Published online 24 March 2012

Nucleic Acids Research, 2012, Vol. 40, No. 13 ***e97***
doi:10.1093/nar/gks262

Mixture of differentially tagged *Tol2* transposons accelerates conditional disruption of a broad spectrum of genes in mouse embryonic stem cells

N. Ika Mayasari[1,2], Keiko Mukougawa[1,3], Toshiaki Shigeoka[1], Koichi Kawakami[4,5], Masashi Kawaichi[1] and Yasumasa Ishida[1,3,*]

[1]Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0192, Japan, [2]Faculty of Veterinary Medicine, Bogor Agricultural University (IPB), Jalan Agatis Kampus IPB Darmaga, Bogor, West Java 16680, Indonesia, [3]The National Bio-Resource Project (NBRP), [4]Division of Molecular and Developmental Biology, National Institute of Genetics and [5]Department of Genetics, Graduate University for Advanced Studies (SOKENDAI), 1111 Yata, Mishima-shi, Shizuoka 411-8540, Japan

Received January 28, 2012; Revised February 17, 2012; Accepted March 8, 2012

ABSTRACT

Among the insertional mutagenesis techniques used in the current international knockout mouse project (KOMP) on the inactivation of all mouse genes in embryonic stem (ES) cells, random gene trapping has been playing a major role. Gene-targeting experiments have also been performed to individually and conditionally knockout the remaining 'difficult-to-trap' genes. Here, we show that transcriptionally silent genes in ES cells are severely underrepresented among the randomly trapped genes in KOMP. Our conditional poly(A)-trapping vector with a common retroviral backbone also has a strong bias to be integrated into constitutively transcribed genome loci. Most importantly, conditional gene disruption could not be successfully accomplished by using the retrovirus vector because of the frequent development of intra-vector deletions/rearrangements. We found that one of the cut and paste-type DNA transposons, *Tol2*, can serve as an ideal platform for gene-trap vectors that ensures identification and conditional disruption of a broad spectrum of genes in ES cells. We also solved a long-standing problem associated with multiple vector integration into the genome of a single cell by incorporating a mixture of differentially tagged *Tol2* transposons. We believe our strategy indicates a straightforward approach to mass-production of conditionally disrupted alleles for genes in the target cells.

INTRODUCTION

Since the completion of the mouse genome-sequencing project, our research communities have been seeking ways to rapidly and efficiently elucidate physiological functions in mice of the vast number of newly discovered genes and gene candidates.

An international collaborative endeavor called the knockout mouse project (KOMP) has been carried out to inactivate all mouse genes in embryonic stem (ES) cells using a combination of random and targeted insertional mutagenesis techniques and to make the created cell lines freely available among researchers (1). To disrupt as many genes in ES cells as possible within a short period of time, gene trapping has been used because it is simple, rapid, and cost-effective (2). The international gene-trap consortium (IGTC) (3), established by gene-trapping research groups, has been collecting, analyzing and distributing all the publically available gene-trapped ES-cell clones and their accompanying information (the IGTC database, http://www.genetrap.org/).

One of the most commonly used gene-trap methods is promoter trapping which involves a gene-trap vector containing a promoterless selectable-marker cassette (4). Although promoter trapping is effective at inactivating genes, transcriptionally silent loci in the target cells can not be identified using this technique. To capture a broader spectrum of genes including those not expressed in the target cells, poly(A)-trap vectors have been developed in which a constitutive promoter drives the expression of a selectable-marker gene lacking a poly(A)-addition signal (5–8). In this strategy, the mRNA of the selectable-marker gene can be stabilized upon trapping of

*To whom correspondence should be addressed. Tel: +81 743 72 5531; Fax: +81 743 72 5539; Email: ishiday@bs.naist.jp
Present address:
Toshiaki Shigeoka, Department of Stem Cell and Developmental Biology, Mie University Graduate School of Medicine, 2-174 Edobashi, Tsu-shi, Mie 514-8507, Japan.

a poly(A) signal of an endogenous gene regardless of its expression status in the target cell.

We previously showed that despite the broader spectrum of its potential targets, poly(A) trapping inevitably selects for the vector integration into the last intron of a trapped gene, resulting in the deletion of only a limited carboxyl-terminal portion of the protein encoded by the last exon of the gene (9). We presented evidence that this remarkable skewing is created by the degradation of a selectable-marker mRNA used for poly(A) trapping via an mRNA-surveillance mechanism called nonsense-mediated mRNA decay (NMD) (9). We also developed a novel poly(A)-trapping strategy, UPATrap, in which an internal ribosome entry site (IRES) sequence inserted downstream of the authentic translation-termination codon of a selectable-marker mRNA prevents the molecule from undergoing NMD, and made it possible to trap both transcriptionally active and silent genes without a bias in the intragenic vector-integration pattern (9).

The UPATrap strategy has been employed in a large-scale gene-trapping effort termed the Centre for Modeling Human Disease (CMHD; a Canadian wing of IGTC) (10) to disrupt a broader spectrum of genes including those not expressed in mouse undifferentiated ES cells (11,12). As shown below, however, transcriptionally silent genes in ES cells still remain relatively unexplored in the international gene-trap endeavor, and genes incapable of being captured by current gene-trap techniques have already been subjected to the more elaborate gene-targeting processes in KOMP (13).

When we try to establish a knockout mouse line based on the ES-cell technologies, a broad range of straight gene-knockout effects (e.g. embryonic lethality) may hamper identification of fine and minute phenotypes that would have appeared in restricted developmental stages and/or anatomical locations of the mutant mice (14,15). Conditional gene disruption, in which gene inactivation is attained in a spatially or temporarily restricted manner, could be an ideal solution that alleviates the disadvantages of straight gene inactivation (16). Conditional gene-targeting experiments have been widely performed since the first introduction of the Cre-*lox*P (derived from the bacteriophage P1) and Flp-*FRT* (yeast-derived) site-specific DNA-recombination systems into the field of genetic manipulation in mouse ES cells (17,18). Recently, these techniques have been employed to perform conditional gene disruption in random gene trapping (promoter trapping in particular) with mouse ES cells (19–21).

Here, we show that conditional gene disruption using the UPATrap strategy can not be successfully accomplished on the basis a retrovirus, the most commonly used backbone of gene-trap vectors in the current IGTC effort, because of the frequent development of intra-vector deletions/rearrangements. We also present evidence that a pivotal advantage of the poly(A)-trapping strategy (i.e. its capability of identifying silent genes in target cells) can be offset by a property of retroviruses (i.e. their preferential integration into transcriptionally active genome loci). We found that one of the cut and paste-type DNA transposons, *Tol2* (22), can be an ideal alternative as a backbone of gene-trap vectors that has none of the disadvantages of retroviruses. We also overcome the only problem of the *Tol2* system (or DNA transposons in general) that had been associated with multiple vector integration by incorporating a mixture of differentially tagged transposons into our experiments. We therefore believe our UPATrap-*Tol2* strategy is a straightforward approach to mass-production of conditionally disrupted alleles for a broad spectrum of genes and gene candidates in the target cells.

MATERIALS AND METHODS

Random sampling of mouse UniGene clusters

By using the RAND and RANK functions of the Excel spreadsheet software (Microsoft), 7811 UniGene clusters were randomly chosen out of all the mouse 79 202 entries at the time of analysis (January 2011), and those without reference-sequence (Refseq) information for proteins [the UniGene clusters classified as 'transcribed loci' (5509 clusters), cDNAs with unknown function (224 clusters), predicted genes (131 clusters), hypothetical proteins (3 clusters) and others (107 clusters)] were excluded. The remaining 1837 clusters for classical protein-coding genes were subjected to further analysis (Supplementary Table S1). The expression level of each gene in undifferentiated ES cells was assessed expediently by using the NCBI (National Center for Biotechnology Information) dbEST libraries #1882, #2512, #10023, #14556, #15703 #17907, and #21037, and the HiCEP database as described in the main text. URLs of the NCBI libraries are shown in Supplementary Table S2.

Vectors for gene trapping

Inverted pairs of the *lox*P, *lox*5171, *FRT* and *F3* sequences, a poly(A)-addition signal of the human growth-hormone gene [as the second poly(A)-addition signal for complete transcriptional termination of trapped genes], synthetic double-stranded (ds) oligonucleotides for the annealing of 3′-RACE primers (RACE), and splinkerette genome-PCR primers (SPL) were inserted into the UPATrap-EGFP retrovirus vector (9) as shown in Figures 2 and 3A to create the conditional retrovirus vector, pCRV2. Internal (non-retrovirus-derived) components of pCRV2 (the 5.73-kb XhoI–NotI fragment) and synthetic SPL oligonucleotides (ds) were cloned into the XhoI–BglII site of a *Tol2*-transposon plasmid pT2AL200R175G (23) to create CTP2F, a *Tol2* version of the conditional UPATrap vectors (Figure 3B). Each one of the *Tol2* vectors for transposon-mixture experiments was constructed by ligating the 5.73-kb XhoI–NotI fragment of pCRV2, synthetic SPL oligonucleotides (ds), one of the CC-in-poly(AT) (for the TMat vectors used in the latest gene-trap rounds TM4, TM5 and TM6) or AA-in-poly(TT) (for the TMtt vectors used in the former gene-trap rounds TM1 and TM2) oligonucleotides (ds), one of the corresponding ID oligonucleotides (ds) (SEQ-01–15) and synthetic Term oligonucleotides (ds) into the XhoI–BglII site of pT2AL200R175G (Figure 6A).

