 14-fold increase in the relative EGFP mRNA level. Similar results were obtained from zebrafish embryos microinjected with these constructs (Fig. S2). Thus, the poly(A) signal is necessary for the stabilization of EGFP mRNA and the inclusion of an ARE element markedly destabilized the SD read-through EGFP mRNA.

Next, we addressed whether the EGFP expression in HeLa cells and developing embryos results from the proper splicing of the SD sequence with the SA2 in the exon 2 of pSPL3/Finding(intron). Total mRNA was isolated from HeLa cells transfected with the pSPL3/Finding(intron). As shown in Fig. 3, two potential transcript variants I and II were analyzed using reverse transcription PCR (RT-PCR). A 1.5-kb band was mainly amplified from cDNAs of pSPL3-Finding(intron)-transfected HeLa cells and a 270 bp band was found in the RT-PCR products from the same cDNA sample. Sequencing results indicate that the 1.5-kb band represents the splice variant I containing EGFP, IRES, SD and exon 2, and that the 270 bp band is derived from the proper splicing of exon 1 with exon 2 in pSPL3. Similar results were obtained in zebrafish embryos microinjected with the pSPL3/Finding(intron) or pSPL3 vector (data not shown). Therefore, the modified SD signal in the gene-finding cassette is able to efficiently direct the proper splicing of the EGFP gene transcript with a downstream exon in a trapped gene.

Taken together, the expression of reporter gene in the gene-finding cassette of our poly(A)-trap vector requires a downstream poly(A) signal and the combination of SD and ARE elements can efficiently reduce the SD read-through background events.

3.3. Activity of the gene-finding cassette in an exon

To examine the activity of the gene-finding cassette after integration into an exon of trapped genes, an exon (exon 3) from carp β-actin gene (Liu et al., 1990) was subcloned into HeLa cells at 80% confluence were transfected with the pSPL3/Finding(intron) or pSPL3 plasmids. Potential splice variants (I and II) and RT-PCR primers are shown in the upper panel and lower left panels. RT-PCR analysis of transcripts was shown in the lower right panel. Sequencing results indicate the 1.5-kb band is derived from the splice variant I and the 270 bp band from splice variant II.
the MCS of pSPL3 to generate a pSPL3-E3 vector (Fig. 4). The gene-finding cassette was then inserted at the Sac II/Xho I site of pSPL3-E3 vector to generate the pSPL3-E3/Finding(exon) for transient transfection of HeLa cells at 80% confluence or microinjection of zebrafish embryos at one-cell stage. Transfected cells were imaged under a Nikon TE2000 fluorescent microscope at 36 h after transfection. Injected embryos at 24 hpf were imaged under a SteReo Lumar V12 microscope from Zeiss.

To investigate whether EGFP expression resulted from the proper splicing of the modified SD with a downstream exon, total mRNA was isolated from pSPL3-E3/Finding(exon)-transfected HeLa cells and three potential transcript variants (I + III, I + II and IV) were analyzed using RT-PCR. As shown in Fig. 5, two bands (2090 bp and 1500 bp) were amplified using the primer pair (F1 + R1) and three bands (2661 bp, 2105 bp and 270 bp) using the primer pair (F1 + R1) from the cDNAs of pSPL3-E3/Finding(exon)-transfected HeLa cells. In addition, two PCR products (411 bp and 270 bp) were obtained from the pSPL3-E3-transfected HeLa cells. Sequencing results indicate that the 2090 bp and 2661 bp bands are derived from the fusion splice variant I + III, the 1500 bp and 2105 bp bands from the fusion transcript I + II, the 411 bp band from the proper splicing of exon 1, exon 3 and exon 2 in pSPL3-E3, and the 270 bp band from the proper splicing of exon 1 and exon 2 in pSPL3. Similar results were obtained from zebrafish embryos injected with pSPL3-E3/Finding(exon) and pSPL3-E3 vectors (data not shown). Thus, the modified SD signal is able to properly splice with a downstream SA in a trapped gene after insertion into an exon. Moreover, the insertion of the gene-finding cassette completely truncated the exon at the insertion site.
Taken together, these data indicate that the gene-finding cassette integrated into an endogenous exon is able to efficiently utilize the poly(A) signal of trapped genes for reporter gene expression.

