Development 140, 4287-4295 (2013) doi:10.1242/dev.096081

Efficient site-specific transgenesis and enhancer activity tests in medaka using PhiC31 integrase

Stephan Kirchmaier[1,*], Burkhard Höckendorf[1,2,*], Eva Katharina Möller[1], Dorothee Bornhorst[1], Francois Spitz[3] and Joachim Wittbrodt[1,2,‡]

SUMMARY

Established transgenesis methods for fish model systems allow efficient genomic integration of transgenes. However, thus far a way of controlling copy number and integration sites has not been available, leading to variable transgene expression caused by position effects. The integration of transgenes at predefined genomic positions enables the direct comparison of different transgenes, thereby improving time and cost efficiency. Here, we report an efficient PhiC31-based site-specific transgenesis system for medaka. This system includes features that allow the pre-selection of successfully targeted integrations early on in the injected generation. Pre-selected embryos transmit the correctly integrated transgene through the germline with high efficiency. The landing site design enables a variety of applications, such as reporter and enhancer switch, in addition to the integration of any insert. Importantly, this allows assaying of enhancer activity in a site-specific manner without requiring germline transmission, thus speeding up large-scale analyses of regulatory elements.

KEY WORDS: PhiC31 integrase, Fish, Regulatory DNA, *In vivo* analysis, Position effects

INTRODUCTION

There is a general interest in identifying putative regulatory elements on a large scale (e.g. Bernstein et al., 2012). Thus, there is an immediate need for techniques that allow the reliable characterization and validation of newly discovered regulatory elements. Fish model systems are particularly appealing for the study of regulatory elements, because the expression of fluorescent reporters can be scored with high resolution *in vivo* and throughout embryogenesis (Bernstein et al., 2012; Eichenlaub and Ettwiller, 2011; Ellingsen et al., 2005; Mongin et al., 2011; Ritter et al., 2012; Shakes et al., 2012). Current strategies for analyzing regulatory activity in fish build upon transgenesis tools that integrate an uncontrolled number of transgene copies into random genomic sites (e.g. Mongin et al., 2011). As the vast majority of the genome has gene regulatory properties (Bernstein et al., 2012), position effects and copy number variations can considerably interfere with transgene expression. Consequently, transgenic approaches that assess the gene regulatory potential of candidate sequences require the comparison of multiple independent lines. Reporter expression patterns observed in the majority of animals with independent transgenic insertions are defined as specific for a given enhancer. This is a powerful strategy for determining the consensus enhancer activity displayed by a regulatory element in most genomic loci. However, it has several limitations. First, it requires costly efforts to generate and maintain the multiple lines needed for each transgene of interest. Second, it does not allow precise and unambiguous comparisons of regulatory elements, a feature especially important for the study of polymorphic regulatory variants (Maurano et al., 2012; Pastinen and Hudson, 2004).

An obstacle to the detailed locus-specific analysis of cis-regulatory elements has been the technical requirement to integrate single copy transgenes into predefined genomic loci. Targeting loci by homologous recombination is technically demanding and, furthermore, unavailable for many model organisms. The streptomyces phage PhiC31 integrase has been used to circumvent this problem. It catalyzes the specific unidirectional recombination between the attachment motifs attB and attP and has been used in several invertebrate (Groth et al., 2004; Venken et al., 2006; Venken et al., 2011; Yonemura et al., 2012), vertebrate (Allen and Weeks, 2005; Belteki et al., 2003; Groth et al., 2000) and plant (Kapusi et al., 2012; Thomson et al., 2010) species.

In zebrafish, enzymatic activity of PhiC31 integrase in somatic and germline cells has been demonstrated (Hu et al., 2011; Lister, 2010; Lu et al., 2011). Recently, a site-specific transgenesis system with characterized landing sites has been reported for zebrafish (Mosimann et al., 2013). Here, we present a PhiC31 integrase-based system to mediate site-specific transgenesis and to allow molecular complementation for the analysis of regulatory elements in medaka. We established transgenic lines each harboring a single landing site and we demonstrate efficient integration and germline transmission of targeting vectors. The system employs integration sensors for efficient pre-selection of putative transgenics. Furthermore, we established a new assay for rapid testing of enhancer and promoter elements in a fully comparable genomic setting.

MATERIALS AND METHODS

Fish maintenance

Medaka (*Oryzias latipes*) stocks were maintained as previously described (Koster et al., 1997). Lines used in the study were generated in the medaka wild-type Cab background. All fish are housed in the fish facility built according to the local animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr. 1) and in accordance with European Union animal welfare guidelines. The facility is under the supervision of the local representative of the animal welfare agency.