The TMat vectors contain additional copies of the mouse and human poly(A)-addition signals. The GenBank/EMBL/DDBJ accession numbers of the gene-trap vectors are shown in Supplementary Table S10.

Cell culture and gene trapping

The V6.4 ES cells (24) were cultured as previously described (8). The ES cells were grown on a layer of mitomycin C-treated SNL–STO cells (25) that had been stably super-transfected with an expression vector pSRα-mLIF-IRES-Puro^r-poly(A) for bi-cistronic expression of the mouse leukemia inhibitory factor (LIF) and the puromycin-resistance gene product (SLPN cells, unpublished).

The recombinant retrovirus was produced using the Plat-E packaging cell line (26). ES cells were infected with the recombinant retrovirus and selected under 200 μg/ml of G418 (Nacalai) for 7–10 days as previously described (8,9). Drug-resistant colonies were isolated manually into 12-well plates, and the high molecular-weight (HMW) genomic DNA and the total cellular RNA were extracted from the expanded cells using a standard procedure (27). For transposon experiments, 2.5×10^5 ES cells were co-transfected with 2.27 μg of pCAGGS-TP (an expression vector for the *Tol2* transposase) (23,28) and 0.23 μg of either pCTP2F or a mixture of differentially tagged *Tol2* plasmids (#01–#15) using the TransFast reagent (Promega). The subsequent steps were carried out as described above for gene trapping using the retrovirus vector.

Availability of ES-cell clones

Detailed information about the ES-cell clones shown in Supplementary Tables S3 and S4 has been transferred to the IGTC database (http://www.genetrap.org/). The RIKEN BioResource Center (Tsukuba, Japan) (http://www.brc.riken.jp/inf/en/index.shtml) distributes the ES-cell clones upon request with minimum shipping charges.

Detection of intra-vector deletions

We assessed the integrity of two different regions inside each genome-integrated vector by genomic PCR (Figure 3). The Lr and Sr regions of the conditional UPATrap-Moloney retrovirus vector were amplified using the SPL-1 and RN2 primers and the RNApol2-F1 and U5 R1 primers, respectively. The Lt and St regions of the conditional UPATrap-*Tol2* transposon vector were amplified using the 5FRT-F1 and RN2 primers and the RNApol2-F1 and R-term primers, respectively. The nucleotide sequences of the primers and the PCR conditions are shown in Supplementary Tables S11 and S12.

Conditional mutagenesis in ES cells

For the first step, the ES-cell clones were transiently transfected with pCAGGS-FLPo-IRES-Puro^r-poly(A), and 24 h after transfection, they were subjected to 48 h of brief selection with puromycin (1 μg/ml). Then, limiting dilution of transfected cells was carried out on a layer of mitomycin C-treated SLPN cells and the culture was maintained for 6–8 days. Colonies were manually isolated and transferred into 12-well plates in duplicates for the G418-sensitivity test, with one set of plates containing the standard ES-cell medium and the other set of plates supplemented with 200 μg/ml of G418 (Nacalai). The genomic DNA was extracted from the unselected group of cells. Structure of both 5′- and 3′-portions of genome-integrated vectors was analyzed as indicated in Supplementary Figures S4 and S5. For the second step, the FLPo-generated six subclones (Figure 5) were transiently transfected with pMC1-Cre-PGK-Puro^r-poly(A), and the subsequent steps were carried out as described above for the FLPo experiment (Supplementary Figures S4 and S5), but the G418-sensitivity test was not performed for the Cre-generated cells.

The Cre-generated six daughter subclones were chosen (Figure 5) and, together with the FLPo-generated six subclones, subjected to the analysis of the efficiency of conditional regulation of trapped-gene expression. For this purpose, the original V6.4 cells, parental 1TP-84 and TP-32 cells, FLPo-generated six subclones and Cre-generated six daughter subclones were depleted of residual mitomycin C-treated SLPN cells by using a standard separation procedure (29). The total cellular RNA was extracted from the feeder-depleted ES cells, and after synthesis of the first strand cDNA using SuperScript II RT (Invitrogen) and the oligo(dT)$_{12–18}$ primer (GE Healthcare), expression of *Atp6ap2* and *Ctps2* was assessed by PCR using the ATP-Ex7-F (located on the sense strand of exon 7) and ATP-Ex9-R (located on the anti-sense strand of exon 9) primers (for *Atp6ap2*), and the CTPS2-F2 (located on the sense strand of exon 12) and CTPS2-R1 (located on the anti-sense strand of exon 17) primers (for *Ctps2*). Disruptive-splicing events (Figures 2 and 5) were detected using the ATP-Ex7-F and Bcl2-R (located on the anti-sense strand of the splice-acceptor component in the gene-trap vectors) primers (for *Atp6ap2*), and the CTPS2-F1 and Bcl2-R primers (for *Ctps2*). The expression level of the β-actin mRNA, which serves as an internal control, was monitored using the β-actin-F and β-actin-R primers in RT-PCR. The nucleotide sequences of the primers and the PCR conditions are shown in Supplementary Tables S11 and S12.

Analyses of the number and IDs of genome-integrated vectors

By using the genomic DNA extracted from the ES-cell clones generated in the transposon-mixture experiments as a template, the PCR was performed with the Phusion DNA polymerase (Finnzymes) and the New-RACE-0.9 and R-term primers. Nucleotide sequences were determined using RS-F4 as a primer. Confirmation of genome integration of each tagged vector was carried out using the PCR Master mix (Promega), the F-int primer and one of the tag-specific reverse primers (R-01–15). See Supplementary Figure S6 for details. The nucleotide sequences of the primers and the PCR conditions are shown in Supplementary Tables S11 and S12.

3′-RACE

To identify trapped genes and predict vector-integration sites, the 3′-RACE PCR and direct sequencing of the PCR products were performed as described (8), but using a different set of primers and slightly modified conditions (Supplementary Tables S11 and S12). Sequence tags obtained were analyzed with the Blat genome-alignment program (http://genome.ucsc.edu/cgi-bin/hgBlat/) based on the NCBI37/mm9 assembly of the mouse genome (July 2007).

Splinkerette genome PCR

For ES-cell clones generated using CRV2, the HMW genomic DNA was digested with HaeIII (New England BioLabs) and, after heat inactivation of the enzyme at 80°C for 20 min, the digested DNA was ligated with the splinkerette SplT-BLT/SplB-BLT linker using T4 DNA ligase (Takara). The linker-ligated DNA was digested with PvuII (New England BioLabs) to avoid amplification of internal vector components. The PvuII-digested DNA served as a template for the first round of PCR (Supplementary Table S12) in which the SPL-1 and P1 primers and the Advantage-GC2 polymerase mix (Clontech) were involved. The second round of PCR was performed as described for the first round using the 1/10 diluted first-round PCR product and the SPL-2 and P2 primers. Direct sequencing was carried out with the New-Spl2.3 primer.

For ES-cell clones generated using the *Tol2*-transposon vector (CTP2F) or the mixture of differentially tagged *Tol2*-transposon vectors, the genomic DNA was digested with HaeIII, TaqI or MspI (New England BioLabs), and, after inactivation of the enzyme, the digested DNA was ligated with a compatible splinkerette-type linker SplT-BLT/SplB-BLT, SplT-Msp/SplB-Msp or SplT-Taq/SplB-Taq, respectively. For CTP2F, the splinkerette PCR amplification and sequencing of amplified products were carried out in the same reaction conditions as described for the ES-cell clones generated using CRV2, but using a different set of primers. The first and second PCRs for CTP2F involved the New T-Spl1 and P1 primers and the New T-Spl2 and P2 primers, respectively. New T-Spl3 was used as the sequencing primer. The PCR and sequencing primers for the ES-cell clones generated using the mixture of the differentially tagged transposons vary, depending on the number and IDs of genome-integrated vectors. See Supplementary Figure S7 for details. The nucleotide sequences of the primers/linkers and the PCR conditions are shown in Supplementary Tables S11 and S12.

RESULTS

Expression level and trapping efficiency of a gene in mouse undifferentiated ES cells are positively correlated

To understand what proportion of protein-coding genes are constitutively expressed in mouse undifferentiated ES cells, we first randomly selected ~10% of total mouse UniGene clusters (7811 out of 79 202 entries) (the UniGene database, http://www.ncbi.nlm.nih.gov/unigene/) and then excluded those for which the protein-coding capability has not been proven (see 'Materials and Methods' for details). For each of the remaining 1837 clusters representing classical protein-coding genes, we tried to determine: (i) if the gene is expressed in undifferentiated ES cells; and (ii) if the gene has already been disrupted in the IGTC effort (Supplementary Table S1).

