Nucleic Acids Research, 2005, Vol. 33, No. 4 ***e44***
doi:10.1093/nar/gni045

Tagging genes with cassette-exchange sites

Gilda Cobellis[1], Giancarlo Nicolaus[2], Mariangela Iovino[2], Antonio Romito[1], Emanuele Marra[2], Manlio Barbarisi[1], Marco Sardiello[1], Francesco P. Di Giorgio[2], Nicola Iovino[2], Massimo Zollo[1], Andrea Ballabio[1,3,*] and Riccardo Cortese[2]

[1]Telethon Institute of Genetics and Medicine, Via P. Castellino, 111, 80131 Naples, Italy, [2]Istituto di Ricerche di Biologia Molecolare P. Angeletti, Via Pontina km 30,600, 00040 Pomezia, Rome, Italy and [3]Medical Genetics, Department of Pediatrics, Federico II University, Via S. Pansini, 5, 80131 Naples, Italy

Received December 28, 2004; Revised and Accepted February 15, 2005

ABSTRACT

In an effort to make transgenesis more flexible and reproducible, we developed a system based on novel 5′ and 3′ 'gene trap' vectors containing heterospecific Flp recognition target sites and the corresponding 'exchange' vectors allowing the insertion of any DNA sequence of interest into the trapped locus. Flp-recombinase-mediated cassette exchange was demonstrated to be highly efficient in our system, even in the absence of locus-specific selection. The feasibility of constructing a library of ES cell clones using our gene trap vectors was tested and a thousand insertion sites were characterized, following electroporation in ES cells, by RACE–PCR and sequencing. We validated the system *in vivo* for two trapped loci in transgenic mice and demonstrated that the reporter transgenes inserted into the trapped loci have an expression pattern identical to the endogenous genes. We believe that this system will facilitate *in vivo* studies of gene function and large-scale generation of mouse models of human diseases, caused by not only loss but also gain of function alleles.

INTRODUCTION

The study of gene function *in vivo* has been facilitated by the insertion of transgenes in the mouse genome, either at random sites or at preselected sites by homologous recombination (1). When a transgene is injected in zygote pronuclei, there is no control over where it is introduced into the genome. This affects the reproducibility of the transgene expression pattern, a problem which was only partially solved by BAC-based transgenesis (2). On the other hand, homologous recombination while ideally suited for gene replacements and knock-outs is still somewhat laborious and time consuming, despite the recent development of high-throughput protocols.

Gene trapping strategies have been developed to partially circumvent these limitations, because they allow the selection of ES cell clones carrying the DNA insertion and a rapid identification of the trapped gene (3–5). The ultimate aim of gene trapping approaches is to generate a collection of ES 'gene trapped' clones that encompass all genes in the genome, as already proposed by Hicks *et al*. (6). This is a very ambitious project that will require the effort of several laboratories over several years. While published reports in this field appear to have adopted the gene trapping approach exclusively or primarily to generate loss of function alleles, it would be relevant to take advantage of these 'insertion sites' into genes for other purposes, such as the creation of allelic series. Along these lines, an additional feature of the vectors used for gene trapping strategies would be that of allowing to exchange the initial reporter cassette with any desired DNA sequence. This additional feature would greatly expand the usefulness, versatility and flexibility of a large collection of ES clones. For example, if the trapped gene is relevant for a human disease, multiple different alleles can be generated by cassette exchange. In addition, it is possible to exchange the trapped cassette with *trans*-regulation functions (cre recombinase, tet-dependent repressors and activators, etc.), which can be expressed in specific cells and tissues, taking advantage of the pattern of expression of the trapped gene.

Cassette exchange has been successfully achieved in cell culture by exploiting the property of Cre or Flp recombinases and their recognition sequences [Lox and Flp recognition target site (FRT), respectively]. For example, Bode *et al*. (7) have shown that a cassette exchange can be readily obtained

*To whom correspondence should be addressed. Tel: +39 081 6132207; Fax: +39 081 5790919; Email: ballabio@tigem.it
Corresponding may also be addressed to Riccardo Cortese. Tel: +39 06 910932200; Fax: +39 06 91093200; Email: riccardo_cortese@merck.com

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

if the recipient and the donor cassette are both flanked by matched heterospecific FRT sites. In addition, some of the public databases (Sanger Institute Gene Trap resource, Centre of Modeling Human Disease and German Gene Trap Consortium) report the sequence of trapped genes employing vectors carrying flanking heterospecific FRT or loxP sites, however, without experimental details describing their use.

In this work, we describe novel vectors that afford flp-based cassette exchange, and, more importantly, present evidence that with these reagents and method we can generate mice strains, which carry at the same locus different cassette obtained by flp-mediated exchange.

