Student ID No. 1192030

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

異なるタグで標識された Tol2 トランスポゾンを混合したものは、
マウス ES 細胞中で多種多様な遺伝子の条件的破壊を促進する

Ni Luh Putu Ika Mayasari

Nara Institute of Science and Technology
Graduate School of Biological Sciences
Professor Masashi Kawaichi

January 24, 2012

博士論文番号：1192030

異なるタグで標識されたTol2トランスポゾンを混合したものは、マウスES細胞中で多種多様な遺伝子の条件的破壊を促進する

ニルプトウ　イカ　マヤサリ

奈良先端科学技術大学院大学
バイオサイエンス研究科 動物遺伝子機能学講座
(川市 正史 教授)

平成24年1月24日提出

Student ID No. 1192030

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

Ni Luh Putu Ika Mayasari

Nara Institute of Science and Technology

Graduate School of Biological Sciences

Professor Masashi Kawaichi

January 24, 2012

Student ID No. 1192030

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

A Doctoral Dissertation
Submitted to Graduate School of Biological Sciences
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy (PhD) in Biological Sciences

Ni Luh Putu Ika Mayasari

Advisory committee:

Professor Masashi Kawaichi	(Supervisor)
Professor Jun-ya Kato	(Member)
Professor Narutoku Sato	(Member)
Dr. Kohsuke Kataoka	(Member)
Dr. Yasumasa Ishida	(Member)

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 because it is simple, rapid, and cost-effective. 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 despite exploitation of poly(A)-trapping strategies that were developed to identify or disrupt transcriptionally silent as well as active genes in the target cells.

We confirmed that 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 or rearrangements.

We found that one of the cut and paste-type DNA transposons, *Tol2*, has none of the above disadvantages of retroviruses and can serve as an ideal platform for gene-trap vectors that ensures identification and disruption of a broad spectrum of genes and gene candidates in ES cells. As for the conditional disruption of trapped gene, we were able to attain the high efficiency of intra-vector recombination induction and the tight regulation of gene expression at the RNA level in ES cells.

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. 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 *Tol2* transposons should be one of the most potent and versatile gene-trapping strategies aiming at the production of conditionally disruptive alleles for a broad spectrum of genes in the target cells.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

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 (Austin et al., 2004). 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 (Stanford et al., 2001). The international gene-trap consortium (IGTC; http://www.genetrap.org/; Nord et al., 2006), 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.

1.1 Gene trapping technology

Gene trap insertional mutagenesis in mouse ES cells is an important component of a comprehensive approach to functional annotation of the mouse genome, providing the possibility of trapping a large proportion of the genome (Standford et al., 2001). Gene trapping has been proven to be highly efficient method to generate large number of mutations, and it has been successfully applied by several large-scale projects to generate libraries of gene-trapped ES-cell clones (Friedel and Soriano, 2010).

Two main strategies of gene-trap technology are promoter trapping (Fig. 1A) and poly(A) trapping (Fig. 1B). One of the most commonly used gene-trap methods is promoter trapping which involves a gene-trap vector containing a promoterless selectable-marker

cassette and a 5′ splice acceptor that splices to the upstream exon of the trapped gene (Grossler et al., 1989). The mRNA of selectable-marker gene can be transcribed only when the gene-trap vector is placed under the control of an active promoter of trapped gene (Fig. 1A). Although promoter trapping is effective at inactivating genes, transcriptionally silent loci in the target cells cannot 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 and followed by a splice donor (Fig. 1B; Niwa et al., 1993; Zambrowicz et al., 1998; Ishida et al., 1999; Matsuda et al., 2004). In this strategy, the mRNA of the selectable-marker gene can be stabilized upon trapping of a poly(A) signal of an endogenous gene regardless of its expression status in the target cell. A 5′ splice acceptor with reporter gene is usually included in poly(A) trap vector to monitor the expression of the trapped gene (Fig. 1B). The poly(A) trap strategy was developed for mutation of genes independent of their expression level in mouse ES cells.

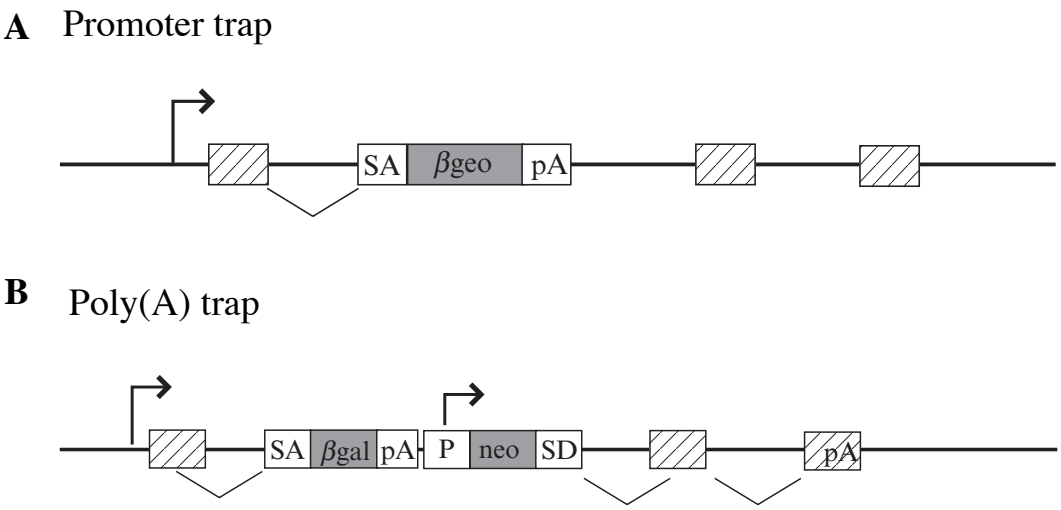

Figure 1. Schematic diagrams of gene trap strategies (Friedel and Soriano, 2010). (**A**) A promoter trap vector contains 5′ splice acceptor (SA), selection gene (*β*geo; *β*-galactosidase/neomycin) and poly(A) addition signal. (**B**) A poly(A) trap vector has two components: an upstream element with 5′ splice acceptor (SA), selection gene (*β*gal; *β*-galactosidase) and poly(A) addition signal to report gene expression, and a downstream element with promoter (P), selection gene (neo; neomycin) and splice donor (SD) to select for drug-resistant intergenic insertions.

1.2 Nonsense-mediated mRNA decay (NMD)

Although the poly(A) trap design should allow inactivation of genes regardless their expression in mouse ES-cell level, we previously showed that 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 (Shigeoka et al., 2005). 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) (Shigeoka et al., 2005).

Nonsense-mediated mRNA decay is a post-transcriptional mechanism that eliminates mRNAs containing premature translation termination codons (PTCs) and code for nonfunctional or even harmful polypeptides (Maquat & Carmichael, 2001). PTC is a UAA, UAG or UGA codon that is located within an mRNA upstream of the normal site of the

translation termination (Fig. 2A). The PTC directs the premature termination of translation. The premature termination of translation causes some exon-exon junction complexes (EJCs) remain in the mRNA, and the interaction between PTC and remaining EJCs goes on to recruit the NMD factors such as up-frameshift (UPF) proteins which are important for NMD activation (Fig. 2B) (Maquat, 2004).

Basically, a cell must distinguish a normal termination codon (TC) from premature one. In mammalian cells, the distinction between TC and PTC depends on the position of the termination codon relative to the 3′-most exon-exon junction. Only those termination codons located more than 50-55 nucleotides upstream of the 3′-most exon-exon junction (measured after splicing) are responsible for induction of NMD (Fig. 3) (Nagy & Maquat, 1998).

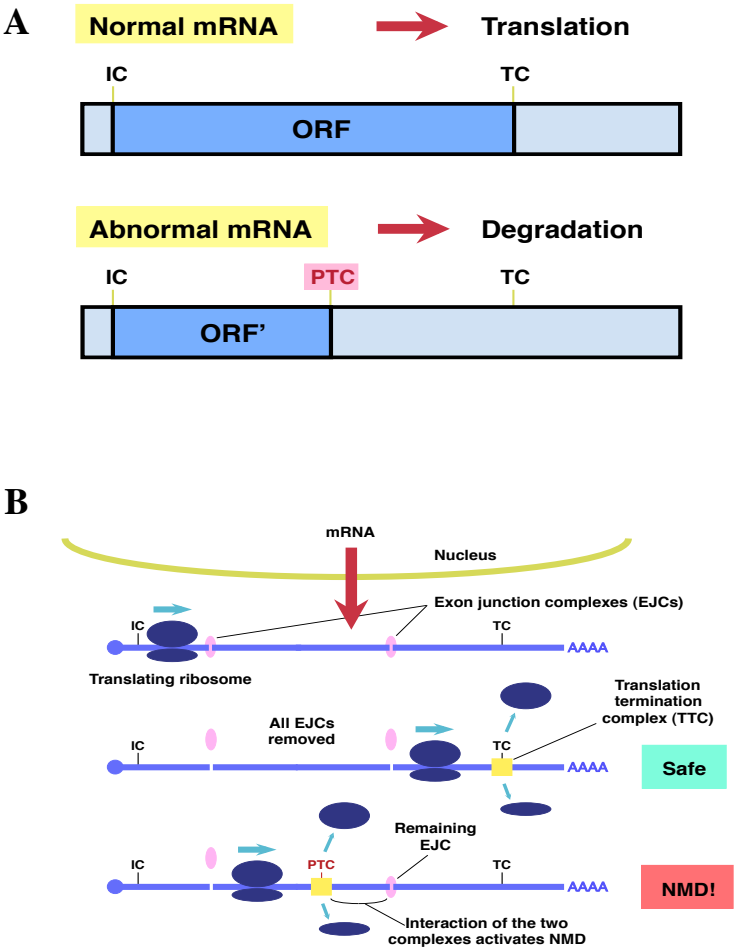

Figure 2. Schematic representation of the molecular mechanism(s) involved in NMD induction. **(A)** PTC located within mRNA upstream of TC. The PTC causes premature termination of translation in abnormal mRNA, and this mRNA will be degraded through mRNA surveillance. (**B**) The role of the exon junction complex in NMD activation. After the splicing, transcript deposits the EJCs, and that EJCs will be removed during translation. In the normal transcript, all EJCs are removed, but in the abnormal transcript containing PTC, the EJC still remain and the interaction between PTC and EJC elicits NMD. IC, initiation codon; TC, normal termination codon; EJC, exon junction complex; PTC, premature termination codon.

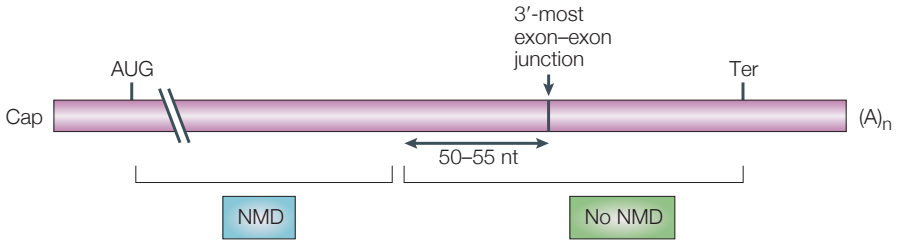

Figure 3. NMD and the "position of an exon-exon junction" rule (Maquat, 2004)**.** Only the 3′-most exon-exon junction within a generic mammalian mRNA is shown. A termination codon that is located in the region indicated in blue, which is followed by an exon-exon junction more than 50-55 nucleotides (nt) downstream (premature termination codon; PTC), elicits nonsense-mediated mRNA decay (NMD), where as termination codon that is located in the region indicated in green fails to elicit NMD. The normal termination codon (Ter) usually resides within the 3′-most exon (Nagy & Maquat, 1998).

1.3 The poly(A)-trapping strategy based on NMD suppression

The first generation of our poly(A)-trapping vector, removable exon trap (RET; Fig. 4A; Ishida and Leder, 1999) inevitably selects for the vector integration into the last intron of a trapped gene, and this strong bias in vector integration was suspected due to NMD (Shigeoka et al. 2005). Insertion of a gene trap vector into the last intron makes the chance of producing a null allele small because such an event results in the deletion of a limited C-terminal portion of the protein encoded by the last exon of the trapped gene. On 2005, we reported that we were able to trap both transcriptionally active and silent genes without a bias in the intragenic vector-integration pattern (Shigeoka et al., 2005). We 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 (Fig. 4B). We assume, insertion of an IRES sequence induces the internal translation toward the end of NEO fusion transcript, which removes EJCs from downstream exon-exon junctions. Other possibility is the IRES sequence generates a highly complex secondary structure at RNA level, and it might hamper the initial steps for NMD by interfering with the interaction between components of the PTC

and downstream EJCs (Shigeoka et al., 2005).

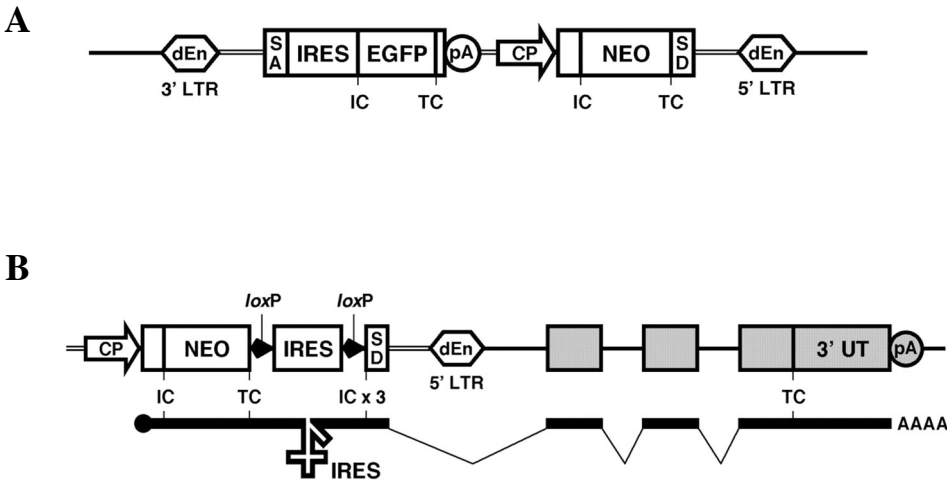

Figure 4. The generation of UPATrap strategy. **(A)** Sructure of RET provirus integrated into the genome of infected cell. **(B)** Structure of UPATrap provirus (3′ half) integrated into the genome of infected cell. The critical elements of the UPATrap vector are the IRES sequence flanked by two tandem *lox*P signals and three ICs (IC × 3) were inserted between the NEO TC and SD sequence of the RET vector. IC, initiation codon; TC, termination codon; dEn, enhancer deletion; LTR, long terminal repeat; SA, splice acceptor; EGFP, enhanced green fluorescence protein; pA, poly-A-addition signal; CP, constitutive promoter; SD, splice donor (Ishida and Leder, 1999; Shigeoka et al., 2005).

1.4 The conditional gene disruption

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 (Deng et al., 1994; Di Cristofano et al., 1998). 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 (Rajewsky et al., 1996). 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 (Gu et al., 1993; Branda et al., 2004). Recently, these techniques have been employed to perform conditional gene disruption in random gene trapping (promoter trapping in particular) with mouse ES cells (Schnütgen et al., 2003; Xin et al., 2005; Schnütgen et al., 2005).

In conditional gene disruption, site-specific recombinases mediated DNA rearrangements by breaking and joining DNA molecules at two specific sites called recombination target. The tyrosine recombinases, Cre and FLPe, are the recombinases most widely used in ES cells at present. Cre recombinase was found in the *E. coli* P1 phage (Sternberg et al., 1986) and FLP recombinase was found in the yeast 2-micron circle (Buchholz et al., 1996; Buchholz et al., 1998). For each of the site-specific recombinases has its own recombination target site. For Cre and FLP, the recombination target sites are termed *lox*P and *FRT*, respectively. They are both composed of inverted 13-bp enzyme-binding sites flanking an 8-bp spacer (Fig. 5). Four mutants of *lox*P (*lox*71, *lox* 66, *lox*511 and *lox* 5171) carry mutations in different places. *lox*71 and *lox*66 carry mutation in the left or in the right 13-bp enzyme-binding sites, respectively. *lox*511 and *lox*5171 have mutation in the spacer sequences. The most commonly used mutant *FRT* is *F3*, which has mutation in spacer sequence (Fig. 5).

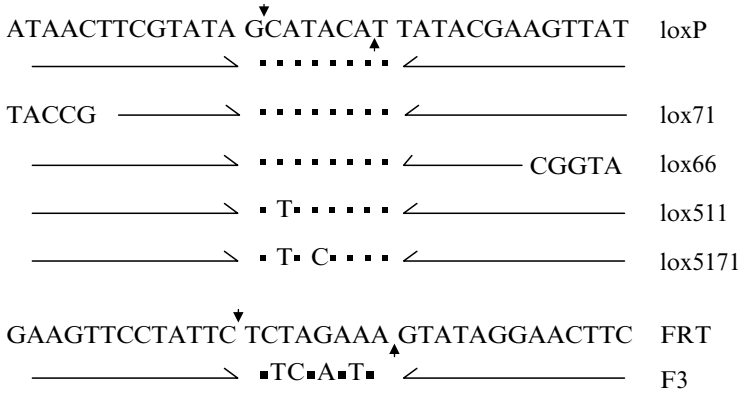

Figure 5. Characteristic of Cre and FLP recombinase target sites (Schnütgen *et al*., 2006**).** Cre and FLP recombinase target signals known as *lox*P and *FRT*, respectively. The 13-bp inverted binding sites are illustrated by inverted arrows flanking the spacer sequences, which contain the region of recombination. The recombination region is denoted by arrowheads above and below the sequences at each end of the spacer. For the variant recombinase target sites, only the sequence differences are shown.

Recombination occurs between two identical recombination target sites. When two recombinase target sites are placed in a DNA molecule so that their spacer are arranged as direct repeats, the DNA interval between recombinase target sites is excised by the appropriate recombinase and released as a circular form (Fig. 6A). When two recombinase target signals are placed in a DNA molecule in an inverted orientation, recombination results in the inversion of the DNA interval between the two RTs (Fig. 6B). The reaction is reversible by the appropriate induction of recombinase. The combination of two or more of recombination target signals in indirect repeat can induce an irreversible recombination (Fig. 7). Schnütgen et al. (2003) developed a strategy called the FLEx switch by using two different recombinase target signals (*lox*P and *lox*511) in alternate orientation, and showed that this strategy can be used to generate a conditional gene disruption.