In order to predict the expression level of each gene in undifferentiated ES cells, we examined the number of corresponding expressed sequence tags (ESTs) in the seven NCBI dbEST libraries that were constructed using mRNAs derived from mouse undifferentiated ES cells (Supplementary Tables S1 and S2). The total number of ESTs included in the seven libraries is 143 423. For each of the selected UniGene clusters, we also inferred the mRNA expression in undifferentiated ES cells by looking for the presence or absence of the corresponding sequence tags in another database for the ES cell-derived transcripts that were created using a highly sensitive PCR-based technology termed HiCEP (the HiCEP database, http://hicepweb.nirs.go.jp/english/index.html) (Supplementary Table S1) (30).

Among 1837 UniGene clusters, 830 (45.2%) contained neither the undifferentiated ES cell-derived ESTs nor HiCEP sequences, and therefore the corresponding genes were regarded as transcriptionally silent in undifferentiated ES cells (Figure 1A). One hundred and ninety three (10.5%) contained the undifferentiated ES cell-derived HiCEP sequences, but not NCBI-ESTs, suggesting that their expression levels in ES cells should be relatively low. 814 (44.3%) contained the undifferentiated ES cell-derived ESTs and therefore were considered to be expressed in the cells (Figure 1A).

In the case of such undoubtedly 'expressed' genes in undifferentiated ES cells, 80.8% had already been disrupted by random gene trapping in the IGTC effort (the IGTC database, http://www.genetrap.org/) (Figure 1A and Supplementary Table S1). In contrast, only 21.8% of the potentially silent genes were found in the IGTC database at the time of the analysis (July 2011). This strongly suggests that expressed genes are more preferentially disrupted in the IGTC laboratories than are silent genes in undifferentiated ES cells. The results shown in Figure 1B also support this conclusion because the number of the ES cell-derived ESTs in the above NCBI libraries and that of the gene-trapped ES-cell clones in the IGTC database appear to be positively correlated, at least with regard to the UniGene clusters that contain less than 15 corresponding ESTs in the seven NCBI libraries (1695 out of 1837). A large fraction of the 'difficult-to-trap' (mostly transcriptionally silent) genes in undifferentiated ES cells have already been disrupted individually by elaborate gene targeting in KOMP [the international knockout mouse consortium (IKMC) database, http://www.knockoutmouse.org/] (Supplementary Table S1) (13).

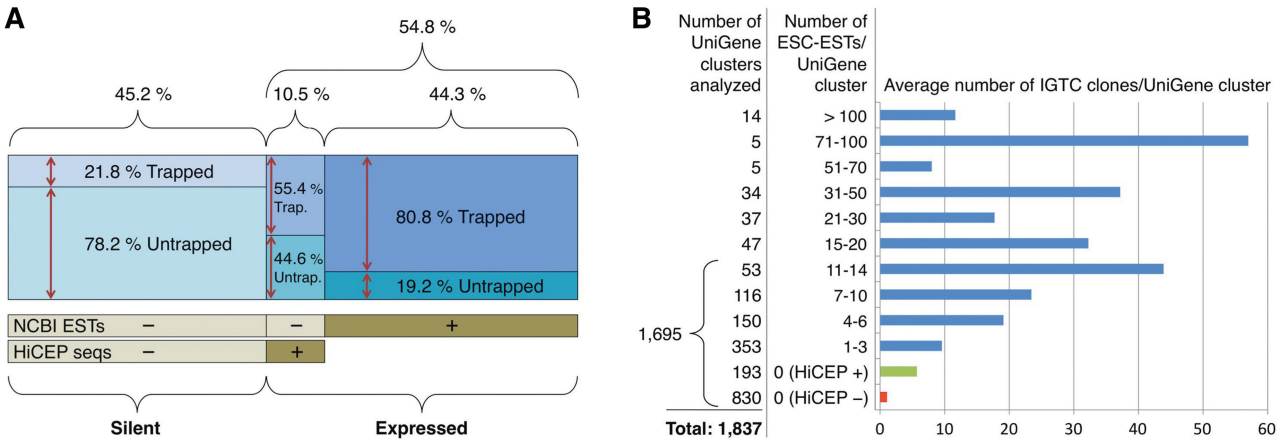

Figure 1. Transcriptionally active genes in mouse undifferentiated ES cells are trapped preferentially in the IGTC effort. (**A**) A crude analysis showing the correlation between the mRNA expression and the trapping efficiency in mouse ES cells of the randomly sampled 1837 protein-coding genes (UniGene clusters). A given gene was considered to be 'expressed' if it has either the corresponding NCBI-ESTs or HiCEP sequences derived from mouse undifferentiated ES cells. Likewise, a given gene was considered to be transcriptionally 'silent' if it has neither the corresponding NCBI-ESTs nor HiCEP sequences derived from mouse undifferentiated ES cells. (**B**) A fine analysis showing the positive correlation between the predicted expression levels and the trapping frequency in mouse ES cells of the majority [1695 (92.3%)] of the randomly sampled 1837 protein-coding genes (UniGene clusters). The expression level of a given gene was assessed by the number of corresponding NCBI-ESTs derived from undifferentiated mouse ES cells. The UniGene clusters that contain no ES cell-derived NCBI-ESTs were further classified into two groups: (i) those also devoid of the ES cell-derived HiCEP sequences [830 clusters (45.2% of 1837) shown as a red bar] and (ii) those containing the ES cell-derived HiCEP sequences [193 clusters (10.5% of 1837) shown as a green bar].

A strategy for conditional gene disruption using random poly(A) trapping

Beside the disruption of transcriptionally silent genes in the target cells, another challenge for random gene trapping has been the conditional inactivation of identified genes (16,19–21). To achieve this in poly(A) trapping, we assembled critical components of a gene-trap vector, as indicated in Figure 2. The first half represents a gene-terminator cassette containing a promoterless enhanced green fluorescent protein (EGFP) cDNA for monitoring the expression of trapped genes in living cells (7–9) and two or four copies of poly(A)-addition signals for the complete transcriptional termination. The second half represents a poly(A)-trapping cassette of the UPATrap type from which a constitutive promoter drives transcription of the NMD-resistant selectable-marker mRNA that plays an essential role in abolishing the extreme bias in the intragenic vector-integration pattern (9). The FLEx methodology (19) conferred the capability of conditional gene disruption on our system.

Upon expression of the Flp recombinase, regions 1 and 3, and central region 2 in the diagram are to be deleted and inverted, respectively, to generate a non-disruptive allele of a trapped gene (Figure 2; see Supplementary Figure S1 for details). The second recombination would be induced in mice by expressing the Cre recombinase in a spatially or temporally restricted manner (Figure 2).

A *Tol2*-trasposon version of the conditional UPATrap vector rarely suffers from intra-vector deletions/ rearrangements

We created a conditional variant of the UPATrap vector on the basis of the Moloney murine leukemia virus (MMLV) (Figure 3A) (31) and performed gene-trap experiments with mouse ES cells. When we examined the integrity of the genome-integrated proviruses by PCR, we immediately noticed that 78.5% of the ES-cell clones either produced shorter bands than expected or did not show any amplification, suggesting that they potentially contain some forms of intra-vector deletions or rearrangements (Figure 3A). We then tried to confirm the presence of deleted/rearranged regions and found that the first and second halves of the provirus molecules carry various forms of deletions (Supplementary Figure S2). Although data are not shown, we found that the standard (i.e. non-conditional) retroviral UPATrap vector also generates intra-vector deletions/rearrangements with high frequency in the target cells. Such alterations inside the vectors severely hamper the conditional poly(A)-trapping because even a tiny deletion covering one of the eight recombinase-target signals distributed throughout the vector (Figure 2) would make it impossible to induce regulated inversion/deletion of the vector components for conditional gene disruption.

We suspected that the deletions/rearrangements inside our retrovirus vectors were created during the reverse-transcription step immediately after infection of the target cell. Therefore, we transferred the essential components for conditional poly(A)-trapping (Figure 2) from an MMLV vector into a *Tol2* transposon (Figure 3B) (23,28). Since *Tol2* is a cut and paste-type DNA transposon devoid of single-stranded nucleic acid steps in its life cycle, the chance of generating intra-vector deletions or rearrangements was expected to be negligible. As a matter of fact, we detected potential deletions/rearrangements associated with the genome-integrated *Tol2* vectors only in 2.3% of the ES-cell clones that contained single-vector integration (Figure 3B).

