Published online 15 March 2018

Nucleic Acids Research, 2018, Vol. 46, No. 10 ***e63***
doi: 10.1093/nar/gky183

Collection of homozygous mutant mouse embryonic stem cells arising from autodiploidization during haploid gene trap mutagenesis

Ayako Yamanishi[1], Atsushi Matsuba[2], Ryohei Kondo[2], Rie Akamatsu[2], Sachiyo Tanaka[1], Masahiro Tokunaga[1], Kyoji Horie[3], Chikara Kokubu[1], Yasumasa Ishida[2,*] and Junji Takeda[1,*]

[1]Department of Genome Biology, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan, [2]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan and [3]Department of Physiology II, Nara Medical University, Kashihara, Nara 634-8521, Japan

Received December 21, 2017; Revised February 27, 2018; Editorial Decision February 27, 2018; Accepted March 09, 2018

ABSTRACT

Haploid mouse embryonic stem cells (ESCs), in which a single hit mutation is sufficient to produce loss-of-function phenotypes, have provided a powerful tool for forward genetic screening. This strategy, however, can be hampered by undesired autodiploidization of haploid ESCs. To overcome this obstacle, we designed a new methodology that facilitates enrichment of homozygous mutant ESC clones arising from autodiploidization during haploid gene trap mutagenesis. Haploid mouse ESCs were purified by fluorescence-activated cell sorting to maintain their haploid property and then transfected with the *Tol2* transposon-based biallelically polyA-trapping (BPATrap) vector that carries an invertible G418 plus puromycin double selection cassette. G418 plus puromycin double selection enriched biallelic mutant clones that had undergone autodiploidization following a single vector insertion into the haploid genome. Using this method, we successfully generated 222 homozygous mutant ESCs from 2208 clones by excluding heterozygous ESCs and ESCs with multiple vector insertions. This relatively low efficiency of generating homozygous mutant ESCs was partially overcome by cell sorting of haploid ESCs after *Tol2* BPATrap transfection. These results demonstrate the feasibility of our approach to provide an efficient platform for mutagenesis of ESCs and functional analysis of the mammalian genome.

INTRODUCTION

The effectiveness of loss-of-function mutagenesis screens in mammalian cells has been hampered by the diploid property of the genome. Even if one allele of an autosomal gene is disrupted, the existence of another allele usually masks its phenotypic manifestation. As a solution, we have previously developed a homozygous mutant mouse embryonic stem cell (ESC) bank in which genome-wide heterozygous mutants were converted into homozygotes by an increased rate of loss of heterozygosity on the Bloom-deficient background (1–3).

Alternatively and more straightforwardly, haploid mouse ESCs (4,5) and near haploid human cells (6,7) have been established as powerful tools for genetic screening of loss-of-function phenotypes (8–13). However, forward genetic screenings of these cell lines have been more or less hampered by inherent cellular autodiploidization during culture. Indeed, undesirably diploidized ESCs are much more efficiently transfected with insertional mutagenesis vectors than their original haploid cells (14). To overcome this drawback, we designed a new methodology that allows efficient enrichment of homozygous mutant clones arising from haploid ESCs mutagenized by gene trap insertional mutagenesis.

Genome-wide gene trap mutagenesis typically uses either of two types of insertional vectors. One is a promoter trap vector that consists of a promoterless splicing acceptor and selectable marker cassette (15). The selectable marker gene is transcribed only when the promoter trap vector inserts within a transcriptionally active gene. Another is a polyA trap vector that consists of a promoter driving expression of a selectable marker gene and a splicing donor lacking a polyA signal (16–19). The selectable marker gene can be stabilized upon trapping of a polyA signal, regardless of whether the target gene is transcriptionally active or silent. Here, we used a polyA trap vector carried by the medaka *Tol2* transposon to mutagenize haploid mouse ESCs randomly. The highly elaborate design of our vector system allowed efficient enrichment of homozygous mutant ESCs

*To whom correspondence should be addressed. Tel: +81 743 72 5531; Fax: +81 743 72 5539; Email: ishiday@bs.naist.jp
Correspondence may also be addressed to Junji Takeda. Tel: +81 6 6879 3262; Fax: +81 6 6879 3266; Email: takeda@mr-envi.med.osaka-u.ac.jp

arising via autodiploidization of their parental haploid mutant ESCs, resulting in the generation of a homozygous mutant mouse ESCs applicable to forward loss-of-function screening of the mammalian genome.

MATERIALS AND METHODS

Vector construction

The neomycin resistance gene (*neo*) was ligated to a fusion gene of puromycin resistance and herpes simplex virus thymidine kinase genes (*puroΔtk*) in a tail-to-tail manner. This tail-to-tail marker cassette was flanked by inversely oriented *lox2272* sequences (3) and cloned into the differentially tagged UPATrap-*Tol2* transposon vector (20). The resulting vector was designated as biallelically polyA trapping (BPATrap). Its sequence has been deposited in GenBank/EMBL/DDBJ (Accession Number: 6648).

Cell culture and sorting

H129-2 haploid mouse ESCs (kindly provided by Martin Leeb, the University of Vienna) and their derivatives were cultured on puromycin-sensitive mouse embryonic fibroblast (MEF) feeder cells in chemically defined two inhibitor (2i) medium plus leukemia inhibitory factor (LIF) or embryonic stem cell medium (ESM) containing 20% fetal calf serum (11). To sort haploid cells, ESCs were stained with 15 μg/ml Hoechst 33342 (Life Technologies) at 37°C for 30 min, and cells in the 1N peak were purified using a BD FACSAria II Cell-sorting System at 375 nm (BD Biosciences). The DNA content was analyzed by a FACSCanto II (BD Biosciences). Data analyses were performed using BD FACSDiva software (BD Biosciences) or FlowJo software (TreeStar).

Genome editing in haploid mouse ESCs

Five micrograms of linearized targeting vector pROSA26-F3-neoW-ERT2-iCre-ERT2 (3) was electroporated into 1×10^6 H129-2 haploid mouse ESCs, together with 5 μg of zinc-finger nuclease (ZFN) mRNA, which targets the *Rosa26* locus (M4574, Sigma-Aldrich), using the Neon transfection system (Invitrogen) with conditions of 1400 V, 20 ms and one pulse. The transfected cells were selected by G418 (200 μg/ml) for 6 days. Individual clones were screened by polymerase chain reaction (PCR) for homologous recombination using primers ROSA26-F-long and ROSA-SA-long. Correct targeting at the Rosa26 locus was confirmed by competitive PCR using primers ROSA26-Xba-U, ROSA26-Xba-L, and ROSA-SA-long2, and the DNA content of cells was analyzed by the FACSCanto II. To remove the neo cassette flanked with F3 [mutant of the flippase (Flpo) target site], *Rosa26*-targeted clones were transfected with pCAGGS-Flpo-IRESpuro (3) using the Neon transfection system under the conditions described above. The transfected cells were transiently selected with puromycin (1 μg/ml) in 2i medium plus LIF for 2 days, and then 2000 cells were plated on a 10-cm dish in 2i medium plus LIF. Individual clones were picked and screened for neo cassette removal by PCR using primers pBig5-F, NeoW-R and ERT2iCre-R. The resulting ESCs were analyzed for their DNA content by the FACSCanto II. Correct targeting was confirmed by Southern blotting. The primer sequences are listed in Supplementary Table S1.