[1]Centre for Organismal Studies, Heidelberg University, 69120 Heidelberg, Germany. [2]Institute for Toxicology and Genetics, KIT, Karlsruhe, Germany. [3]Developmental Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany.

*These authors contributed equally to this work
‡Author for correspondence (jochen.wittbrodt@cos.uni-heidelberg.de)

Accepted 29 July 2013

Molecular biological reagents

The landing site vector and TagRFP-T targeting vector were designed with Geneious version 5 and 6 (Biomatters) and synthesized by GeneArt (Life Technologies). Utilized attB and attP sequences are listed in supplementary material Table S2. Subsequent versions of the targeting vector are derivatives of the synthesized version. SB100x (Mátés et al., 2009), PhiC31o (Raymond and Soriano, 2007) and CreNLS mRNA was prepared using the SP6 mMessage mMachine Kit (Ambion). Enhancers are derived from published sequences (Bernstein et al., 2012; Eichenlaub and Ettwiller, 2011; Visel et al., 2009a). The concatemeric version of the Fam44b enhancer was cloned using the Golden GATEway cloning kit (S.K., K. Lust and J.W., unpublished).

Embryo injection

One-cell-stage medaka embryos were injected as described (Rembold et al., 2006). We co-injected Sleeping Beauty 100x mRNA (7 ng/μl) with the landing site vector (10 ng/μl). For the PhiC31 targeting reaction, we co-injected 10 ng/μl targeting vector DNA together with 8 ng/μl PhiC31o mRNA. For the Cre-mediated cleanup, CreNLS mRNA was injected (10 ng/μl) in 1× Yamamoto buffer. Nucleic acid concentrations were determined using Nanodrop (Peqman; Thermo Scientific).

Generation of transgenic PhiC31 landing site lines

Embryos injected for the generation of landing site lines were screened for enhanced green fluorescent protein (EGFP) expression in the heart muscle and raised to maturity. We genotyped single fish via fin-clip PCR to identify insertion site and copy number (Grabher and Wittbrodt, 2007) after the third outcross generation. For fin clip, fish were anesthetized with tricaine and a small fin biopsy was taken for DNA extraction using the innuPREP DNA Mini Kit (Analytik Jena). The insertion sites were determined using linker-mediated PCR (Wu et al., 2003). Targeted loci were identified by sequence analysis blasting the medaka genome (Assembly HdrR, Version 70.1). The loci were verified using standard genotyping PCR with primers homologous to sequences flanking the mapped site. The presence of single copy landing sites was determined by 50% Mendelian segregation of the landing site in an outcross and by Southern blot analysis.

Southern blot analysis

Tissue of defined single landing site fish was digested overnight at 37°C in Ten9 buffer (100 mM Tris HCl pH 8.5, 10 mM EDTA, 200 mM NaCl, 1% SDS) with 0.9 mg/ml Proteinase K (Roche). The DNA was subsequently purified by phenol chloroform isoamyl alcohol (PCI) extraction. Genomic DNA (13 μg) was digested overnight using *Nco*I (Thermo Fisher) in a volume of 30 μl. *Nco*I cuts once in the landing site within the polyA sequence of the selection cassette. A digoxygenin (DIG)-labeled DNA probe against the EGFP gene was generated using a PCR DIG labeling kit (Roche) according to the manufacturer's instruction. A standard Southern blot procedure was followed using CDP Star detection (Roche) with AP-conjugated anti-Dig Fab fragments (Roche). Hybond-N+ membranes (GE Healthcare) were used and the staining was developed using an Intas ChemoCam Imager.