3.4. Activity of the gene-breaking cassette in an intron

We subcloned the gene-breaking of our poly(A)-trap vector into the Sac II/Xho I site of pSPL3 vector to produce pSPL3/Breaking(intron) (Fig. 6A), which was then utilized for HeLa cells transfection and zebrafish embryos microinjection. Total mRNA from HeLa cells or zebrafish embryos at 24 hpf was isolated for RT-PCR analysis of two potential transcript variants I and II (Fig. 6A). As shown in Fig. 6B, a 169 bp PCR band was mainly obtained using primers (F1 and R) and a 270 bp band was sometimes amplified using primers (F1 and R1) from pSPL3/Breaking(intron)-transfected HeLa cells. Sequencing results indicate that the 169 bp band represents the splice variant containing exon 1 and partial SA-exon, and the 270 bp product results from the weak splicing of exon 1 with exon 2 in pSPL3/Breaking(intron) and pSPL3. Similar results were obtained in zebrafish embryos injected with the pSPL3/Breaking(intron) or pSPL3 vector (data not shown). These data indicate that the gene-breaking cassette in our poly(A)-trap vector is able to efficiently disrupt the expression of trapped gene once inserted into an intron.

3.5. Activity of the gene-breaking cassette in an exon

To examine further the ability of the gene-breaking cassette in disruption of trapped gene expression, qRT-PCR assays were performed to determine the copy numbers of exon 1 and exon 2 after reverse transcription of total mRNA from pSPL3/Breaking(intron)-transfected HeLa cells. As shown in Fig. 6C, the copy number of exon 2 transcripts was about 20 times lower than that of exon 1 in pSPL3/Breaking(intron)-transfected HeLa cells (P < 0.01), but the copy numbers of exon 1 and exon 2 were nearly identical in pSPL3-transfected cells. Similar results were obtained from zebrafish embryos injected with the pSPL3/Breaking(intron) or pSPL3 vector (data not shown). These data indicate that the gene-breaking cassette in our poly(A)-trap vector is able to efficiently disrupt the expression of trapped gene once inserted into an intron.
three bands (834 bp, 675 bp and 270 bp) using the primer pair (F1 + R1) from pSPL3-E3/Breaking(exon)-transfected HeLa cells. In addition, two PCR products (411 bp and 270 bp) were obtained from the pSPL3-transfected HeLa cells. Sequencing data indicate that the 328 bp and 834 bp bands are derived from the fusion splice variant I + III, the 169 bp and 675 bp bands from the fusion transcript II + III, the 411 bp band from the proper splicing of exon 1, exon 3 and exon 2 in pSPL3-E3, and the 270 bp band from the proper splicing of exon 1 and exon 2 in pSPL3 and pSPL3-E3. Similar results were obtained in zebrafish embryos injected with the pSPL3-breaking(exon) or pSPL3-E3 vector (data not shown). Thus, insertion of the gene-breaking cassette into an exon can efficiently abrogate the proper splicing of downstream exons in trapped genes in addition to truncation of the trapped exon.