Vector transfection

One million *Rosa26*^{E2iCreE2}haploid ESCs were co-transfected with 0.1 μg of a mixture of 15 differentially tagged BPATrap vectors and 1 μg pCAGGS-mT2TP transposase plasmid using the Neon transfection system (1400 V, 10 ms and three pulses).

Screening for single copy vector insertion

Single copy BPATrap insertion per haploid genome per cell was screened by PCR amplification of the 15 differentially tagged regions of the vector using primers n-RACE and R-term, followed by Sanger sequencing using the primer RS-F4. The primer sequences are listed in Supplementary Table S1.

G418 plus puromycin double selection

ESC clones were transferred into a gelatin-coated 12-well plate in ESM and treated with 1 μM 4-hydroxytamoxifen (4-HT) for 16 h to activate Cre-*lox2272* recombination. From the next day onward, the ESCs were double selected with G418 (300 μg/ml) and puromycin (2 μg/ml) for 5 days or more. Genomic DNA was extracted from the ESCs by a standard phenol/chloroform method.

Splinkerette PCR

Approximately 100 ng genomic DNA was digested with HaeIII or AluI (New England BioLabs) and ligated with the splinkerette Spl-top/SplB-BLT linker (3) using T4 DNA ligase (Takara). Vector insertion sites were amplified by nested PCR with primer sets Spl-P1/SPL-1 (first round) and Spl-P2/SPL-2m (second round) (3,20). The resulting amplicons were sequenced with the primer SPL-seq to map insertion sites to the mouse reference genome (mm10, Dec 2011) using BLAT (http://genome.ucsc.edu/). Primers sequences are listed in Supplementary Table S1.

Quantitative PCR (qPCR) analysis

Locus-specific primers were designed based on the flanking sequences of each insertion site of BPATrap obtained by splinkerette PCR. Quantitative PCR (qPCR) was performed using ESC clone-derived genomic DNA to genotype all BPATrap insertion loci by measuring the relative copy number of gene trap and wild-type alleles. GoTaq qPCR Master Mix (Promega) and the Applied Biosystems 7900HT Fast Real-Time PCR System were used under the following conditions: denaturation at 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15 s and extension at 60°C for 1 min, and finally a dissociation stage of 95°C for 15 s, 60°C for 15 s, and 95°C for 15 s. The relative amounts of amplification products were calculated using *Gapdh* (glyceraldehyde-6-phosphate dehydrogenase) as a standard. Primers sequences are listed in Supplementary Table S2.

Karyotyping

The karyotypes of ESC clones were analyzed as described previously (3). A total of 100 metaphase cells were analyzed for each clone.

Isolation of revertant clones

To remove the mutagenic vector sequences flanked with F3, 1×10^6 homozygous mutant cells (H03-D12: *Psph*$^{m/m}$ and H03-G08: *Zic3*$^{m/m}$) were transfected with pCAG-Flpo-IRESBsd using the Neon transfection system (1400 V, 10 ms and three pulses). The transfected cells were transiently selected with blasticidin S (10 μg/ml) in ESM for 2 days, and then 1000 cells were plated on a 10-cm dish in ESM. Individual clones were picked and screened by PCR with primer sets SA-R/H03-D12-F/H03-D12-R (H03-D12: *Psph*$^{r/r}$) or SA-R/H03-G08-F/H03-G08-R H03-G08: *Zic3*$^{r/r}$). The primer sequences are listed in Supplementary Table S1.

Cell lysis and western blotting

ESCs were suspended in ESM and plated on gelatin-coated culture dishes for 30 min to remove MEFs. Non-attached ESCs were resuspended in sodium dodecyl sulphate (SDS) sample buffer [50 mM Tris–HCl, pH 6.8, 2% (wt/vol) SDS, 10% (vol/vol) glycerol, 5% (vol/vol) 2-mercaptoethanol and 0.01% bromophenol blue], and cell lysates were separated on SDS-polyacrylamide gels. Proteins were transferred to nitrocellulose membranes (Millipore) and detected using primary antibodies recognizing phosphoserine phosphatase (PSPH) (Abcam, ab156414, 1:100), ZIC3 (Abcam, ab215063, 1:1000) or β-actin (Sigma, clone AC15, 1:2000). Membranes were further incubated with appropriate secondary antibodies, and proteins were visualized using enhanced chemiluminescence (ECL) reagents (GE Healthcare).

RESULTS

Basic principle of the strategy to generate a collection of homozygous mutant mouse ESCs

To generate homozygous mutant mouse ESCs, we exploited the inherent property of haploid mouse ESCs that undergo autodiploidization during culture (4,5,10,21,22). Figure 1 shows an outline of the strategy that focuses on the structure and function of gene trap vectors, together with the ploidy of target chromosomes under drug selection pressure. The efficiency of the integration of exogenous vectors is lower in the haploid ESC genome than the autodiploidized ESC genome for undefined reasons (14). Therefore, the key challenge in this approach is how to eliminate single allelic (heterozygous) gene trap events that occur following autodiploidization of haploid ESCs. We constructed a gene trap vector in which neomycin (N) and puromycin (P) resistance genes were located downstream of the promoter (Pr) in a tail-to-tail orientation and flanked by inversely oriented *lox* sequences. Cre-lox site-specific recombination inverts the lox-flanked (floxed) DNA cassette, resulting in the induction of resistance against either G418 or puromycin. When only a single copy of the gene trap vector is inserted into a haploid or diploid ESC genome, the host ESCs will be killed by G418 plus puromycin double selection, regardless of the orientation of the floxed DNA cassette (Figure 1A and C). In contrast, when host ESCs undergo autodiploidization following completion of a single locus vector insertion at the haploid phase, the resulting ESCs should contain two copies of the floxed cassettes of which can be inverted by Cre-mediated recombination to confer double resistance against both G418 and puromycin (Figure 1B). Moreover, diploid ESCs with a single locus vector insertion were occasionally converted to tetraploid ESCs, which conferred resistance to G418 plus puromycin double selection (Figure 1D).