Ligation-mediated (LM) PCR to map single insertions

Genomic DNA from fin clips was extracted using the innuPrep DNA Mini Kit (Analytik Jena). Genomic DNA (1 μg) was digested using 10 U Hsp92II (Promega) at 37°C for 4 hours. The restriction enzyme was inactivated at 80°C for 20 minutes. The resulting genomic fragments were then ligated to a linker cassette corresponding to the overhangs created by digestion. Two different linker cassettes were used: linker A (oligonucleotides 5′-us-Hsp92II and 5′-ls) and linker B (oligonucleotides 3′-us-Hsp92II and 3′-ls). To generate the linker cassettes, complementary oligonucleotides were annealed as follows: 25 μl 100 μM upper strand (us), 25 μl 100 μM lower strand (ls) and 0.5 μl 5 M NaCl were mixed, denatured for 5 minutes at 95°C and then placed at room temperature for 5 minutes. The linker cassettes were stored at −20°C. Two separate ligation reactions were performed for every sample, one reaction for each linker cassette using 15 U T4 ligase (30 U T4 ligase HC from Thermo Scientific). The ligation reactions were incubated at 16°C overnight. Two PCR reactions were performed for each sample. Each PCR used one primer specific to one end of the PhiC31-docking site (SB recognition sites) and the other primer specific to one of the linker cassettes. For samples with linker A primers LM_SB-left2 and oLM-5rev were used and for samples with linker B the primers LM_SB-right1 and oLM-3rev were used. PCRs were performed with 1.25 U Platinum TAQ polymerase (Life Technologies). One microliter of the primary PCR was used as template for the nested PCR. The nested primers LM_SB-left1 and oLM-5rev-nest were used for samples with linker A and the nested primers LM_SB-right2 and oLM-3rev-nest for samples with linker B. Primer sequences are provided in supplementary material Table S1. The resulting bands were sequenced directly or gel-purified and subcloned using the pGem-T-easy vector system (Promega) as per the manufacturer's instructions. DNA was sequenced using Sanger sequencing (MWG Eurofins) and sequences were analyzed using Geneious Software (Biomatters).

Heatshock to activate *hsp70* promoter

Fish were kept at 28°C. For the heatshock, embryos were transferred to 37°C for 30 minutes in a low volume of embryo-rearing medium (1 g/l NaCl, 30 mg/l KCl, 40 mg/l $CaCl_2 \cdot 2H_2O$, 163 mg/l $MgSO_4 \cdot 7H_2O$, adjusted to pH 6.0, afterwards addition of 170 ml/l 1 M HEPES, pH 7.3). Subsequently, 42°C embryo-rearing medium was added and the embryos were kept at 37°C for 2 hours. Afterwards, the fish were transferred back to 28°C.

Imaging

We used the Olympus MVX10 fluorescence-screening binocular to screen for transgenic landing site lines and to score the recombination efficiency. All images were acquired with a Nikon AZ100, with exception of Fig. 7D, which was created using a Leica TCS SPE confocal laser scanning microscope. The image consists of four stacks that were stitched using software described previously (Preibisch et al., 2009). Embryos were imaged in agarose-based injection molds *in vivo* (Rembold et al., 2006).

Reagent availability

Plasmids used in this study will be made available via Addgene. Lines 1, 2 and 3 are available upon request.

RESULTS

Design of the PhiC31 landing site

We designed a PhiC31 integrase-based transgenesis system consisting of two components (Fig. 1): the genomic landing site, in which EGFP is driven in the heart muscle by the zebrafish *cardiac myosin light chain 2* (*cmlc2*; also known as *myl7*) promoter (Auman et al., 2007; Huang et al., 2003); and the targeting vector, containing an insert of interest and an additional TagRFP-T marker that lacks both promoter and start codon. The PhiC31 attP (landing site) and attB (targeting vector) sites (Rausch and Lehmann, 1991) (supplementary material Table S2) were designed and positioned such that PhiC31-mediated recombination displaces the EGFP from the *cmlc2* promoter and corresponding start codon by the in-frame insertion of TagRFP-T encoded in the targeting vector (Fig. 1C). A fluorescence color switch in the myocardium from EGFP to TagRFP-T thus indicates proper vector integration. Alternatively, the displaced EGFP of the genomic landing site may be placed under transcriptional control of any selected enhancer/promoter elements provided with the targeting vector (Figs 5, 6). Both the landing site and the targeting vector contain heterotypic loxP and lox2272 sites (Branda and Dymecki, 2004) to facilitate Cre-mediated excision of vector and reporter sequences (Fig. 1D). To establish randomly placed, single-copy landing sites, we used the well-characterized Sleeping Beauty transposon system 100x (SB100x) (Grabher and Wittbrodt, 2007; Ivics et al., 1997; Mátés et al., 2009).