MATERIALS AND METHODS

Vector construction

The pFLIPa1 vector is a modification of the pGT1.8bgeo (8), and pFLIPb1 and FLIPb2 are variants of the RET gene trap vector (9), respectively. These vectors have been engineered and modified through the introduction of two heterospecific Flp-recognition sites (F5/F1) (7) and buffer regions derived from the pBS $SK^{+/-}$ plasmid. Two double-stranded oligonucleotides containing the mutated and wild-type FRT sequences (indicated as F5 and F1 in Figure 1a) were ligated to the 5′ and 3′ ends of gene trapping cassettes derived from the pGT1.8bgeo

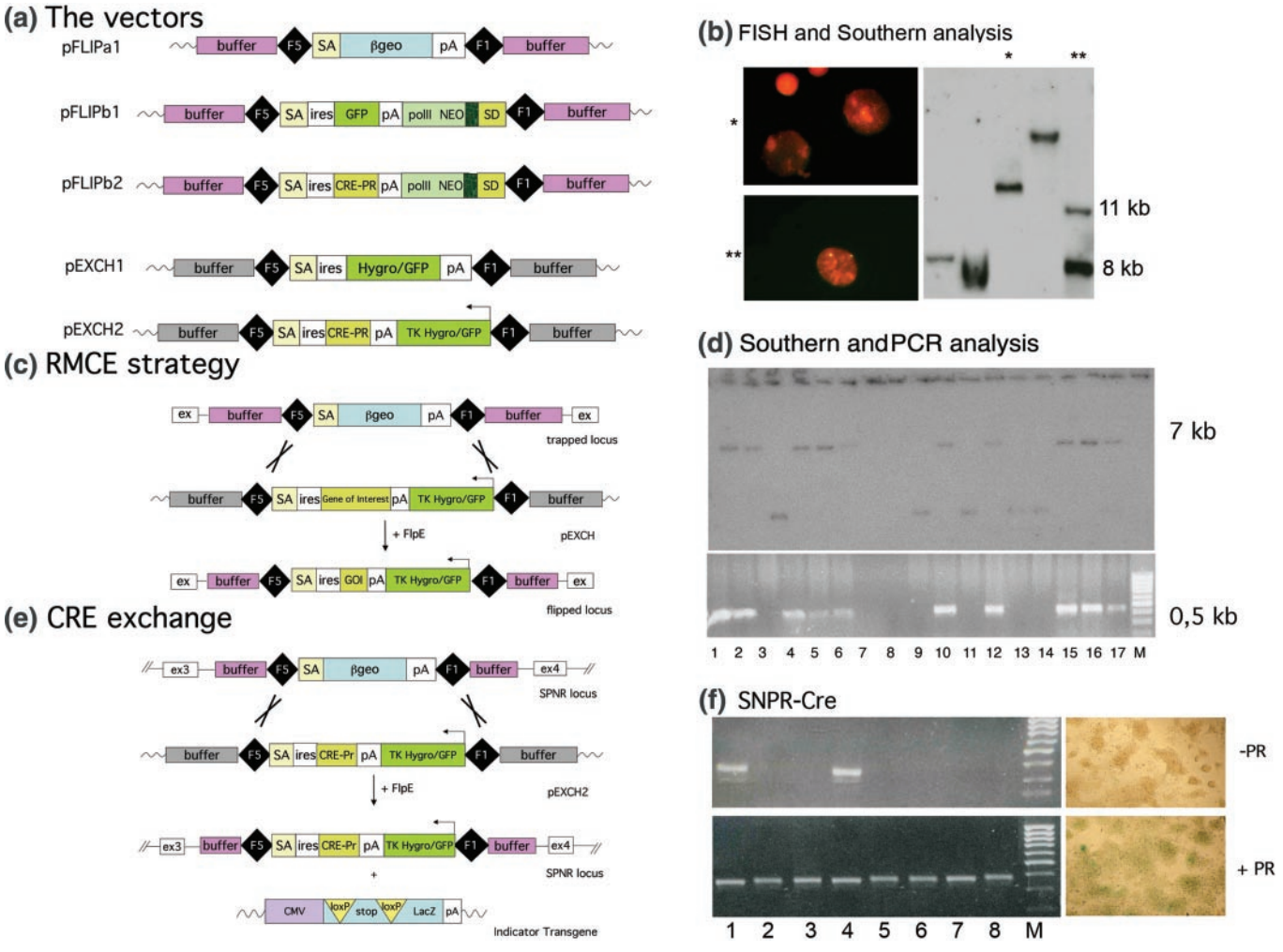

Figure 1. (**a**) Schematic representation of the novel trapping vector, pFLIPa1, pFLIPb1 and pFLIPb2, and the exchange vectors pEXCH1 and pEXCH2. The salient features of these vectors are described in the text. (**b**) FISH and Southern analysis. Fluorescent hybridization, performed with the entire vector sequence as probe, visualized a single spot in those clones, in which a single copy integration occurred (indicated by asterisk). A double fluorescent spot is visualized in those clones where double integration occurred (indicated by double asterisk). The clones were then analyzed by Southern blot analysis. Single-copy integrants show single band of different length, while the double integrant is visualized by the presence of two bands. (**c**) Schematic diagram of the RMCE strategy with the pEXCH2 vector. A tagged clone is subjected to a co-transfection with an exchange vector, containing a GOI, and the FlpE recombinase. The DNA cassette to be exchanged must be flanked by the same set of heterospecific FRT sites, present in the trapping vector. (**d**) Molecular characterization of recombinant clones of the *Rex2* gene. The PCR product specific to FlpE-mediated cassette exchange gave the expected 500 bp product from all the candidate clones diagnosed by the Southern blot analysis, in which a 7 kb band is observed. (**e**) Introduction of *Cre-PR* fusion gene into the *SPNR* locus. The clone is co-transfected with the pEXCH2 vector in the presence of pCAGGS-FlpE. The indicator LacZ transgene was used to test the functionality of CRE recombinase. (**f**) PCR analysis of 8 out of 30 clones to find the corrected exchanged ones at *SPNR* locus, using buffer-specific primer. Control PCR was performed on the same DNA sample using beta-actin oligonucleotides. *SPNR-Cre* expressing cell line was infected with an indicator vector containing a loxP-disrupted lacZ transgene in the absence and in the presence of the progesterone. The βgal activity is visualized only in the presence of progesterone.

and RET vectors. The modified cassette was ligated to SalI-digested BS $SK^{+/-}$; the buffer sequences, flanking the cassette, are the segment of the pBS plasmid derived from the linearization of the vector with ScaI present in the ampicillin gene.