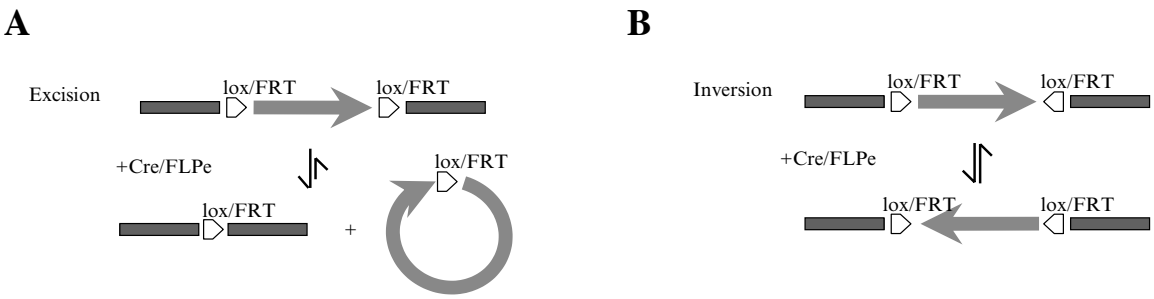

Figure 6. Schematic representation of recombination between two recombinase target signals (Schnütgen et al., 2006). (**A**) Excision recombination between two recombinase target signals in a direct repeat results a covalent circular DNA and a linear DNA. (**B**) Inversion occurs when two recombinase target signals are in an inverted orientation.

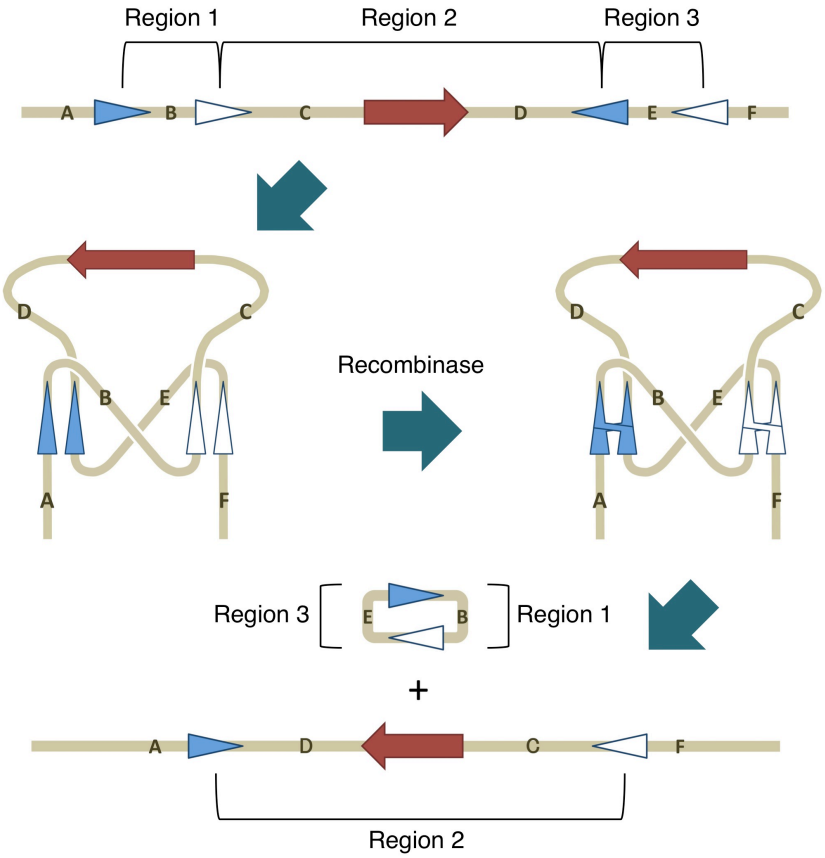

Figure 7. The FLEx-type of coupled homologous recombination and its potential intermediates. A blue arrowhead represents a target sequence of a recombinase (*e.g., lox*P for Cre, or *FRT* for Flp) that has a proper orientation, and a white arrowhead represents another target sequence of the same enzyme (*e.g. lox*5171 for Cre, or *F3* for Flp) that also has a proper orientation. In the presence of the recombinase, two blue signals recombine, and two white signals also recombine. However, a bule signal and a white signal never do so. In order to induce the FLEx-type of homologous recombination, two inverted blue signals and two inverted white signals are alternately placed as shown in the top diagram. A red arrow indicates the orientation of the middle portion. The DNA molecule containing four recombinase-target signals (top) is supposed to form an intermediate structure as shown in the middle-left diagram. In the presence of the recombinase, homologous recombination is induced between the two blue signals and between the two white signals, probably almost simultaneously (middle-right). These paralleled reactions are resolved to form two separated DNA molecules: a circular DNA containing regions B and E, together with a blue signal and a white one; and a linear molecule containing inverted regions C, red arrow and D, as shown in the bottom diagram. Importantly, the final linear product contains only one blue signal and only one white signal that will never recombine to each other, implicating that the recombination reaction is terminated when the two separated DNA molecules are formed as shown in the bottom. In short, region 2 is inverted, and regions 1 and 3 are deleted from the initial DNA molecule during the FLEx-type of coupled homologous recombination.

CHAPTER II

MATERIALS AND METHODS

2.1 Random sampling of mouse UniGene clusters

The UniGene clusters were selected randomly for expression analysis of protein-coding genes. 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 the reference sequence (RefSeq) information for proteins in UniGene database [the UniGene clusters classified as 'transcribe loci' (5509), cDNA with unknown function (224), predicted genes (131), hypothetical protein (3) and others (107)] were excluded. The remaining 1837 clusters for classical protein-coding genes were subjected to further analysis.

The expression level of each gene in undifferentiated ES cells was examined by using NCBI (the National Center for Biotechnology Information) dbESTs libraries #1882, #2512, #10023, #14556, #15703, #17907 and #21037, that constructed using mRNA from undifferentiated ES cells. URLs of the NCBI dbESTs libraries are shown in Table 1. By using NCBI dbESTs libraries, I assessed the number of ES cells-derived expressed sequence tags (ESTs) for 1837 UniGene clusters. I also look in to the presence or absence of corresponding sequence tags in another database for ES cell-derived transcript termed HiCEP (High-Coverage Expression Profilling) database. HiCEP database had been created by using a highly sensitive PCR-based technology (Fukumura et al., 2003; the HiCEP database, http://hicepweb.nirs.go.jp/english/index.html). The analysis by using HiCEP database was performed only for UniGene clusters that do not possess ES cells-derived ESTs. UniGene clusters that have neither ESTs nor HiCEP sequences regarded as transcriptionally silent genes in ES cells. The expression of UniGene clusters that do not have ESTs but possessed

HiCEP sequences are considered low in ES cells. UniGene clusters with ESTs are regarded as being expressed in ES cells.

Other than the expression level of each gene in ES cells, I also performed analysis for the gene-inactivation status by using IGTC database for gene disruption by random gene trapping and IKMC (the International Knockout Mouse Consortium; http://www.knockoutmouse.org/) database for individual gene inactivation by gene targeting.

Table 1. URLs of the NCBI dbEST libraries for mRNAs derived from mouse undifferentiated ES cells

Library 1882 (dbEST ID) (9,615 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=1882

Library 2512 (dbEST ID) (19,923 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=2512

Library 10023 (dbEST ID) (6,623 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=10023

Library 14556 (dbEST ID) (3,908 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=14556

Library 15703 (dbEST ID) (6,670 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=15703

Library 17907 (dbEST ID) (1,118 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=17907

Library 21037 (dbEST ID) (95,566 EST sequences)
http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Mm&LID=21037

2.2 Vectors for gene trapping

The 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 to ensure complete transcriptional termination), the synthetic double-stranded (ds) oligonucleotides for the annealing of the 3′ RACE primers (RACE; Rapid Amplification of cDNA Ends) and the splinkerette genome-PCR primers (SPL) were inserted into the UPATrap-EGFP retrovirus vector as shown in Figure 8A to create the conditional UPATrap-Moloney retrovirus vector, pCRV2. The internal (*i.e.* non-retrovirus-derived) components of pCRV2 (the 5.73-kb *Xho*I-*Not*I fragment) and the synthetic SPL oligonucleotides (ds) were transplanted into the *Xho*I-*Bgl*II site of a *Tol2*-transposon plasmid pT2AL200R175G (Urasaki et al., 2006) to create pCTP2F and pCTP2R, the *Tol2* versions of the UPATrap vectors (Fig. 8B). Each one of the *Tol2* vectors for the mixture experiments was constructed by ligating the 5.73-kb *Xho*I-*Not*I fragment of pCRV2, the synthetic SPL oligonucleotides (ds), one of the synthetic CC-in-poly(AT) (the gene-trap series TM4, TM5 and TM6) or AA-in-poly(TT) (the gene-trap series TM1 and TM2) oligonucleotides (ds), the corresponding synthetic 30mer oligonucleotides (ds) (*i.e.* one of the SEQ-01 ~ 15 sequences) and the synthetic Term oligonucleotides (ds) into the *Xho*I-*Bgl*II site of pT2AL200R175G (Fig. 8C). 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 Table 2.

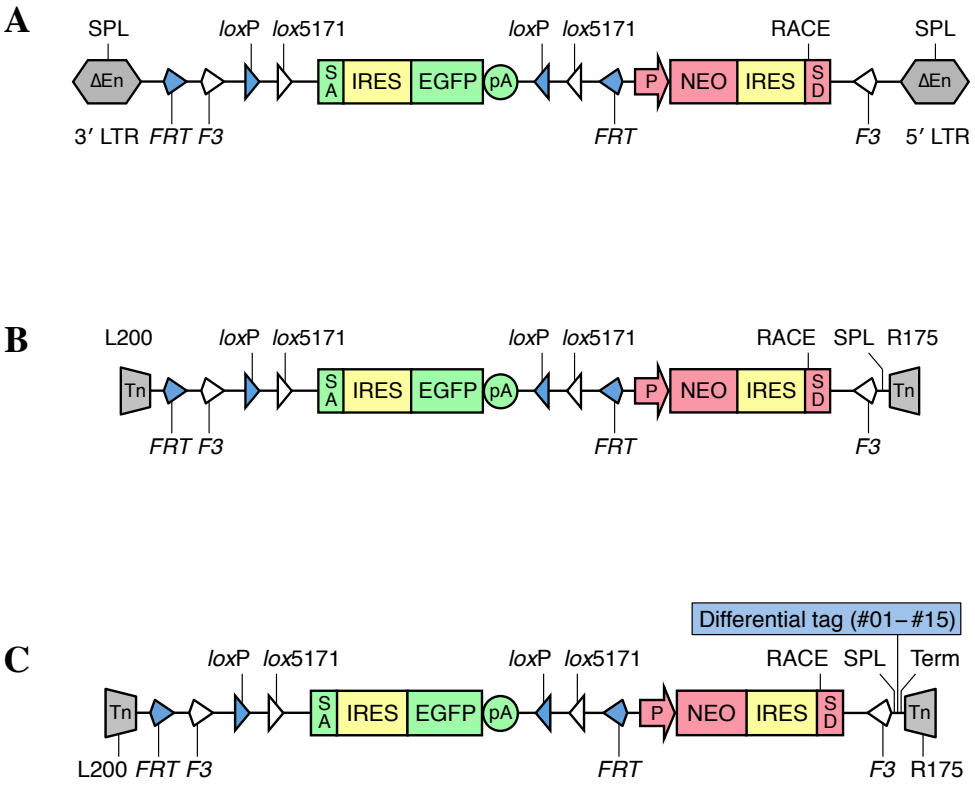

Figure 8. Structure of conditional UPATrap vectors. (**A**) The conditional UPATrap-Moloney retrovirus vector pCRV2. (**B**) The conditional UPATrap-*Tol2* transposon vector pCTP2R. (**C**) The conditional UPATrap-*Tol2* transposon mixture vector pTMtt/pTMat. Orientation of the triangular and diamond-shaped arrowheads represents those recombinase-target signals *lox*P, *lox*5171, *FRT* and *F3*. SA, the splice acceptor sequence of the human Bcl-2 gene (the intron 2-modified exon 3 portion); SD, the splice donor sequence of 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, four tandemly connected poly(A)-addition signals derived from the mouse and human growth-hormone genes; LTR, the long terminal repeat of the MMLV; ΔEn, enhancer deletion (Soriano et al., 1991); Tn, terminal essential sequences (L200 and R175) of *Tol2* transposon (Urasaki et al., 2006); RACE, the synthetic nucleotide sequence (90mer) that facilitates 3′ RACE; SPL, the synthetic nucleotide sequence (90mer) that facilitates splinkerette genome PCR. 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.

Table 2. The GenBank/EMBL/DDBJ accession numbers of the gene-trap vectors

pCRV2	AB673329
pCTP2F	AB673330
pTMtt-01	AB673331
pTMtt-02	AB673332
pTMtt-03	AB673333
pTMtt-04	AB673334
pTMtt-05	AB673335
pTMtt-06	AB673336
pTMtt-07	AB673337
pTMtt-08	AB673338
pTMtt-09	AB673339
pTMtt-10	AB673340
pTMtt-11	AB673341
pTMtt-12	AB673342
pTMtt-13	AB673343
pTMtt-14	AB673344
pTMtt-15	AB673345
pTMat-01	AB673346
pTMat-02	AB673347
pTMat-03	AB673348
pTMat-04	AB673349
pTMat-05	AB673350
pTMat-06	AB673351
pTMat-07	AB673352
pTMat-08	AB673353
pTMat-09	AB673354
pTMat-10	AB673355
pTMat-11	AB673356
pTMat-12	AB673357
pTMat-13	AB673358
pTMat-14	AB673359
pTMat-15	AB673360

2.3 ES-cell culture and gene trapping

The V6.4 ES cells (You et al., 1998) were cultured in Dulbecco's modified Eagle's medium supplemented with 15% fetal calf serum, 2 mM L-glutamine, 50 U/ml penicillin, 50 μg/ml streptomycin, 1×10^{-4} M 2-mercaptoethanol and 1×10^{-4} M non-essential amino acids (Matsuda et al., 2004). The ES cells were grown on a layer of mitomycin C-treated SNL-STO cells (McMahon and Bradley, 1990) 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 (the SLPN cells, unpublished).

The recombinant retrovirus was produced using the Plat-E packaging cell line as previously described (Morita et al., 2000) and transfected with retrovirus vector by using FuGENE®6 Transfection Reagent (Roche) according to the manufacturer's protocol. The culture supernatant was harvested 48 hours after transfection and filtered through a 0.45 μm filter. ES cells were infected with the recombinant retrovirus in the presence of 8 μg/ml of Polybrene (Sigma) for 2 hours at 37°C and selected under 200 μg/ml of G418 (Nacalai) for 7−10 days as previously described (Matsuda et al., 2004; Shigeoka et al., 2005). Drug-resistant colonies were isolated manually into 12-well plates and, after making frozen-cell vials, the high molecular-weight (HMW) genomic DNA and the total cellular RNA were extracted using a standard procedure from the expanded cells (Sambrook and Russel, 2001).

For *Tol2*-transposon experiment, 2.5×10^5 ES cells were co-transfected with 2.27 μg of pCAGGS-TP (a constitutive expression vector for the *Tol2* transposase) (Kawakami et al., 2000; Urasaki et al., 2006) and either 0.23 μg of one of the above *Tol2* gene-trap plasmids or 0.23 μg of a mixture of the differentially tagged *Tol2* gene-trap plasmids (#01~#15) by using the TransFast reagent (Promega) according to the manufacturer's protocol. The transfected

cells were plated into one well of a 24-well plate. The G418 selection and subsequent steps were carried out as described above for gene trapping using the retrovirus vector.

2.4 Detection of intra-vector deletions

I assessed the integrity of the two different regions inside each genome-integrated vector by the genomic PCR. The Lr and Sr regions of the conditional UPATrap-Moloney retrovirus vector (Fig. 9A) 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 (Fig. 9B) were amplified using the 5FRT-F1 and RN2 primers and the RNApol2-F1 and R-term primers, respectively. The PCR conditions were as follows: (*i*) 94°C, 3 min, (*ii*) 30 × (94°C, 30 sec; 60°C, 30 sec and 72°C, 3 min) and (*iii*) 72°C, 5 min. The amplification size was confirmed by agarose gel electrophoresis (2% agarose gel in 1× TAE buffer). The nucleotide sequences of the primers are shown in Table 3.

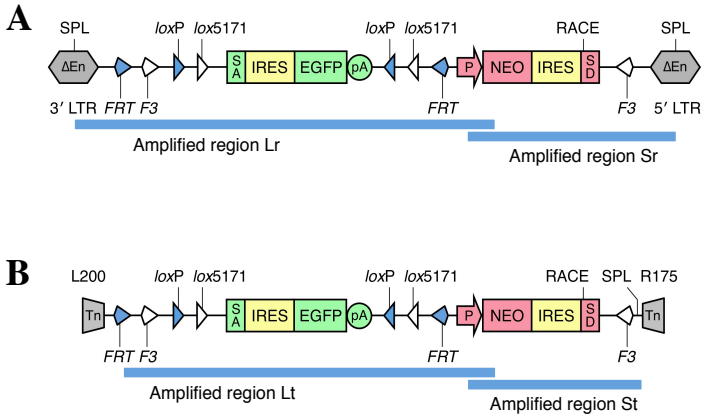

Figure 9. Two different regions of intra-vector deletion analysis. (**A**) Amplification region of conditional UPATrap-Moloney retrovirus vector. Lr, long region of retrovirus vector; Sr, short region of retrovirus vector. (**B**) Amplification region of conditional UPATrap *Tol2* transposon vector. Lt, long region of *Tol2* transposon vector; St, short region of *Tol2* transposon vector.

2.5 Conditional mutagenesis in ES cells

The two ES-cell clones 1TP-84 and TP-32 with vector integration into the X-chromosomal genes *Atp6ap2* and *Ctps2*, respectively, were tested for the efficiencies of intra-vector recombination induction and conditional inactivation of the trapped genes.

For the first step, the ES-cell clones were transiently transfected with pCAGGS-FLPo-IRES-Puro^r-poly(A), and 24 hours after transfection, they were subjected to 48 hours of brief selection with puromycin (1 μg/ml). Then, limiting dilution of the transfected cells was carried out on a layer of the 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 plate containing the standard ES-cell medium and the other one supplemented with 200 μg/ml of G418 (Nacalai). The genomic DNA was extracted from the unselected group of cells after several days of expansion. Structure of both 5′ and 3′ portions of genome-integrated vectors was analyzed as indicated in Figures 10 and 11.

For the second step, six of the FLPo-generated subclones (1TP84/003F, 1TP84/014F and 1TP84/028F for the parental clone 1TP-84, and TP-32/003F, TP-32/014F and TP-32/028F for the parental clone TP-32) were transiently transfected with pMC1-Cre-PGK-Puro^r-poly(A). The following steps including the brief selection with puromycin, the limiting dilution of the Cre-transfected cells and the analyses of the vector structures at the 5′ and 3′ ends were carried out as described above for the FLPo experiment (Figs. 10 and 11), but the G418-sensitivity test was not performed for the Cre-generated cells.