Establishment of defined transgenic landing site lines for site-specific transgenesis

We selected and further characterized landing site lines that are well suited for transgenesis and the characterization of regulatory

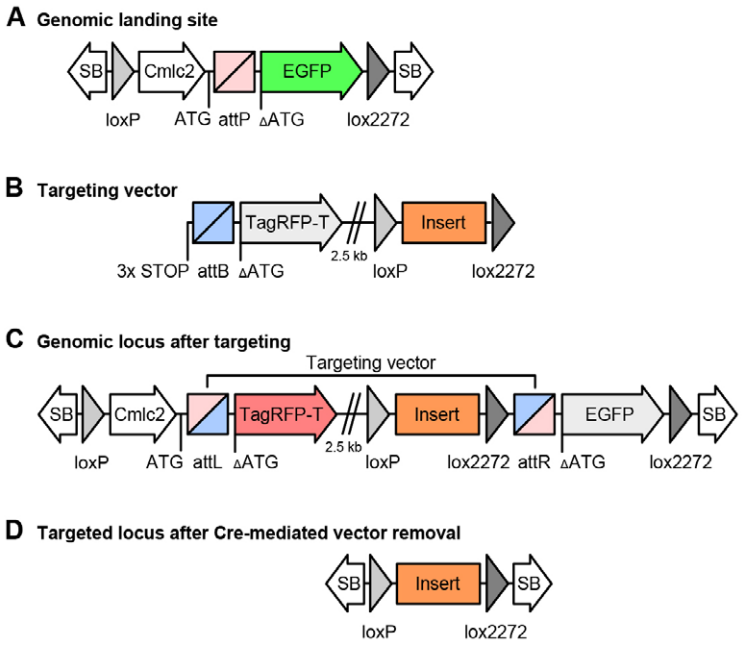

Fig. 1. Design of the PhiC31 transgenesis system. (**A**) Landing site with *cmlc2* promoter (Auman et al., 2007; Huang et al., 2003) driving EGFP in the myocardium. The EGFP open reading frame (ORF) contains an in-frame attP site. This cassette is flanked by heterotypic lox sites (loxP and lox2272) as well as Sleeping Beauty transposon IR/DRs (SB). (**B**) The targeting vector contains an attB site and heterotypic lox sites corresponding to the respective sites of the landing site. Additionally, the vector contains a TagRFP-T coding sequence lacking both promoter and start codon. This coding sequence is preceded by the attB site. To prevent unscheduled expression of TagRFP-T by protein trapping, the attB site is furthermore preceded by stop codons in all forward reading frames. (**C**) Site-specific recombination between attP and attB (colored triangles/squares) is initiated upon microinjection of targeting vector and PhiC31o integrase mRNA into a transgenic line containing a genomic landing site. Specific targeting of the landing site is reported by targeting sensors that are activated through molecular complementation of targeting vector and landing site. Site-specific integration disrupts the attP-EGFP ORF of the landing site by replacing the EGFP with the TagRFP-T of the integration vector. This displacement occurs in-frame, thereby generating functional *cmlc2*::attL-TagRFP-T resulting in a fluorescence color switch in the myocardium. (**D**) The targeted genomic landing site initially contains the entire targeting vector plasmid (C). The complementation of heterotypic lox sites of targeting vector and landing site allows the removal of these undesired sequences through Cre-mediated locus cleanup.

elements. These lines carry a unique, intergenic landing site (Fig. 2A). Single-copy integration was confirmed by mapping, Mendelian segregation and Southern blot analysis (Fig. 2D). The integrations are located on chromosome 18 (line 1) and the so far unassembled scaffold 1518 (line 2). The closest annotated genes to landing site 1 are at a distance of 78 kb and 89 kb, respectively (including gaps in the assembly). No genes are annotated on scaffold 1518 (total size 29 kb). Thus, the landing sites are located in gene deserts.

Next, we investigated whether the landing sites are influenced by activating or repressive regulatory elements in their respective genomic loci (i.e. position effects). For this, we employed the zebrafish *hsp70* promoter (Blechinger et al., 2002), which has a known sensitivity to regulatory influences and is therefore routinely used for heterologous enhancer tests (Bernstein et al., 2012; Eichenlaub and Ettwiller, 2011; Mongin et al., 2011). The *hsp70* promoter autonomously drives expression of basal EGFP levels in the lens. Figure 2B shows fish transgenic for a targeting vector containing the zebrafish *hsp70* promoter driving EGFP expression. Our data show that both landing sites in line 1 and line 2 are free of position effects detectable with the *hsp70* promoter.

The loci of landing site lines 1 and 2 are accessible for transcriptional activation throughout development. This is indicated by the activation of the *hsp70* promoter by a heatshock at different developmental stages. In both lines, this results in ubiquitous reporter gene expression. Figure 3 shows representative pictures from line 2. Taken together, these results show that the landing sites are located in genomic regions that are neutral for the *hsp70* promoter and fully accessible throughout development. Homozygous fish display no obvious phenotype concerning viability, health status and fecundity.