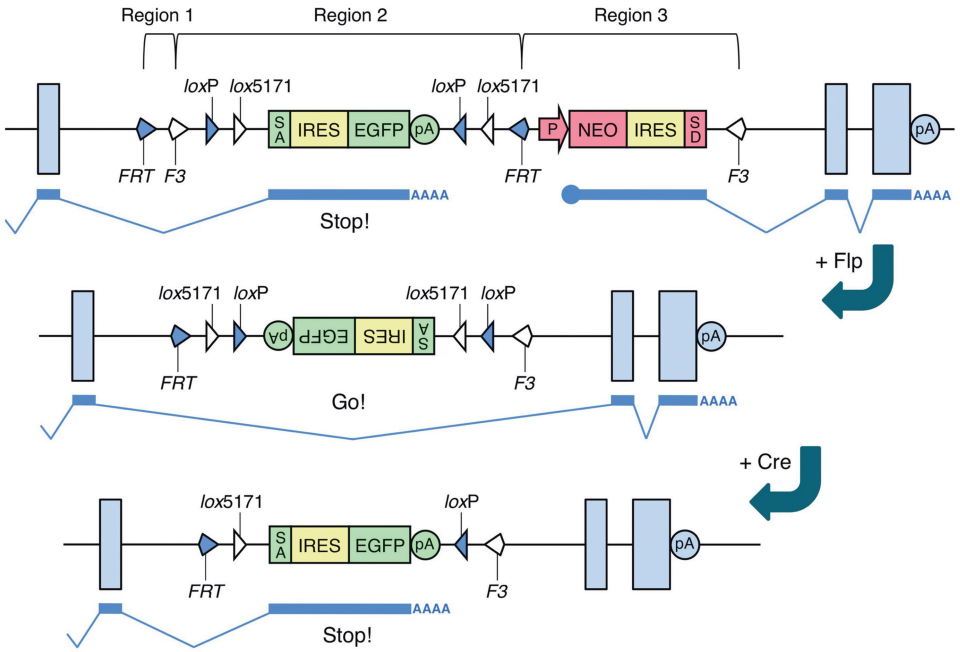

Figure 2. A strategy of conditional poly(A) trapping based on the NMD-suppressing UPATrap technology. Orientations of the triangular and diamond-shaped arrowheads represent those of the recombinase-target signals *lox*P, *lox*5171, *FRT* and *F3*. Light-blue rectangles are exons of an endogenous gene. Thick and thin blue lines represent exonic and intronic portions of pre-mRNAs, respectively. SA, the splice-acceptor sequence of the human Bcl-2 gene (the intron 2−modified exon 3 portion); SD, the splice donor sequence of the mouse Hprt gene (the modified exon 8−intron 8 portion); P, a constitutive promoter of the mouse RNA polymerase II (the RPB1 subunit) gene; pA inside the vector, two or four copies of poly(A)-addition signals derived from the mouse and human growth-hormone genes. pA next to the last light-blue rectangle, the poly(A)-addition signal of an endogenous gene.

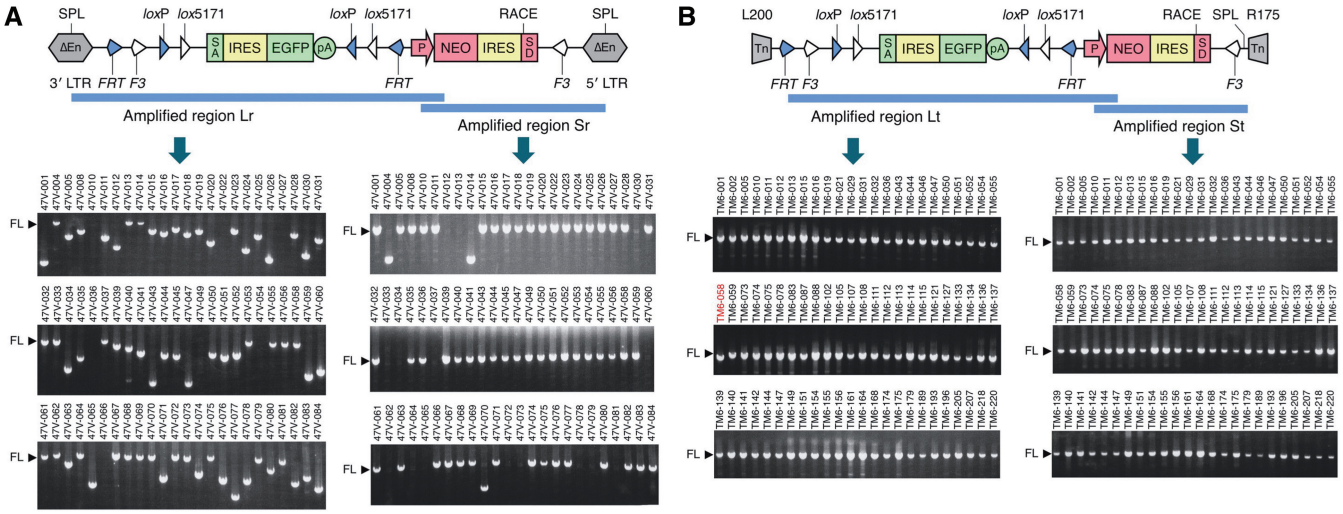

Figure 3. Structure and integrity of the genome-integrated conditional gene-trap vectors. (**A**) Structure and low integrity of the conditional UPATrap-Moloney retrovirus vector in the target cells. Seventy two independent gene-trapped clones were randomly chosen from the ES cells infected with the conditional UPATrap retrovirus, and the integrity of the introduced vectors was analyzed by genomic PCR for the regions Lr and Sr. (**B**) Structure and high integrity of the conditional UPATrap-*Tol2* transposon vector in the target cells. Seventy two independent gene-trapped clones with single-vector integration (see Figure 6 for details) were randomly chosen from the ES cells introduced with the conditional UPATrap-*Tol2* transposon, and the integrity of the genome-integrated vectors was analyzed by genomic PCR for the regions Lt and St. Only one clone TM6-058 (indicated by red letters) showed to possess a smaller Lt portion than the other clones. LTR, the long terminal repeat of the MMLV; ΔEn, enhancer deletion (31); Tn, terminal essential sequences (L200 and R175) of *Tol2* transposon (23); RACE, the synthetic nucleotide sequence (90mer) that facilitates 3′-RACE; SPL, the synthetic nucleotide sequence (90mer) that facilitates splinkerette genome PCR; FL, full length. Both of the RACE and SPL sequences are devoid of the GT (potential splice donor), AG (potential splice acceptor) and AATAAA/ATTAAA [potential poly(A)-addition] sequences in both sense and antisense strands.