Six of the Cre-generated daughter subclones (1TP84/003F/012C, 1TP84/014F/012C and 1TP84/028F/012C for the parental clone 1TP-84, and TP-32/003F/012C, TP-32/014F/012C and TP-32/028F/012C for the parental clone TP-32) were chosen and, together with six of the FLPo-generated subclones, subjected to the analyses of the efficiency of

conditional regulation of the trapped-gene expression. For the studies of expression of the trapped genes at the mRNA level, the original V6.4 cells, the parental 1TP-84 and TP-32 cells, the FLPo-generated six subclones and the Cre-generated six daughter subclones were expanded and depleted of the residual mitomycin C-treated SLPN cells by using the standard separation procedure (Nagy et al., 2003). The total cellular RNA was extracted from the feeder-depleted ES cells, and after the synthesis of the first strand cDNA using the SuperScript II RT (Invitrogen) and the oligo(dT)$_{12\text{-}18}$ primer (GE Healthcare), expression of *Atp6ap2* and *Ctps2* was assessed at the mRNA level 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 (*Atp6ap2*), and using the CTPS2-F2 (located on the sense strand of exon 12) and CTPS2-R1 (located on the anti-sense strand of exon 17) primers (*Ctps2*). The disruptive-splicing events were detected using the ATP-Ex7-F and Bcl2-R primers for *Atp6ap2*, and using 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 PCR conditions were as follows: (*i*) 94°C, 3 min, (*ii*) 30 × (94°C, 30 sec; 60°C, 30 sec and 72°C, 1 min) and (*iii*) 72°C, 3 min. The amplification size was confirmed by agarose gel electrophoresis (2% agarose gel in 1 × TAE buffer). The nucleotide sequences of the primers are shown in Table 3.

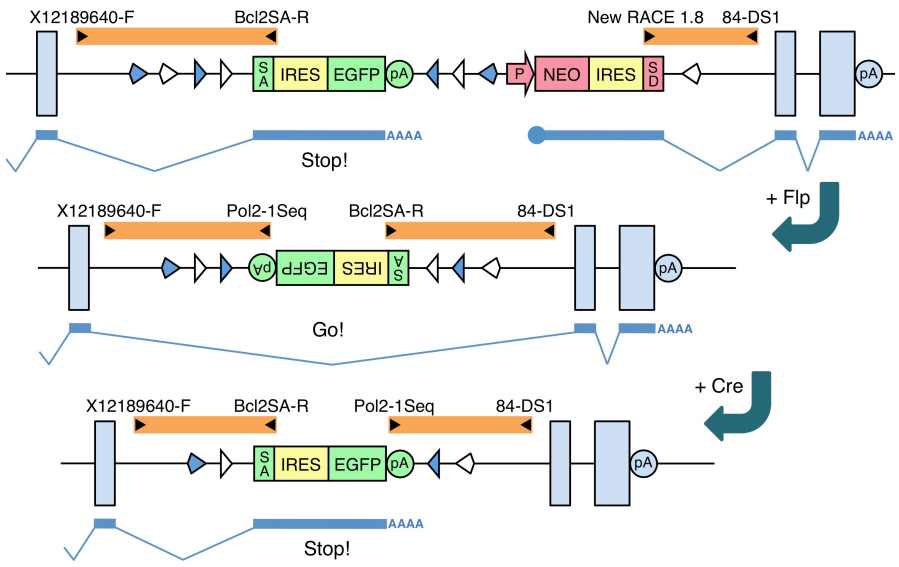

Figure 10. Assessment of the efficiency of the FLPo-mediated and Cre-mediated coupled homologous recombination in the conditional UPATrap-*Tol2* transposon vector integrated into *Atp6ap2* on the mouse X chromosome. The ES-cell clone 1TP-84, in which the conditional UPATrap-*Tol2* transposon vector had been integrated in a forward orientation into the intron 8 of *Atp6ap2* on the X chromosome, was transiently transfected with a plasmid for the expression of FLPo (driven by the CAG promoter). The inversion/deletion status of the internal components of the integrated vectors was analyzed by PCR as shown in the diagram. Then, three subclones, in which the FLPo-mediated coupled recombinations (inversion and deletion) had been successfully induced, were chosen, transiently transfected with a plasmid for the expression of Cre (driven by the MC1 promoter). The inversion/deletion status of the internal components of the integrated vectors was analyzed by PCR as shown in the diagram. Black arrowheads and horizontal orange bars represent the PCR primers and the PCR products, respectively. The actual sizes and numbers of exons and introns of *Atp6ap2* and vector-integration site are not reflected exactly in the diagram.

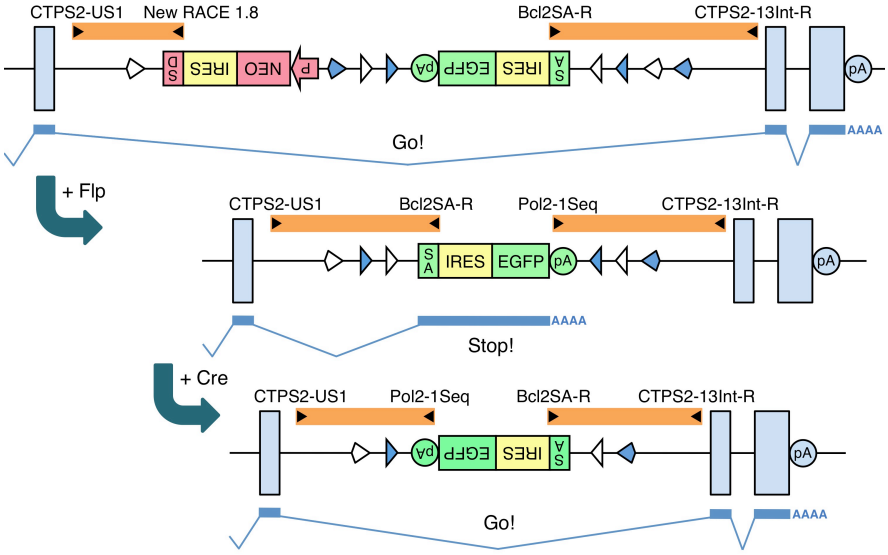

Figure 11. Assessment of the efficiency of the FLPo-mediated and Cre-mediated coupled homologous recombination in the conditional UPATrap-*Tol2* transposon vector integrated into *Ctps2* on the mouse X chromosome. The ES-cell clone TP-32, in which the conditional UPATrap-*Tol2* transposon vector had been integrated in a reverse orientation into the intron 13 of *Ctps2* on the X chromosome, was transiently transfected with a plasmid for the expression of FLPo (driven by the CAG promoter). The inversion/deletion status of the internal components of the integrated vectors was analyzed by PCR as shown in the diagram. Then, three subclones, in which the FLPo-mediated coupled recombinations (inversion and deletion) had been successfully induced, were chosen, transiently transfected with a plasmid for the expression of Cre (driven by the MC1 promoter). The inversion/deletion status of the internal components of the integrated vectors was analyzed by PCR as shown in the diagram. Black arrowheads and horizontal orange bars represent the PCR primers and the PCR products, respectively. The actual sizes and numbers of exons and introns of *Ctps2* and vector-integration site are not reflected exactly in the diagram.

2.6 PCR analysis and sequencing for trapped gene

2.6.1 3′ RACE PCR

Total cellular RNA was extracted from gene-trapped ES cells grown in 6-well plates by using Sepasol reagent (Nacalai Tesque) according to the manufacturer's instructions. The Moloney murine leukemia virus reverse transcriptase (MMLV-RT) (Invitrogen) was used to synthesize cDNA strands in 20 μl reaction (0.1−3.0 μg of total RNA) and the AD8-20dT primer (10 μM). Two rounds of 3′ RACE PCR were performed in 10 μl reaction by using Advantage-GC2 polymerase mix (Clontech) and 10 μM for each primer. The first round, 1 μl of the above cDNA mix was amplified with primer pairs New-RACE-0.9 and AD8 in 10 μl reaction. The PCR conditions were as follows: (*i*) 94°C, 3 min, (*ii*) 2× (94°C, 30 sec; 72°C, 4 min), (*iii*) 2× (94°C, 30 sec; 70°C, 2 min; 72°C, 2 min), (*iv*) 2× (94°C, 30 sec; 68°C, 2 min; 72°C, 2 min), (*v*) 36× (94°C, 30 sec; 66°C, 2 min; 72°C, 2 min) and (*vi*) 72°C, 5 min. For the second round, I used 1 μl of the first PCR products as a template under the similar condition using the New-RACE-1.8 and AD8-4dT primers. After confirming the successful amplification of the DNA fragment in 2% agarose gel electrophoresis, the second round of 3′ RACE PCR product was purified by using ExoSAP-IT (USB Corporation). Nucleotide sequences of the purified-PCR fragments were determined by direct sequencing using the New-RACE-Seq primer (10 μM) and the BigDye Terminator V3.1 cycle sequencing kit (Applied Biosystem) according to the manufacturer's instructions. The thermal cycled reactions are (*i*) 96°C for 2 min, (*ii*) 25× (96°C, 20 sec; 50°C, 15 sec and 60°C, 4 min). The sequencing reaction is cleaned-up using the BigDye X-Terminator purification kit (Applied Biosystems) according to the manufacturer's protocol, and the sequencing results were obtained by using the ABI PRISM 3100 genetic analyzer (Applied Biosystems). The 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 mouse genome (July 2007). The nucleotide sequences of the primers are shown in Table 3.

2.6.2 Splinkerette genomic PCR

For the ES-cell clones generated using the retrovirus vector, the HMW genomic DNA extracted from individual gene-trapped ES-cell clones that had been expanded in 12-well plates was digested with *Hae*III (New England BioLabs) and, after the 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) in 20 ml reaction volume. Then, the linker-ligated DNA was digested with *Pvu*II (New England BioLabs) to avoid amplification of the internal vector components. The *Pvu*II-digested DNA was used as a template for the first round of PCR with the primer pairs SPL-1 and P1 and the Advantage-GC2 polymerase mix (Clontech) in a reaction volume of 10 μl. The PCR conditions were as follows: (*i*) 94°C, 2 min, (*ii*) 30× (94°C, 30 sec; 60°C, 30 sec; 72°C, 2 min) and (*iii*) 72°C, 7 min. The second round of PCR was performed using 1 μl of the 1/10 diluted first-round PCR product and the primer pairs SPL-2 and P2 in the same reaction condition described above in the first-round PCR. Direct sequencing was carried out as described for 3′ RACE PCR with the New-Spl2.3 primer.

For the ES-cell clones generated using the *Tol2*-transposon vector (CTP2F) or the mixture of the differentially tagged *Tol2*-transposon vectors, the genomic DNA was digested with *Hae*III, *Taq*I or *Msp*I (New England BioLabs), and, after the 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 the amplified products were carried out in the same reaction conditions as described for the ES-cell clones generated using the retrovirus vector,

by 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. The New T-Spl3 was used as the sequencing primer. The PCR and sequencing primers used in the splinkerette-type genome analyses for the ES-cell clones generated using the mixture of the differentially tagged transposons vary, depending on the number and identities of the genome-integrated vectors (Fig. 12). The nucleotide sequences of the primers and linkers are shown in Table 3.

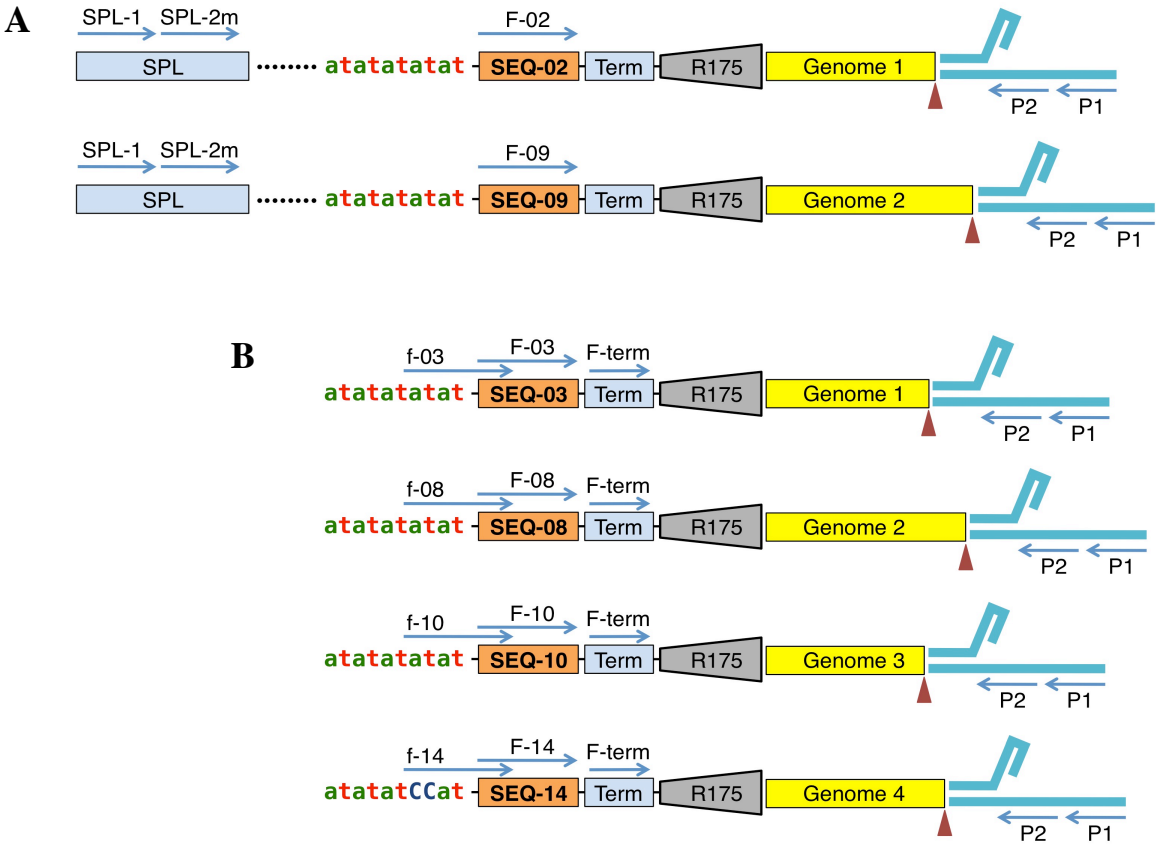

Figure 12. Splinkerette genome PCRs for differentially tagged transposons. (**A**) Tag-specific sequencing of the mixed splinkerette PCR products. The genomic DNA extracted from ES-cell clones containing one or two vectors per cell is digested with a restriction endonuclease (a red vertical arrowhead) and ligated with a splinkerette-type linker (a pair of the straight and bent blue lines). Then, the nested PCR is performed using combinations of the common primers and the linker-specific primers: SPL-1 and P1 (for the first round); and SPL-2m and P2 (for the second round). For ES-cell clones with single-vector integration, a common internal primer, such as F-term shown in the panel B, is used as a sequencing primer. On the other hand, the tag-specific primers are used for direct sequencing for ES-cell clones containing two vectors per cell. A schematic example of an ES-cell clone integrated with the vectors #2 and #9 is shown. (**B**) Standard sequencing of the DNA fragments that are independently generated through the tag-directed splinkerette PCR. The genomic DNA extracted from ES-cell clones containing more than two vectors per cell is digested with a restriction endonuclease and ligated with a splinkerette-type linker. Then, the nested PCR is performed using combinations of the tag-specific primers and the linker-specific primers: e.g. f-08 and P1 (for the first round); and F-08 and P2 (for the second round). A common internal primer F-term is used for direct sequencing of each splinkerette-PCR product. A schematic example of an ES-cell clone integrated with the vectors #3, #8, #10 and #14 is shown.

2.7 Analysis of the number and identities 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. The PCR conditions were as follows: (*i*) 94°C, 3 min, (*ii*) 30× (94°C, 30 sec; 60°C, 30 sec; 72°C, 1 min) and (*iii*) 72°C, 5 min. Amplification of DNA fragment was confirmed using agarose gel electrophoresis (1% agarose gel in 1× TAE buffer). After purification of the PCR products by using ExoSAP-IT® (USB Corporation), their nucleotide sequences were determined in the same direct-sequencing reaction as described for the 3′ RACE products using the RS-F4 oligo as a primer (Fig. 13A).

The confirmation of the 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 (Fig. 13B). The PCR conditions were as follows: (*i*) 94°C, 3 min, (*ii*) 30× (94°C, 30 sec, 60°C, 30 sec; 72°C, 1 min) and (*iii*) 72°C, 3 min. Amplification of DNA fragment was confirmed by using agarose gel electrophoresis (2% agarose gel in 1× TAE buffer). The nucleotide sequences of the primers are shown in Table 3.

Figure 13. Determination of the number and IDs of differentially tagged transposons integrated into the genome of a target ES-cell clone. (**A**) A simple and straightforward determination strategy for the ES-cell clones containing one, two or three vectors per cell. The genomic DNA is extracted from all of the gene-trapped ES-cell clones, and the differential-tag portions are amplified by using the PCR primers New-RACE-0.9 and R-term. Then, the nucleotide sequences of the CC-in-poly(AT) portions are determined through the direct sequencing of the PCR products by using RS-F4 as a primer. (**B**) A reliable determination strategy for the ES-cell clones containing more than three vectors per cell. The simple and straightforward method shown in **A** cannot be reliably applied for the ES-cell clones containing more than three vectors per cell. This is because the CC signals in the sequencing chromatograms become weaker, and distinction of the CC signals from the background noise turns more difficult as the number of the genome-integrated vectors increases, especially beyond three. To solve this problem, a set of 15 independent tag-specific PCRs are performed (but no sequencing reactions are involved here) for each of the gene-trapped ES-cell clones that are suspected to contain more than three vectors from the results of the initial analysis as shown in **A**, and the presence or absence of the vectors is examined individually. The diagram represents the successful confirmation of the presence of the vectors #04, #06, #09 and #14 in a gene-trapped ES-cell clone.