Germline transmission of the targeting vector

Specific integration of targeting vectors into genomic landing sites can be screened for by a number of targeting sensors. They are designed to yield specific fluorescence protein expression only upon proper integration of the targeting vector. The targeting sensors are used to efficiently preselect candidate founder fish for stable transgenic lines. We employed three distinct sensors: (1) displacement of the EGFP in the genomic landing site by a TagRFP-T from the targeting vector (Fig. 1C); (2) reactivation of EGFP in the genomic landing site by providing the appropriate promoter and start codon with the targeting vector (Fig. 5A; Fig. 6A); and (3) activation in heart muscle of the *hsp70* promoter by the *cmlc2* promoter (Fig. 4). This activation is very strong in comparison with the landing-site specific *cmlc2*::attP-EGFP transgene. The targeting efficiencies scored with either targeting sensor were fully comparable.

Targeting sensors enable the identification of transgenic founder fish from the injected generation. Co-injection of targeting vector and mRNA encoding PhiC31o recombinase resulted in site-specific

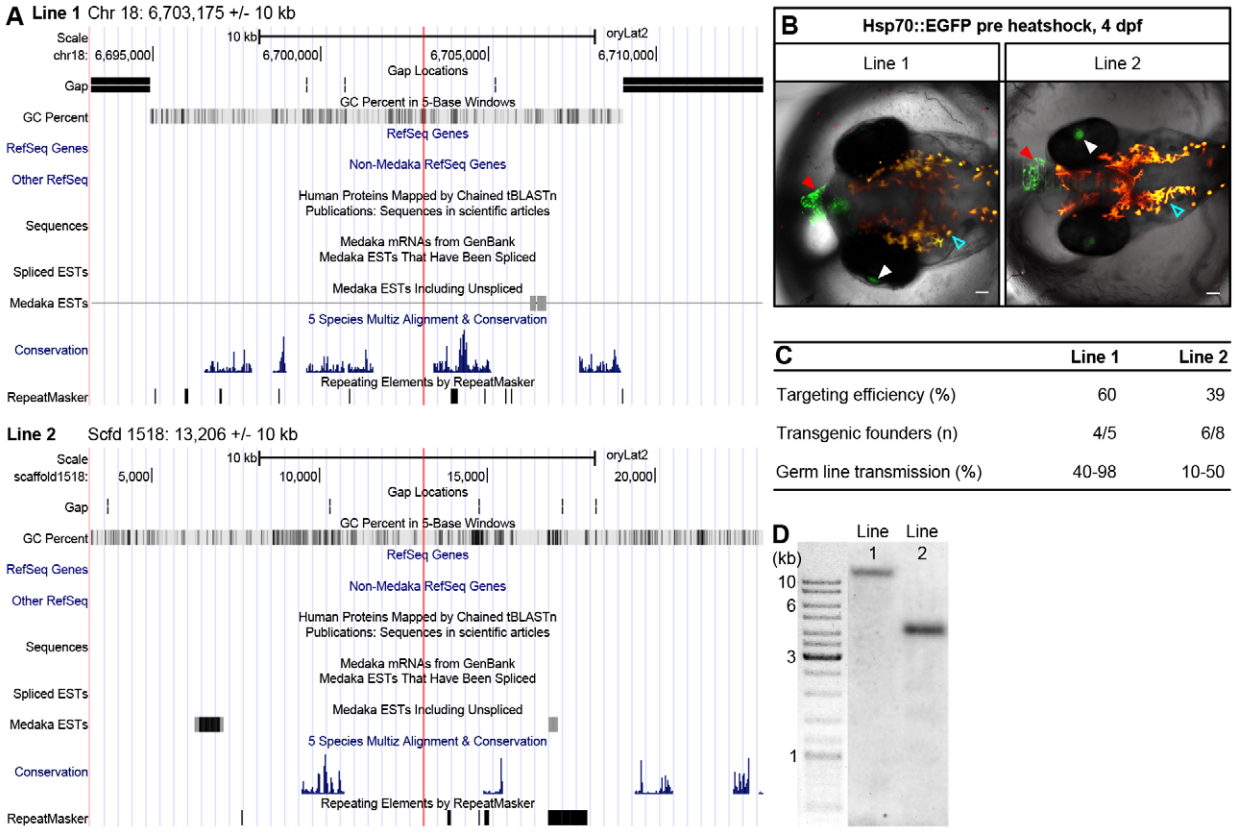

Fig. 2. Two genomic landing site lines. (**A**) Genomic locations of the landing sites of lines 1 (top) and 2 (bottom). Both sites are located in gene deserts. The closest neighboring genes for line 1 are a distance of 78 kb and 89 kb (including gaps in the assembly). No genes are annotated on scaffold 1518 that landing site 2 maps to (size of the scaffold is 29 kb). (**B**) Representative embryos stably targeted with *hsp70*::EGFP. Strong EGFP expression in lens (white arrowheads) reflects the autonomous activity of this promoter. EGFP expression in the myocardium (red arrowheads) occurs as the result of an interaction of *hsp70* with the *cmlc2* promoter in the targeting vector. No other EGFP fluorescence is detectable. Pigment cells are autofluorescent in both red and green color spectra (open cyan arrowheads; see also supplementary material Fig. S5). Scale bars: 50 μm. (**C**) Targeting efficiency and germline transmission of targeting vector. The targeting efficiency was calculated as the weighted average of three independent experiments, and expressed as percentage of the maximum 50% positive embryos in a heterozygous outcross. Using targeting sensors, candidate fish were preselected to establish stable transgenic lines. Four out of five (line 1) and six out of eight (line 2) of the candidates transmitted the targeting vector through the germline. (**D**) Southern blot analysis to confirm the presence of single copy genomic landing site integration in both lines. Genomic DNA was digested with an enzyme cutting once in the landing site (*NcoI*). An EGFP probe was used for hybridization. The ladder is the GeneRuler DNA Ladder Mix (Thermo Fisher).