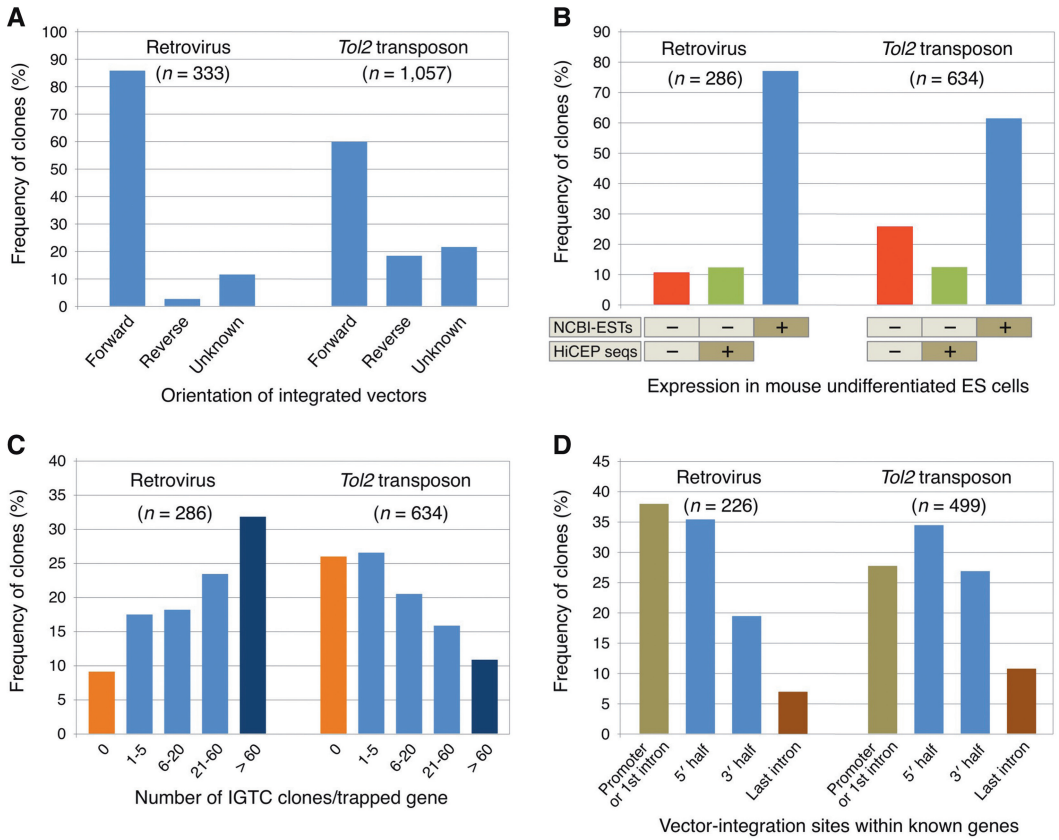

Figure 4. Nature of genes and gene candidates identified by using the conditional UPATrap vectors. (**A**) Orientation of vector integration relative to that of transcription of trapped genes. The orientation of an integrated vector is regarded as forward when the transcriptional orientation of the EGFP and NEO cassettes of the gene-trap vector and that of the trapped known gene are the same. Likewise, when their orientations are opposite to each other, the vector insertion is regarded as reverse. In the cases of vector integration into unknown genes, the orientation of vector integration is marked unknown. (**B**) Transcriptional status of genes identified by using the conditional UPATrap vectors. Transcriptional status of known genes trapped in a forward orientation was classified into three groups: (i) NCBI-ESTs −/HiCEP seqs −; (ii) NCBI-ESTs −/HiCEP seqs +; and (iii) NCBI-ESTs +. See Figure 1 for details about this classification. (**C**) Number of the mutant ES-cell clones already registered in the IGTC database for each known gene trapped in a forward orientation by using the conditional UPATrap vectors. (**D**) Distribution of the vector-integration sites around known genes trapped in a forward orientation by using the conditional UPATrap vectors. The vector-integration sites were predicted from the nucleotide sequences of the 3′-RACE fragments. Events of vector integration into the introns of genes consisting of 1–4 exons and the right-middle introns of genes with even numbers of exons were excluded from the analysis.

Genes identified using the conditional UPATrap-*Tol2* transposon vector

In addition to the frequency of the generation of intra-vector deletions/rearrangements, the nature of genes and gene candidates identified through poly(A) trapping based on the NMD-suppressing technology was also significantly different between the MMLV and *Tol2* vectors (Supplementary Tables S3 and S4). For unknown reasons, the frequency of trapping the antisense strands of 'known genes' [in which the non-redundant (NR) genes and the genome regions associated with the corresponding ESTs are included] or trapping 'unknown genes' (from which no ESTs have thus far been identified) was higher for the *Tol2* vector (18.4 and 21.6%, respectively) than for the retrovirus counterpart (2.7 and 11.4%, respectively) (Figure 4A).

As for the expression status, only 10.8% of the genes trapped using the conditional retrovirus vector were considered to be transcriptionally silent in mouse undifferentiated ES cells based on the criteria shown in Figure 1A (Figure 4B and Supplementary Table S3). This is consistent with the previous reports showing that MMLV possesses a strong preference to be integrated into transcriptionally active genome regions (32). In contrast, the frequency of trapping potentially silent genes using the *Tol2* counterpart was 25.9%, ~2.4 times higher than that of the retrovirus version (Figure 4B and Supplementary Table S4).

It is also worth noting that the frequencies of identifying genes that had never been trapped were 26.0 and 9.1% for the *Tol2* and retrovirus vectors, respectively (Figure 4C, orange bars), while those of identifying genes that had already been trapped more than 60 times in the IGTC endeavor were 10.9 and 31.8% for the *Tol2* and retrovirus vectors, respectively (Figure 4C, navy bars). This indicates that the spectrum of genes identified by gene trapping with *Tol2* vectors is quite different from that of retrovirus vectors.

We found that the conditional retrovirus vector tends to be preferentially inserted into the promoter regions (located 5′ to the first exons) or into the first introns of trapped genes as has already been shown for a number of

MMLV vectors (Figure 4D, beige bars; Supplementary Figure S3) (33). The *Tol2*-transposon vector, on the other hand, did not show strong preference for particular insertion sites around a gene (Figure 4D). We did not observe a strong integration-site bias toward the last introns of trapped genes (7–9) for either the retrovirus or *Tol2*-transposon vector (Figure 4D, brown bars). This indicates that unbiased poly(A) trapping was indeed attained with our vectors that were constructed by using the NMD-suppressing UPATrap technology (9).

Conditional disruption of trapped genes

In order to confirm the conditionality of gene disruption with our vector, two mutant-cell clones 1TP-84 and TP-32, in which the gene-trap vector had been integrated into the X-chromosomal genes *Atp6ap2* and *Ctps2*, respectively, in a male-derived ES-cell line V6.4 (24), were selected and tested for the recombinase-mediated inversion and deletion of the vector components. In 1TP-84, the vector was integrated into an intron of *Atp6ap2* in a forward orientation (Supplementary Figure S4). In contrast, the reverse strand of *Ctps2* was trapped in TP32 (Supplementary Figure S5).

Atp6ap2 was constitutively expressed in undifferentiated ES cells, but expression of the 3′-portion of *Atp6ap2* (located downstream of the vector-integration point) was completely disrupted in the parental clone 1TP-84 (Figure 5A). Transient expression of the FLPo gene, a codon-optimized version of the FLPe gene (the one for a thermostable variant of the Flp recombinase) (34,35), caused deletion of the NEO cassette and inversion of the gene-terminator cassette with high efficiency (98.9%) in the ES-cell subclones examined (Supplementary Figure S4 and Supplementary Table S5). After the FLPo-mediated first recombination, the expression of *Atp6ap2* fully recovered as expected (Figure 5A).

Next, three 1TP-84 subclones in which the FLPo-mediated recombination had been successfully completed were selected and transiently transfected with an expression vector for the Cre recombinase. In the overwhelming majority (87.5%) of the daughter subclones examined, the gene-terminator cassette was successfully re-inverted to create a disruptive allele for *Atp6ap2* (Supplementary Figure S4 and Supplementary Table S6), and no leakiness of expression of the disrupted 3′-portion of *Atp6ap2* was detected (Figure 5A).

For the second X-chromosomal gene *Ctps2*, we were also able to induce deletion and inversion of the vector components efficiently and obtain a tightly regulated pattern of conditional gene disruption, although the orientation of the vector integration inside *Ctps2* was opposite to that of *Atp6ap2* (Figure 5B, Supplementary Figure S5, and Supplementary Tables S7 and S8).

A transposon-mixture strategy permits straightforward analyses of multiple vector-integration sites

As shown above, the conditional UPATrap-*Tol2* transposon vector has several significant advantages over its retroviral counterpart. The only disadvantage of *Tol2* (or DNA transposons in general), however, is the difficulty in stringently controlling the number of genome-integrated vectors in a target cell. For a gene-trapped ES-cell clone in which multiple copies of a uniform vector are integrated into the genome, precise analysis of the vector-integration sites is not a simple task, and many gene-trapping researchers tend to abandon their newly generated ES-cell clones when they fail to obtain clear results about the vector-integration sites, and the involvement of the multiply genome-inserted vectors is suspected as the cause of their failure.

To overcome this problem, we developed a strategy using a mixture of differentially tagged *Tol2* transposons. Each of the 15 different synthetic tags consisted of two parts: (i) a diagnostic CC-in-poly(AT) part; and (ii) a vector-identification (ID) part (Figure 6A). Each tag is flanked by the common sequences SPL and Term. For the first diagnostic part, the position of the CC dinucleotides in the poly(AT) background is determined according to the ID of each differential tag. For the second part, we designed 15 different vector-ID sequences (30mers) that

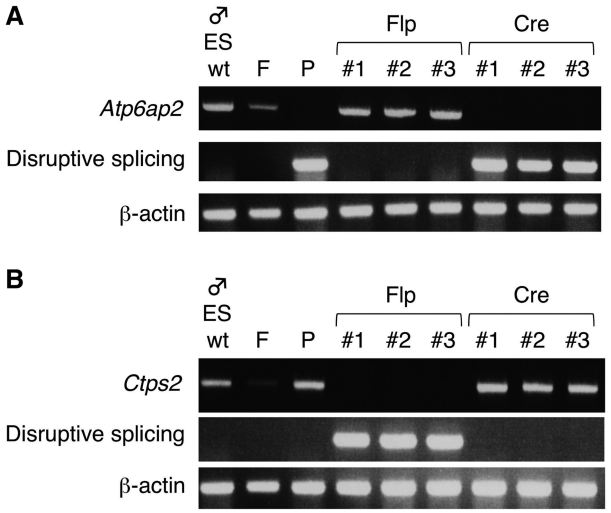

Figure 5. Conditional disruption of the trapped genes. (**A**) Conditional disruption of *Atp6ap2*. (**B**) Conditional disruption of *Ctps2*. The sense and antisense strands of the X-chromosomal genes *Atp6ap2* and *Ctps2*, respectively, were trapped in a male-derived ES-cell line by using the UPATrap-*Tol2* vector. Expression of the *Atp6ap2* and *Ctps2* mRNAs was detected by RT-PCR with the primers located on the exons flanking the introns into which the gene-trap vector was integrated. Disruptive-splicing stands for the splicing of the pre-mRNAs between the upstream exons of the trapped gene and the SA element of the EGFP cassette in the gene-trap vector. Flp-#1, #2, and #3 represent subclones generated after the transient transfection of the parental (P) gene-trapped ES-cell clones 1TP-84 (A) and TP-32 (B) with the Flp-expression vector. Cre-#1, #2 and #3 represent daughter subclones generated after the transient transfection of the Flp-#1, #2 and #3 subclones with the Cre-expression vector. Before the RNA extraction, ES cells were completely depleted of the feeder cells. Identity of the Flp-generated subclones and Cre-generated daughter subclones is as follows: Flp #1, 1TP84/003F; Flp #2, 1TP84/014F; Flp #3, 1TP84/028F; Cre #1, 1TP84/003F/012C; Cre #2, 1TP84/014F/012C; Cre #3, 1TP84/028F/012C in **A**. Flp #1, TP-32/003F; Flp #2, TP-32/014F; Flp #3, TP-32/028F; Cre #1, TP-32/003F/012C; Cre #2, TP-32/014F/012C; Cre #3, TP-32/028F/012C in B. See Supplementary Tables S5–S8 for the derivation of these subclones and daughter subclones. β-actin served as an internal control. F, the mitomycin-C-treated SLPN feeder cells without ES cells.