Table 3. List of primers and linkers, and their nucleotide sequences

Primers used for genome PCR to examine intra-vector deletion	
SPL-1	5′-CATGCATCATATCCATCGCAATCGCATCC-3′
RN2	5′-CCGCTTCAGTGACAACGTCGAGCACAG-3′
RNApol2-F1	5′-GGGCCTCTCCACTAATAACGGCTACTTCCA-3′
U5-R1	5′-GGGTCTCCTCTGAGTGATTGACTACCCGTC-3′
5FRT-F1	5′-GAAAGTATAGGAACTTCGGCATCGGCATC-3′
R-term	5′-GATCCTTTATCGCAATGGATGCGATGATCC-3′
Primers used for 3′ RACE PCR	
AD8-20dT	5′-GCGAGGCTCCTAGCTCTAGACTGCACTCAAC(T)20-3′
New-RACE-0.9	5′-ACGCCATGGCCAATCGCATCATGGATTCCATC-3′
AD8	5′-GCGAGGCTCCTAGCTCTAGACTGCACTCAAC-3′
New-RACE-1.8	5′-GATGCGAATCGCATCCATGGATGCGCCAATC-3′
AD8-4dT	5′-GCGAGGCTCCTAGCTCTAGACTGCACTCAACTTTT-3′
New-RACE-Seq	5′-CGGATTCCATGCGATGCCATTTCCGATGCG-3′
Linkers used for splinkerette genome PCR	
SplT-BLT	5′-CGAATCGTAACCGTTCGTACGAGAATTCGTACGAGAATCGCTGTCCTCTCCAACGAGCCAAGG-3′
SplB-BLT	5′-CCTTGGCTCGTTTTTTTTTGCAAAAA-3′
SplT-Msp	5′-CGAATCGTAACCGTTCGTACGAGAATTCGTACGAGAATCGCTGTCCTCTCCAACGAGCCAAGGC-3′
SplB-Msp	5′-CGGCCTTGGCTCGTTTTTTTTTGCAAAAA-3′
SplT-Taq	5′-CGAATCGTAACCGTTCGTACGAGAATTCGTACGAGAATCGCTGTCCTCTCCAACGAGCCAAGGT-3′
SplB-Taq	5′-CGACCTTGGCTCGTTTTTTTTTGCAAAAA-3′
Primers used for splinkerette genome PCR in the experiments with CRV2 or CTP2F	
P1	5′-CGAATCGTAACCGTTCGTACGAGAA-3′
P2	5′-TCGTACGAGAATCGCTGTCCTCTCC-3′
SPL-1	5′-CATGCATCATATCCATCGCAATCGCATCC-3′
SPL-2	5′-CCATTCAATTCGCATCGCATATCCGATGCC-3′
SPL-2m	5′-CCATTCAATTCACATCGCATATCCGATGCC-3′
New-Spl2.3	5′-CATCGCATATCCGATGCCGATATGGCATC-3′
New T-Spl1	5′-CGGATATGCGATGTGAATTG-3′
New T-Spl2	5′-GATGTGAATTGAATGGGGATG-3′
New T-Spl3	5′-CGATTGCGATGGATATGATG-3′
F-term	5′-GGATCATCGCATCCATTGCGATAAAGG-3′
Primers used for confirmation of FLPo and Cre-mediated recombination in 1TP-84	
New-RACE 1.8	5′-GATGCGAATCGCATCCATGGATGCGCCAATC-3′
Bcl2SA-R	5′-GAGGGGATGCATATTATTTCTACTGCTTTAGTG-3′
84-DS1	5′-CCGAATGTCTGGTTTCGTCTCCAGCATC-3′
X12189640-F	5′-CATCGACTCTGTGAGATTGGAACTAACAC-3′
Pol2-1Seq	5′-ATGAGGAAATTGCATCGCATTGTC-3′

Table 3. (continued)

Primers used for confirmation of FLPo and Cre-mediated recombination in TP-32	
New-RACE 1.8	5′-GATGCGAATCGCATCCATGGATGCGCCAATC-3′
Bcl2SA-R	5′-GAGGGGATGCATATTATTTCTACTGCTTTAGTG-3′
CTPS2-US1	5′-TAGGGACTCCAGATTGACTTCAGAAGTCTG-3′
CTPS2-13int-R	5′-ACGAGGACCAGGAAGGAGTCCAAAGGTAGCCT-3′
Pol2-1Seq	5′-ATGAGGAAATTGCATCGCATTGTC-3′
Primer used for analysis of expression of trapped gene at the mRNA level	
ATP-Ex7-F	5′-GGAAGACTCTGAACAGTTCAGGGATGCTTC-3′
ATP-Ex9-R	5′-CCAAACCTGTTGTCTAAGTCTAGCATAAGG-3′
CTPS2-F1	5′-AAGCTTGGCAGAAGCTGTGCTTAGCTGATG-3′
CTPS2-R1	5′-GTATGGCAGCTTGTTCATCTGCTGCAGATG-3′
Bcl2-R	5′-AGGCCGCATGCTGGGGCCTTACAGTTCCACAAAG-3′
β-actin-F	5′-CTGGCACCACACCTTCTACAATGAGCTGCG-3′
β-actin-R	5′-CCAATAGTGATGACCTGGCCGTCAGGCAGC-3′
Primers used to determine the number and IDs of genome-integrated transposons	
New-RACE 0.9	5′-ACGCCATGGCCAATCGCATCATGGATTCCATC-3′
R-term	5′-GATCCTTTATCGCAATGGATGCGATGATCC-3′
RS-F4	5′-TTGACCCGACTGATGGTTCCCATTAGTCA-3′
F-int	5′-GGTTCCCATTAGTCACATAAAGCTGTAGTCAAG-3′
R-01	5′-CGGATTGCAATTCGCGCAAATTGGATGGCGATCAT-3′
R-02	5′-CGGATGGATTTTGATTGATCATCGCGCAATGCCCA-3′
R-03	5′-CGGATCGCCATATTCCGATCCCGATCATGCGATCA-3′
R-04	5′-CGGATTTGGCGAATGCGCGAATGCCATGCGATTTG-3′
R-05	5′-CGGATTTCATCGCCAATCGCGCATGCCCGATGATA-3′
R-06	5′-CGGATGGAATCGCGGAATTTGGAATCGGAATTGGA-3′
R-07	5′-CGGATCCGATTGCCCAATGGATTCGCGATTATGGC-3′
R-08	5′-CGGATGGATGCATTGGAATGGCATATCCCGCATTT-3′
R-09	5′-CGGATTGGCAATTTCCCATGATCATGCGCGATCGG-3′
R-10	5′-CGGATGATGGAATCCGCATTTGCCCGATTCCAATA-3′
R-11	5′-CGGATCGCAATCGCCATTTGCCATCCATTCCATAT-3′
R-12	5′-CGGATGCCCGATGATATGAATATGGGATTGCATGG-3′
R-13	5′-CGGATTGGGCATTCCATCCGATAATGGCGCAATAT-3′
R-14	5′-CGGATATGGCATGATCGGGGATGATGCAATGGAAA-3′
R-15	5′-CGGATCGCATTATGCATGCGATTCGCATATCGCCG-3′

Table 3. (continued)

Primers used for splinkerette genome PCR in the transposon-mixture experiments	
P1	5′-CGAATCGTAACCGTTCGTACGAGAA-3′
P2	5′-TCGTACGAGAATCGCTGTCCTCTCC-3′
SPL-1	5′-CATGCATCATATCCATCGCAATCGCATCC-3′
SPL-2m	5′-CCATTCAATTCACATCGCATATCCGATGCC-3′
F-term	5′-GGATCATCGCATCCATTGCGATAAAGG-3′
F-01	5′-ATGATCGCCATCCAATTTGCGCGAATTGCA-3′
F-02	5′-TGGGCATTGCGCGATGATCAATCAAAATCC-3′
F-03	5′-TGATCGCATGATCGGGATCGGAATATGGCG-3′
F-04	5′-CAAATCGCATGGCATTCGCGCATTCGCCAA-3′
F-05	5′-TATCATCGGGCATGCGCGATTGGCGATGAA-3′
F-06	5′-TCCAATTCCGATTCCAAATTCCGCGATTCC-3′
F-07	5′-GCCATAATCGCGAATCCATTGGGCAATCGG-3′
F-08	5′-AAATGCGGGATATGCCATTCCAATGCATCC-3′
F-09	5′-CCGATCGCGCATGATCATGGGAAATTGCCA-3′
F-10	5′-TATTGGAATCGGGCAAATGCGGATTCCATC-3′
F-11	5′-ATATGGAATGGATGGCAAATGGCGATTGCG-3′
F-12	5′-CCATGCAATCCCATATTCATATCATCGGGC-3′
F-13	5′-ATATTGCGCCATTATCGGATGGAATGCCCA-3′
F-14	5′-TTTCCATTGCATCATCCCCGATCATGCCAT-3′
F-15	5′-CGGCGATATGCGAATCGCATGCATAATGCG-3′
f-01	5′-ATATATATATTATGATCGCCATCCAATTTG-3′
f-02	5′-TATATATATTTGGGCATTGCGCGATGATCA-3′
f-03	5′-ATATATATATATTTGATCGCATGATCGGGA-3′
f-04	5′-ATATATATATATTCAAATCGCATGGCATTC-3′
f-05	5′-ATATATATATTTATCATCGGGCATGCGCGA-3′
f-06	5′-TATATATATATTTCCAATTCCGATTCCAAA-3′
f-07	5′-TATATATATATTGCCATAATCGCGAATCCA-3′
f-08	5′-TATATATATTAAATGCGGGATATGCCATTC-3′
f-09	5′-ATATATATATATTCCGATCGCGCATGATCA-3′
f-10	5′-ATATATATATTTATTGGAATCGGGCAAATG-3′
f-11	5′-CATATATATATTATATGGAATGGATGGCAA-3′
f-12	5′-TCCATATATATTCCATGCAATCCCATATTC-3′
f-13	5′-TATCCATATATTATATTGCGCCATTATCGG-3′
f-14	5′-ATATATCCATATTTTTCCATTGCATCATCC-3′
f-15	5′-ATATATCCATTCGGCGATATGCGAATCGCA-3′

CHAPTER III

RESULTS

3.1 Expression level and trapping efficiency of a gene in mouse undifferentiated ES cells

To understand what proportion of protein-coding genes are constitutively expressed in mouse undifferentiated ES cells, I first randomly selected ~10% of total mouse UniGene clusters [7811 out of 79 202 entries (January 2011); the Unigene database, http://www.ncbi.nlm.nih.gov/unigene/] and then excluded those of transcribed loci (5509), cDNAs with unknown function (224), predicted genes (131), hypothetical proteins (3) and others (107) for which the protein-coding capability has not been proven. For each of the remaining 1837 clusters representing classical protein-coding genes, I analyzed (*i*) if the gene is expressed in undifferentiated ES cells and (*ii*) if the gene has already been disrupted in the IGTC gene-trap effort.

In order to predict the expression level of each gene in undifferentiated ES cells, I examined the number of corresponding ESTs in the NCBI dbEST libraries #1882, #2512, #10023, #14556, #15703, #17907 and #21037 that had been constructed by using mRNAs derived from mouse undifferentiated ES cells. The total number of ESTs included in the above seven libraries is 143 423. For each of the selected UniGene clusters, I also inferred the mRNA expression in undifferentiated ES cells by examining the presence or absence of the corresponding sequence tags in a database for the ES cell-derived transcripts in HiCEP database.

Among randomly selected 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 (Fig. 14A). 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 (Fig. 14A). Eight hundred and fourteen (44.3%) contained the undifferentiated ES cell-derived ESTs and therefore were considered to be expressed in the cells (Fig. 14A).

In the case of such undoubtedly 'expressed' genes in undifferentiated ES cells, 80.8% (659 out of 814) had already been disrupted by random gene trapping in the IGTC effort (Fig. 14A). In contrast, only 21.8% (181 out of 830) of the potentially silent genes are found in the IGTC database at the time of the analysis (July 2011) (Fig. 14A). 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 14B 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).

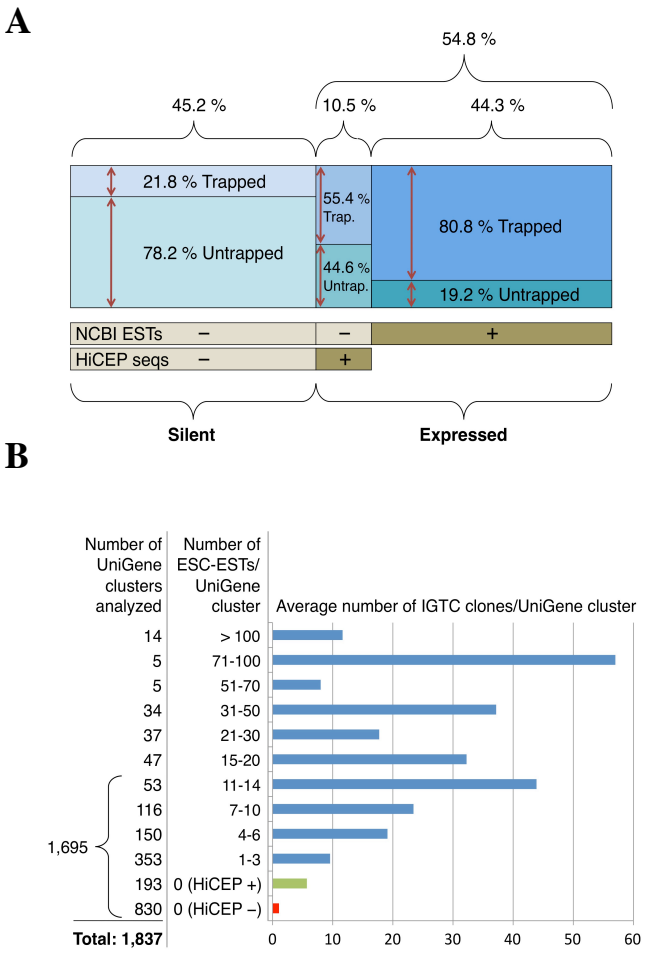

Figure 14. Transcriptionally active genes in mouse undifferentiated ES cells are trapped preferentially in the IGTC effort. (**A**) A crude analysis showing the correlation between the mRNA expression and the trapping efficiency in mouse ES cells of the randomly sampled 1837 protein-coding genes (UniGene 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].

Although 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 (Skarnes et al., 2011; the international knockout mouse consortium (IKMC) database, http://www.knockoutmouse.org/), 44.1% of the randomly selected silent genes (366 out of 830) had yet to be trapped or targeted by the time of the analysis.

3.2 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 (Rajewsky et al., 1996; Schnütgen et al., 2003; Xin et al., 2005; Schnütgen et al., 2005). To achieve this in poly(A) trapping, we assembled critical components of a gene-trap vector, as indicated in Figure 15. 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 (Ishida et al., 1999; Matsuda et al., 2004; Shigeoka et al., 2005) and 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 (Shigeoka et al., 2005). The FLEx methodology (Schnütgen et al., 2003) 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 (Fig. 15). The second recombination would be induced in mice by expressing the Cre recombinase in a spatially or temporally restricted manner (Fig. 15).

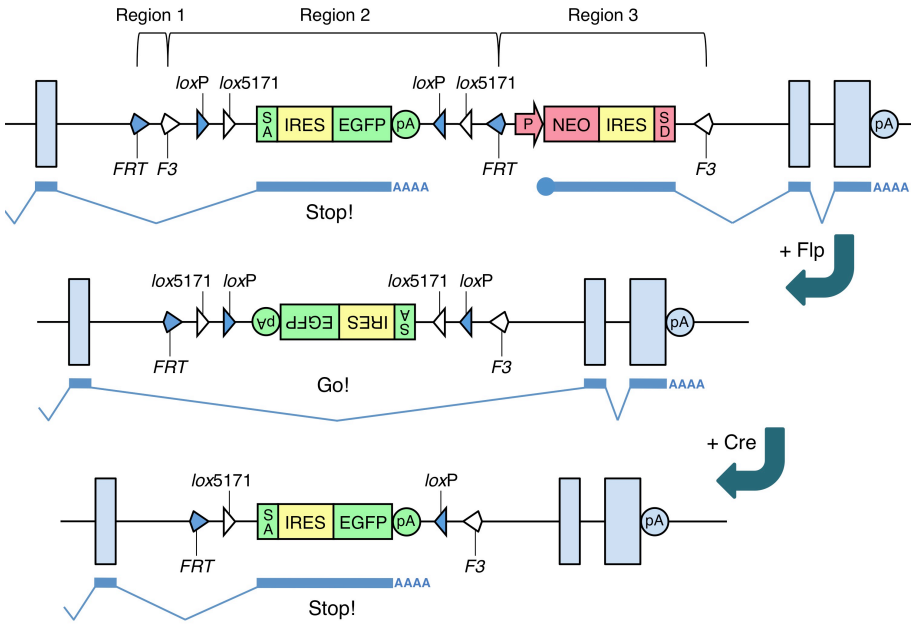

Figure 15. A strategy of conditional poly(A) trapping based on the NMD-suppressing UPATrap technology. Orientations of the triangular and diamond-shaped arrowheads represent those of the recombinase-target signals *lox*P, *lox*5171, *FRT*, and *F3*. Light-blue rectangles are exons of an endogenous gene. 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, four tandemly connected 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.

3.3 The *Tol2*-transposon version of the conditional UPATrap vector hardly suffers from deletions or rearrangements

We created a conditional variant of the UPATrap vector employing the FLEx methodology as shown on Figure 15 on the basis of the standard Moloney Murine Leukemia Virus (MMLV) (Soriano et al., 1991) (Fig. 8A) and performed gene-trap experiments in mouse ES cells. When I examined the integrity of the genome-integrated proviruses by PCR, I 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 are potentially contain some forms of intra-vector deletions or rearrangements (Fig. 16A). I then tried to confirm the presence of deleted or rearranged regions and found that both upstream and downstream portions of the NEO cassette tend to suffer from a high frequency of deletions (Fig. 16B). Although data are not shown, we found that the standard (*i.e.* non-conditional) retroviral UPATrap vector also generates intra-vector deletions or rearrangements with high frequency in the target cells. Such alterations inside the vectors severely hamper the conditional poly(A)-trapping strategy because even a tiny single deletion covering one of the eight recombinase-target signals distributed throughout the vector would make it impossible to induce regulated inversion or deletion for conditional gene disruption.