vector targeting as validated by sequencing (supplementary material Fig. S1). We determined the efficiency of PhiC31-mediated targeting for each landing site line by counting successful targeting relative to all injected embryos that contain a landing site (Fig. 2C). The average targeting efficiencies in a heterozygous outcross are 60% and 39% for lines 1 and 2, respectively. We selected a number of candidate founder fish to establish transgenic lines. The majority of the candidates transmitted the targeting vector through the germline (Fig. 2C). This high efficiency enables the straightforward selection of stable transgenic founders from the injected generation using the targeting sensors. Importantly, we observed no toxicity of PhiC31 mRNA injection. Only between 2% and 5% of the injected embryos died or showed developmental defects. It is interesting to note that targeting efficiency and germline transmission rate vary for the different lines. Locus-specific effects on the efficiency of PhiC31-mediated recombination were previously reported in *Drosophila* (Bischof et al., 2007) and in cell lines (Thyagarajan et al., 2001).

For a site-specific transgenesis system, it is important to consider the possibility of integrations outside the landing side. Illegitimate integrations are indicated by the presence of transgene-specific expression features and absence of targeting sensors. For the *hsp70* promoter driving EGFP, illegitimate integrations are characterized by the absence of both TagRFP-T and strong EGFP expression in the heart. Illegitimate insertions of the targeting vector were restricted to a subset of founder fish derived from line 2 (three out of eight) and occur with low frequency. The features of our system allowed the analysis of correct transgene integrations within a pool of fish containing uncharacterized landing sites, yielding similar results. Nineteen out of 26 (65%) pre-selected fish transmitted the correctly integrated construct to the next generation (supplementary material Fig. S3).

The designed targeting sensors faithfully report correct integrations of the targeting vector into genomic landing sites. They enable a pre-selection of candidate founders for stable transgenic lines with high confidence, thereby reducing the number of animals required to raise a transgenic line. Taken together, this demonstrates a strong preference for specific targeting of the landing site and a high efficiency of germline transmission of the integrated vector with PhiC31 integrase in medaka.

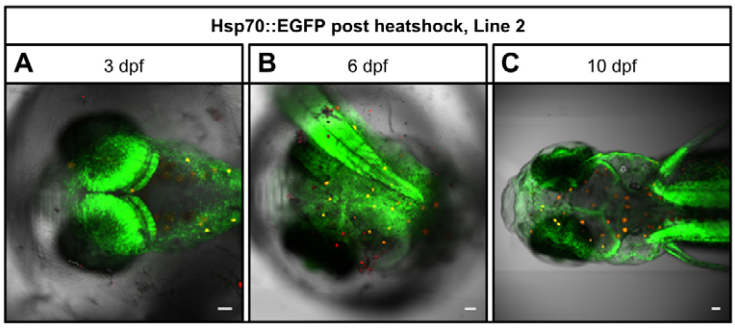

Fig. 3. Transcriptional accessibility of the landing site during development. (**A-C**) Landing site line 2, stably transgenic with *hsp70*::EGFP was heatshocked at different stages of development. At all tested stages, the heatshock resulted in ubiquitous EGFP expression (for pre-heatshock expression compare with Fig. 2B), highlighting full transcriptional accessibility of the locus during development. Identical results were obtained for line 1 (data not shown). Live embryos, imaged through the chorion (A,B) are shown; dorsal views, anterior to the left. Scale bars: 50 µm.