The exchange vector pEXCH1 was generated by the substitution of the bgeo cassette of pIRESbgeo vector (10) with the Hygromicin–green fluorescent protein (GFP) fusion gene (NheI/ClaI fragment) derived from pTK-HygEGFP vector (Clontech). The pEXCH2 vector was created by the insertion of the SAiresCre-PR cassette, in front of the tk/Hygro/GFP cassette of the commercial vector. As buffer sequence, at 3′ of the F1 site in the backbone of the vector, we cloned at SpeI site a 1 kb region containing segments of noncoding DNA of human origin (stuffer), derived from adenoviral vector (C4HSU) described by Sandig *et al*. (11) amplified with the following oligonucleotides: C4HSU up, 5′-GGACACTCGC-TTTCTGCTCTC-3′; and C4HSU low, 5′-TAAACCAAGTT-CTTCCCAGAC-3′.

The pCAGGS-FlpE vector was a kind gift from Dr F. Stewart. The indicator transgene (loxP-disrupted LacZ) was a kind gift from Dr F. Graham.

ES cell cultures

The feeder-independent mouse ES E14 Tg2A.4 cell line was a kind gift from Dr W. Skarnes. We used standard culture conditions to grow and maintain the ES cell lines. ES cells were electoporated with ScaI-linearized pFLIPa1, pFLIPb1 and pFLIPb2 vectors and selected with G418 as described previously. Briefly, 10^8 ES cells were electroporated with 150 μg of linearized vector DNA in 1 ml phosphate-buffered saline (PBS), by applying single pulse at 0.8 mV, 3 μF in a Bio-Rad Gene Pulser. After electroporation, cells were seeded on gelatin-coated plates at 3×10^6 cells/100 mm dish. Selection with G418 (Gibco-BRL) at the recommended concentration started 48 h after electroporation. After 7–10 days of selection, single G418-resistant colonies were isolated, then picked and replicated in 96-well dishes. A set of master plates were frozen at −80°C using standard conditions.

The βgal activity of the clones generated with pFLIPa1 vector was assayed by Xgal staining (12), and βgal positive cell lines were cultured and duplicated for total RNA purification.

5′ and 3′ RACE–PCR and direct sequencing

The pFLIPa1 insertions were characterized by 5′ RACE–PCR, while the pFLIPb1 and pFLIPb2 insertions by 3′ RACE–PCR. In detail, total RNA from ES cell lines was prepared using SV 96 total RNA isolation system (Promega) with the BIOMEK 2000 robot (Beckman Coulter). The 5′ RACE–PCRs were performed by using 5′ RACE kit (Invitrogen) on total RNA extracted from ES cells. Two rounds of nested PCR were carried out on cDNA template, synthesized using the Superscript RT II (Invitrogen). The following LacZ-specific oligonucleotides were used: LacZrt 5′-TGGCGAAAGGGGGATGTG-3′ for the first strand cDNA, L232 5′-GATGTGCTGCAAGGC-GATTA-3′ for the first PCR and LacZ 5′-CCAGGGTTTT-CCCAGTCACG-3′ for nested PCR. The PCR products were directly sequenced using automated ABI377 DNA sequencer and the Byg Dye Terminator V.2.0 Cycle Sequencing kit (Applied Biosystem) with SA3 oligonucleotide (5′-AACTC-AGCCTTGAGCCTCTG-3′). We performed 3′ RACE as described previously (9). Two rounds of nested PCR were carried out on cDNA template, as described previously. The neo-specific oligonucleotides were used: NEO2.8, 5′-TCGCC-TTCTTGACGAGTTCTTCTGACC-3′; and NEO3, 5′-GCG-TCCACCTTTGTTGTTGGATATTGCC-3′.

To map the genomic insertion of the vector at the B56 and SPNR loci, a cloning strategy using inverse PCR is used (12).

The chromatogram data are interpreted with the base-calling program Phred. The high-quality sequences are then processed with Repeatmasker to eliminate repetitive sequences and then mapped against the murine genome using the BLAT algorithm at Santa Cruz server (13). Candidate sequences that are at least 95% identical to exonic sequence, over at least 90% of the query sequence, are annotated with the name of the gene. Unidentified sequences are those sequences that do not find match in the genome.

All gene trap insertions identified in this study have been submitted to the NCBI dbGSS database with the following accession numbers: CL982785–983254 and CW020243–020581.

Fluorescent *in situ* hybridization (FISH) analysis

Standard fluorescence *in situ* hybridization procedure using the sequence of pFLIPa1 vector as probe, labeled by nick translation with the digoxigenin-11-dUTP (Roche Diagnostics), was used for the screening of the single- or multiple-copy integrants (14).