3.6. Activity of poly(A)-trap inserts in HeLa cells

To evaluate the activity of our poly(A)-trap vector in trapping and mutagenesis of endogenous genes, HeLa cells growing at 32°C were transfected with the poly(A)-trap vector and transfected cells were then incubated at 37°C for 2 days before G418 selection. By this way, we obtained plenty of single-cell colonies for further analysis of trap insertion site and remobilization (Fig. S3). The transposition events were analyzed in some of G418-resistant single-cell colonies by using splinkerette-PCR. Blasting the human genome in the ENSEMBL database indicated the majority of poly(A)-traps land in an intron of endogenous genes and one of them in
an exon (Table 1). In addition, 88% (21 out of 24) of trapping events from the ARE-containing vector integrated in the sense strand of trapped genes, while deletion of the ARE resulted in only 14% (2 out of 14) of trapping events inserting in the sense strand of trapped genes (Table S2). The inclusion of ARE element in the poly(A)-trap vector markedly improved the enrichment of authentic poly(A)-trapping events by reduction of the SD read-through background. These data suggest an advantage of our ARE-containing poly(A)-trap vector over the non-ARE-containing one and other poly(A)-trap vectors with a reverse insertion preference (Chen et al., 2004). Furthermore, trap inserts appeared to randomly integrate into targeted genes since insertions generated by our vector did not exhibit the severe 3' bias observed with the widely used poly(A)-trap vectors (Ruley et al., 2005; Shigeoka et al., 2005). Thus, the employment of an IRES element in our poly(A) trapping vector is able to reduce the 3' bias.

To examine the effects of poly(A)-trap inserts on the transcription of trapped endogenous genes, RT-PCR was performed with mRNA samples isolated from four single-cell colonies. One trapped gene in each colony was selected and analyzed using two pairs of primers (F1 + R1 and F2 + R2). As shown in Fig. 8B, a small DNA fragment (~200 bp) was detected with primers (F1 + R1) and a large DNA fragment (~800 bp) was obtained with primers (F2 + R2) from four of randomly selected cell colonies. Sequencing results indicate that these PCR products resulted from the proper splicing of the SA signal with an upstream exon in trapped genes. To further check whether the transcriptional expression...
of four trapped genes was affected by the poly(A)-trap insertion, qRT-PCR assays were conducted by using mRNA samples from normal HeLa cells (N) and cell colonies (C1—C4). As shown in Fig. 8C, the relative mRNA levels of four trapped genes was reduced to 25%—40% of those in normal HeLa cells and was significantly lower than those in other cell colonies (P < 0.01 or <0.05 in all cases). Therefore, the novel poly(A)-trap system is suitable for the capture and interruption of endogenous genes.

3.7. Remobilization of poly(A)-trap inserts

A growing body of evidence indicates that the efficiency of transposon-based vectors can be improved by inducible remobilization of integrated trap cassettes to new genomic sites. Here, we tested the possibility to improve the poly(A)-trap efficiency through remobilization of trap inserts in the genome of G418-resistant cell colonies. SB11 expression under the regulation of the TiHsp70 promoter (which is active at 37°C but not 32°C) was induced in six cell colonies to allow remobilization of our trap cassettes. Total genomic DNA was subjected to excision assays as previously described (Yant and Kay, 2003; Liu et al., 2004), using a reverse primer (R) against the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Sequencing results indicate that these PCR products from the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Genomic DNA-transposon junction sequences were amplified by using the splinkerette-PCR. Partial genome sequences were shown in italic form on the left of TA and partial trap cassette sequences in regular form on the right of TA.

3.7. Remobilization of poly(A)-trap inserts

A growing body of evidence indicates that the efficiency of transposon-based vectors can be improved by inducible remobilization of integrated trap cassettes to new genomic sites. Here, we tested the possibility to improve the poly(A)-trap efficiency through remobilization of trap inserts in the genome of G418-resistant cell colonies. SB11 expression under the regulation of the TiHsp70 promoter (which is active at 37°C but not 32°C) was induced in six cell colonies to allow remobilization of our trap cassettes. Total genomic DNA was subjected to excision assays as previously described (Yant and Kay, 2003; Liu et al., 2004), using a reverse primer (R) against the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Sequencing results indicate that these PCR products from the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Genomic DNA-transposon junction sequences were amplified by using the splinkerette-PCR. Partial genome sequences were shown in italic form on the left of TA and partial trap cassette sequences in regular form on the right of TA.