Cre-mediated locus cleanup

In order to visualize the locus cleanup, we employ a specific landing site line (line 3) that harbors two genomic landing sites at the *alg2* locus (Thiel et al., 2003) on chromosome 16 (Fig. 4). One of the two landing sites was mapped to exon 3 of the *alg2* gene. Targeting the landing sites using the vector containing *hsp70*::EGFP revealed a specific enhancer trap effect in the brain. This effect was only uncovered by the *hsp70* promoter. Notably, the observed position effect occurred reproducibly in all of the targeted fish from that line in different experiments (Fig. 6), demonstrating the capability of our system to place transgenes under a constant positional bias. Position effects are not unexpected in light of the density of regulatory elements in the genome (Bernstein et al., 2012) and the long-range action of enhancers (Ellingsen et al., 2005; Ruf et al., 2011).

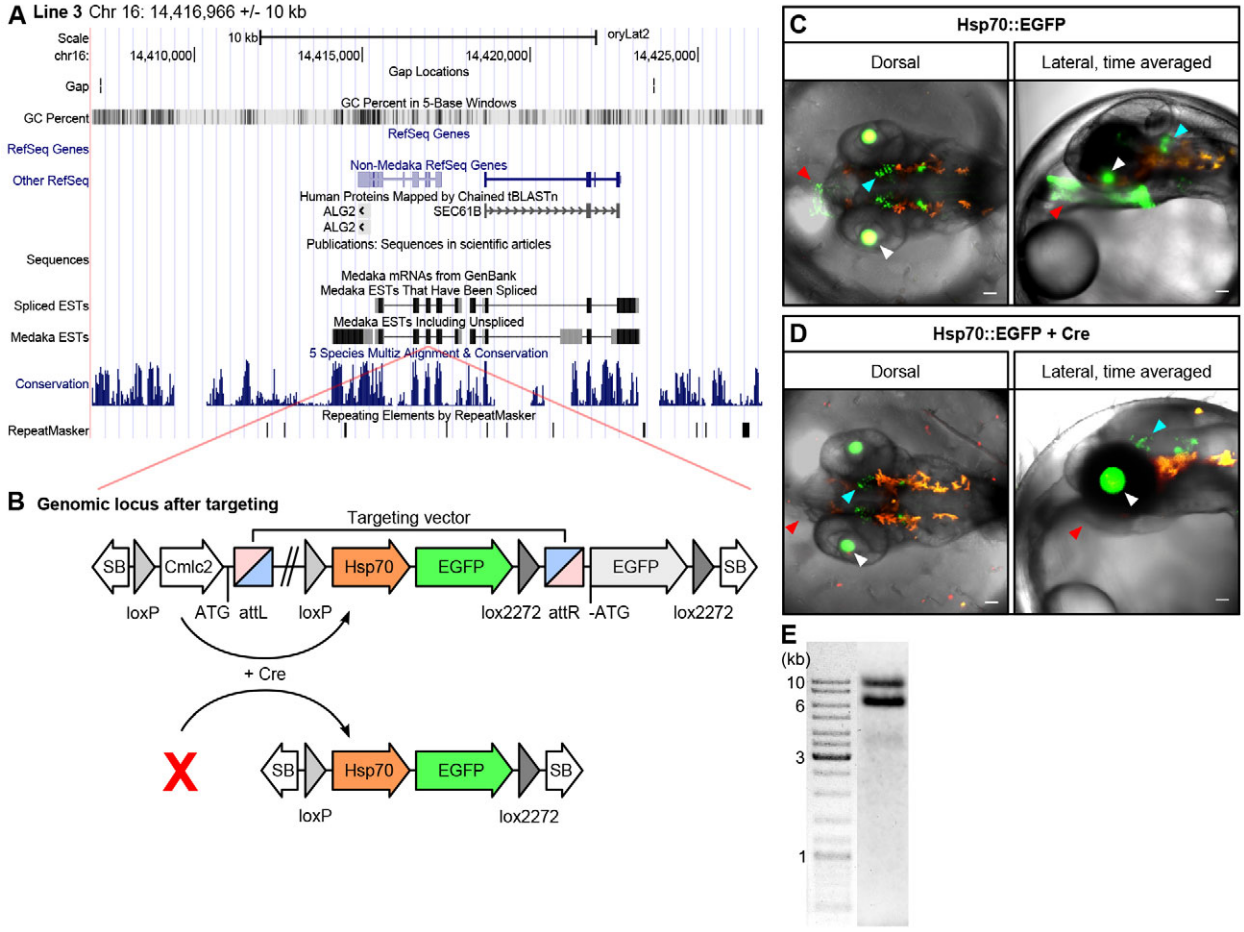

Fig. 4. Cre-mediated locus cleanup. (**A**) The *alg2* locus. Two landing site integrations are contained within the locus, one of them within exon 3. Note the high number of conserved non-coding sequences around the locus. (**B**) Genomic locus after targeting, before and after Cre-mediated removal of vector sequences. The Cre-cleanup removes the *cmlc2* promoter from the locus, thereby disrupting its interaction with the *hsp70* promoter. (**C**) Landing site line 3 stably transgenic for *hsp70*::EGFP; dorsal view (left) and time-averaged lateral view (right). In addition to the *hsp70*-autonomous lens expression (white arrowheads), and the heart expression resulting from its interaction with *cmlc2* (red arrowheads), an enhancer trap in the brain is apparent (cyan arrowheads). (**D**) Landing site line 3 stable transgenic for *hsp70*::EGFP after Cre-mediated locus cleanup. This results in the loss of EGFP fluorescence in the heart (red arrowheads) but does not affect the brain-specific enhancer trap (cyan arrowheads), demonstrating that it is the result of the regulatory landscape of the *alg2* locus. The *hsp70*-autonomous EGFP expression in the lens is retained (white arrowheads). (**E**) Southern blot analysis showing multiple landing site integrations in the line. Scale bars: 50 µm.