Construction of the BPATrap vector system

For random mutagenesis, polyA trapping has a strong advantage in mutagenizing transcriptionally silent and active genes. We have previously shown that insertion of an internal ribosome entry site (IRES) sequence in a polyA trap vector suppresses a nonsense-mediated mRNA decay (NMD) mechanism, allowing unbiased mutagenesis of endogenous genes independent of the position of vector insertion within target gene loci (23). In this study, we created the *Tol2* transposon-based BPATrap vector consisting of three functional components: a gene terminator cassette, invertible selection marker cassette and differential tag sequence (Figure 2A). First, the gene terminator cassette consisted of promoterless enhanced green fluorescent protein cDNA and polyA signals, which can monitor the expression of trapped genes in cells and terminate their endogenous transcription, respectively (20). Second, the invertible selection marker cassette contained a *Pgk* (phosphoglycerate kinase-1) promoter to drive a conditionally invertible *neo-puroΔtk* segment (3). The *neo-puroΔtk* segment flanked by inversely oriented *lox2272* sequences was fused to an IRES sequence to suppress the undesired NMD mechanism as mentioned above. Third, the differential tag sequence, which contained any of the 15 variant barcodes, consisted of a diagnostic CC-in-poly(AT) portion and vector identification (ID) portion. As described in our previous study (20), PCR-based direct sequencing of the barcode (referred hereafter as 'direct PCR seq') allows straightforward estimation of the number of integrated vector copies by counting the overlapping dinucleotide peaks within a sequence chromatogram of the CC-in-poly(AT) segment (Figure 2B; Supplementary Figure S1A and B). Therefore, based on the results of direct PCR seq, ESC clones with a single copy BPATrap insertion per haploid genome per cell (i.e. ESC clones with a single locus BPATrap insertion) could be selected.

In parallel, we engineered host haploid mouse ESCs in which Cre activity was inducible by 4-HT administration. An ERT2-iCre-ERT2 cassette (3), which is the fusion of codon-optimized Cre (iCre) and two copies of the mutant-type estrogen receptor ligand-binding domain (ERT2), was targeted in the endogenous *Rosa26* locus of haploid mouse ESCs via ZFN-mediated homologous recombination (Figure 2C and Supplementary Figure S2A, see details in 'Materials and Methods' section). Among 42 transfected clones,

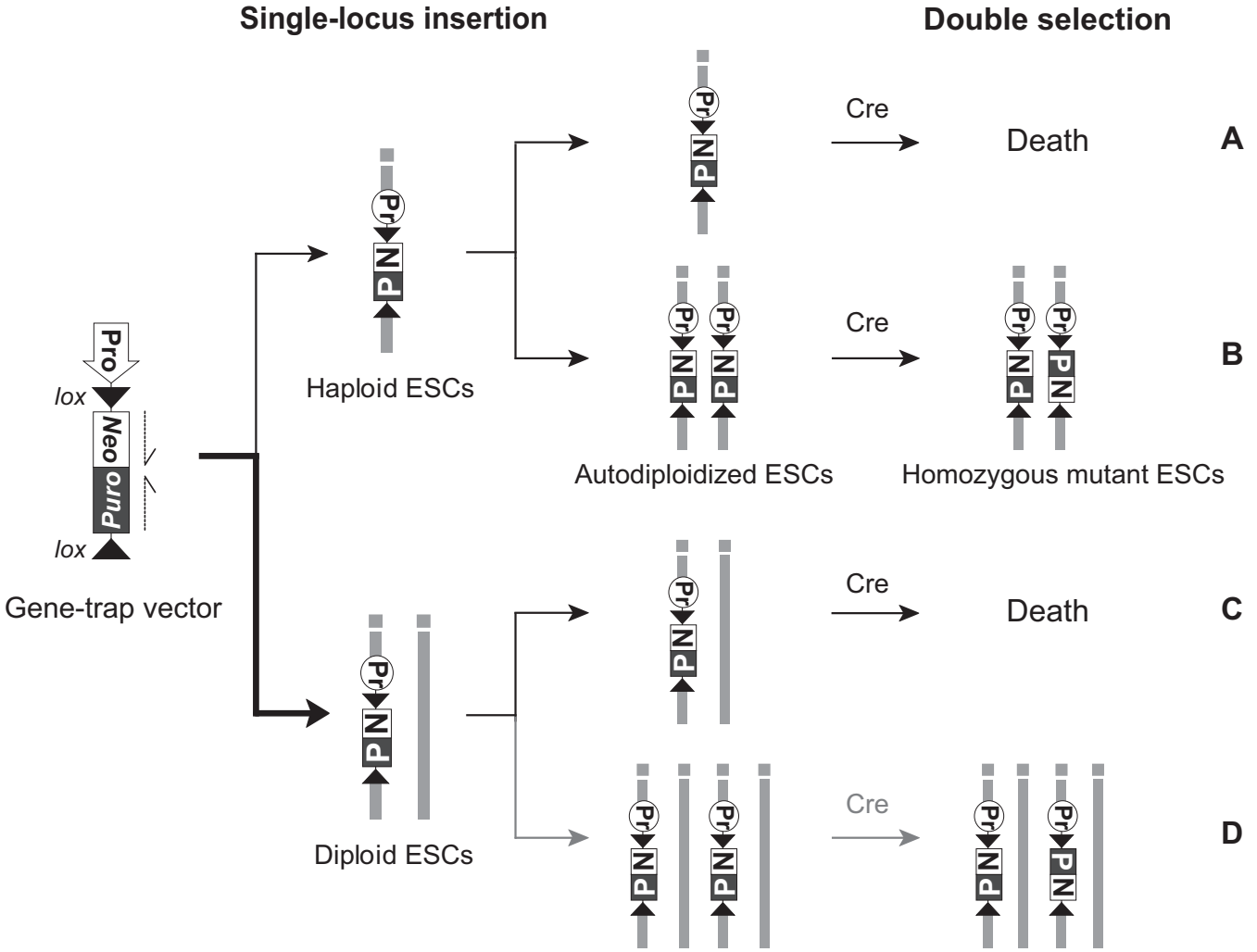

Figure 1. Experimental design. The gene trap vector consisting of a promoter (Pr), tail-to-tail-oriented neomycin (N) and puromycin (P) resistance genes, and two inversely oriented *lox* sequences (solid triangles). Dotted arrows indicate the orientation of coding regions. The thickness of solid arrows represents the integration efficiency of the vector in mouse ESCs. (**A**) Vector insertion at the haploid phase. (**B**) Autodiploidization following vector insertion. (**C**) Vector insertion at the diploid phase. (**D**) Conversion to the tetraploid phase following vector insertion. See text for details.

23 clones had homologous integration in the Rosa26 locus as judged by homologous PCR. However, 21 clones had the wild-type *Rosa26* allele as shown by competitive PCR screening, suggesting that the targeting occurred after diploidization of haploid ESCs. The ploidy of the remaining two clones was analyzed, and we found that one clone had lost the haploid peak (Supplementary Figure S2B, middle). Thus, only one clone out of 42 clones was correctly targeted at the haploid stage of ESCs (Supplementary Figure S2B, lower). Correct targeting of ERT2-iCre-ERT2 in the *Rosa26* locus was confirmed by Southern blotting (Supplementary Figure S2C), resulting in establishment of the *Rosa26*^{E2iCreE2} haploid mouse ESC line.