3.7. Remobilization of poly(A)-trap inserts

A growing body of evidence indicates that the efficiency of transposon-based vectors can be improved by inducible remobilization of integrated trap cassettes to new genomic sites. Here, we tested the possibility to improve the poly(A)-trap efficiency through remobilization of trap inserts in the genome of G418-resistant cell colonies. SB11 expression under the regulation of the TiHsp70 promoter (which is active at 37°C but not 32°C) was induced in six cell colonies to allow remobilization of our trap cassettes. Total genomic DNA was subjected to excision assays as previously described (Yant and Kay, 2003; Liu et al., 2004), using a reverse primer (R) against the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Sequencing results indicate that these PCR products from the target genomic DNA and a forward primer (F) containing the typical SB footprints TAC(A/T)GTA at its 3'-terminus and the target genome sequence (Fig. 9A). The excision of a SB transposon from its original integration site gave rise to a PCR product in cells at 37°C, but not in cells at 32°C (Fig. 9B).

Genomic DNA-transposon junction sequences were amplified by using the splinkerette-PCR. Partial genome sequences were shown in italic form on the left of TA and partial trap cassette sequences in regular form on the right of TA.

Table 1
Trapped endogenous genes from geneticin-resistant HeLa cells transfected with ARE-containing vector

<table>
<thead>
<tr>
<th>Junction sequence</th>
<th>Gene name</th>
<th>GenBank accession No.</th>
<th>Chromosome name</th>
<th>Exon number</th>
<th>Insertion site</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>tttgtttttttTAcatgttgaag</td>
<td>SPP2</td>
<td>NG_008668.1</td>
<td>Chr. 2</td>
<td>8</td>
<td>Intron 4</td>
<td>F</td>
</tr>
<tr>
<td>cggattgagacccATcaggttgaag</td>
<td>ADAMTS19</td>
<td>NC_000005.9</td>
<td>Chr. 5</td>
<td>23</td>
<td>Intron 14</td>
<td>F</td>
</tr>
<tr>
<td>gttctagctgtcTAcaggttgaag</td>
<td>HECW1</td>
<td>NM_015052.3</td>
<td>Chr. 7</td>
<td>30</td>
<td>Intron 2</td>
<td>F</td>
</tr>
<tr>
<td>atccctctcgtgATcaggttgaag</td>
<td>EFHC2</td>
<td>NG_021391.1</td>
<td>Chr. X</td>
<td>15</td>
<td>Intron 11</td>
<td>F</td>
</tr>
<tr>
<td>ctatttggaagATcaggttgaag</td>
<td>CHST9</td>
<td>NC_000018.9</td>
<td>Chr. 18</td>
<td>6</td>
<td>Intron 3</td>
<td>F</td>
</tr>
<tr>
<td>gatttagggaacATcaggttgaag</td>
<td>SPTBN1</td>
<td>NC_000002.11</td>
<td>Chr. 2</td>
<td>31</td>
<td>Intron 26</td>
<td>R</td>
</tr>
<tr>
<td>taggggaagactATcaggttgaag</td>
<td>RP11-425A23.1</td>
<td>NA</td>
<td>Chr. 4</td>
<td>2</td>
<td>Intron 1</td>
<td>F</td>
</tr>
<tr>