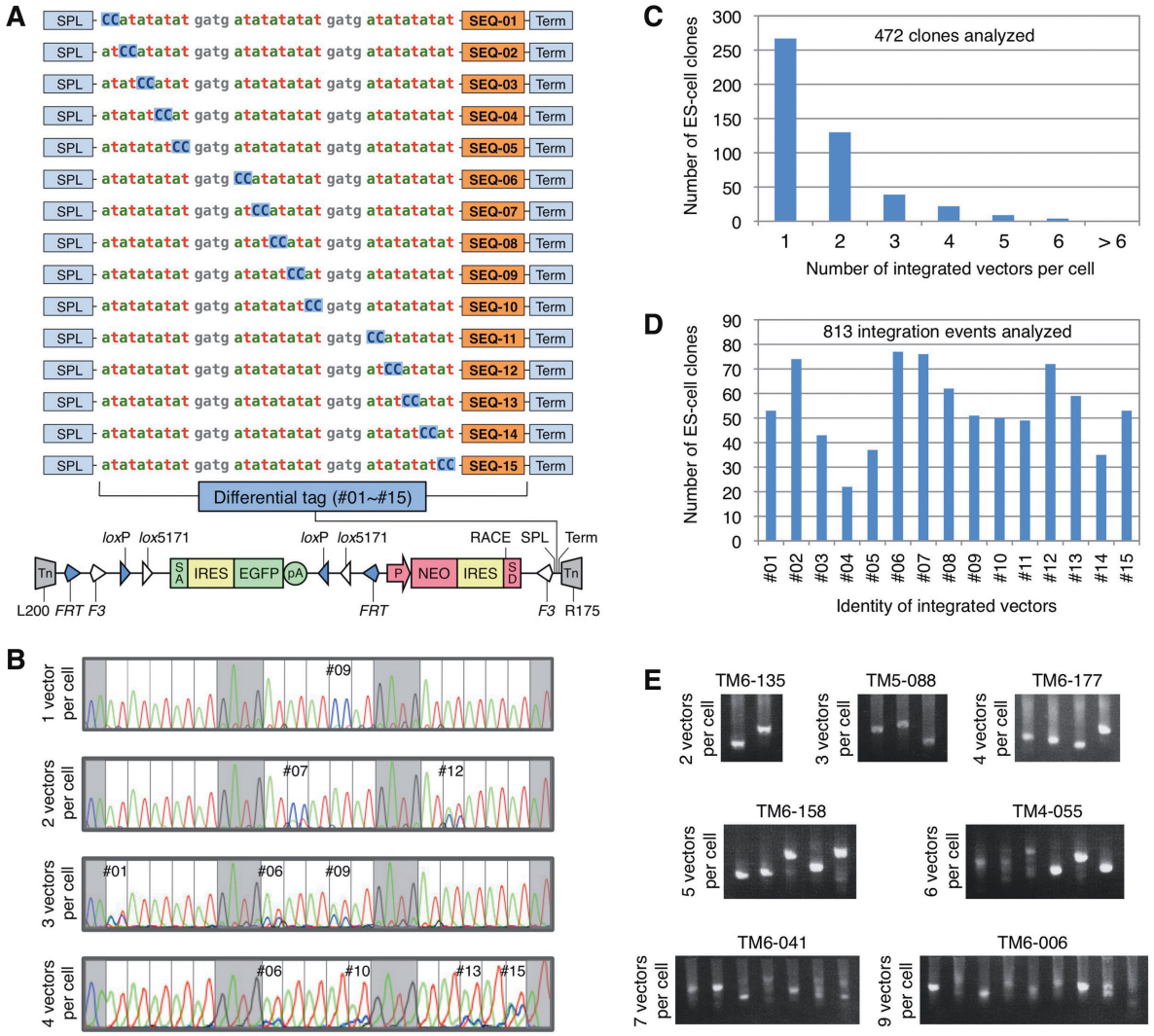

Figure 6. A gene-trap strategy based on the mixture of differentially tagged UPATrap-*Tol2* transposons. (**A**) Structure of the 15 differential tags located between the common SPL and Term sequences near the 3′-ends of the gene-trap vectors. SEQ-01–15 are the synthetic ID nucleotides (30mers) with similar G/C:A/T composition that were designed to serve as the base sequences for the annealing of the PCR and sequencing primers. All of the CC-in-poly(AT) and SEQ-01–15 portions are devoid of the GT (potential splice donor), AG (potential splice acceptor), and AATAAA/ATTAAA [potential poly(A)-addition] sequences in both sense and antisense strands. (**B**) Representative results of the analyses of the number and IDs of the integrated vectors based on the PCR amplification and direct sequencing of the differential-tag portions of the gene-trap vectors integrated into the genome of ES-cell clones. (**C**) Distribution of the number of integrated vectors in an ES-cell clone. The number of integrated vectors was confirmed by the tag-specific PCR when three or more vectors were suspected to be integrated into the genome of an ES-cell clone. See Supplemental Figure S6 for details. (**D**) Usage of the fifteen differentially tagged gene-trap vectors in the transposon-mixture experiments. (**E**) Amplification of different genome portions adjacent to the 3′-ends of multiple integrated vectors from a single ES-cell clone by the tag-directed splinkerette PCR. See Supplementary Figure S7 for details.

are able to serve as the base sites for the annealing of specific primers in both forward and reverse orientations (SEQ-01–15 in Figure 6A). We inserted these differential-tag sequences near the 3′-ends of the conditional UPATrap-*Tol2* vectors and created an equimolar mixture of the 15 differentially tagged transposons.

After we obtain gene-trapped ES-cell clones with the transposon mixture, we first extract the genomic DNA from the cells and amplify the differential tags by PCR. Then, we perform direct sequencing of the amplified tags to learn the number and IDs of the genome-integrated vectors (Supplementary Figure S6A). The results in Figure 6B show examples for one, two, three and four-vector integration events. As the number of the integrated vectors per cell increases, the intensity of the CC-dinucleotide signals becomes weaker. However, by performing PCR-based analyses as shown in Supplementary Figure S6B, we were able to determine the number and IDs of the vectors reproducibly, even for the ES-cell clones containing more than three transposons per cell (Figure 6C). Among the 15 differentially tagged transposons, we observed weak bias in the usage of the particular vector(s) (Figure 6D), but it did not hamper our analyses on the genome-integrated vectors.

Once the number and IDs of the transposons within an ES-cell clone are determined, we analyze the nucleotide

sequences of the vector-integration sites by performing either tag-specific sequencing of the mixed splinkerette PCR products (36,37) or standard sequencing of the DNA fragments that are independently generated through the tag-directed splinkerette PCR, depending upon the number of vectors involved (Supplementary Figure S7). With this strategy, we were able to determine the nucleotide sequences of up to six different vector-integrated sites within an ES-cell clone reproducibly, without performing complicated separation or subcloning procedures (Figure 6E and Supplementary Table S9).

DISCUSSION

We previously developed a revised version of the poly(A)-trapping technology termed UPATrap, and made it possible to create an unbiased pattern of vector integration into endogenous genes by suppressing the adverse effect of NMD (9). Here, we tried to render the conditional gene-disruption capability to the original retrovirus version of UPATrap by incorporating the elaborate FLEx technique (19,21), but we frequently experienced broadly distributed intra-vector deletions/rearrangements that should have deleterious effects on the Flp- and Cre-mediated DNA recombination in the FLEx-type of conditional gene regulation (Figure 3).

In an attempt to elucidate the molecular mechanism(s), we found that the majority of such structural alterations occur around the IRES sequences inside the genome-integrated vectors (Supplementary Figure S2). The IRES sequence of the encephalomyocarditis virus (EMCV), which is one of the most crucial components of the UPATrap strategy (9), is known to form a highly complex secondary structure at the RNA level (38). We suspected that, upon reverse transcription of the retroviral RNA in infected cells, the highly structured portions in the IRES sequences could induce abnormal transfers ('jumps') of the minus-strand cDNA, resulting in generation of deletions/rearrangements in the genome-integrated proviruses as previously observed for some of the retrovirus constructs containing the EMCV-IRES sequences (39,40).