Generation of homozygous mutant ESCs by gene trapping in haploid mouse ESCs

Using the *Tol2* transposon-based BPATrap vector, we performed random gene trap mutagenesis of haploid mouse ESCs and generated a collection of homozygous mutant ESCs. The overall procedure is shown in Figure 3A.

First, a haploid population was enriched by cell sorting of the *Rosa26*^{E2iCreE2} haploid mouse ESC line that was cultured in chemically defined 2i medium plus LIF on puromycin-sensitive MEF feeder cells (day 7, Figure 3A) (11). To direct a high rate of single copy insertion of the transposon-based vector into the haploid ESC genome, preliminary experiments showed that the optimal mass ratio of the transposon and transposase vectors was 1:10 in our transfection conditions (Supplementary Figure S3A). Accordingly, we co-transfected haploid ESCs with the mixture of 15 differentially tagged BPATrap vectors (0.1 μg, transposon) and pCAGGS-mT2TP (1.0 μg, transposase) at a haploid cell proportion of 89% (Figure 3B, middle of top). The transfected cells were plated on MEF feeder cells, cultured for 4 days and then selected by G418 (200 μg/ml). After 5 or 6 days of G418 selection, 2208 G418-resistant (G418^{R}) colonies were individually isolated in MEF-coated

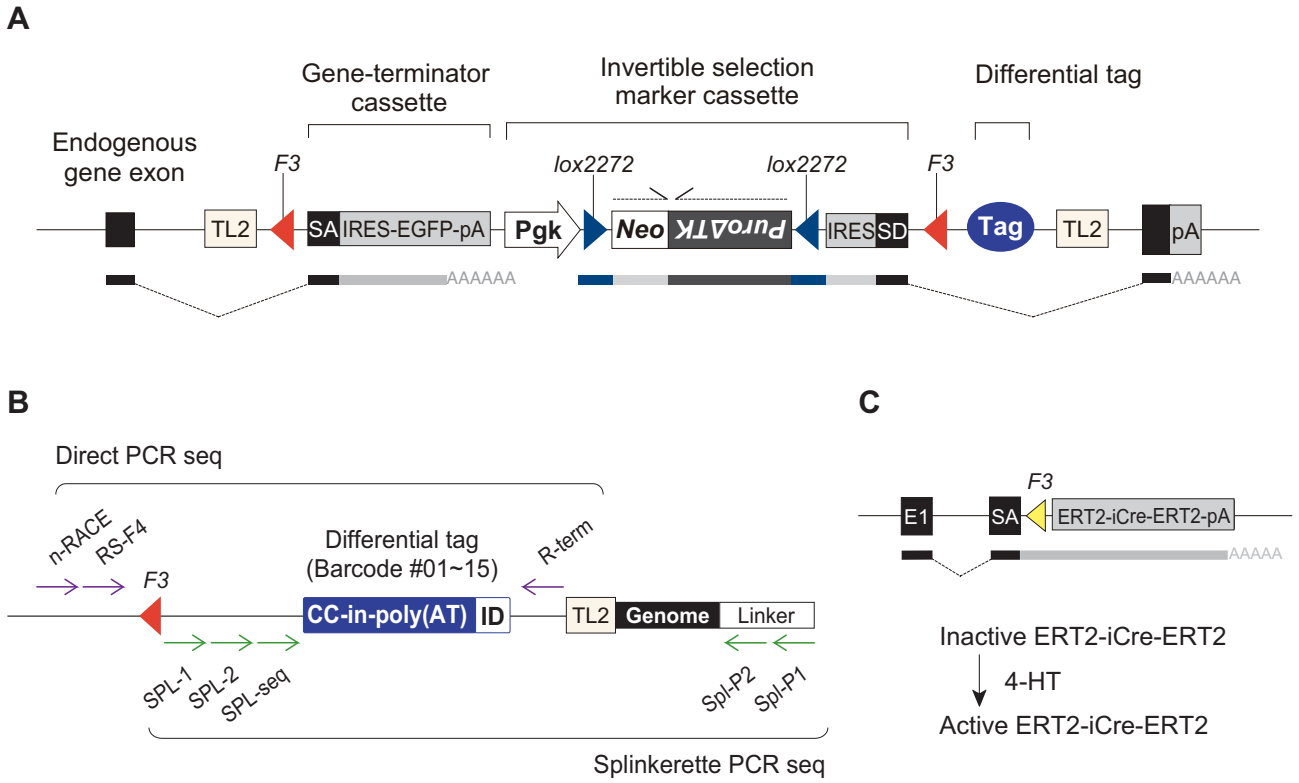

Figure 2. Schematic diagram of the biallelically polyA-trapping (BPATrap) vector system. (**A**) A BPATrap cassette carried in the Tol2 transposon can be inserted into introns of endogenous genes. TL2, *Tol2* transposon terminal inverted repeats; *F3*, mutant of the Flpo recombinase target site; *lox2272*, mutant 2272 of the Cre recombinase target site; SA, splicing acceptor site; SD, splicing donor site; IRES, internal ribosomal entry site; EGFP, enhanced green fluorescent protein; pA, polyadenylation signal; Pgk, phosphoglycerate kinase-1 promoter; Neo, neomycin resistance gene; PuroΔtk, puromycin-delta thymidine kinase fusion gene; Tag, region encoding any of the 15 differential barcodes. Dotted arrows indicate the orientation of coding regions. (**B**) Use of the Tag region within the BPATrap vector facilitates screening and re-confirmation of ESC clones with a single locus integration of the vector. (Top) Direct PCR sequencing of the diagnostic CC-in-poly(AT) region allows estimation of the number of integrated vector copies by counting dinucleotide peaks within a sequence chromatogram (20). Primers for amplification (n-RACE and R-term) and sequencing (RS-F4) are shown. (Bottom) Sequencing reads obtained by splinkerette PCR of the transposon-genome junctions include the barcode sequence, allowing re-confirmation of a single locus insertion of the vector. Primers specific for the vector (SPL-1, Spl-P1) and splinkerette linker (SPL-2m, Spl-P2), and a sequencing primer (SPL-seq) are shown. (**C**) An ERT2-iCre-ERT2 fusion gene cassette was knocked into the *Rosa26* locus of haploid mouse ESCs using ZFN-mediated homologous recombination. The Cre recombinase activity derived from the fusion protein is dependent on the presence of 4-HT.

96-well plates (plate numbers H01–H23). Among them, 1369 ESC clones were categorized as single locus BPATrap insertion clones based on direct PCR seq of the vector barcodes (day 20, Figure 3A). The distribution of each barcode was random (Supplementary Figure S3B).

Next, the 1369 ESC clones were individually plated and cultured in gelatin-coated 12-well plates. During culture, a substantial fraction of ESCs was expected to change their ploidy from haploid to diploid spontaneously (autodiploidization). The ESC clones were then treated with 1 μM 4-HT to induce Cre-mediated inversion of either allele of the duplicated floxed *neo-puroΔtk* cassette, which conferred G418 plus puromycin double resistance (day 28, Figure 3A). After 5 days of double selection by G418 (300 μg/ml) and puromycin (2 μg/ml), double-resistant ($G418^R Puro^R$) clones were obtained from 964 ESC clones, while the other 405 clones were sensitive to the double selection (day 33, Figure 3A).