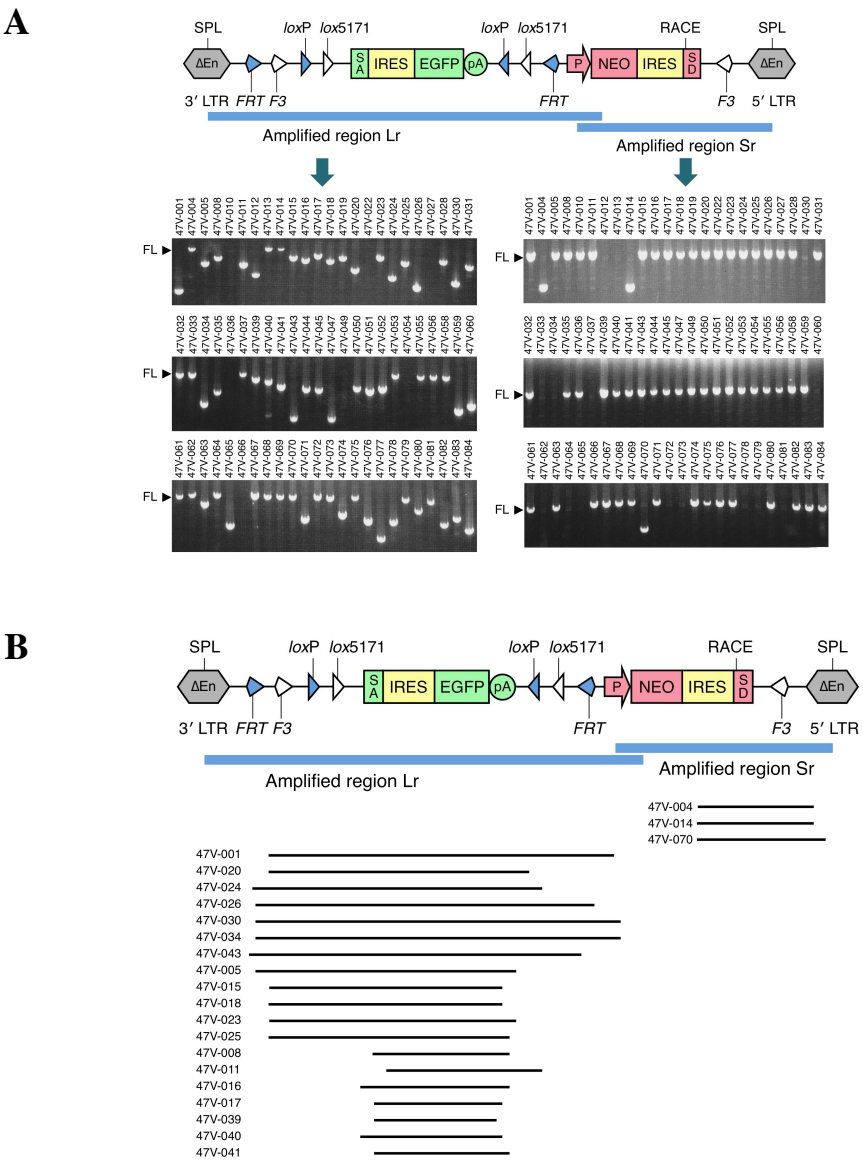

Figure 16. Intra-vector deletion in the genome-integrated conditional UPATrap-Moloney retrovirus vectors. (**A**) Frequent generation of intra-vector deletions/rearrangements in the target cells. Seventy two independent gene-trapped clones were randomly chosen from the ES cells infected with the conditional UPATrap retrovirus, and the integrity of the introduced vectors was analyzed by genomic PCR for the regions Lr and Sr. (**B**) Deleted regions in the genome-integrated vectors. After PCR amplification of the Lr and Sr portions, a variety of DNA fragments that were shorter than the full-length molecules were detected from the genome of the gene-trapped ES-cell clones that had been generated using the conditional UPATrap-Moloney retrovirus vector (panel A). These fragments were subjected to direct sequencing to determine the deleted regions, which are indicated as horizontal bars. Although data are not shown, some of the ES-cell clones contained deletions spanning the LTR portions or complex rearrangements inside the genome-integrated retrovirus vectors. In panel A, the majority of such clones did not give rise to any amplified DNA bands.

We suspected that the frequent deletions or 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 from the retrovirus vector into the *Tol2* transposon (Fig. 8B). Since *Tol2* is a cut and paste-type DNA transposon devoid of the 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, I detected potential deletions or rearrangements associated with the genome-integrated *Tol2* vectors only in 2.3% of the ES-cell clones that contained single-vector integration (6 out of 263 clones examined) (Fig. 17).

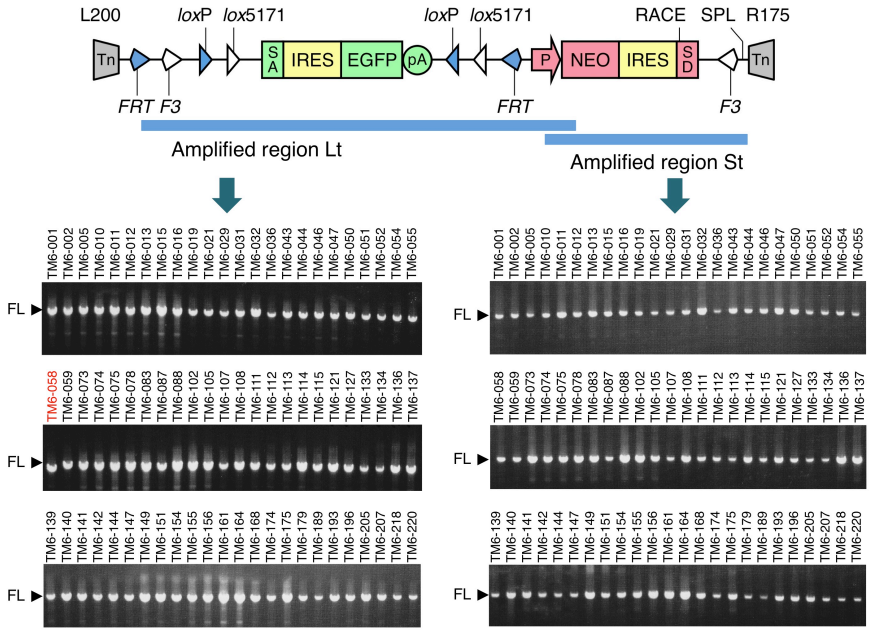

Figure 17. Structure and high integrity of the conditional *Tol2*-transposon vector in the target cells. Seventy two independent gene-trapped clones with single-vector integration (see Fig. 13 for details) were randomly chosen from the ES cells introduced with the conditional UPATrap-*Tol2* transposon, 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.

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

In addition to the frequency of the generation of intra-vector deletions or 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. 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) (Fig.18).

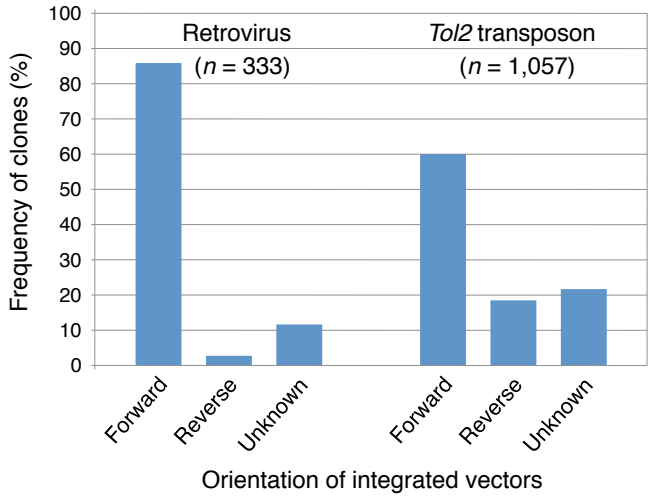

Figure 18. Orientation of vector integration relative to that of transcription of trapped genes. The orientation of an integrated vector is regarded as forward when the transcriptional orientation of the EGFP and NEO cassettes of the gene-trap vector and that of the trapped known gene are the same. Likewise, 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.

As for the expressions status in the undifferentiated ES cells, 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 criteria shown in Figure 14A (Fig. 19). This is consistent with the previous report showing that MMLV possesses a strong preference to be integrated into transcriptionally active genome regions (Scherdin et al., 1990). In contrast, the frequency of trapping potentially silent genes using the *Tol2*-transposon counterpart was 25.9%, ~2.4 times higher than that of the retrovirus version (Fig. 19).

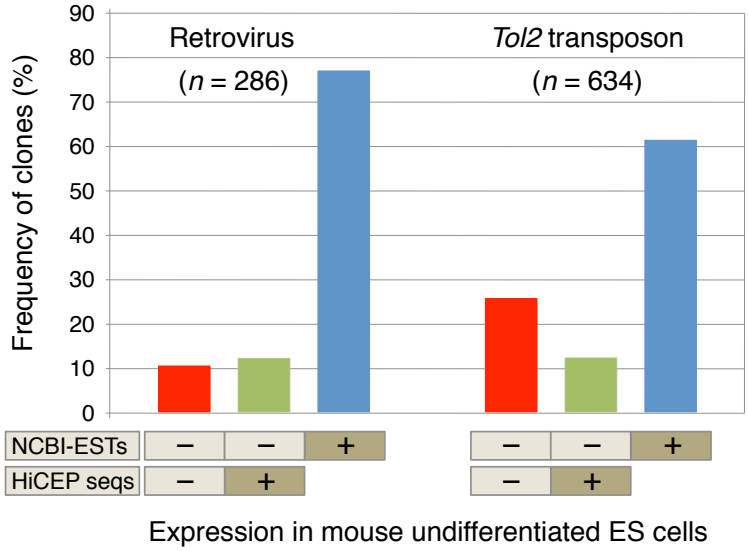

Figure 19. Transcriptional status of genes identified by using the conditional UPATrap vectors. Transcriptional status of known genes trapped in a forward orientation was classified into three groups: (*i*) NCBI-ESTs –/HiCEP seqs –; (*ii*) NCBI-ESTs –/HiCEP seqs +; and (*iii*) NCBI-ESTs +. See Figure 14A for details about this classification.

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 (Fig. 20, 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 (Fig. 20, navy bars). This indicates that the spectrum of genes identified by gene trapping with *Tol2* vectors is quite different from that of retrovirus vectors.

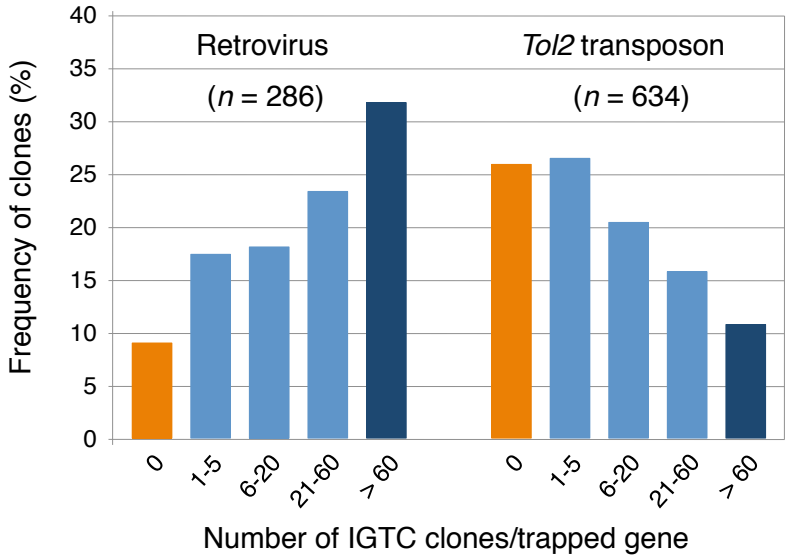

Figure 20. Number of the mutant ES-cell clones already registered in the IGTC database for each known gene trapped in a forward orientation by using the conditional UPATrap vectors. Orange bars represent ES-cell clones that have never been trapped in IGTC database; blue and navy bars represent ES cell clones that had been trapped multiple times in IGTC database

The results shown in Figure 21 indicate the distribution of the vector insertion sites inside trapped genes. I found that the conditional retrovirus vector tends to be more preferentially inserted in the promoter regions (which is located 5′ to the first exon) or in the first introns of trapped genes in comparison with the *Tol2* counterpart (Fig. 21**,** beige bars), as has already been shown for a number of MMLV vectors (Wu et al., 2003). The *Tol2*-transposon vector, on the other hand, did not show strong preference for particular insertion sites inside a gene (Fig. 21). I did not observe the strong integration-site bias toward the last intron of trapped genes (Ishida et al., 1999; Matsuda et al., 2004; Shigeoka et al., 2005) for either the retrovirus or *Tol2*-transposon vector (Fig. 21**,** 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 (Shigeoka et al., 2005).

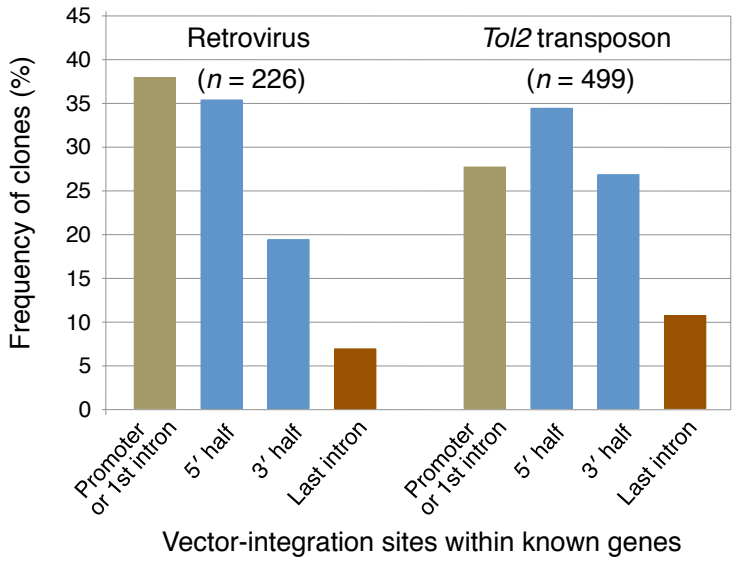

Figure 21. Distribution of the vector-integration sites around known genes trapped in forward orientation by using the conditional UPATrap vectors. The vector-integration sites were predicted from the nucleotide sequences of the 3′ RACE fragments. Events of vector integration into the introns of genes consisting of 1−4 exons and the right-middle introns of genes with even numbers of exons were excluded from the analysis.

The 3′ RACE analysis alone cannot tell the exact vector-integration sites, especially when the vectors are integrated into the promoter regions or introns 1 of the endogenous genes. Since the standard first exons of cellular genes do not possess splice acceptors, both of the NMD-suppressing UPATrap vectors integrated into the first intron (Fig. 22A, top) and the promoter region (Fig. 22A, bottom) produce exactly the same 3′ RACE products consisting of the nucleotide sequences of exons 2, 3, 4 and so on. To distinguish the vector integration into the promoter regions from that into the first introns, I collected ES-cell clones with the 3′ RACE products consisting of the nucleotide sequences of exons 2, 3, 4 and so on, and performed the splinkerette genome PCR (followed by direct sequencing of the PCR products) to determine the precise vector-integration points. The results indicate that both of the conditional UPATrap vectors are more preferentially integrated into the first introns than into the promoter regions, and the preference appears to be slightly stronger for the retrovirus vector than for the *Tol2* counterpart (Fig. 22B).

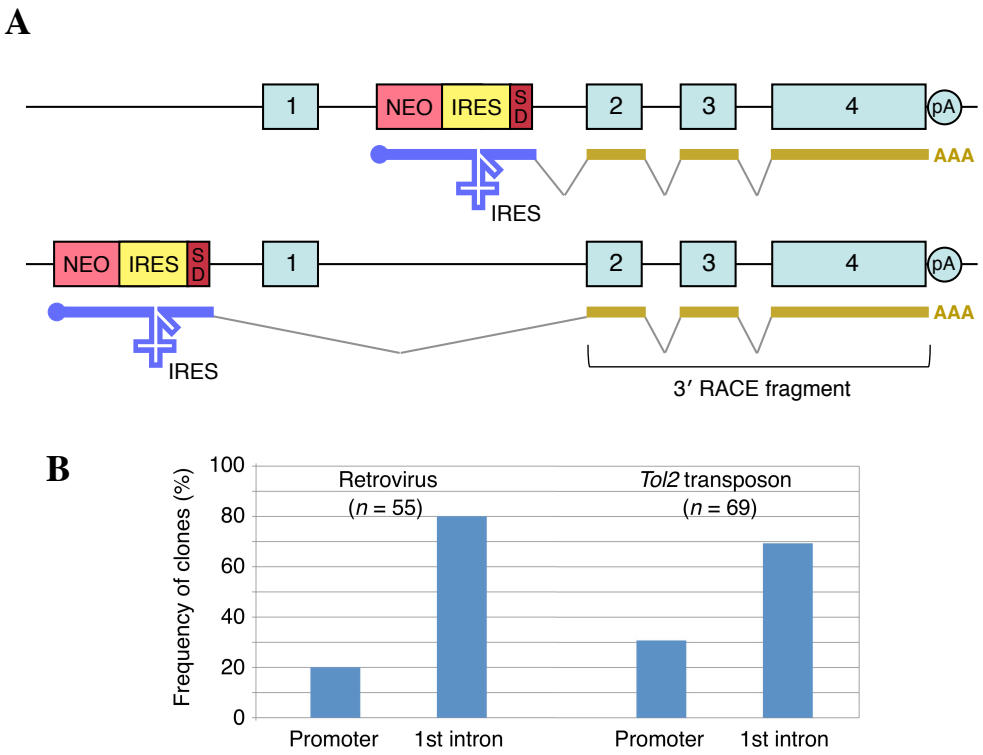

Figure 22. The 3′ RACE analysis alone cannot determine whether the vector is integrated into the promoter region or the first intron of a cellular gene. **(A)** Schematic representation of 3′ RACE result of UPATrap vectors integrated into the first intron (top) and the promoter region (bottom). Only the poly(A)-trapping NEO cassettes of the UPATrap vector are shown here. (**B**) The results of splinkerette genome PCR analysis show precisely the vector integration site. Fifty-five and sixty-nine ES-cell clones selected and analyzed for the retrovirus and *Tol2*-transposon vectors, respectively.

3.5 Conditional disruption of the trapped genes

In order to confirm the conditionality of gene disruption with our vector, *in vitro* experiments were carried out. 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 (You et al., 1998) 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 the gene *Atp6ap2* in a forward orientation (*i.e*. the orientation of transcription was the same between endogenous *Atp6ap2* and the vector components EGFP and NEO). In contrast, the reverse strand of *Ctps2* was trapped in TP32.

Atp6ap2 is constitutively expressed in undifferentiated ES cells (lane wt; wild type in Fig. 23A), but the expression of the 3′ portion of *Atp6ap2* that is located downstream of the vector-integration point was disrupted in the parental clone 1TP-84 because the transcription of *Atp6ap2* is terminated at the efficient poly(A) addition signals in the EGFP cassette of the vector (lane P; parental clone in Fig. 23A). The first step of recombination was mediated by the transient expression of FLPo, a codon-optimized version of the FLPe (a thermostable variant of the FLP recombinase) gene (Buchholz et al., 1998; Raymond and Soriano, 2007), in 1TP-84. As shown in Figure 10, introduction of FLPo caused a deletion of the NEO cassette and an inversion of the gene-terminator cassette. Efficiency of the FLPo-mediated recombination for this clone was as assessed by a simple transfection and limiting-dilution method, and the recombination events is 98.9% (89 out of 90) in the ES-cell subclones examined (Table 4)**.** After the FLPo-mediated first recombination, the expression of *Atp6ap2* was fully recovered as expected (Fig. 23A).