We therefore cloned the conditional UPATrap elements into a cut and paste-type DNA transposon, *Tol2* (22,23,28), and succeeded in suppressing the frequent development of deleterious intra-vector alterations (Figure 3B). Consequently, it became feasible for us to perform unbiased poly(A) trapping in a conditional manner, especially with high reliability (Figure 5). The high stability of the *Tol2* vectors has already been demonstrated in the context of the genomes of cultured ES cells (41) and transgenic mice (42). Since a large fraction of the ~455 thousand mutant ES-cell clones in the current IGTC repository (as of November 2011) have been generated using retrovirus vectors (the IGTC database, http://www.genetrap.org/), we need to be cautious about the integrity of the proviruses (especially those containing the EMCV-IRES sequences) in the genome of the deposited ES-cell clones. The use of the UPATrap-*Tol2* transposons also turned out to be advantageous for identifying/disrupting transcriptionally silent genes in mouse undifferentiated ES cells (Figure 4B), and the chance of trapping genes that have never been captured in the current IGTC effort is significantly higher with the *Tol2*-transposon vector than with the retrovirus counterpart (Figure 4C).

In IGTC, the majority of research groups have been engaged in promoter trapping that was originally developed for the disruption of constitutively expressed genes in the target cells (http://www.genetrap.org/). Interestingly, Friedel *et al*. demonstrated that the expression levels of genes in ES cells required for successful promoter trapping (and targeted promoter trapping as well) is quite low (i.e. higher than 1–5% of the expression level of the transferrin-receptor gene) (43). On the other hand, however, they also showed that the gene-expression levels affect the efficiency of promoter trapping/targeted promoter trapping (43), and our findings shown in Figure 1 are basically consistent with their observations. In addition to conventional promoter trapping, the poly(A)-trapping strategies including original UPATrap (9,11,12) have also been used in a large scale in the IGTC effort in order to capture transcriptionally silent as well as active genes in the target cells. Nevertheless, transcriptionally silent genes in undifferentiated ES cells still remain largely unexplored, as shown in Figure 1. This should probably be at least in part due to the strong preference of retroviruses (the most popular backbone of gene-trap vectors) to be integrated into transcriptionally active genome loci (32,33), and this propensity of retroviruses appears to have been neutralizing the pivotal advantage of poly(A) trapping (i.e. its capability of identifying silent genes).

Although we found that the UPATrap-*Tol2* transposon vector shows a weaker preference to be integrated into transcriptionally active genes than does the retrovirus counterpart (Figure 4B), this does not mean that *Tol2* is completely 'bias-free' in terms of the selection of integration sites. The results of Figure 1 suggest that, among all protein-coding genes, 45.2% would be transcriptionally silent in undifferentiated ES cells, but the frequency of trapping silent genes using our *Tol2* vector was 25.9%, indicating that *Tol2* still has a mild preference to be integrated into transcriptionally active genes (Figure 4B). Among DNA transposons other than *Tol2*, *Sleeping Beauty* (*SB*) and *piggyBac* have been well-characterized and are widely used in the context of mammalian cells (44–46), and a recent investigation suggested that *SB* does not have strong preference to be integrated into transcriptionally active loci (47). To conduct a large-scale random insertional mutagenesis of both transcriptionally silent and active genes in the target cells, it might be reasonable to use *SB* in combination with *Tol2* as the backbone of gene-trapping vectors.

The only disadvantage of the *Tol2*-based gene-trap strategy was the difficulty in stringently regulating the copy number of genome-integrated vectors. To overcome this problem, we generated differentially tagged *Tol2* transposons and subjected their mixture to the random gene-trap experiments, thereby permitting straightforward analyses of multiple vector-integration sites, instead of attempting to obtain only the ES-cell

clones with single-vector integration (Figure 6). Precise information about the multiple vector-integration sites obtained from a single ES-cell clone would allow us to analyze the function(s) of the trapped gene of interest by creating the ES cell-derived mice and segregating the focused allele from the others through mouse crossing. We therefore believe the generation and application of a mixture of differentially tagged UPATrap-*Tol2* transposons should be one of the most potent and versatile gene-trapping strategies aiming at the production of conditionally disrupted alleles for a broad spectrum of genes in the target cells.

As for the current progress of KOMP, the initial target (i.e. conditional disruption of the majority of protein-coding genes in mouse ES cells) appears to be approaching its completion (13). However, because of the elaborate (albeit highly efficient) nature of the procedures involved, the gene-targeting wing of KOMP had to pre-select (or limit) its focus to be almost exclusively on the 'difficult-to-trap' (mostly transcriptionally silent) protein-coding genes (13). In the case of random gene trapping, on the other hand, we do not have to pre-determine our target on the basis of already available knowledge, and a broad spectrum of genes including those without the protein-coding capability (48,49) can be identified and disrupted using limited time, effort, and budget. Besides conventional mouse ES cells, we also have additional candidate cell lines with which we could perform large-scale insertional-mutagenesis experiments [e.g. rat and human ES cells, induced pluripotent stem (iPS) cells, tissue-specific stem cells, and some of the human cancer-cell lines]. The recent derivation of mouse haploid ES-cell lines (50,51) would certainly increase the chance of conducting insertional-mutagenesis experiments based on the phenotypic screening at the individual-laboratory level. The gene-trapping strategy using a mixture of conditional UPATrap-*Tol*2 transposons described in this article should have a lot to contribute to these potential future analyses.

ACCESSION NUMBERS

The GenBank/EMBL/DDBJ accession numbers of the gene-trap vectors are: AB673329, AB673330, AB673331, AB673332, AB673333, AB673334, AB673335, AB673336, AB673337, AB673338, AB673339, AB673340, AB673341, AB673342, AB673343, AB673344, AB673345, AB673346, AB673347, AB673348, AB673349, AB673350, AB673351, AB673352, AB673353, AB673354, AB673355, AB673356, AB673357, AB673358, AB673359 and AB673360.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–7 and Supplementary Tables 1–12.

ACKNOWLEDGEMENTS

We thank J. Takeda (Osaka University, Japan) and A.F. Stewart (TU Dresden, Germany) for kindly providing us with pMC1-Cre-PGK-Puro^r-poly(A) and pCAGGS-FLPo-IRES-Puro^r-poly(A), respectively. We are grateful to N. Yoshida (NAIST, Japan) for his advice on the data retrieval and to J. Mashima (DDBJ, National Institute of Genetics, Japan) for his skillful handling and deposition with public databases of the vector-sequence information. We thank Y. Kawakami, Y. Kimura, A. Ozaki, A. Murakami and A. Oshitani for their technical assistance. We also thank the Iida Foundation for the Iida International Student Scholarship to N.I.M.

FUNDING

The Japan Society for the Promotion of Science (19310130 to Y.I., 21310128 to Y.I., 23241063 to K.K.); and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (Fundamental Technologies Upgrading Program [H19-20 and H21-22] of the National Bio-Resource Project [NBRP] to Y.I.). Funding for open access charge: the NBRP grant from MEXT, Japan.

Conflict of interest statement. None declared.

REFERENCES

1. Austin,C.P., Battey,J.F., Bradley,A., Bucan,M., Capecchi,M., Collins,F.S., Dove,W.F., Duyk,G., Dymecki,S., Eppig,J.T. *et al.* (2004) The Knockout Mouse Project. *Nat. Genet.*, **36**, 921–924.
2. Stanford,W.L., Cohn,J.B. and Cordes,S.P. (2001) Gene-trap mutagenesis: past, present and beyond. *Nat. Rev. Genet.*, **2**, 756–768.
3. Nord,A.S., Chang,P.J., Conklin,B.R., Cox,A.V., Harper,C.A., Hicks,G.G., Huang,C.C., Johns,S.J., Kawamoto,M., Liu,S. *et al.* (2006) The International Gene Trap Consortium Website: a portal to all publicly available gene trap cell lines in mouse. *Nucleic Acids Res.*, **34**, D642–D648.
4. Gossler,A., Joyner,A.L., Rossant,J. and Skarnes,W.C. (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science*, **244**, 463–465.
5. Niwa,H., Araki,K., Kimura,S., Taniguchi,S., Wakasugi,S. and Yamamura,K. (1993) An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J. Biochem.*, **113**, 343–349.
6. Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L., Person,C. and Sands,A.T. (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature*, **392**, 608–611.
7. 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.
8. Matsuda,E., Shigeoka,T., Iida,R., Yamanaka,S., Kawaichi,M. and Ishida,Y. (2004) Expression profiling with arrays of randomly disrupted genes in mouse embryonic stem cells leads to in vivo functional analysis. *Proc. Natl Acad. Sci. USA*, **101**, 4170–4174.
9. Shigeoka,T., Kawaichi,M. and Ishida,Y. (2005) Suppression of nonsense-mediated mRNA decay permits unbiased gene trapping in mouse embryonic stem cells. *Nucleic Acids Res.*, **33**, e20.
10. To,C., Epp,T., Reid,T., Lan,Q., Yu,M., Li,C.Y., Ohishi,M., Hant,P., Tsao,N., Casallo,G. *et al.* (2004) The Centre for Modeling Human Disease gene trap resource. *Nucleic Acids Res.*, **32**, D557–D559.
11. Stanford,W.L., Epp,T., Reid,T. and Rossant,J. (2006) Gene trapping in embryonic stem cells. *Meth. Enzymol.*, **420**, 136–162.
12. Brickman,J.M., Tsakiridis,A., To,C. and Stanford,W.L. (2010) A wider context for gene trap mutagenesis. *Methods Enzymol.*, **477**, 271–295.
13. Skarnes,W.C., Rosen,B., West,A.P., Koutsourakis,M., Bushell,W., Iyer,V., Mujica,A.O., Thomas,M., Harrow,J., Cox,T. *et al.* (2011)