RMCE

For the recombination mediated cassette exchange, 10 μg of pCAGGS-Flpe and 30 μg of pEXCH vector were co-electroporated using standard conditions into cell clones containing one integrated copy of pFLIPa or pFLIPb vector. Hygromicin selection (200 μg/ml) was applied 48 h after the electroporation. Clones were isolated, picked after 9–10 days of the selection and cultured for Southern and PCR analyses.

Southern and PCR analysis

To evaluate the copy number of the vector, genomic DNA from 12 independent clones was extracted, cut with ClaI at 37°C overnight and run on 0.8% agarose gel in 1× TAE. The gel was blotted on nylon membrane (Hybond N) and hybridized with a probe spanning the neomycin gene (15).

To evaluate the correct exchange, genomic DNA was extracted from hygromycin-resistant clones, cut with BamHI overnight at 37°C and run on 0.8% agarose gel in 1× TAE. The gel was blotted on nylon membrane (Hybond N) and hybridized with a probe spanning the hygromycin gene, following standard procedures (15). PCR analysis was performed using the following oligonucleotides: 5′-TCACTCTCGGC-ATGGACGAG-3′ (in the GFP gene) and the 5′-TCACAC-AGGAAACAGCTATGAC-3′ (in the buffer region) for the reaction with pEXCH1. To analyze the corrected exchanged clones with the pEXCH2 vector, the PCR analysis was performed using the following oligonucleotides: EXCH2, 5′-TTCCCCAATGTCAAGCACTTC-3′; and Buffer, 5′-TC-ACACAGGAAACAGCTATGAC-3′.

ES cell injections, breeding and genotyping

The mice are maintained in appropriate environment according to the Italian laws.

The ES cell lines were injected into C57BL/6 blastocyst and chimeras were obtained (16). The resulting male chimeras were bred to C57BL/6 females to test for germline transmission. Agouti offspring were genotyped by PCR on DNA extracted from tail biopsies, according to the standard procedures (12). Animals heterozygous for the insertion were crossed to C57BL/6 animals for embryos collection.

In situ hybridization analysis and immunofluorescence

The experiments with animals were performed in compliance with the institutional guidelines and approved by the Local Committee for 'Good animal experimental activities'. The embryos were removed by cesarean section and immersed in 4% paraformaldehyde in PBS, pH 7.4 overnight. Then, the embryos were dehydrated in 10, 20 and 30% sucrose and embedded in O.C.T. compound (Tissue Tek). Cryostat sections (16 μm) were cut. *In situ* hybridization was performed as described previously. The DNA fragments used as probes were obtained by PCR using the following oligonucleotides: SPNR for 5′-TGATGATCGCCATGTTATGG-3′ and rev 5′-TGCCAAACACTCCATTACTC-3′; and B56 for 5′-GGAGGCTGGTTTCTTGCCTTA-3′ and B56 rev 5′-CGGCCACCTGGAAGTGAGAG-3′. The immunofluorescence was performed using the standard protocol (12), and the anti-GFP antibodies were purchased from Molecular Probes. X-Gal staining of embryos and/or sections was performed using the standard protocol (12). Photographs were taken using a fluorescence microscope, Zeiss Axioplan 2.

RESULTS

The vectors

5′ tagging vectors. We constructed a 5′ tagging vector, pFLIPa1, in which the expression of a selectable marker (in this case neo) depends on the capture of an active transcriptional promoter in the target cells (Figure 1a). The salient features of this vector are (i) the buffer regions, which are segments added to avoid loss of informative sequences from both ends before integration in the genome; (ii) the mutated FRT recognition sequence (F5) and the wild-type FRT sequence (F1), which are unable to recombine with each other (17) and allow Flp-mediated cassette exchange (Flp-RMCE); (iii) the mouse *engrailed 2* splice acceptor sequence (8); (iv) the promoterless coding sequence for the *beta-galactosidase-neo* fusion gene (8,18); and (v) the SV40 poly(A) addition signal.

3′ tagging vectors. We also constructed a novel 3′ tagging vector, pFLIPb (Figure 1a), in which the selectable marker lacks a poly(A) signal, whereby if its insertion occurs upstream of an endogenous poly(A) signal, it will be expressed independently on the expression of the trapped gene (9). The salient features of this vector are (i) the 'protective' buffer regions; (ii) the F5 and F1 sequences (17); (iii) the mouse *bcl-2* splice acceptor sequence; (iv) the mouse RNA polII gene promoter, which drives the selectable marker neomycin; and (v) the mouse HPRT splice donor sequence, followed by the mRNA instability signal derived from the human granulocyte/macrophage colony-stimulating factor gene (9).

An additional feature of the 3′ trapping vector is the possibility to carry any promoterless gene of interest (GOI), upstream of the selectable marker, and whose transcription should be driven by the endogenous promoter of the trapped locus. As an example, we constructed pFLIPb1, in which the promoterless GOI is *GFP*, and FLIPb2 in which the promoterless GOI is a progesterone-regulated *Cre* gene (Figure 1a) (19).