<td>gtaggggggtgacctATcaggttgaag</td>
<td>C1GALT1</td>
<td>NC_000007.13</td>
<td>Chr. 7</td>
<td>4</td>
<td>Intron 1</td>
<td>F</td>
</tr>
<tr>
<td>gtttcctggttagATcaggttgaag</td>
<td>SMR3</td>
<td>NC_000004.11</td>
<td>Chr. 4</td>
<td>3</td>
<td>5'UTR</td>
<td>F</td>
</tr>
<tr>
<td>gtaatgaggggaggATcaggttgaag</td>
<td>VAPB</td>
<td>NC_000020.10</td>
<td>Chr. 20</td>
<td>6</td>
<td>Intron 4</td>
<td>F</td>
</tr>
<tr>
<td>gttgaggggtgacctATcaggttgaag</td>
<td>FTX</td>
<td>NM_015052.11</td>
<td>Chr. 6</td>
<td>6</td>
<td>Intron 5</td>
<td>F</td>
</tr>
<tr>
<td>ctgttgctgagcctATcaggttgaag</td>
<td>AIFM3</td>
<td>NC_000022.10</td>
<td>Chr. 20</td>
<td>20</td>
<td>Intron 14</td>
<td>R</td>
</tr>
<tr>
<td>cattgctaggttgaagATcaggttgaag</td>
<td>SPOPL</td>
<td>NC_000002.11</td>
<td>Chr. 2</td>
<td>11</td>
<td>Intron 1</td>
<td>F</td>
</tr>
<tr>
<td>gtttcctggttagATcaggttgaag</td>
<td>APPL1</td>
<td>NC_000003.11</td>
<td>Chr. 3</td>
<td>22</td>
<td>Intron 7</td>
<td>F</td>
</tr>
<tr>
<td>acgccgccgagcATcaggttgaag</td>
<td>ARMC4</td>
<td>NC_000010.10</td>
<td>Chr. 10</td>
<td>20</td>
<td>Intron 17</td>
<td>F</td>
</tr>
<tr>
<td>gctagaggttggcctATcaggttgaag</td>
<td>TXNDC16</td>
<td>NC_000014.8</td>
<td>Chr. 14</td>
<td>21</td>
<td>Intron 16</td>
<td>F</td>
</tr>
<tr>
<td>ttctcttggaataATcaggttgaag</td>
<td>CCDC93</td>
<td>NC_000002.11</td>
<td>Chr. 2</td>
<td>24</td>
<td>Intron 11</td>
<td>F</td>
</tr>
<tr>
<td>atctagctgttcatgATcaggttgaag</td>
<td>NHSL2</td>
<td>NC_000023.10</td>
<td>Chr. X</td>
<td>8</td>
<td>Intron 1</td>
<td>F</td>
</tr>
<tr>
<td>cattacattataATcaggttgaag</td>
<td>KCNK2</td>
<td>NC_000001.10</td>
<td>Chr. 1</td>
<td>7</td>
<td>Intron 5</td>
<td>R</td>
</tr>
<tr>
<td>ggcagcatttcacgATcaggttgaag</td>
<td>ERC1</td>
<td>NC_000012.11</td>
<td>Chr. 12</td>
<td>18</td>
<td>Intron 9</td>
<td>F</td>
</tr>
<tr>
<td>gacatgagacacaATcaggttgaag</td>
<td>MIR587</td>
<td>NR_030314.1</td>
<td>Chr. 6</td>
<td>1</td>
<td>Exon 1</td>
<td>F</td>
</tr>
<tr>
<td>agacccgcgggccATcaggttgaag</td>
<td>TIMM17A</td>
<td>NC_000001.10</td>
<td>Chr. 1</td>
<td>6</td>
<td>Intron 1</td>
<td>F</td>
</tr>
<tr>
<td>tggatttagaccaATcaggttgaag</td>
<td>APAF1</td>
<td>NG_029094.1</td>
<td>Chr. 12</td>
<td>27</td>
<td>Intron 16</td>
<td>F</td>
</tr>
<tr>
<td>atctctcttctatgATcaggttgaag</td>
<td>ABCB11</td>
<td>NG_007374.1</td>
<td>Chr. 2</td>
<td>15</td>
<td>Intron 5</td>
<td>F</td>
</tr>
<tr>
<td>gttctagctgtcTAcaggttgaag</td>
<td>RP11-115J16.2</td>
<td>NA</td>
<td>Chr. 8</td>
<td>4</td>
<td>Intron 3</td>
<td>F</td>
</tr>
</tbody>
</table>
Fig. 8. Integration of the SB-based poly(A)-trap vector efficiently disrupts the endogenous gene transcript in HeLa cells.