The next step is transient expression of Cre recombinase in the 1TP-84 subclones in which the FLPo-mediated deletion and inversion had already been successfully completed. Three of such 1TP-84 subclones were selected and transiently transfected with an expression

vector for the Cre recombinase, and the Cre-generated daughter subclones were analyzed for inversion and deletion of the key components inside the vector. In the overwhelming majority (87.5%; 63 out of 72 subclones examined) of the Cre-generated daughter subclones examine (Table 5), the gene-terminator cassette was successfully re-inverted to create a disruptive allele for *Atp6ap2*, and no leakiness of the expression of the disrupted 3′-portion of *Atp6ap2* was detected (Fig. 23A).

For the second X-chromosomal gene *Ctps2*, I was also able to induce deletion and inversion of the vector components efficiently and obtain a tightly regulated pattern of conditional gene disruption (Tables 6 and 7), although the orientation of the vector integration inside *Ctps2* was opposite to that of *Atp6ap2* (Fig. 23B).

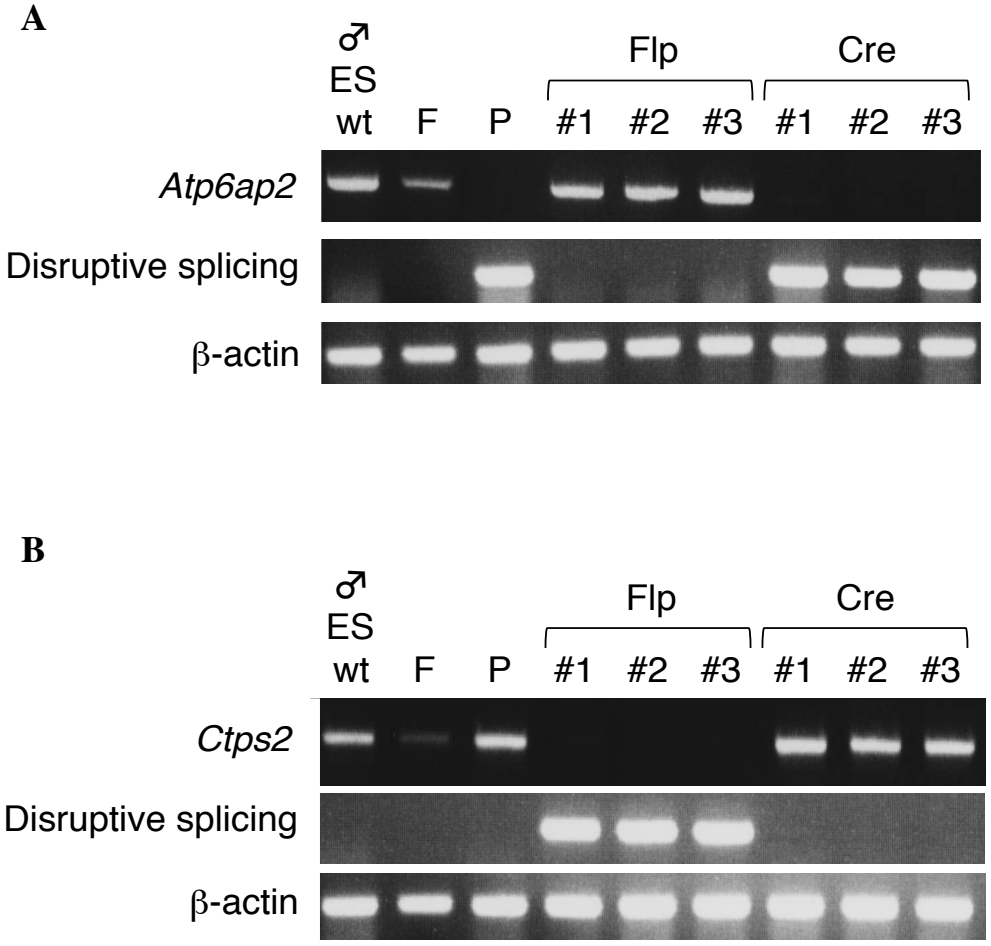

Figure 23. Conditional disruption of the trapped genes. (**A**) Conditional disruption of *Atp6ap2*. (**B**) Conditional disruption of *Ctps2*. The sense and antisense strands of the X-chromosomal genes *Atp6ap2* and *Ctps2*, respectively, were trapped in a male-derived ES-cell line by using the UPATrap-*Tol2* 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. See Tables 4-7 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.

Table 4. Assessment of the efficiency of FLPo-mediated coupled homologous recombination inside the gene-trap vector integrated into *Atp6ap2* (1TP-84)

Clone number	New Race 1.8/ 84-DS1 (Parental)[a]	X12189640-F/ Pol2-1Seq (5' portion)[b]	Bcl2SA-R/ 84-DS1 (3' portion)[c]	G418 sensitivity test
1TP-84	+	–	–	R
1TP-84/003F	–	+	+	S
1TP-84/007F	–	+	+	S
1TP-84/014F	–	+	+	S
1TP-84/017F	–	+	+	S
1TP-84/026F	–	+	+	S
1TP-84/028F	–	+	+	S
1TP-84/037F	–	+	+	S
1TP-84/038F	–	+	+	S
1TP-84/039F	–	+	+	S
1TP-84/040F	–	+	+	S
1TP-84/041F	–	+	+	S
1TP-84/042F	–	+	+	S
1TP-84/043F	–	+	+	S
1TP-84/044F	–	+	+	S
1TP-84/045F	–	+	+	S
1TP-84/046F	–	+	+	S
1TP-84/047F	–	+	+	S
1TP-84/048F	–	+	+	S
1TP-84/049F	–	+	+	S
1TP-84/050F	–	+	+	S
1TP-84/051F	–	+	+	S
1TP-84/052F	–	+	+	S
1TP-84/053F	–	+	+	S
1TP-84/054F	–	+	+	S
1TP-84/055F	–	+	+	S
1TP-84/056F	–	+	+	S
1TP-84/057F	–	+	+	S
1TP-84/058F	–	+	+	S
1TP-84/059F	–	+	+	S
1TP-84/060F	–	+	+	S
1TP-84/061F	–	+	+	S
1TP-84/062F	–	+	+	S
1TP-84/063F	–	+	+	S
1TP-84/064F	–	+	+	S
1TP-84/065F	–	+	+	S
1TP-84/066F	–	+	+	S
1TP-84/067F	–	+	+	S
1TP-84/068F	–	+	+	S
1TP-84/069F	–	+	+	S
1TP-84/070F	–	+	+	S
1TP-84/071F	–	+	+	S
1TP-84/072F	–	+	+	S
1TP-84/073F	–	+	+	S
1TP-84/074F	–	+	+	S
(continued)				
1TP-84/075F	–	+	+	S
1TP-84/076F	–	+	+	S
1TP-84/077F	–	+	+	S
1TP-84/078F	–	+	+	S
1TP-84/079F	–	+	+	S
1TP-84/080F	–	+	+	S
1TP-84/081F	–	+	+	S
1TP-84/082F	–	+	+	S
1TP-84/083F	–	+	+	S
1TP-84/084F	–	+	+	S
1TP-84/085F	–	+	+	S
1TP-84/086F	–	+	+	S
1TP-84/087F	–	+	+	S
1TP-84/088F	–	+	+	S
1TP-84/089F	–	–	+	S
1TP-84/090F	–	+	+	S
1TP-84/091F	–	+	+	S
1TP-84/092F	–	+	+	S
1TP-84/093F	–	+	+	S
1TP-84/094F	–	+	+	S
1TP-84/095F	–	+	+	S
1TP-84/096F	–	+	+	S
1TP-84/097F	–	+	+	S
1TP-84/098F	–	+	+	S
1TP-84/099F	–	+	+	S
1TP-84/100F	–	+	+	S
1TP-84/101F	–	+	+	S
1TP-84/102F	–	+	+	S
1TP-84/103F	–	+	+	S
1TP-84/104F	–	+	+	S
1TP-84/105F	–	+	+	S
1TP-84/106F	–	+	+	S
1TP-84/107F	–	+	+	S
1TP-84/108F	–	+	+	S
1TP-84/109F	–	+	+	S
1TP-84/110F	–	+	+	S
1TP-84/111F	–	+	+	S
1TP-84/112F	–	+	+	S
1TP-84/113F	–	+	+	S
1TP-84/114F	–	+	+	S
1TP-84/115F	–	+	+	S
1TP-84/116F	–	+	+	S
1TP-84/117F	–	+	+	S
1TP-84/118F	–	+	+	S
1TP-84/119F	–	+	+	S
1TP-84/120F	–	+	+	S

[a]A set of primers for the detection of the initial vector configuration in the parental ES-cell clone.

[b]A set of primers for the detection of the 5' portion of the FLPo-mediated inversion/deletion events.

[c]A set of primers for the detection of the 3' portion of the FLPo-mediated inversion/deletion events.

(+), presence of an amplified DNA fragment with a correct size; (–), absence of an amplified DNA fragment; (R), resistant; (S), sensitive.

See Figure 10 for details.

Table 5. Assessment of the efficiency of Cre-mediated coupled homologous recombination inside the gene-trap vector integrated into *Atp6ap2* (1TP-84)

	1TP-84/003F					1TP-84/ 014F					1TP-84/028F			
Clone number	New RACE 1.8/ 84-DS1 (Parental)[a]	Bcl2SA-R/ 84-DS1 (FLPo)[b]	X12189640/ Bcl2SA-R (5' portion)[c]	Pol2-1Seq/ 84-DS1 (3' portion)[d]	Clone number	New RACE 1.8/ 84-DS1 (Parental)[a]	Bcl2SA-R/ 84-DS1 (FLPo)[b]	X12189640/ Bcl2SA-R (5' portion)[c]	Pol2-1Seq/ 84-DS1 (3' portion)[d]	Clone number	New RACE 1.8/ 84-DS1 (Parental)[a]	Bcl2SA-R/ 84-DS1 (FLPo)[b]	X12189640/ Bcl2SA-R (5' portion)[c]	Pol2-1Seq/ 84-DS1 (3' portion)[d]
1TP-84	+	–	+	+	1TP-84	+	–	+	+	1TP-84	+	–	+	+
1TP-84/003F	–	+	–	–	1TP-84/014F	–	+	–	–	1TP-84/028F	–	+	–	–
1TP-84/003F/01C	–	–	+	+	1TP-84/014F/01C	–	–	+	+	1TP-84/028F/01C	–	–	+	+
1TP-84/003F/02C	–	–	+	+	1TP-84/014F/02C	–	+	–	–	1TP-84/028F/02C	–	–	+	+
1TP-84/003F/03C	–	–	+	+	1TP-84/014F/03C	–	+	–	–	1TP-84/028F/03C	–	?	–	–
1TP-84/003F/04C	–	–	+	+	1TP-84/014F/04C	–	–	+	+	1TP-84/028F/04C	–	–	+	+
1TP-84/003F/05C	–	–	+	+	1TP-84/014F/05C	–	–	+	+	1TP-84/028F/05C	–	–	+	+
1TP-84/003F/06C	–	–	+	+	1TP-84/014F/06C	–	–	+	+	1TP-84/028F/06C	–	–	+	+
1TP-84/003F/07C	–	–	+	+	1TP-84/014F/07C	–	–	+	+	1TP-84/028F/07C	–	–	+	+
1TP-84/003F/08C	–	–	+	+	1TP-84/014F/08C	–	?	–	–	1TP-84/028F/08C	–	–	+	+
1TP-84/003F/09C	–	+	–	–	1TP-84/014F/09C	–	–	+	+	1TP-84/028F/09C	–	–	+	+
1TP-84/003F/10C	–	–	+	+	1TP-84/014F/10C	–	–	+	+	1TP-84/028F/10C	–	–	+	+
1TP-84/003F/11C	–	–	+	+	1TP-84/014F/11C	–	–	+	?	1TP-84/028F/11C	–	–	+	+
1TP-84/003F/12C	–	–	+	+	1TP-84/014F/12C	–	–	+	+	1TP-84/028F/12C	–	–	+	+
1TP-84/003F/13C	–	–	+	+	1TP-84/014F/13C	–	–	+	+	1TP-84/028F/13C	–	–	+	+
1TP-84/003F/14C	–	–	+	+	1TP-84/014F/14C	–	–	+	+	1TP-84/028F/14C	–	–	+	+
1TP-84/003F/15C	–	–	+	+	1TP-84/014F/15C	–	+	–	–	1TP-84/028F/15C	–	+	+	+
1TP-84/003F/16C	–	–	+	+	1TP-84/014F/16C	–	–	?	+	1TP-84/028F/16C	–	–	+	+
1TP-84/003F/17C	–	–	+	+	1TP-84/014F/17C	–	–	+	+	1TP-84/028F/17C	–	–	+	+
1TP-84/003F/18C	–	–	+	+	1TP-84/014F/18C	–	–	+	+	1TP-84/028F/18C	–	–	+	+
1TP-84/003F/19C	–	–	+	+	1TP-84/014F/19C	–	–	+	+	1TP-84/028F/19C	–	–	+	+
1TP-84/003F/20C	–	–	+	+	1TP-84/014F/20C	–	–	+	+	1TP-84/028F/20C	–	–	+	+
1TP-84/003F/21C	–	–	+	+	1TP-84/014F/21C	–	–	+	+	1TP-84/028F/21C	–	–	+	+
1TP-84/003F/22C	–	–	+	+	1TP-84/014F/22C	–	–	+	+	1TP-84/028F/22C	–	–	+	+
1TP-84/003F/23C	–	–	+	+	1TP-84/014F/23C	–	–	+	+	1TP-84/028F/23C	–	–	+	+
1TP-84/003F/24C	–	–	+	+	1TP-84/014F/24C	–	–	+	+	1TP-84/028F/24C	–	–	+	+

[a]A set of primers for the detection of the initial vector configuration in the parental ES-cell clone.

[b]A set of primers for the detection of the 3' portion of the FLPo-mediated inversion/deletion events.

[c]A set of primers for the detection of the 5' portion of the Cre-mediated inversion/deletion events.

[d]A set of primers for the detection of the 3' portion of the Cre-mediated inversion/deletion events.

(+), presence of an amplified DNA fragment with a correct size; (–), absence of an amplified DNA fragment; (?), presence of an amplified DNA fragment with an incorrect size.

See Figure 10 for details.

Table 6. Assessment of the efficiency of FLPo-mediated coupled homologous recombination inside the gene-trap vector integrated into *Ctps2* (TP-32)

Clone number	New Race 1.8/ CTPS2-US1 (Parental)[a]	CTPS2-US1/ Bcl2SA-R (5' portion)[b]	Pol2-1Seq/ CTPS2-13int-R (3' portion)[c]	G418 sensitivity test
TP-32	+	–	–	R
TP-32/002F	–	+	+	S
TP-32/003F	–	+	+	S
TP-32/005F	–	+	+	S
TP-32/006F	–	+	+	S
TP-32/012F	–	+	+	S
TP-32/013F	+	–	–	R
TP-32/014F	–	+	+	S
TP-32/015F	–	+	+	S
TP-32/016F	–	+	+	S
TP-32/017F	–	+	+	S
TP-32/018F	–	+	+	S
TP-32/022F	–	+	+	S
TP-32/024F	–	+	+	S
TP-32/025F	+	–	–	R
TP-32/027F	–	+	+	S
TP-32/028F	–	+	+	S
TP-32/030F	–	+	+	S
TP-32/031F	–	+	+	S
TP-32/032F	–	+	+	S
TP-32/033F	–	+	+	S
TP-32/035F	–	+	+	S
TP-32/037F	–	+	+	S
TP-32/038F	–	+	+	S
TP-32/040F	–	+	+	S
TP-32/041F	–	+	+	S
TP-32/042F	–	+	+	S
TP-32/044F	–	+	+	S
(continued)				
TP-32/045F	–	+	+	S
TP-32/046F	–	+	+	S
TP-32/047F	–	+	+	S
TP-32/050F	–	+	+	S
TP-32/054F	–	+	+	S
TP-32/055F	–	+	+	S
TP-32/057F	–	+	+	S
TP-32/058F	–	+	+	S
TP-32/059F	–	+	+	S
TP-32/061F	–	+	+	S
TP-32/062F	–	+	+	S
TP-32/063F	+	–	–	R
TP-32/064F	–	+	+	S
TP-32/066F	–	+	+	S
TP-32/067F	–	+	+	S
TP-32/068F	+	–	–	R
TP-32/070F	–	+	+	S
TP-32/071F	–	+	+	S
TP-32/072F	–	+	+	S
TP-32/073F	–	+	+	S
TP-32/077F	–	+	+	S
TP-32/078F	–	+	+	S
TP-32/079F	–	+	+	S
TP-32/080F	–	+	+	S
TP-32/081F	–	+	+	S
TP-32/082F	–	+	+	S
TP-32/083F	–	+	+	S
TP-32/084F	–	+	+	S

[a]A set of primers for the detection of the initial vector configuration in the parental ES-cell clone.

[b]A set of primers for the detection of the 5' portion of the FLPo-mediated inversion/deletion events.

[c]A set of primers for the detection of the 3' portion of the FLPo-mediated inversion/deletion events.

(+), presence of an amplified DNA fragment with a correct size; (–), absence of an amplified DNA fragment; (R), resistant; (S), sensitive.

See Figure 11 for details.