Cassette-exchange vectors. In order to exchange the tag at the trapped locus, the exchange vectors must contain similarly oriented FRT sequences, and different buffer sequences. For this purpose, we constructed pEXCH1 and pEXCH2 (Figure 1a) with the following features: (i) the same set of heterospecific FRT sites (F5 and F1), in the same orientation of the corresponding FLIPa and FLIPb vectors, flanking the gene cassette; (ii) the 'protective' buffer sequences different from those added to the FLIP vectors in order to design discriminative primers between the vectors (see Materials and Methods); (iii) the mouse *engrailed 2* splice acceptor sequence; and (iv) the SV40 poly(A) addition signal (Figure 1a).

Gene tagging

To disseminate various tags (βgal, GFP or Cre) into the ES cells genome, the same experimental procedure was applied to the 5′ and 3′ trapping vectors. In all our experiments, a feeder-independent ES cell line (E14Tg2A.4) was used.

After electroporation with pFLIPa1, the cells were subjected to G418 selection and independent clones were analyzed for βgal activity. In our experimental conditions, an average of 30% (1012 out of 3360) of the clones expressed detectable βgal. To identify the integration sites, we performed the 5′ RACE–PCR on total RNA extracted from βgal-positive cells. One-fifth (203 out of 1012) of the clones did not yield any PCR product. The remaining clones (809) were analyzed by direct PCR-sequencing and their sequences compared with the public mouse genome databases using the BLAST algorithm. This analysis revealed that in 95% (768) of the clones splicing between endogenous splice sites and the vector occurred as expected. Five percent (41) of the clones, in which there was no sign of a splicing event, were discarded. Sequence analysis of the remaining clones revealed that 60% (461) of them matched known genes, 34% (261) matched expressed sequence tags (ESTs). The remaining 6% (46) did not match any sequence in the public sequence databases. Although most of the insertions were unique (42%), the same integration site was found in more than one clone, probably due to integration hot-spots, as already reported in other gene trapping studies (5).

A list of unique trapped genes with pFLIPa1 is reported in Supplementary Table 1.

We introduced the pFLIPb1 and pFLIPb2 into ES cells and 48 h later G418 selection was initiated to select for integration events. A total of 4 × 96 well plates were arrayed and the 3′ RACE–PCR was used to detect the insertion sites. Direct PCR sequencing followed by the BLAST analysis revealed that the splicing properly occurred between the vector and the endogenous splice sites in 91% (262) of the clones with the pFLIPb1 vector. Bioinformatic analysis revealed that 74% (213) of them matched with known genes, 26% (68) with ESTs.

The analysis of the clones obtained with the pFLIPb2 vector shows that the splicing occurred in the 80% (77) of the clones with a similar distribution of matches, 66% (51) with known

genes and 33% (25) with ESTs. A list of the trapped genes with pFLIPb1 and FLIPb2 is shown in the Supplementary Table 2.

To test the ability of the pFLIPb2 vector to trap genes independently of their expression status, we used the progesterone-regulated Cre system and an indicator vector containing a loxP-disrupted lacZ transgene. Using this system, we found that in 60% of these clones Cre was able to induce βgal activity in undifferentiated ES cells, while in an additional 30% of the clones the induction of βgal activity was observed only in differentiated ES cells, obtained after LIF deprivation. The remaining 10% of the clones did not induce βgal activity (data not shown).

To estimate the fraction of ES clones with multiple insertions, we performed FISH analysis on 12 clones, using the entire plasmid vector as probe. A single fluorescent spot was observed in 11 out of 12 analyzed clones. In the 12th clone, we observed a double fluorescent spot (Figure 1b).

Southern blot analysis was also performed on the same clones. Hybridization with the neo probe revealed that those clones with a single integration show a single band of different length, whereas the clone with double integration shows a double band, of 8 kb and 11 kb, respectively, confirming the results of the FISH (Figure 1b).

Cassette exchange

The cassette exchange mediated by the Flp recombinase is based on the observation that a DNA segment flanked by two heterospecific FRT sites can be exchanged with another DNA cassette flanked by the same set of heterospecific FRT sites (7).

To find out whether the Flp-RMCE takes place at the tagged loci and how efficient is this reaction, we designed a set of experiments to evaluate the efficiency of the exchange with or without a locus-specific selection. The RMCE strategy is schematically shown in Figure 1c.

Six different tagged clones (*B56*/*PP2A*γ, *SPNR*, *Otub1*, *Pep4*, *Rex2* and *Usp9y*) were transfected with the exchange vector pEXCH1 (Figure 1a), which contains the promoterless Hygro/GFP cassette, which will be driven by the endogenous promoter of the tagged locus, flanked by the two heterospecific FRT sites in the same orientation as the genomic tag. The *in vitro* enhanced FlpE recombinase was transfected in an expression vector (pCAAGS/FlpE) by electroporation, and 20 independent hygro-resistant ES clones were subjected to further analysis. Hybridization with the hygro probe revealed a 7 kb fragment on Southern blot analysis in the subclones in which RMCE had taken place (Figure 1d). The data were also confirmed by PCR using buffer-specific primers (Figure 1d). We evaluated the frequency of the site-specific recombination occurred in the different clones, ranging from 92 to 56%, with an average frequency of 75%. Thus, the results demonstrated that the locus-specific recombination event mediated by FlpE is highly favored over random insertion.