A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, **474**, 337–342.
14. Deng,C.X., Wynshaw-Boris,A., Shen,M.M., Daugherty,C., Ornitz,D.M. and Leder,P. (1994) Murine FGFR-1 is required for early postimplantation growth and axial organization. *Genes Dev.*, **8**, 3045–3057.
15. Di Cristofano,A., Pesce,B., Cordon-Cardo,C. and Pandolfi,P.P. (1998) *Pten* is essential for embryonic development and tumour suppression. *Nat. Genet.*, **19**, 348–355.
16. Rajewsky,K., Gu,H., Kühn,R., Betz,U.A., Müller,W., Roes,J. and Schwenk,F. (1996) Conditional gene targeting. *J. Clin. Invest.*, **98**, 600–603.
17. Gu,H., Zou,Y.R. and Rajewsky,K. (1993) Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-*loxP*-mediated gene targeting. *Cell*, **73**, 1155–1164.
18. Branda,C.S. and Dymecki,S.M. (2004) Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell*, **6**, 7–28.
19. Schnütgen,F., Doerflinger,N., Calléja,C., Wendling,O., Chambon,P. and Ghyselinck,N.B. (2003) A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat. Biotechnol.*, **21**, 562–565.
20. Xin,H.B., Deng,K.Y., Shui,B., Qu,S., Sun,Q., Lee,J., Greene,K.S., Wilson,J., Yu,Y. and Feldman,M. (2005) Gene trap and gene inversion methods for conditional gene inactivation in the mouse. *Nucleic Acids Res.*, **33**, e14.
21. Schnütgen,F., De-Zolt,S., Van Sloun,P., Hollatz,M., Floss,T., Hansen,J., Altschmied,J., Seisenberger,C., Ghyselinck,N.B., Ruiz,P. *et al.* (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc. Natl Acad. Sci. USA*, **102**, 7221–7226.
22. Koga,A., Suzuki,M., Inagaki,H., Bessho,Y. and Hori,H. (1996) Transposable element in fish. *Nature*, **383**, 30.
23. Urasaki,A., Morvan,G. and Kawakami,K. (2006) Functional dissection of the *Tol2* transposable element identified the minimal *cis*-Sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics*, **174**, 639–649.
24. You,Y., Bersgtram,R., Klemm,M., Nelson,H., Jaenisch,R. and Schimenti,J. (1998) Utility of C57BL/6J x 129/Sv^{Jae} embryonic stem cells for generating chromosomal deletions: tolerance to γ radiation and microsatellite polymorphism. *Mamm. Genome*, **9**, 232–234.
25. McMahon,A.P. and Bradley,A. (1990) The *Wnt*-1 (*int*-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell*, **62**, 1073–1085.
26. Morita,S., Kojima,T. and Kitamura,T. (2000) Plat-E: an efficient and stable system for transient packaging of retroviruses. *Gene Ther.*, **7**, 1063–1066.
27. Sambrook,J. and Russel,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Kawakami,K., Shima,A. and Kawakami,N. (2000) Identification of a functional transposase of the *Tol2* element, an *Ac*-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. USA*, **97**, 11403–11408.
29. Nagy,A., Gertsenstein,M., Vintersten,K. and Behringer,R. (2003) *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Fukumura,R., Takahashi,H., Saito,T., Tsutsumi,Y., Fujimori,A., Sato,S., Tatsumi,K., Araki,R. and Abe,M. (2003) A sensitive transcriptome analysis method that can detect unknown transcripts. *Nucleic Acids Res.*, **31**, e94.
31. Soriano,P., Friedrich,G. and Lawinger,P. (1991) Promoter interactions in retrovirus vectors introduced into fibroblasts and embryonic stem cells. *J. Virol.*, **65**, 2314–2319.
32. Scherdin,U., Rhodes,K. and Breindl,M. (1990) Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.*, **64**, 907–912.
33. Wu,X., Li,Y., Crise,B. and Burgess,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.
34. Buchholz,F., Angrand,P.O. and Stewart,A.F. (1998) Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.*, **16**, 657–662.
35. Raymond,C.S. and Soriano,P. (2007) High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells. *PLoS One*, **2**, e162.
36. Devon,R.S., Porteous,D.J. and Brookes,A.J. (1995) Splinkerettes–improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.*, **23**, 1644–1645.
37. Horn,C., Hansen,J., Schnütgen,F., Seisenberger,C., Floss,T., Irgang,M., De-Zolt,S., Wurst,W., von Melchner,H. and Noppinger,P.R. (2007) Splinkerette PCR for more efficient characterization of gene trap events. *Nat. Genet.*, **39**, 933–934.
38. Pilipenko,E.V., Blinov,V.M., Chernov,B.K., Dmitrieva,T.M. and Agol,V.I. (1989) Conservation of the secondary structure elements of the 5′-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res.*, **17**, 5701–5711.
39. Negroni,M. and Buc,H. (2001) Mechanisms of retroviral recombination. *Annu. Rev. Genet.*, **35**, 275–302.
40. Duch,M., Carrasco,M.L., Jespersen,T., Aagaard,L. and Pedersen,F.S. (2004) An RNA secondary structure bias for non-homologous reverse transcriptase-mediated deletions *in vivo*. *Nucleic Acids Res.*, **32**, 2039–2048.
41. Yagita,K., Horie,K., Koinuma,S., Nakamura,W., Yamanaka,I., Urasaki,A., Shigeyoshi,Y., Kawakami,K., Shimada,S. and Takeda,J. (2010) Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proc. Natl Acad. Sci. USA*, **107**, 3846–3851.
42. Sumiyama,K., Kawakami,K. and Yagita,K. (2010) A simple and highly efficient transgenesis method in mice with the *Tol2* transposon system and cytoplasmic microinjection. *Genomics*, **95**, 306–311.
43. Friedel,R.H., Plump,A., Lu,X., Spilker,K., Jolicoeur,C., Wong,K., Venkatesh,T.R., Yaron,A., Hynes,M., Chen,B. *et al.* (2005) Gene targeting using a promoterless gene trap vector ("targeted trapping") is an efficient method to mutate a large fraction of genes. *Proc. Natl Acad. Sci. USA*, **102**, 13188–13193.
44. Ivics,Z., Li,M.A., Mátés,L., Boeke,J.D., Nagy,A., Bradley,A. and Izsvak,Z. (2009) Transposon-mediated genome manipulation in vertebrates. *Nat. Methods*, **6**, 415–422.
45. Horie,K., Kokubu,C. and Takeda,J. (2010) Functional genomics in the mouse using the Sleeping Beauty transposon system. *Meth. Enzymol.*, **477**, 71–89.
46. Li,M.A., Pettitt,S.J., Yusa,K. and Bradley,A. (2010) Genome-wide forward genetic screens in mouse ES cells. *Methods Enzymol.*, **477**, 217–242.
47. Huang,X., Guo,H., Tammana,S., Jung,Y.C., Mellgren,E., Bassi,P., Cao,Q., Tu,Z.J., Kim,Y.C., Ekker,S.C. *et al.* (2010) Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. *Mol. Ther.*, **18**, 1803–1813.
48. The FANTOM consortium. (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
49. Guttman,M., Donaghey,J., Carey,B.W., Garber,M., Grenier,J.K., Munson,G., Young,G., Lucas,A.B., Ach,R., Bruhn,L. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.
50. Leeb,M. and Wutz,A. (2011) Derivation of haploid embryonic stem cells from mouse embryos. *Nature*, **479**, 131–134.
51. Elling,U., Taubenschmid,J., Wirnsberger,G., O'Malley,R., Simon-Pierre,D., Vanhaelen,Q., Shukalyuk,A.I., Schmauss,G., Schramek,D., Schnuetgen,F. *et al.* (2011) Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell*, **9**, 563–574.