A: schematic representation of a poly(A)-trapping insertion in an endogenous gene and potential transcripts in HeLa cells. Endogenous exons are boxed and arrows indicate the positions of primers used for transcript analysis. B: RT-PCR analysis of transcripts from a trapped endogenous gene (ADAMTS19, CHST9, EFHC2 and SPP2) in four cell colonies. N: normal HeLa cells; C: cell colonies; ET: Endogenous transcript; FT: Fusion transcript. The GAPDH is used as the control for equal amount of cDNA template in PCR reactions. C: the mutagenicity of poly(A)-trap insertions. Total RNA was isolated from each cell colony and subjected to qRT-PCR analysis. The mRNA levels of an endogenous gene in the normal HeLa cells and two other colonies without insertion at the gene loci (the controls) were compared to that in a cell colony containing an insertion at the corresponding gene loci. N represents normal HeLa cells; C1 represents ADAMTS19 gene; C2 represents CHST9 gene; C3 represents EFHC2 gene; C4 represents SPP2 gene. Data are given as means ± standard deviation (n = 3). ** and * indicate $P < 0.01$ and $P < 0.05$ versus the controls, respectively.
Fig. 9. Remobilization of trap inserts in the HeLa cell genome.

A: A schematic diagram of trap cassette remobilization and a canonical footprint left in the original insertion site. Arrows indicate the primer containing the footprint (F) and a gene-specific primer (R).

B: Individual cell colony containing a trap insertion at shown gene loci (ADAMTS19, CHST9, EFHC2, SPP2, SPTBN1, and KCNK2) was cultured at 32°C in two 35 mm dishes wells. Cells on one dish (T) were subjected to heat induction at 37°C for 5 days and cells on the other dish (N) were kept at 32°C. Total genome DNA was isolated and subjected to PCR analysis using gene-specific primers (F + R) in Table S1.

C: The results of Southern blotting. Neo probes were used to detect the copy number of transposons in the genome of HeLa cells with un-heat treatment (−) and after excised from their insertion sites with heat shock treatment (+) at 37°C for 48 h and recovered at 32°C for 5 days. The new DNA bands (black arrows pointed) were generated after the heat shock treatment (+).

D: The results of Southern blotting. Southern blotting with Neo probe indicated that the DNA size reduced (black arrows pointed) in the samples expressing Cre recombinase (+).
To avoid the remobilization of trap inserts that may lead to the normal expression of trapped genes and the biosafety concerns, the Cre recombinase was transiently expressed in G418-selected cell colonies to delete the LoxP-TiHsp70-SB11-polyA-LoxP cassette in our poly(A)-trap vector. As shown in Fig. 9D, Southern blotting with the Neo probe indicated that the DNA size reduced with the transient expression of Cre recombinase. These data indicate that the SB11 cassette can be conditionally deleted without affecting the phenotypes of cell or animal mutants.

4. DISCUSSION

Conventional gene trapping has been widely used for insertional mutagenesis across a target genome. This approach is effective in the capture of expressing genes, but is not suitable for the search of genes poorly expressed or not expressed in target cells (von Melchner et al., 2005; Schnutgen et al., 2008). On the other hand, poly(A)-trap vectors have been developed and successfully utilized for the identification of differentially expressed genes (Zambrowicz et al., 2003; Hirashima et al., 2004; Tsakiridis et al., 2009). However, most of these vectors are still suffering from the 3′-bias caused by the NMD mechanism, the SD read-through background, and no mutagenicity or uncertain mutagenic response (Zambrowicz et al., 2003; Hirashima et al., 2004; Sivasubbu et al., 2006). In addition, large-scale applications of transposon-based trap vectors in mutagenesis of undifferentiated cells and animals are severely limited by their relatively low activity in transposition. In this study, we demonstrated the development of a novel expression-independent poly(A)-trap vector that is highly effective in finding a broad range of genes and disrupting gene expression in HeLa cells. This poly(A)-trap vector contains three functional cassettes that improve the effectiveness of trapping and increases randomness of the detected integration sites. In addition, the usage of a Hsp70 promoter from the tilapia genome can drive the expression of SB11 transposase gene in an inducible manner.