We then evaluated whether the FlpE enzyme is able to drive a genetic element in the tagged locus in the absence of a locus-specific selection. To this aim, we generated an exchange vector (pEXCH2, Figure 1a) containing (i) a promoterless progesterone-regulated form of the *Cre* gene (CRE-PR) (19) that will be driven by the endogenous promoter of the tagged locus and (ii) the *Hygro*/*GFP* fusion gene, under the control of the tk promoter that is necessary for the selection of transfectants but not for the selection of the exchange event at the trapped locus. The pEXCH2 vector was introduced with FlpE by electroporation in the *SPNR*-tagged clone (Figure 1e).

After transfectants selection, 30 hygro-resistant clones were analyzed by PCR, showing that a correct exchange occurred in 2 clones (Figure 1f, 7%). Taken together, these results indicate that our cassette-exchange system does not require locus-specific selection and that the process is frequent enough to identify the exchanged clone by screening a relatively small number of clones. We also demonstrated that the progesterone-regulated Cre was functional in our clones by infecting the exchanged clones with an indicator vector containing a loxP-disrupted lacZ transgene (Figure 1e). We found that βgal activity was strictly dependent on Cre-mediated recombination and the production of an active Cre is regulated by the progesterone (Figure 1f).

In vivo studies

Of crucial importance was to verify that the ES cells that had undergone multiple passages were still capable of germline transmission and that the transcription of the various tags exchanged by RMCE in the tagged loci show the identical pattern of expression of the endogenous gene. To this end, we generated transgenic animals using ES cell clones before and after the cassette exchange. For this analysis, we selected the ES cell clones bearing the insertions in the *B56*/*PP2A*γ isoform and *SPNR* genes, respectively (20,21).

Both the original trapped ES cell clone and the corresponding one after the cassette exchanges were injected into blastocysts to generate mouse chimeras. The resulting chimeras were used to produce mouse lines bearing either the original trap cassette (βgal) or the exchanged GFP cassette. We studied the expression patterns of βgal and GFP in heterozygous animals, and compared them with those of the corresponding endogenous genes. The *B56*/*PP2A*γ isoform transcript showed at 14.5 d.p.c. a major expression site in the developing heart. Figure 2b (top left) shows the expression pattern of the endogenous *B56*/*PP2A*γ isoform gene in the heart by *in situ* hybridization. Analysis of βgal activity (trap cassette) and GFP fluorescence (exchanged cassette) revealed an identical expression pattern (Figure 2b). We confirmed that the protein co-localizes in the same tissue by immunofluorescence using anti-GFP antibodies. In conclusion, we observed that the expression pattern of the reporter genes faithfully mimicked that of the endogenous gene.

We also studied the *B56*/*PP2A*γ isoform gene expression in adult mice. In this case, a specific pattern of expression of the endogenous gene is observed in the cerebellum and hippocampus. We confined the analysis to the $B56^{+/GFP}$ mice and we demonstrated by *in situ* hybridization that the GFP transgene reproduced the same expression pattern, as shown in Figure 2c.

A similar analysis was also performed for the SPNR mice carrying the GFP cassette after the RMCE (Figure 3a). We observed that the transgene had the same pattern of expression of the endogenous *SPNR* gene, as visualized by *in situ* hybridization with both SPNR and GFP probes and by direct GFP fluorescence (Figure 3b). Taken together, these data provide experimental evidence that ES clones that have undergone the cassette exchange treatment retain the capacity of germ-line

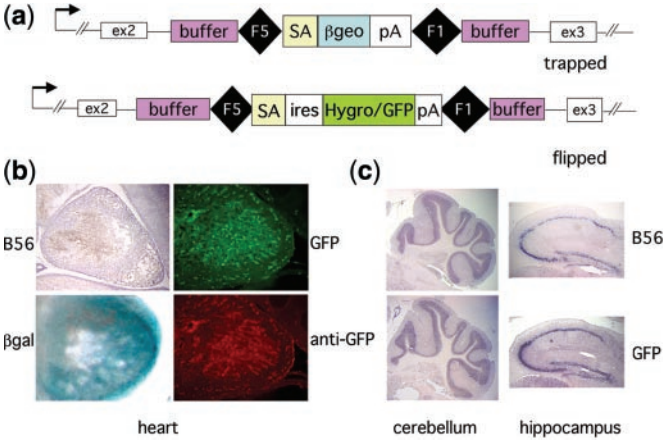

Figure 2. Analysis of expression of the *B56*/*PP2A* gamma isoform gene. (**a**) Schematic diagram of the insertion in the B56 locus occurred in the second intron. (**b**) Expression of the *B56*/*PP2A* gene in the heart on sagittal sections of heterozygous embryos at 14.5 d.p.c. The signal is visualized throughout the heart with DIG-labeled wild-type probe, as well as with an X-Gal staining on $B56^{+/gal}$ embryos (top and bottom left panels). Sagittal sections of heterozygous $B56^{+/GFP}$ embryos were analyzed by direct GFP visualization and immunofluorescence with the anti-GFP antibodies (top and bottom right panels). (**c**) Expression of *B56*/*PP2A* gamma isoform gene in the brain of adult mice. Sagittal sections of brain of heterozygous mice $B56^{+/GFP}$ were hybridized with the DIG-labeled *B56*/*PP2A* probe, as well as with the DIG-labeled GFP probe, indicates a specific signal in the cerebellum and in the hippocampus.