The genomic insertion sites of the BPATrap vectors were analyzed by splinkerette PCR sequencing (seq) (3) in each of the 964 $G418^R Puro^R$ ESC clones. In the analysis, amplicons encompassing transposon-genome junctions also included the region of differential barcodes, which conveniently allowed re-confirmation of a single locus insertion of the BPATrap vector (Figure 2B). As a result, 645 ESC clones were confirmed as single locus BPATrap insertion clones, whereas 273 ESC clones appeared to have unexpected multiple locus insertions of the BPATrap vector (day 33, Figure 3A). The 273 clones containing multiple locus BPATrap insertions were further classified into two types: multiple and single barcode types (Supplementary Figure S4A). The former type included 27 ESC clones in which a mixture of at least two different barcodes was detected by splinkerette PCR seq. In contrast, the latter type included 246 ESC clones in which only a single barcode sequence was detected, although at least two different insertion sites were determined by splinkerette PCR seq. The unexpectedly high number of multiple locus insertions with a single barcode prompted us to analyze the integration sites of the transposon by splinkerette PCR seq, in which the two PCR products were separately gel purified and sequenced. We analyzed 74 of 246 clones, and two insertion sites of 29

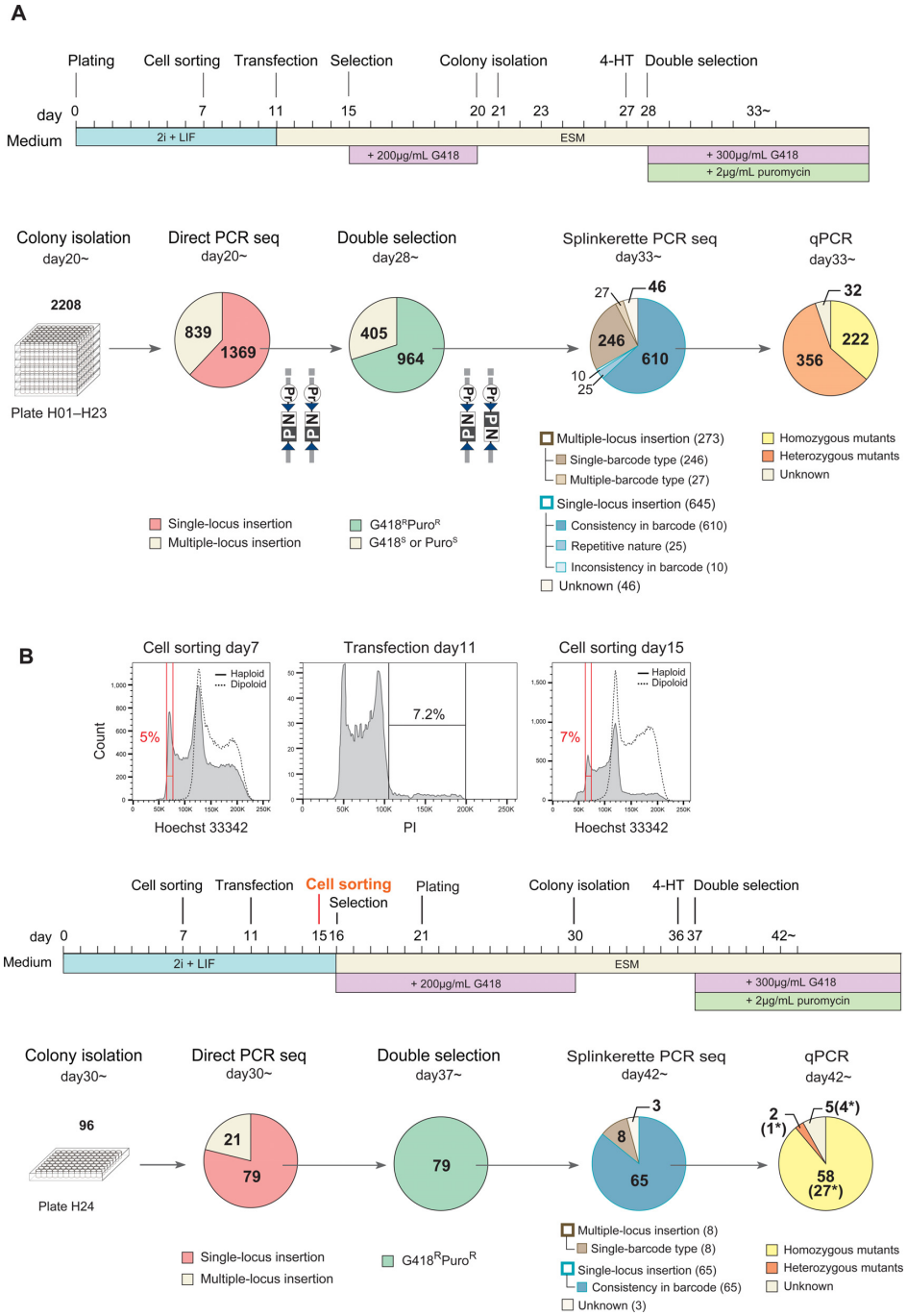

Figure 3. Generation of a homozygous mutant ESCs by gene trapping in haploid mouse ESCs. (A) Haploid mouse ESCs were enriched by cell sorting and then co-transfected with BPATrap and pCAGGS-mT2Tp transposase vectors by electroporation (day 11). ESCs harboring BPATrap insertions within endogenous gene loci were selected using G418 (day 15). A total of 2208 $G418^R$ ESC colonies were isolated in 96-well plates (days 20 and 21). Genomic DNA was extracted from each ESC clone and subjected to direct PCR sequencing (seq) of vector barcodes to identify clones with a single locus insertion of the vector (day 22). The resulting 1369 clones with a single locus BPATrap insertion were seeded on gelatin-coated 12-well plates, treated with 4-HT (1 μM) for 16 h, and then subjected to G418 plus puromycin double selection (day 27), resulting in the isolation of 964 $G418^R Puro^R$ clones (day 33). Out of the 964 clones, 645 clones contained single locus BPATrap insertions that were re-confirmed by splinkerette PCR-based sequencing (splinkerette PCR seq) of the vector-genome junctions (day 33). Among the 645 ESC clones, 610 clones were genotyped by qPCR using specific primers for each insertion junction, resulting in the isolation of 222 ESC clones homozygous for BPATrap insertional mutations (day 34). (**B**) To remove diploidized ESCs, the haploid ESCs were enriched by cell sorting after transfection (day 15). The sorted ESCs were selected with G418 (day 16) and plated at 1000 cells per 10-mm dish (day 21). The flow cytometric profiles before and after transfection with the BPATrap vector are shown (Left and right of top). The sorted population is indicated by the vertical red line (5 or 7%). Solid and dashed lines show the profiles of haploid and diploid ESCs, respectively. And the DNA profile at 4 days after sorting was shown (Middle of top). The population of haploid ESCs is 89%. A total of 96 $G418^R$ ESC colonies were isolated in a 96-well plate (day 30). Out of the 96 clones, 79 clones were classified as single locus BPATrap insertion by direct PCR seq (day 31). All clones were resistant to G418 plus puromycin double selection (day 42). Among them, 65 clones with single locus BPATrap insertions were re-confirmed by splinkerette PCR seq (day 43). Out of the 65 clones, 58 homozygous mutants were analyzed by qPCR (day 44). Asterisk indicates the number of independent clones.

clones were successfully determined. Sixteen clones had two different integration sites in the same chromosome (Supplementary Table S3). Because transposition of a transposon prefers the same chromosome, which is called 'local hopping' (24), the high number of clones with multiple locus insertions could be partly explained by this local hopping after initial integration of the transposon.