Poly(A)-trap vectors have the potential to capture a broader spectrum of genes including those not expressed in undifferentiated cells, but trapped cell clones and mutant strains created by these vectors in the International Gene Trap Consortium database have so far been limited. To resolve this problem, efforts have been made by combinations of different promoters and SD signals (Ruley et al., 2005; Shigeoka et al., 2005; Tsakiridis et al., 2009) or insertion of a synthetic intron within the selectable marker gene (Lin et al., 2006). Recently, an efficient poly(A)-trap cassette was developed by placing an ARE from the human GM-CSF gene into the second intron of rabbit β-globin gene and the usage of this ARE led to a 7-fold enrichment in properly spliced trap transcripts by reducing the incidence of background SD read-through events (Tsakiridis et al., 2009). In this study, the addition of an ARE from the human GM-CSF gene into the intron1 sequence of a modified SD from carp β-actin gene markedly reduced the SD read-through background and enriched the poly(A)-trap events. Second, genomic analysis of trap inserts from early poly(A)-trap vectors indicates that an insertion preference for the 3′ ends of endogenous genes often occurs as a result of the activation of NMD mechanism for degradation of mRNAs containing trap inserts in the 5′ exons (Baker and Parker, 2004). To get rid of the 3′-bias for insertion, the NAISTrap group developed a novel vector known as UPATrap (Shigeoka et al., 2005) by inserting a floxed IRES sequence between the selectable marker gene and the SD sequence of the conventional RET poly(A)-trap vector (Ishida and Leder, 1999). The IRES sequence prevented the activation of NMD, allowing the trap of transcriptionally silent genes without a severe 3′-bias. Further modification of this strategy was recently performed to efficiently capture and disrupt endogenous genes (Tsakiridis et al., 2009). Additionally, it has been shown that the introduction of an IRES into a retrovirus vector has resulted in a slight 5′-bias (Hansen et al., 2003; von Melchner et al., 2005). In this study, the insertion of an EMCV/IRES downstream of the selectable marker gene appears to result in random integrations of our poly(A)-trap cassettes into endogenous genes in HeLa cells; however, more mega-data are needed to address the random insertion of our vector in a target genome. Third, the mutagenic abilities of available poly(A)-vectors remain to be evaluated since partial functions of trapped genes may be retained by the fusion protein and reduced levels of wild-type transcript may be generated, depending on the site of insertion and alternative splicing. The incomplete mutation of endogenous genes may lead to subtle phenotypes that are useful for dissection of normal gene functions, but careful characterization is required.

To improve the mutagenic efficiency of gene trapping, attempts have been made by the combination of a SA-gene-breaking cassette with a SD-gene-finding cassette (Jenkins et al., 2005; Largaespada et al., 2005; Clark et al., 2011). SB transposon allows the inclusion of gene-breaking cassette in our trap vector, which can efficiently interrupt the proper splicing of trapped genes both in vitro and in vivo.