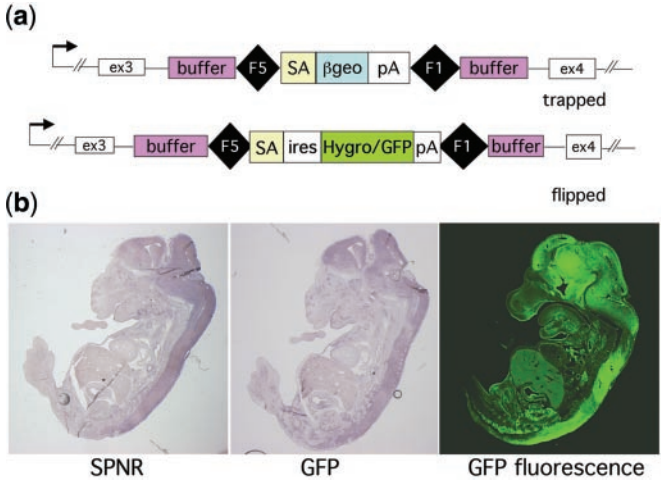

Figure 3. Analysis of expression of the *SPNR* gene. (**a**) Diagram of the insertion in the *SPNR* locus, occurred in the third intron. (**b**) Sagittal sections of whole embryos at 13.5 d.p.c. of heterozygous $SPNR^{+/GFP}$ mice, hybridized with a DIG-labeled probe of the endogenous gene and the GFP gene, indicate a specific signal in the telencephalon, mesencephalon and spinal cord. Direct GFP visualization of the adjacent sections shows the same pattern of expression of the transgene.

transmission and that the RMCE does not significantly interfere with the regulation of the endogenous promoter.

DISCUSSION

In the majority of the previously published work on gene trap methodologies, the genomic modifications mediated by the gene trap vectors are essentially irreversible and the trapped loci cannot be further manipulated. In some cases, loxP and FRT sites were used to perform RMCE (22,23), but these studies were confined to *in vitro* experiments. The present work sought to refine the design of the RMCE vectors and to demonstrate the flexibility of the strategy for *in vivo* applications, i.e. the construction of mouse strains carrying exchanged cassettes.

The cassette-exchange system has important applications in the development of animal models, especially for those models of human genetic diseases that are characterized by allelic heterogeneity. In 'recessive' diseases, the presence of 'hypomorphic', rather than null mutations may result in milder phenotypes, while in 'dominant' diseases different gain of function mutations at the same locus may result in completely different phenotypes. In these situations, a simple 'knock-out' of the disease gene does not provide a good animal model to study the phenotype of different mutations. Conversely, a disease locus trapped by our cassette-exchange system can be easily manipulated multiple times with mutated cDNAs, thus allowing the substitution of the wild-type allele with a variety of disease-causing mutant alleles.

Transgenic mice are typically generated by using cDNAs flanked by 'surrogate' promoters driving their expression either in a ubiquitous or in a tissue-specific manner. This approach has several limitations. First, the pattern of expression of the transgene might differ (sometimes significantly) from the pattern of expression of the endogenous gene, both in the amount and tissue distribution. Second, the expression pattern will be influenced by the site of integration into the mouse genome, thus limiting the reproducibility of the expression pattern even when the same promoter is used multiple times. Third, the choice of the promoter that mimics as much as possible the expression of the favored gene is limited by the 'repertoire' of characterized promoters.

Ideally, transgenic animals should express the GOI with the same regulation of the endogenous one. In this case, labor-intensive gene-targeting experiments must be performed to insert an extra-copy of the transgene under the control of the endogenous regulatory sequences.

Our approach has the potential of overcoming these difficulties using the power of gene trapping with the flexibility and reproducibility of the cassette-exchange system.

Our system can be used to create a large collection of ES cells, each carrying a tag whose expression is driven by a different promoter. These tags may potentially become 'standard' insertion sites for the expression of any transgene in specific tissues and in a reproducible way. Among the potential applications of this strategy are the construction of a collection of ES cells that can be tagged at the same locus with expression reporter genes (beta-gal, gfp, etc.) with site-specific recombinase (for example cre, as shown in this work), or transactivators (e.g. tTA, rtTA, etc.), for drug-induced and reversible gene expression in different tissues, or with toxin-encoding genes (e.g. tk, DTA) for cell ablation studies.

We validated *in vivo* the efficacy of our system by generating two sets of transgenic mice carrying the tag in two different loci. In both the cases, expression pattern of transgenes, before and after the cassette exchange, was identical to those of the endogenous-tagged genes. These data provide clear evidence that RMCE does not interfere with the regulation of gene expression.

In conclusion, we demonstrated that our vectors can efficiently and reliably trap genes in ES cells. We isolated several hundreds of ES cell clones carrying insertions in genes expressed and not expressed in ES cells. A comprehensive collection of ES cells bearing cassette-exchange tags will be a useful resource for functional genomics and disease studies.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank W. Wurst and P. Gruss for advices and reagents, M. Studer for technical assistance and G. Diez-Roux for critical reading of the manuscript. The authors are grateful to Dr Lasgura for FISH analysis, Dr Carotenuto for *in situ* hybridization advices and Dr Roma for bioinformatic analysis. This work was supported by Telethon Foundation (TABP15TELB) and by *Fondo per gli investimenti della Ricerca di Base* (FIRB RBNE01BBYW_003). Funding to pay the Open Access publication charges for this article was provided by Telethon Foundation.