From the 645 clones categorized as single locus insertion clones (day33, Figure 3A), 25 clones were removed because their insertion sites were unmappable because of the repetitive nature of their flanking sequences. In addition, 10 clones were removed because of inconsistency in barcode sequence reads between direct PCR seq and splinkerette PCR seq (Supplementary Figure S4B). For the remaining 610 clones, we performed qPCR analysis to distinguish homozygous and heterozygous mutations using primer pairs specific for each insertion site of the BPATrap vector (Supplementary Figure S5A). As a result, 222 ESC clones were homozygous for the BPATrap insertion, whereas 356 ESC clones were heterozygous for the insertional and wild-type alleles (day 33, Figure 3A).

A possible explanation for the carryover of the 356 heterozygous clones is that a significant fraction of diploidized ESCs, which had received a single allelic insertion of the BPATrap vector following an autodiploidization event, changed spontaneously into tetraploid cells containing two insertions and two wild-type alleles per cell (Supplementary Figure S5A). Indeed, the average ratio of insertion versus wild-type alleles in the 356 ESC clones was 0.52 ± 0.29 ($n = 356$) as determined by qPCR analysis (data not shown), suggesting that the majority of these 356 'heterozygous' ESC clones were tetraploid (Supplementary Figure S5B and C). We next performed karyotyping of arbitrarily chosen ESC clones: three heterozygous mutants selected with only 200 μg/ml G418 (Heterozygote day 20); three heterozygous mutants selected with 200 μg/ml G418, followed by 300 μg/ml G418 (Heterozygote day 33); three homozygous mutants selected with 200 μg/ml G418, followed by 300 μg/ml G418 (Homozygote day 33); and two puromycin-sensitive ESC clones selected with only 200 μg/ml G418 (Sensitive day 20). One hundred metaphase spreads were analyzed for each clone (Figure 3A and Supplementary Figure S6). The analysis showed the population of tetraploid in heterozygous clones increased from 9–20% at day 20 to 23–57% at day 33, suggesting that the high dose of G418 (300 μg/ml) may have induced tetraploidy, possibly due to aberrant cytokinesis. In contrast, almost all ESCs were diploid in homozygous and puromycin-sensitive clone populations (Supplementary Figure S6). The initial configuration of the cassette conferred G418 resistance and then Cre induced conversion of the cassette to puromycin resistance. However, G418 resistance continued immediately after the conversion. Thus, such a high dose of G418 was chosen to exclude cells bearing a single copy of the selection cassette. Similar high dose G418 has been used to select bi-allelically mutagenized diploid ESCs (3). Overall, we obtained 222 diploid homozygous mutant ESC clones from the 2208 haploid ESC clones mutagenized by the BPATrap vector (Supplementary Table S4).

To exclude these 'heterozygous' clone populations, we developed a new procedure in which haploid ESCs were purified by cell sorting after transposon transfection (Figure 3B, right of top). Among the 96 ESC clones (plate number H24), 79 were categorized as a single locus insertion by direct PCR seq and all of them survived the double selection. The genomic insertion sites of the BPATrap vectors were analyzed by splinkerette PCR seq. As a result, 65 clones were reconfirmed as a single locus insertion. qPCR analysis revealed that 58 homozygous mutants and two heterozygous mutants were obtained. The clones were repeatedly analyzed in this procedure and the number of independent clones was 27 (asterisk in Figure 3B and Supplementary Table S5). Thus, this procedure improved the efficiency of homozygous mutant isolation from 10% (222/2208) to 28% (27/96).

Validation of gene trapping

Psph and *Zic3 genes* are well expressed in ESCs. Revertant clones were obtained with expression of Flpo, and the amount of each protein was evaluated by western blotting (Figure 4A and B). Although we could not determine whether faint bands from trapping clones were due to either contamination of feeder cells or incomplete knockout, gene trap events were essentially successful.

DISCUSSION

In this study, we developed a new method to enrich homozygous mutant ESC clones from a pool of haploid polyA trap mutant ESCs by exploiting the inherent property of autodiploidization in haploid ESCs. Key features of this method include a barcode sequencing-based screening system to select ESCs with a single locus insertion of the differentially tagged polyA trap vector and a double resistance selection system to enrich autodiploidized ESCs. Using this approach, we successfully generated homozygous mutant mouse ESCs that so far include 222 different gene mutant clones.

Establishment of the *Rosa26*^{E2iCreE2} haploid mouse ESC line would be valuable to examine gene functions in not only ESCs but also differentiated cells and organs. Conditional alleles are first introduced at the haploid stage of ESCs. The timing of analyses either during or after ESC differentiation could be controlled by addition of 4-HT.

We previously reported that bi-allelically mutagenized diploid ESCs could be generated under the Bloom gene-deficient condition (3). These mutant cells were identified by the positive selection cassette after Cre-mediated conversion, which was similar to the cassette in this report. Conversion of mono- to bi-allelic mutation was not very high in the previous report. However, the conversion was spontaneous in this report, so the overall procedure was much simpler.