Transposon-based vectors have increasingly attracted attentions due to their potential application in the achievement of saturation mutagenesis (Ivics et al., 2009). However, most transposons demonstrate nonrandom integrations in a target genome because of their characteristic preferences for insertion sites at the primary DNA sequence level. For example, the Harbinger3_DR transposon preferentially inserts into a 15-bp consensus sequence AAAACCCWGGTTT (Singh et al., 2008), the piggyBac transposon targets the tetranucleotide sequence TTAA, and all of known Tc1/mariner transposons, including SB, Frog Prince, Minos and Hsma1, prefer to integrate into the TA dinucleotides (Ivics and Izsvak, 2010). The Tol2 element does not appear to have a pronounced insertion preference for any primary DNA sequence (Grabundzija et al., 2010). In addition, integration of some transposons exhibits hotspots and cold regions on the target chromosomes. For instance, the piggyBac demonstrates a higher preference for integrations in regions surrounding transcriptional start sites and within long terminal repeat elements (Wilson et al., 2007), and the Tol2 transposon shows a pronounced preference for integration close to transcriptional start sites (Grabundzija et al., 2010). By contrast, Tc1/
mariner elements exhibit no or weak preference for transcription units (Ivics and Izsvak, 2010) and SB has no detectable genomic bias with respect to insertions in genes or intergenic regions (Liang et al., 2009). Moreover, local hopping, a phenomenon of chromosomal transposition in which transposons have a preference for landing into cis-linked sites in the vicinity of the donor locus, limits the chromosomal regions accessible to a transposon jumping out of a given chromosomal site; however, it may be useful for saturation mutagenesis (Takeda et al., 2005), which appears to be a shared feature of cut-and-paste transposons. It has been shown that the majority (83%) of Tol2 reinsertions are mapped on chromosomes other than the transposon donor chromosomes and that 9% of local hopping events mapped less than 300 kb away from the donor loci (Urasaki et al., 2008). The SB transposon seems to have a much larger local transposition interval between 5 and 40 Mb (Izsvak and Ivics, 2005; Largaespada et al., 2005). Therefore, diverse insertion site preferences of available transposon systems need to be carefully considered before construction of efficient transposon-based trapping vectors for large-scale mutagenesis.

Theoretically, the remobilization of trap inserts within a target genome of germline cells could directly create new animal mutants without the needs for mutant ESCs lines. Indeed, transposons can be remobilized from chromosomally resident loci and reintegrated somewhere else in the target genome by providing the transposase transiently (Geurts et al., 2006; Keng et al., 2009). SB transposon has been successfully mobilized in mouse somatic cells at frequencies high enough to induce tumors (Jenkins et al., 2005). Tol2 elements were induced to remobilize from their original insertion sites and thus new mutants were generated in zebrafish (Urasaki et al., 2008) and Xenopus (Yergeau et al., 2010), also piggyBac system have been developed successfully for ligand inducible insertional mutagenesis to overcome the adverse effect of the continued expression of transposase (Kong et al., 2010). In this study, we demonstrate that the SB-mediated poly(A)-trap system seems to randomly integrate into intron and exon of target genes and integrated trap cassettes in the HeLa cell genome can be induced to excise from the original insertion site and generate some new integration sites by inducing the expression of SB11 transposase at 37°C. The close-to-random insertion site distribution and remobilization of SB system appears to meet the urgent needs of genome-wide mutagenesis. However, it may be necessary to utilize multiple alternative trap vectors with distinct insertion site preferences for screen of the large-scale genome-wide mutagenesis.

In summary, our SB-based poly(A)-trap vector can be used as an alternative tool for large-scale mutagenesis in cells and vertebrates.

ACKNOWLEDGEMENTS

We thank all other members of the Cui’s lab for helpful suggestions and technical assistance. This work was supported by the National Natural Science Foundation of China (No. 30871442) to Z. Cui and the National Basic Research Program of China (No. 2012CB944500).

SUPPLEMENTARY DATA

Fig. S1. Standard curves for absolute quantification of exon 1 and exon 2 transcripts from pSPL3-derived vectors.

Fig. S2. In vivo EGFP expression in the finding cassette requires a poly(A) signal and a combination of SD and ARE elements.

Fig. S3. A schematic overview of the experimental procedure for polyA trap analysis in HeLa cells.

Table S1. Primers used in this study.

Table S2. Trapped endogenous genes from geneticin-resistant HeLa cells transfected with non-ARE-containing vector.

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jgg.2012.05.010.

REFERENCES


Grabundzija, I., Irgang, M., Mates, L., Belay, E., Matrai, J., Gogol-Doring, A., Kawakami, K., Chen, W., Ruiz, P., Chua, M.K., VandenDriessche, T.,


Sleeping Beauty transposon systems in human cells. Mol. Ther. 18, 1200–1209.
high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. Nat. Protoc. 4, 789–798.