Table 7. Assessment of the efficiency of Cre-mediated coupled homologous recombination inside the gene-trap vector integrated into *Ctps2* (TP-32)

	TP-32/003F					TP-32/014F					TP-32/028F			
Clone number	CTPS2-US1/ New RACE 1.8 (Parental)[a]	CTPS2-US1/ Bcl2SA-R (FLPo)[b]	CTPS2-US1/ Pol2-1Seq (5' portion)[c]	Bcl2SA-R/ CTPS2-13int-R (3' portion)[d]	Clone number	CTPS2-US1/ New RACE 1.8 (Parental)[a]	CTPS2-US1/ Bcl2SA-R (FLPo)[b]	CTPS2-US1/ Pol2-1Seq (5' portion)[c]	Bcl2SA-R/ CTPS2-13int-R (3' portion)[d]	Clone number	CTPS2-US1/ New RACE 1.8 (Parental)[a]	CTPS2-US1/ Bcl2SA-R (FLPo)[b]	CTPS2-US1/ Pol2-1Seq (5' portion)[c]	Bcl2SA-R/ CTPS2-13int-R (3' portion)[d]
TP-32	+	–	+	+	TP-32	+	–	+	+	TP-32	+	–	+	+
TP-32/003F	–	+	–	–	TP-32/014F	–	+	–	–	TP-32/028F	–	+	–	–
TP-32/003F/01C	–	–	+	+	TP-32/014F/01C	–	–	+	+	TP-32/028F/01C	–	+	–	–
TP-32/003F/02C	–	–	+	+	TP-32/014F/02C	–	–	+	+	TP-32/028F/02C	–	–	+	+
TP-32/003F/03C	–	–	+	+	TP-32/014F/03C	–	–	+	+	TP-32/028F/03C	–	+	–	–
TP-32/003F/04C	–	–	+	+	TP-32/014F/04C	–	–	?	+	TP-32/028F/04C	–	–	+	+
TP-32/003F/05C	–	–	+	+	TP-32/014F/05C	–	–	+	+	TP-32/028F/05C	–	–	+	+
TP-32/003F/06C	–	–	+	+	TP-32/014F/06C	–	–	+	+	TP-32/028F/06C	–	–	?	+
TP-32/003F/07C	–	–	+	+	TP-32/014F/07C	–	–	+	+	TP-32/028F/07C	–	–	+	+
TP-32/003F/08C	–	–	+	+	TP-32/014F/08C	–	–	+	+	TP-32/028F/08C	–	–	+	+
TP-32/003F/09C	–	–	+	+	TP-32/014F/09C	–	–	+	+	TP-32/028F/09C	–	–	+	+
TP-32/003F/10C	–	–	+	+	TP-32/014F/10C	–	–	+	+	TP-32/028F/10C	–	–	?	+
TP-32/003F/11C	–	–	+	+	TP-32/014F/11C	–	–	+	+	TP-32/028F/11C	–	–	+	+
TP-32/003F/12C	–	–	+	+	TP-32/014F/12C	–	–	+	+	TP-32/028F/12C	–	–	+	+
TP-32/003F/13C	–	–	+	+	TP-32/014F/13C	–	–	+	+	TP-32/028F/13C	–	–	+	+
TP-32/003F/14C	–	–	+	+	TP-32/014F/14C	–	–	+	+	TP-32/028F/14C	–	–	+	+
TP-32/003F/15C	–	–	+	+	TP-32/014F/15C	–	–	+	+	TP-32/028F/15C	–	+	+	?
TP-32/003F/16C	–	–	+	+	TP-32/014F/16C	–	–	+	+	TP-32/028F/16C	–	–	+	+
TP-32/003F/17C	–	–	+	+	TP-32/014F/17C	–	–	+	+	TP-32/028F/17C	–	–	+	+
TP-32/003F/18C	–	–	+	+	TP-32/014F/18C	–	+	–	–	TP-32/028F/18C	–	–	+	+
TP-32/003F/19C	–	–	+	+	TP-32/014F/19C	–	–	+	+	TP-32/028F/19C	–	–	+	+
TP-32/003F/20C	–	–	+	+	TP-32/014F/20C	–	–	+	+	TP-32/028F/20C	–	–	+	+
TP-32/003F/21C	–	+	+	+	TP-32/014F/21C	–	–	+	+	TP-32/028F/21C	–	–	+	+
TP-32/003F/22C	–	–	+	+	TP-32/014F/22C	–	–	+	+	TP-32/028F/22C	–	–	+	+
TP-32/003F/23C	–	–	+	+	TP-32/014F/23C	–	–	+	+	TP-32/028F/23C	–	+	–	–
TP-32/003F/24C	–	–	+	+	TP-32/014F/24C	–	–	+	+	TP-32/028F/24C	–	–	+	+

[a]A set of primers for the detection of the initial vector configuration in the parental ES-cell clone.

[b]A set of primers for the detection of the 5' portion of the FLPo-mediated inversion/deletion events.

[c]A set of primers for the detection of the 5' portion of the Cre-mediated inversion/deletion events.

[d]A set of primers for the detection of the 3' portion of the Cre-mediated inversion/deletion events.

(+), presence of an amplified DNA fragment with a correct size; (–), absence of an amplified DNA fragment; (?), presence of an amplified DNA fragment with an incorrect size.

See Figure 11 for details.

3.6 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 the retroviral counterpart. The only disadvantage of *Tol2* (or DNA transposons in general), however, is the difficulty in stringently controlling the number of the integrated vectors in a target cell. In the case of retrovirus, it is easy to find an appropriate infection (*i.e*. vector-transduction) condition for a single-vector integration per cell by adjusting the ratio of virus particles number to that of target cells. 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 issue, we developed a strategy using a mixture of the differentially tagged *Tol2* transposons. Each of the 15 different tag consists of two parts: (*i*) a diagnostic CC dinucleotide-in-poly(AT) part; and (*ii*) a vector-identification (ID) part (Fig. 24). Each tag is flanked by the common synthetic sequences SPL (~90 bp) and Term (~30 bp). For the first diagnostic part, the position of the CC-dinucleotide in the poly(AT) background is determined according to the identity of each differential tag. In the second part, we designed 15 different vector-ID sequences (30mers) that are able to serve as the base sites for the annealing of specific primers in both forward and reverse orientations (SEQ-01 ~ 15 in Fig. 24). We inserted these differential-tag sequences near the 3′-ends of the conditional UPATrap-*Tol2* transposon vectors and created an equimolar mixture of the 15 differentially tagged transposons.

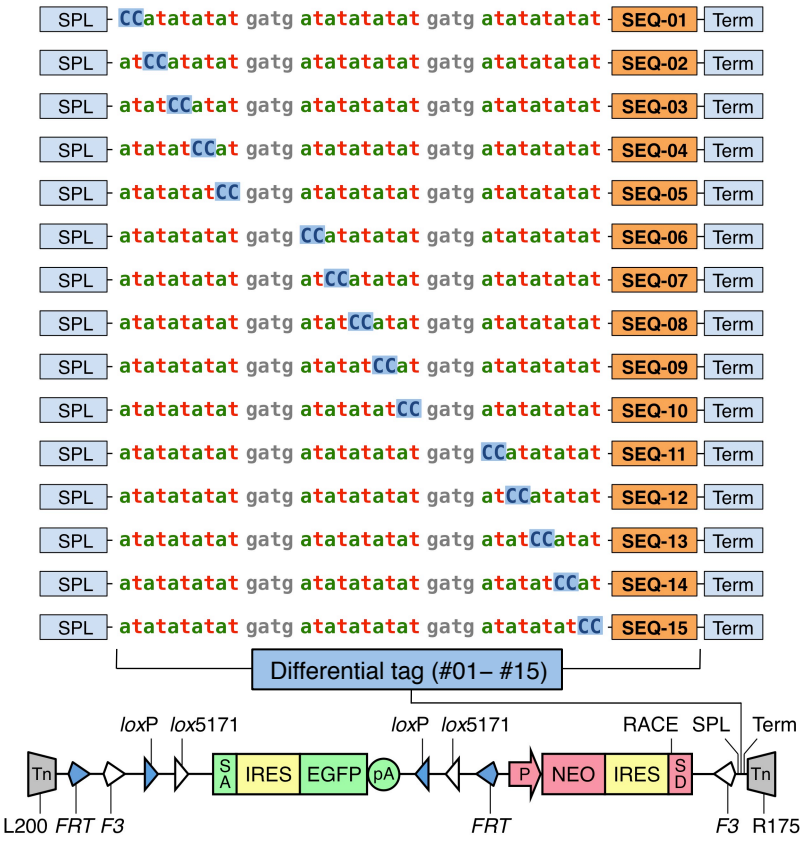

Figure 24. A gene-trap strategy based on the mixture of differentially tagged UPATrap-*Tol2* transposons. Structure of the fifteen differential tags located between the common SPL and Term sequences near the 3′ ends of the gene-trap vectors. SEQ-01~15 are the synthetic ID nucleotides (30mers) with similar 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.

After I obtain gene-trapped ES-cell clones with the transposon mixture, I first extract genomic DNA from the cells and amplify the differential tags by PCR using common sequences flanking the differential tags. Then, I perform direct sequencing of the amplified tags to learn the number and IDs of the genome-integrated vectors (Fig. 13A). The results in Figure 25 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, and the distinction of the CC signals from the background noise becomes more difficult. However, by performing PCR-based analyses as shown in Figure 13B, I was able to determine the number and IDs of the vectors reproducibly even for the ES-cell clones containing more than three transposons per cell (Fig. 26A). According to the both sequencing-based and PCR-based analyses, 56.6% (267 out of 472) of the established ES-cell clones in our current protocol show a single-vector integration, and 27.5% (130 out of 472) had two integrated vectors, indicating that the number and ID of the integrated vectors can be diagnosed by a single round of PCR and sequencing experiments for the majority (84.1%) of the established clones (Fig. 26A). Among 15 differentially tagged transposons, I observed weak bias in the usage of the particular vector(s) (Figure 26B), but it did not hamper our analyses on the genome-integrated vectors.

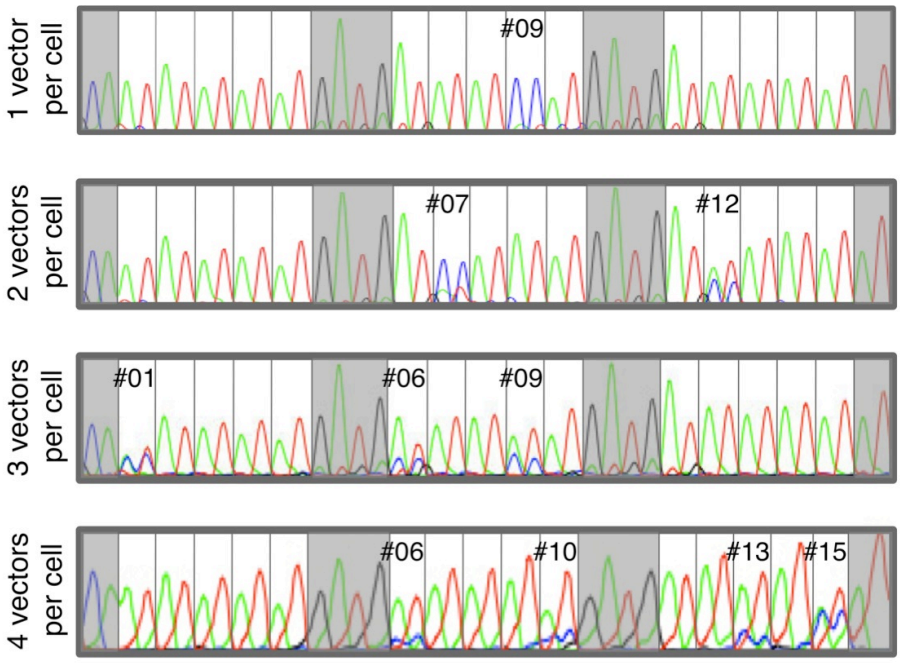

Figure 25. Representative results of the analyses of the number and IDs of the integrated vectors based on the PCR amplification and direct sequencing of the differential-tag portions of the gene-trap vectors integrated into the genome of ES-cell clones. See Figure 13 for details. The chromatogram shows the results for 1, 2, 3 and 4 integrated vectors per ES-cell clone. The CC-dinucleotide signals (blue peaks) became weaker as the number of integrated vector per cell increases.

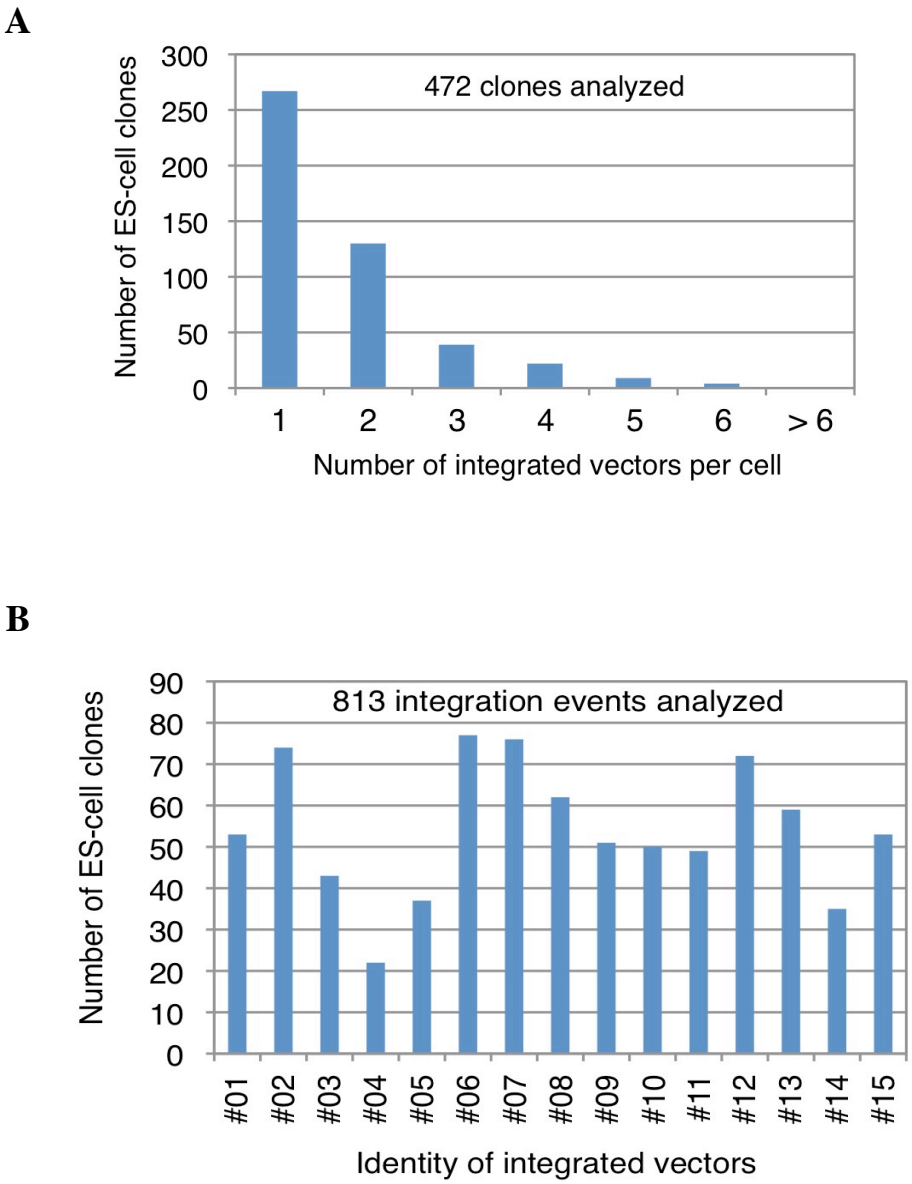

Figure 26. The number and ID of the differential tags of the integrated vectors within ES-cell clones generated by mixture of differentially tagged *Tol2*-transposons vector. (**A**) Distribution of the number of integrated vectors in an ES-cell clone. The number of integrated vectors was confirmed by the tag-specific PCR when three or more vectors were suspected to be integrated into the genome of an ES-cell clone. See Figure 13 for details. (**B**) Usage of the fifteen differentially tagged gene-trap vectors in the transposon-mixture experiments.

Once the number and IDs of the transposons within an ES-cell clone are determined, I analyze the nucleotide sequences of the vector-integration sites by performing either tag-specific sequencing of the mixed splinkerette PCR products (Devon et al., 1995; Horn et al., 2007) or standard sequencing of the DNA fragments that are independently generated through the tag-directed splinkerette PCR, depending upon the number of vectors involved (Fig. 12). With this strategy, I was 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 (Fig. 27; Table 8).

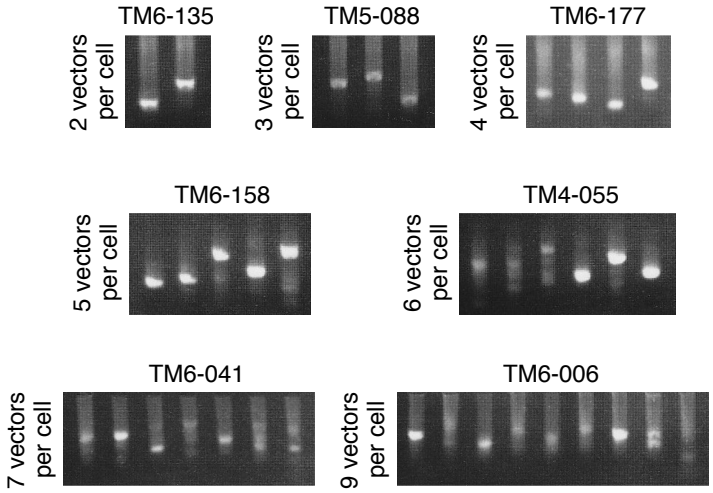

Figure 27. Amplification of different genome portions adjacent to the 3′-ends of multiple integrated vectors from a single ES-cell clone by the tag-directed splinkerette PCR. See Figure 12 for details.