Prior to transfection with the BPATrap vector, haploid ESCs had been enriched by fluorescence-activated cell sorting (FACS), so that ∼90% of cells were haploid and 10% were diploid (11). Accordingly, both haploid (90%) and diploid (10%) ESCs were initially transfected with the vector. The barcode-based screening showed that 1369 out of the 2208 isolated clones had a single locus insertion of the vector, which could be enriched using retroviral or lentiviral

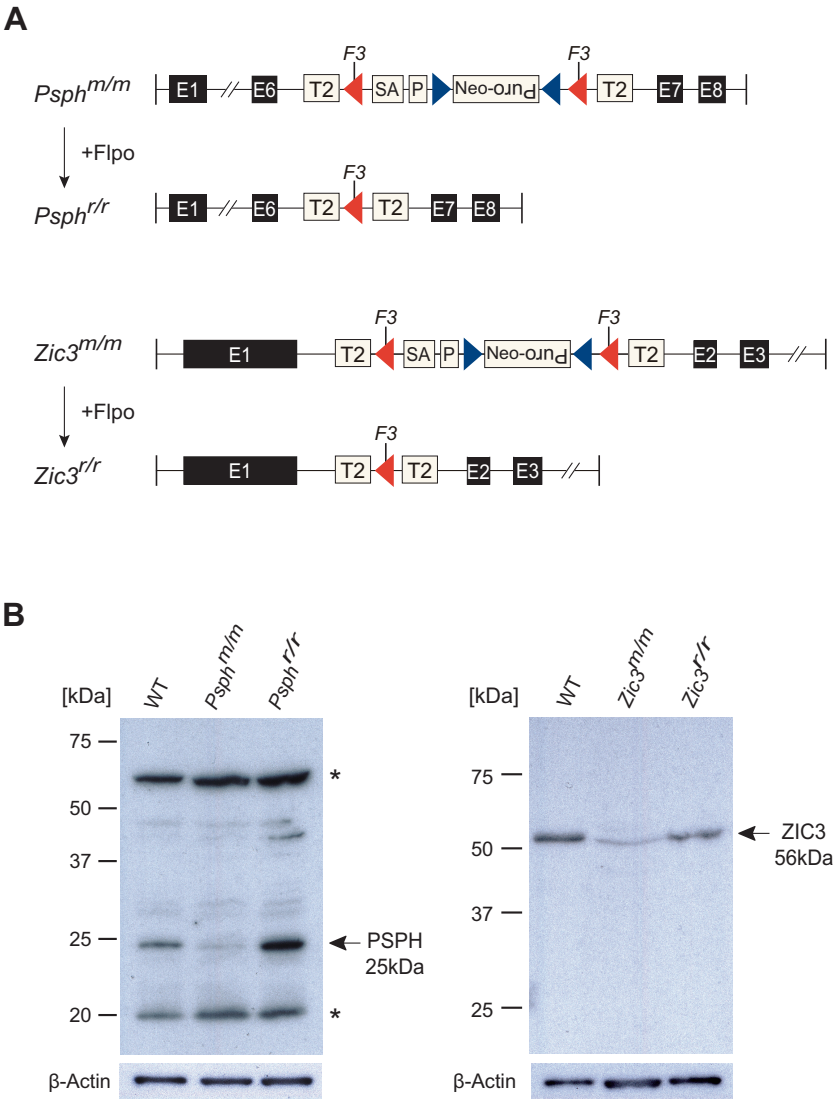

Figure 4. Validation of gene trapping. (**A**) Schematic of homozygous mutants with BPATrap insertions (*Psph*$^{m/m}$ *and Zic3*$^{m/m}$). Each revertant was isolated by removing the mutagenic vector sequence through Flpo-F3 recombination (*Psph*$^{r/r}$ *and Zic3*$^{r/r}$). m, mutant allele; r, revertant allele. (**B**) Western blot analysis of homozygous mutant clones and their revertant clones. β-Actin was used as an internal control. Asterisk indicates nonspecific bands. WT, wild-type of H129-2 haploid mouse ESCs.

vectors for transfection. ESC clones with a single locus vector insertion were then isolated and individually subjected to the double selection assay to select diploidized clones (Figure 3A). Karyotyping showed that, until the beginning of 4-HT administration and double selection, no residual haploid cells were detectable in the cell population (Supplementary Figure S6, clones H16_B02 and H16_F04), indicating that almost all haploid ESCs had been autodiploidized by that time. Out of the 1369 assayed clones, 405 clones (30%) were sensitive to the double selection, suggesting that they were diploid prior to transfection of the vector (Figure 3A). This discrepancy between the estimated initial ratios of the haploid cell population (10 versus 31%) could be the result of the difference in vector integration efficiencies between haploid and diploid ESCs. The efficiency of vector integration may be higher in diploid ESCs than in haploid ESCs, as described previously (14).

Despite screening haploid ESCs with a single locus vector insertion and the double resistance selection for subsequently autodiploidized ESCs, splinkerette PCR-based determination of vector insertion sites revealed 246 clones containing multiple locus insertions of identically tagged BPATrap vectors (Figure 3A and Supplementary Figure S4A). The number of such clones was unexpectedly high. A possible explanation for this result is that more than two copies of the identically tagged BPATrap vector were inserted independently into different genomic loci within a host cell, although an equimolar mixture of 15 differentially tagged BPATrap vectors was used for the transfection. Alternatively, a copy of the *Tol2*-based BPATrap vector integrated into a diploidized ESC could undergo a second transposition in the 4N phase of the cell cycle, resulting in two copy insertion of the vector in different loci of its daughter cells. This second possibility strongly suggested

that two integrations had occurred in the same chromosome (Supplementary Table S3), which is the hallmark of 'local hopping'.

Among the 610 clones in which single locus vector insertions were re-confirmed by qPCR-based genotyping, 356 clones (58.4%) were heterozygous for BPATrap insertional mutations, which was an unexpectedly high number (Figure 3A). This result could be an artifact produced by contamination of feeder cell-derived genomic DNA in the qPCR template. However, it is unlikely because we used puromycin-sensitive MEFs as feeders that should be eliminated during the selective culture. As shown in Supplementary Figure S6, karyotyping suggested that a substantial proportion of the 'heterozygous' cells were tetraploid. Therefore, autotetraploidization could be a major factor limiting this autodiploidization-based screening approach. Pre-removal of tetraploid ESCs by FACS before the double selection step might improve the selection efficiency. Exclusion of 'heterozygous' cells was partially successful by cell sorting after the transposon transfection (Figure 3B). Because we repeatedly obtained the same clones, we should increase the starting ESCs for generation of a large number of mutant clones.

The emergence of clustered regularly interspaced short palindromic repeats (CRISPR) technology has changed the landscape of genome engineering dramatically (25). The CRISPR system is effective enough to introduce biallelic mutations in diploid cells, so that genotype–phenotype correlations can be examined directly at the cellular level (26–28). A major concern of the CRISPR system is potential off-target effects, although improved Cas9 enzymes with reduced off-target effects have been engineered recently (29,30). Thus, the CRISPR system requires validation experiments. In contrast, the BPATrap system allows the generation of revertant cells by Flpo-induced removal of gene trap cassettes to provide strong experimental evidence of genotype–phenotype correlations. In this study, we were able to isolate revertant cells, thus we could simply compare a mutant cell with a revertant cell (Figure 4). Moreover, the CRISPR-based screening approach usually requires *a priori* information about targeted genes to design guide RNAs. Although tiled CRISPR libraries can identify regulatory elements without *a priori* information, these analyses are focused on the targeting regions of the genome, not genome-wide (31,32). In contrast, a gene trap screening approach such as BPATrap targets not only previously annotated genes but also any unknown genetic components, allowing a pure forward genetic approach in a random and genome-wide fashion. In fact, 38% (95/249) of the homologous mutants were inserted into unannotated genes, so we could find new genes with novel functions (Supplementary Tables S4 and 5). Taken together, although still at a limited scale, our BPATrap-based mutagenesis method and its derived mutant ESCs may provide a valuable platform to explore recessive genomic traits in the mammalian genome. In fact, the utility of a haploid ESC bank has been recently shown to clarify novel gene functions by genome-wide screening (33).