REFERENCES

1. Nagy,A., Perrimon,N., Sandmeyer,S. and Plasterk,R. (2003) Tailoring the genome: the power of genetic approaches. *Nature Genet*., **33** (Suppl.), 276–284.
2. Gong,S., Zheng,C., Doughty,M.L., Losos,K., Didkovsky,N., Schambra,U.B., Nowak,N.J., Joyner,A., Leblanc,G., Hatten,M.E. *et al*. (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, **425**, 917–925.
3. Stryke,D., Kawamoto,M., Huang,C.C., Johns,S.J., King,L.A., Harper,C.A., Meng,E.C., Lee,R.E., Yee,A., L'Italien,L. *et al*. (2003) BayGenomics: a resource of insertional mutations in mouse embryonic stem cells. *Nucleic Acids Res*., **31**, 278–281.
4. Zambrowicz,B.P., Abuin,A., Ramirez-Solis,R., Richter,L.J., Piggott,J., BeltrandelRio,H., Buxton,E.C., Edwards,J., Finch,R.A., Friddle,C.J. *et al*. (2003) Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl Acad. Sci. USA*, **100**, 14109–14114.
5. Hansen,J., Floss,T., Van Sloun,P., Fuchtbauer,E.M., Vauti,F., Arnold,H.H., Schnutgen,F., Wurst,W., von Melchner,H. and Ruiz,P. (2003) A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA*, **100**, 9918–9922.
6. Hicks,G.G., Shi,E.G., Li,X.M., Li,C.H., Pawlak,M. and Ruley,H.E. (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet*., **16**, 338–344.
7. Seibler,J., Schubeler,D., Fiering,S., Groudine,M. and Bode,J. (1998) DNA cassette exchange in ES cells mediated by Flp recombinase: an efficient strategy for repeated modification of tagged loci by marker-free constructs. *Biochemistry*, **37**, 6229–6234.
8. Skarnes,W.C., Auerbach,B.A. and Joyner,A.L. (1992) A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev*., **6**, 903–918.
9. Ishida,Y. and Leder,P. (1999) RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucleic Acids Res*., **27**, e35.
10. Bonaldo,P., Chowdhury,K., Stoykova,A., Torres,M. and Gruss,P. (1998) Efficient gene trap screening for novel developmental genes using IRES beta geo vector and *in vitro* preselection. *Exp. Cell Res*., **244**, 125–136.
11. Sandig,V., Youil,R., Bett,A.J., Franlin,L.L., Oshima,M., Maione,D., Wang,F., Metzker,M.L., Savino,R. and Caskey,C.T. (2000) Optimization of the helper-dependent adenovirus system for production and potency *in vivo*. *Proc. Natl Acad. Sci. USA*, **97**, 1002–1007.
12. Matise,M., Auerbach,W. and Joyner,A.L. (2000) *Gene Targeting: A Practical Approach*. Oxford University Press, NY.
13. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol*., **215**, 403–410.
14. Pinkel,D., Gray,J.W., Trask,B., van den Engh,G., Fuscoe,J. and van Dekken,H. (1986) Cytogenetic analysis by *in situ* hybridization with fluorescently labeled nucleic acid probes. *Cold Spring Harb. Symp. Quant. Biol*., **51**, 151–157.
15. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd edn*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16. Auerbach,W., Dunmore,J.H., Fairchild-Huntress,V., Fang,Q., Auerbach,A.B., Huszar,D. and Joyner,A.L. (2000) Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines. *Biotechniques*, **29**, 1024–1028, 1030, 1032.
17. Seibler,J. and Bode,J. (1997) Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. *Biochemistry*, **36**, 1740–1747.
18. Chowdhury,K., Bonaldo,P., Torres,M., Stoykova,A. and Gruss,P. (1997) Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucleic Acids Res*., **25**, 1531–1536.
19. Wunderlich,F.T., Wildner,H., Rajewsky,K. and Edenhofer,F. (2001) New variants of inducible Cre recombinase: a novel mutant of Cre-PR fusion protein exhibits enhanced sensitivity and an expanded range of inducibility. *Nucleic Acids Res*., **29**, e47.
20. Muneer,S., Ramalingam,V., Wyatt,R., Schultz,R.A., Minna,J.D. and Kamibayashi,C. (2002) Genomic organization and mapping of the gene encoding the PP2A B56gamma regulatory subunit. *Genomics*, **79**, 344–348.
21. Pires-daSilva,A., Nayernia,K., Engel,W., Torres,M., Stoykova,A., Chowdhury,K. and Gruss,P. (2001) Mice deficient for spermatid perinuclear RNA-binding protein show neurologic, spermatogenic, and sperm morphological abnormalities. *Dev. Biol*., **233**, 319–328.
22. Araki,K., Imaizumi,T., Sekimoto,T., Yoshinobu,K., Yoshimuta,J., Akizuki,M., Miura,K., Araki,M. and Yamamura,K. (1999) Exchangeable gene trap using the Cre/mutated lox system. *Cell. Mol. Biol. (Noisy-le-grand)*, **45**, 737–750.
23. Hardouin,N. and Nagy,A. (2000) Gene-trap-based target site for cre-mediated transgenic insertion. *Genesis*, **26**, 245–252.