Table 8. Nucleotide sequences of the endogenous genome portions adjacent to the integrated vectors within a single ES-cell clone

Clone number	IDs of integrated transposons	Trapped gene symbol	Chromosome	Orientation[a]	Total number of exons	Vector integration site	Nucleotide sequence of the splinkerette-PCR products
TM6-135	7	Ppp1r14d	02 - 119053822 119053863	forward	4	intron 01	TTTTTACACCTCTGGCTGTCCTGCTATGAGAGTAGAAAGGATGACTGATGGGGCCTTGGCTCGTTGGAGAGGACA
(2 vectors)	12	unknown	14 + 84691807 84691940				TTTTTACACCTCTGACAGAGGATATATATATTCCAAGAATCTAGATATCAACAATGTCTTAAATCAAACAAATAAAGTCAAACAAAATATAGGGTACAGAATAAAACAGAAAATTCTCAACAGAGGAAACTCACGTAGCTAATAAATA
TM5-088	1	Fitm1	14 + 56194955 56195061	forward	2	intron 01	TTTTTACACCTCTGGTGCCGGGGAAGTAAAGGACGGAAGAGGAAAGGTGGACCCTAGTTGCCCTGTGGTCCTGCAAGGAGAAGGGCAACATTGAAATGATAGGGGTTATAAAAAAAAAAAA
(3 vectors)	6	Mtfmt	09 + 65291801 65291948	forward	9	intron 06	TTTTTACACCTCTGATTGTTTCCTTTCTAGGTATATTTATTTAGGAATGAACCTTATGAACTAAACATGAAATTTTCTTAAACTTTAAATTTTAAGTTTTGATATAGCTTTGGAATTCACAAAGTGCTTTCAAATGTACATTCTGCCATATC
	9	repetitive					TTTTTACACCTCTGGTGGTGGCACACCCCTTTAATCCCAGCATTTGGGAGGCAGAGGCAGGTGGATTTCTGAGTTCGAGGCCTT
TM6-177	7	Myef2	02 + 124915087 124915219	reverse	19	intron 16	TTTTTACACCTCTGATTATGACCCCTGGTGACATTACACCTTTAATGAAAATGATACAGTAGCTCACTGTCCAATCTCCTGAATATTTCTCCCTTCTGTTTTTATCTTCTCTTTGCACCTCACAAGAGTAACTATAGTTTCAGTCAG
(4 vectors)	8	repetitive					TTTTTACACCTCTGGAAAGCTGAGGAAGTTCTCTTGGGTGCATTGGGGTCCTTGAACTTCTTTTTGGTCTCCCCTTTATGGGGGGATGTAGGTTTTCATTTCCCTTTCATAATGATCCTTGTCAGCCTTCCACATAACTTCAAAT
	12	Tnc	04 + 63668624 63668677	reverse	26	intron 08	TTTCATAATCTCTACAGGGTTGCAAAATAGGGACATATGCTCCTGTCCTTTCGACCTTGGCTCGTTGGAGAGGACA
	15	unknown	18 + 38304138 38304266				TTTTTACACCTCTGCCTGAAGCCCCCGGTCTAAACCCACCCCACCTCATCTCTATGCCTCCAGCCCCTCTCTCCATCCATCCCAAACAACTCGGCGTTTGCCAGCAGCCTCATTCCTACTTCTTTCCTCCTTCCCCTCTTGTC
TM6-158	6	Pola2	19 + 5946387 5946485	reverse	18	intron 13	TTTTTACACCTCTGACGGCCACGCGGTTTCAGTTTTCCTACTTGTCCTAGAGGAACCTGGAGAGCCCCAGGTTCTGCCCAGCAACTGGAATGCTAACCTTCTCGTTCCTTCGACCTTGGCTCGTTGGAGAGGACA
(5 vectors)	7	Mast2	04 + 116068628 1160687567	reverse	30	intron 03	TTTTTACACCTCTGCATGCTTGTCTGTGTGCTACAGATACACCTAGTGCCATAGAGGCCAGAAAAGGGCTGCAGATCCCCTAGAACAGGAGTTACAGATGGTGGTGATCCACATGCTCATGTCCTCTGGAAGAGTAGCTAGTG
	10	Fkbp5	17 - 28563197 28563328	forward	11	intron 05	TTTTTACACCTCTGGTTGTAAGCCACCATGTGGTTGCTGGGATTTGAACTCAGAACCTTCAGAAGAGCATTCAGTGCTCTTAACCGCTGAGCCATCTCTCCAGCCCCACATTTGTATTTAAAACATGAGTACTAGGCATAGCAG
	11	unknown	01 - 188657949 188658076				TTTTTACACCTCTGCAGGACGCACCCCGGTGAGCGGGAGCAGTTATCCCCAAGACCCGGATCTGAGACTAGTTCCGTTTTTTCCTTTGGGGTGACTAGTCATCTGCACGAGTATACATCTTACCCACACTCGCGCACTCCTG
	15	unknown	01 + 124582681 124582815				TTTTTACACCTCTGTACCTAACATTTGTGCTCTTCTGAAAATATAAGCAAAAATAAAATCTTCCATCTATCCCTTTGGATCCTGTTGTTTTATCACAGCAACAGAAAAACAACTAATACATAACTAAATGTAAAATAGCTGAATAATTT
TM4-055	4	unknown	03 + 18464578 18464709				TTTTTACACCTCTGAAGTAGAAGCAGGTTATGATGAAGGTGGGCTTTAATGTTCCTATGAAGAACAACCAGATAACAAATAACTAAAGGATCAAGACTGCTATTTCAAGCATCAAATTCTGCTTGGACAATGGAGCCAACTCAG
(6 vectors)	5	Elf2	03 - 51082559 51082690	forward	9	intron 03	TTTTTACACCTCTGCTACACACCACACACACCCCTTCGGTTTCCATTTCCTTACTAAATAGAATACATGGAAAATGTGCTACTGGAGCAGAAATTCTGATTTGTTATCGGTTTCATTGATCTGCCTACAGCCATCCAATTACCCTT
	9	unknown	19 - 27703677 27703808				TTTTTACACCTCTGCAACTAGATTTCCTGTTACTTTCCATTTTCAGAAGAAAAACAACAAACAAACAAACAAACAAACCCCAGGGAATTGCTATTATAACACAGTAAATAAATTTGAGGGGGGGAAGAGATTCCTCAATGAGCTT
	12	Ggnbp2	11 + 84651111 84651233	reverse	14	intron 10	TTTTTACACCTCTGTACACAAGTGTGTGCAACTCCAGCTCCAGCAGATCCAACACCCTCACACAGACATCCATGCAGGCAAAACAACAATGCACATAAAATAGAAATGAACGGATAATTAAGAAAAAAAAAAAAAA
	14	Lamc3	02 - 31766511 31766642	reverse	28	intron 07	TTTTTACACCTCTGTATATATATATATATACATGCTTTGCTTGCAATCATGTTTGTGTACTGTGTGCATGTCTGGTGCCTGCAGCATCCAGAGAGGGTGTTGAATTGAATCCCCTGGAACTGGAGTTTCAGATGGTTGTGAGCA
	15	Exoc6	19 - 37680424 37680506	reverse	22	intron 17	TTTTTACACCTCTGAGGGGCACCAAGACCTGACACTATTATTGCTGCTATGATGTGCTTAAAGACAGGAGCCTAGCATGGCTGTCCTCCGAGGGGCCTTGGCTCGTTGGAGAGGACA
TM6-041	1	repetitive					TTTTTACACCTCTGGGCCTAAGACTGATCTTCAATAACAACATAAATAATGGAAAGCCAACATTCACGTGGAAACTGAATAACACTCTTCTCAATGATACCTTGGTCAAGGAAGGAATAAAGAAAGAAATTAAAGACTTTTTAGAG
(7 vectors)	2	repetitive					TTTTTACACCTCTGGGTGGCACACACCTTTAATCTCAGCACTTGGAAGACAGAGGCAGGCAGATTTCTGAGTTCAAGGCCACCTTGCCTACATAGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTTGAA
	4	Bdh1	16 - 31429507 31429605	reverse	7	intron 01	TTTTTACACCTCTGGTGGAGGCCACGGGAGGAGGAGGACGGACGTAGAGCAAGGTCGGGAGATGGGTCGGGACCACGCCGTGCCTTGTTAAGGCTGCATAAGTGGCTTCGACCTTGGCTCGTTGGAGAGGACA
	6	Lrp1	10 - 126977984 126978110	forward	89	exon 82	TTTTTACACCTCTGAGCCCAAGTGCCGTTGCCAGCCCCGTTACACAGGCGATAAGTGTGAGCTGGATCAGTGCTGGGAATACTGTCACAACGGAGGCACCTGTGCGGCTTCCCCATCTGGTATGCCCCCAACCAGGGATGG
	7	unknown	11 - 100519064 100519191				TTTTTACACCTCTGTGGCCTGCTGCTGTGTTTCTCCTGTGTTACCTTGTCTCCAAGATCCCATCTCGGTCTTTGAGATTTGCTTTCAGGCCCTGGGACCCTGTGTCCAACCGGCTCCCCATGCGCAGCCATTTCTGACACTC
	10	unknown	11 - 11678422 11678529				TTTTTACACCTCTGATTACATGTGATTTGAAAGATGGCTGAACAAATGGGTTGGGAACGGGGCTGACCCGATGCTTTGAATGCTTTTTACCCATAGATTCCTGCACTTTATGGTCCCCGAGGACTTGGCTCGTTGGAGAGGAC
	15	poor signal					
TM6-006	2	unknown	02 - 160180372 160180501				TTTTTACACCTCTGGAAGAGACTATTCCCATGTTCTGGATGAGGAAACAGAATCACAGAGAGGTTAAGCAATTTGCCCAAAAATGAACAGCTAGTTAATTGCAGTGGTAGTGTTTGAACTTTGTCAGACTGAGTCTGGAGTCT
(9 vectors)	3	unkonwn	01 + 127524951 127525087				TTTTTACACCTCTGTAGGTCAATGATGTCAGCCATGGGGGTCACGCCATTTCATAAAGCTCCTCCAATGTGATTAAACAATGACAATTCTCCCTAAATAAAGGAACTTGGTCCCCAGAACTAAAACCATTGGCTTCGGGCTTTC
	5	AK078321	05 + 67066999 67067122	reverse	5	intron 01	TTTTTACACCTCTGTTACCCTGGAGGATATATTGCCTAAGGGTAAGTACAAACTGTTCCGAGATGGGCGGGGGGGGGGAGGAGAAAAACCGACAAATATTTATATTCATATTAAACACTACATAAACCTGTTGATTTTGGGCT
	6	unknown	18 + 69911167 69911296				TTTTTACACCTCTGCCTGCTTGTGATTACCATTCATTCTCTGGGCCCTTTCTCACTGGCTGTCCCCATACCTGTCCCCCCCGTACACACACACACCTCCCTCCTCTTCCCCTTCCCATTTCCTCTCCCAACCAGTTCCTTGGTC
	8	poor signal					
	12	poor signal					
	13	Add2	06 + 86019813 86019941	forward	7	intron 01	TTTTTACACCTCTGGAGCTCCTCTGCTAGTGCAGGGCCTAGAGACCCGTCCACTGGTCTGTGTTGGATTTTTAACTGGGTTGATCTTGTGTGGGTAGCCACAGTTGTGTAAGTTGATGTGTACAGCATCCATGTCGTGTCTAG
	14	poor signal					
	15	unknown	02 + 149110978 149111111				TTTTTACACCTCTGGCTTCTTTTATACAAAGGATAAACAGGCTAGTAAAAAAGAATAAGGAAACAACTCCCTTCATAATAGGCACAAATAATATAAAATATCTTGAAGTAACTCTAACCAAACAAGTGAAAGACCAACATGACAATA

[a]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.

CHAPTER 4

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 (Shigeoka et al., 2005). Here, I tried to render the conditional gene-disruption capability to the original retrovirus version of UPATrap by incorporating the elaborate FLEx technique (Schnütgen et al., 2003; Schnütgen et al., 2005), but I frequently experienced broadly distributed intra-vector deletions or rearrangements that should have deleterious effects on the Flp- and Cre-mediated DNA recombination in the FLEx-type of conditional gene regulation (Fig. 16A).

In an attempt to elucidate the molecular mechanism(s), I found that the majority of such structural alterations occur around the IRES sequences inside the genome-integrated vectors (Fig. 16B). The IRES sequence of the encephalomyocarditis virus (EMCV), which is one of the most crucial components of the UPATrap strategy (Shigeoka et al., 2005), is known to form a highly complex secondary structure at the RNA level (Pilipenko et al., 1989). 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 or rearrangements in the genome-integrated proviruses as previously observed for some of the retrovirus constructs containing the EMCV-IRES sequences (Negroni and Buc, 2001; Duch et al., 2004).

We therefore cloned the conditional UPATrap elements into a cut and paste-type DNA transposon, *Tol2* (Koga et al., 1996; Kawakami et al., 2000; Urasaki et al., 2006), and succeeded in suppressing the frequent development of deleterious intra-vector alterations (Fig. 17). Consequently, it became feasible for us to perform unbiased poly(A) trapping in a

conditional manner, especially with high reliability (Fig. 23). The high stability of the *Tol2* vectors has already been demonstrated in the context of the genomes of cultured ES cells (Yagita et al., 2010) and transgenic mice (Sumiyama et al., 2010). 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 (Fig. 19), 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 (Fig. 20).

In IGTC, the majority of research groups have been engaged in promoter trapping originally developed for the disruption of expressed genes in the target cells, and 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) (Friedel et al., 2005). On the other hand, however, they also showed that the gene-expression levels affect the efficiency of promoter trapping/targeted promoter trapping (Friedel et al., 2005), and our finding shown in Figure 14 are basically consistent with their observations. In addition to conventional promoter trapping, the poly(A)-trapping strategies including original UPATrap (Shigeoka et al., 2005; Stanford et al., 2006; Brickman et al., 2010) 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 14. 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 (Scherdin et al., 1990; Wu et al., 2003), 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 I found that the UPATrap-*Tol2* transposon vector shows a weaker preference to be integrated into transcriptionally active genes than does the retrovirus counterpart (Fig. 19), this does not mean that *Tol2* is completely 'bias-free' in terms of the selection of integration sites. The results of Figure 14 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 (Fig. 19). 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 (Ivics et al., 2009; Horie et al., 2010; Li et al., 2010), and a recent investigation suggested that *SB* does not have strong preference to be integrated into transcriptionally active loci (Huang et al., 2010). 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 (Figs. 24-27). 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. I 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 (Skarnes et al., 2011). 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 (Skarnes et al., 2011). 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 (the FANTOM consortium, 2005; Guttman et al., 2011) 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 would be interested in performing 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 (Leeb and Wutz, 2011; Elling et al., 2011) would certainly increase the chance of conducting insertional-mutagenesis experiments based on the phenotypic screening at individual-laboratory level. The gene-trapping strategy using a mixture of conditional UPATrap-*Tol*2 transposons described in this paper should have a lot to contribute to these potential future analyses.

ACKNOWLEDGEMENTS

My sincere homage and gratitude to God for His blessing and grace for the successful completion of this work.

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Masashi Kawaichi for welcoming me into his laboratory, and for his constructive criticism and encouragement throughout this study. I would like to extend my gratitude to Dr. Yasumasa Ishida for his excellent supervision, invaluable discussion, guidance and advice during the course of this work. The majority of this work would not be otherwise possible without his help.

I wish to thank Dr. Chio Oka and Dr. Eishou Matsuda for sharing their experience and knowledge. I also thank to Ms. Akiko Murakami and Mrs. Yumiko Kawakami for their guidance in some technical works. Thanks to all past and present members of Gene Function in Animals Laboratory for their cooperation and warm friendship.

I would like to acknowledge and best regards to my Advisory Committee members: Prof. Jun-ya Kato, Prof. Thomas N. Sato and Dr. Kohsuke Kataoka for their constructive criticism and factual comment on this research.

I would like to give my sincere gratitude to Mrs. Sachiko Iida for providing financial support through the Iida International Student Scholarship during my PhD course, and for her encouragement and thoughtfulness.

I thank to Indonesian Ministry of Education and Bogor Agricultural University for giving me permission to study abroad.

Finally, I thank to my beloved parents, my sister and brother, and my relatives for their blessing, support and keep praying for my success throughout the whole period of my studies.

REFERENCES

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.

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.

Brickman, J.M., Tsakiridis, A., To, C., and Stanford, W.L. (2010). A wider context for gene trap mutagenesis. Methods Enzymol 477, 271-295.

Buchholz, F., Angrand, P.O., and Stewart, A.F. (1996). A simple assay to determine the functionality of Cre or FLP recombinase targets in genomic manipulation constructs. Nucleic Acid Res 24, 3118-3119.

Buchholz, F., Angrand, P.O., and Stewart, A.F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. Nat Biotechnol 16, 657-662.

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.

Devon, R.S., Porteous, D.J., and Brooke, A.J. (1995). Splinkerettes--improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res 23, 1644-1645.

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.

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.

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 genetic through derivation of haploid mouse embryonic stem cells. Cell Stem Cell 9, 563-574.

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.

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.

Friedel, R.H., and Soriano, P. (2010). Gene trap mutagenesis in the mouse. Methods Enzymol 477, 243-269.

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.

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.

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.

Horie, K., Kokubu, C., and Takeda, J. (2010). Functional genomics in the mouse using the Sleeping Beauty transposon system. Methods Enzymol 477, 71-89.

Horn, C., Hansen, J., Schnutgen, 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.

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.

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.

Ivics, Z., Li, M.A., Mates, L., Boeke, J.D., Nagy, A., Bradley, A., and Izsvak, Z. (2009). Transposon-mediated genome manipulation in vertebrates. Nat Methods 6, 415-422.

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 97, 11403-11408.

Koga, A., Suzuki, M., Inagaki, H., Bessho, Y., and Hori, H. (1996). Transposable element in fish. Nature 383, 30.

Leeb, M., and Wutz, A. (2011). Derivation of haploid embryonic stem cells from mouse embryos. Nature 479, 131-134.

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.

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 101, 4170-4174.

Maquat, L.E., and Carmichael, G.G. (2001). Quality Control of mRNA Function. Cell 104, 173-176.

Maquat, L.E. (2004). Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 5, 89-99.

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.

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.

Nagy, E., and Maquat, L.E. (1998). A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci 23, 198-199.

Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2003). Manipulating the Mouse Embryo: A Laboratory Manual, 3rd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Negroni, M., and Buc, H. (2001). Mechanisms of retroviral recombination. Annu Rev Genet 35, 275-302.

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.

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, D642-648.

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.

Rajewsky, K., Gu, H., Kuhn, R., Betz, U.A., Muller, W., Roes, J., and Schwenk. F. (1996). Conditional gene targeting. J Clin Invest 98, 600-603.

Raymond, C.S., and Soriano, P. (2007). High-efficiency FLP and ΦC31 site-specific recombination in mammalian cells. PLoS One 2, e162.

Sambrook, J., and Russel, D.W. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J Virol 64, 907-912.

Schnütgen, F., Doerflinger, N., Calleja, 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.

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 102, 7221-7226.

Schnütgen, F., Stewart, A. F., von Melchner, H., and Anastassiadis, K. (2006) Engineering embryonic stem cells with recombinase systems. Methods Enzymol 420, 100-136.

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.

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.

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.

Stanford, W.L., Cohn, J.B., and Cordes, S.P. (2001). Gene-trap mutagenesis: past, present and beyond. Nat Rev Genet 2, 756-768.

Stanford, W.L., Epp, T., Reid, T., and Rossant, J. (2006). Gene trapping in embryonic stem cells. Methods Enzymol 420, 136-162.

Sternberg, N., Sauer, B., Hoess, R., and Abremski, K. (1986). Bacteriophage P1 Cre gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J. Mol. Biol. 187, 197-212.

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.

The FANTOM consortium. (2005). The transcriptional landscape of the mammalian genome. Science 309, 1559-1563.

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.

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.

Xin, H.B., Deng, K.Y., Shui, B., Qu, S., Sun, Q., Lee, J., Greene, K.S., Wilson, J., Yu, Y., Feldman, M., and Kotlikoff, M.I. (2005). Gene trap and gene inversion methods for conditional gene inactivation in the mouse. Nucleic Acids Res 33, e14.

Yagita, K., Horie, K., Koinuma, S., Nakamura, W., Yamanaka, I., Urasaki, A., Shigeyoshi, Y., Kawakami, K., Shimada, S., Takeda, J., and Uchiyama, Y. (2010). Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. Proc. Natl. Acad. Sci. USA 102, 3846-3851.

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.

Zambrowicz, B.P., Friedrich, G.A., Buxton, E.C., Lilleberg, S.L., Person, C., Sands, A.T. (1998). Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392, 608-611.