DATA AVAILABILITY

GenBank/EMBL/DDBJ (Accession Number: 6648).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Martin Leeb (Max F. Perutz Laboratories University of Vienna, Austria) for providing H129-2 haploid ESC lines, Yasushi Okada (Laboratory for Cell Polarity Regulation, Quantitative Biology Center RIKEN, Japan) for providing ZFN mRNA, Koichi Kawakami (National Institute of Genetics, Japan) for providing pCAGGS-mT2TP, Mitchell Arico from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript and Arihiro Kohara (National Institute of Biomedical Innovation, Health and Nutrition, Japan) for depositing and distributing homozygous mutant ESCs.

FUNDING

Research on Development of New Drugs from the Japan Agency for Medical Research and Development; Japan Society for the Promotion of Science KAKENHI [25290034]. Funding for open access charge: Japan Society for the Promotion of Science KAKENHI [25290034].
Conflict of interest statement. None declared.

REFERENCES

1. Yusa,K., Horie,K., Kondoh,G., Kouno,M., Maeda,Y., Kinoshita,T. and Takeda,J. (2004) Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature*, **429**, 896–899.
2. Guo,G., Wang,W. and Bradley,A. (2004) Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature*, **429**, 891–895.
3. Horie,K., Kokubu,C., Yoshida,J., Akagi,K., Isotani,A., Oshitani,A., Yusa,K., Ikeda,R., Huang,Y., Bradley,A. *et al.* (2011) A homozygous mutant embryonic stem cell bank applicable for phenotype-driven genetic screening. *Nat. Methods*, **8**, 1071–1077.
4. Elling,U., Taubenschmid,J., Wirnsberger,G., O'Malley,R., Demers,S.P., Vanhaelen,Q., Shukalyuk,A.I., Schmauss,G., Schramek,D., Schnuetgen,F. *et al.* (2011) Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell*, **9**, 563–574.
5. Leeb,M. and Wutz,A. (2011) Derivation of haploid embryonic stem cells from mouse embryos. *Nature*, **479**, 131–134.
6. Kotecki,M., Reddy,P.S. and Cochran,B.H. (1999) Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.*, **252**, 273–280.
7. Carette,J.E., Guimaraes,C.P., Varadarajan,M., Park,A.S., Wuethrich,I., Godarova,A., Kotecki,M., Cochran,B.H., Spooner,E., Ploegh,H.L. *et al.* (2009) Haploid genetic screens in human cells identify host factors used by pathogens. *Science*, **326**, 1231–1235.
8. Pettitt,S.J., Rehman,F.L., Bajrami,I., Brough,R., Wallberg,F., Kozarewa,I., Fenwick,K., Assiotis,I., Chen,L., Campbell,J. *et al.* (2013) A genetic screen using the PiggyBac transposon in haploid cells identifies Parp1 as a mediator of olaparib toxicity. *PLoS One*, **8**, e61520.
9. Leeb,M., Dietmann,S., Paramor,M., Niwa,H. and Smith,A. (2014) Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell*, **14**, 385–393.
10. Kokubu,C. and Takeda,J. (2014) When half is better than the whole: advances in haploid embryonic stem cell technology. *Cell Stem Cell*, **14**, 265–267.
11. Tokunaga,M., Kokubu,C., Maeda,Y., Sese,J., Horie,K., Sugimoto,N., Kinoshita,T., Yusa,K. and Takeda,J. (2014) Simulation and estimation of gene number in a biological pathway using almost complete saturation mutagenesis screening of haploid mouse cells. *BMC Genomics*, **15**, 1016.

12. Monfort,A., Di Minin,G., Postlmayr,A., Freimann,R., Arieti,F., Thore,S. and Wutz,A. (2015) Identification of Spen as a crucial factor for Xist function through forward genetic screening in haploid embryonic stem cells. *Cell Rep.*, **12**, 554–561.
13. Bai,M., Wu,Y. and Li,J. (2016) Generation and application of mammalian haploid embryonic stem cells. *J. Intern. Med.*, **280**, 236–245.
14. Pettitt,S.J., Tan,E.P. and Yusa,K. (2015) piggyBac transposon-based insertional mutagenesis in mouse haploid embryonic stem cells. *Methods Mol. Biol.*, **1239**, 15–28.
15. 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.
16. 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.
17. Zambrowicz,B.P., Friedrich,G.A., Buxton,E.C., Lilleberg,S.L., Person,C. and Sands,A.T. (1998) Disruption and sequence identification of 2, 000 genes in mouse embryonic stem cells. *Nature*, **392**, 608–611.
18. 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.
19. 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. U.S.A.*, **101**, 4170–4174.
20. Mayasari,N.I., Mukougawa,K., Shigeoka,T., Kawakami,K., Kawaichi,M. and Ishida,Y. (2012) Mixture of differentially tagged Tol2 transposons accelerates conditional disruption of a broad spectrum of genes in mouse embryonic stem cells. *Nucleic Acids Res.*, **40**, e97.
21. Yang,H., Shi,L., Wang,B.A., Liang,D., Zhong,C., Liu,W., Nie,Y., Liu,J., Zhao,J., Gao,X. *et al.* (2012) Generation of genetically modified mice by oocyte injection of androgenetic haploid embryonic stem cells. *Cell*, **149**, 605–617.
22. Li,W., Shuai,L., Wan,H., Dong,M., Wang,M., Sang,L., Feng,C., Luo,G.Z., Li,T., Li,X. *et al.* (2012) Androgenetic haploid embryonic stem cells produce live transgenic mice. *Nature*, **490**, 407–411.
23. 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.
24. Urasaki,A., Asakawa,K. and Kawakami,K. (2008) Efficient transposition of the Tol2 transposable element from a single-copy donor in zebrafish. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 19827–19832.
25. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
26. Koike-Yusa,H., Li,Y., Tan,E.P., Velasco-Herrera Mdel,C. and Yusa,K. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.*, **32**, 267–273.
27. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
28. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
29. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Joung,J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
30. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.
31. Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 192–198.
32. Simeonov,D.R., Gowen,B.G., Boontanrart,M., Roth,T.L., Gagnon,J.D., Mumbach,M.R., Satpathy,A.T., Lee,Y., Bray,N.L., Chan,A.Y. *et al.* (2017) Discovery of stimulation-responsive immune enhancers with CRISPR activation. *Nature*, **549**, 111–115.
33. Elling,U., Wimmer,R.A., Leibbrandt,A., Burkard,T., Michlits,G., Leopoldi,A., Micheler,T., Abdeen,D., Zhuk,S., Aspalter,I.M. *et al.* (2017) A reversible haploid mouse embryonic stem cell biobank resource for functional genomics. *Nature*, **550**, 114–118.