Our system design includes complementary heterotypic lox sites in the genomic landing site as well as in the targeting vector (Fig. 1C). Use of a Cre-mediated locus cleanup step will retain only the insert, a set of incompatible heterotypic lox sites and the Sleeping Beauty inverted repeat/direct repeats (IR/DRs; Fig. 1D). Upon injection of CreNLS mRNA, we efficiently removed different residual markers from the landing sites, including the *cmlc2* promoter, keeping the *hsp70*::EGFP transgene intact (confirmed sequence in supplementary material Fig. S2). Cleanup of the landing site is visualized by the loss of EGFP expression in the heart muscle, both transiently and in subsequent generations (Fig. 4C,D). By contrast, position-dependent EGFP expression in the brain persists after Cre-mediated locus cleanup, indicating distant interactions with regulatory elements specific for the *alg2* locus. Cre-mediated locus cleanup was performed with fish transgenic for the same construct in line 2 (supplementary material Fig. S4). We raised two fish of both lines injected with Cre recombinase and found 100% germline transmission of the cleaned locus in all tested fish.

Locus-specific testing of promoters and enhancers

Our system allows studying the activity of regulatory DNA in the context of specific loci. As both, readout and targeting sensor, we complement EGFP in the landing site with any promoter of interest inserted with the targeting vector (Figs 5, 6). This molecular complementation avoids background fluorescence caused by random insertion or episomal transcription. As indicated by the *hsp70*::EGFP transgene, the targeted loci of the lines 1 and 2 were accessible during development and do not show position effects. Therefore, in our system, the expression of the reporter gene located in a region neutral for the *hsp70* promoter solely depends on the regulatory input from DNA provided by the targeting construct.

In order to assess quickly the activity of regulatory elements, it is desirable to faithfully recapitulate specific expression patterns in the injected generation. The *hsp70* promoter (Blechinger et al., 2002) was used to assess the achievable uniformity in reporter gene expression after injection. We scored EGFP expression after heatshock in embryos at the ≤50% epiboly stage (Fig. 5B,C). This stage is amenable to determining the uniformity of expression owing to its simple morphology. Importantly, of all the injected fish that show EGFP fluorescence, 35% show highly homogenous expression. Additionally, this highlights that the function of regulatory DNA can be analyzed very early during development. After heatshock of injected embryos at 4 dpf, uniform EGFP expression was detected (Fig. 5D).

After this initial characterization of the assay, we tested the specific regulatory potential of promoter and enhancer elements. To test promoters, we chose the lens crystallin (Emelyanov and Parinov, 2008) and the sonic hedgehog (*shh*) promoter (Neumann

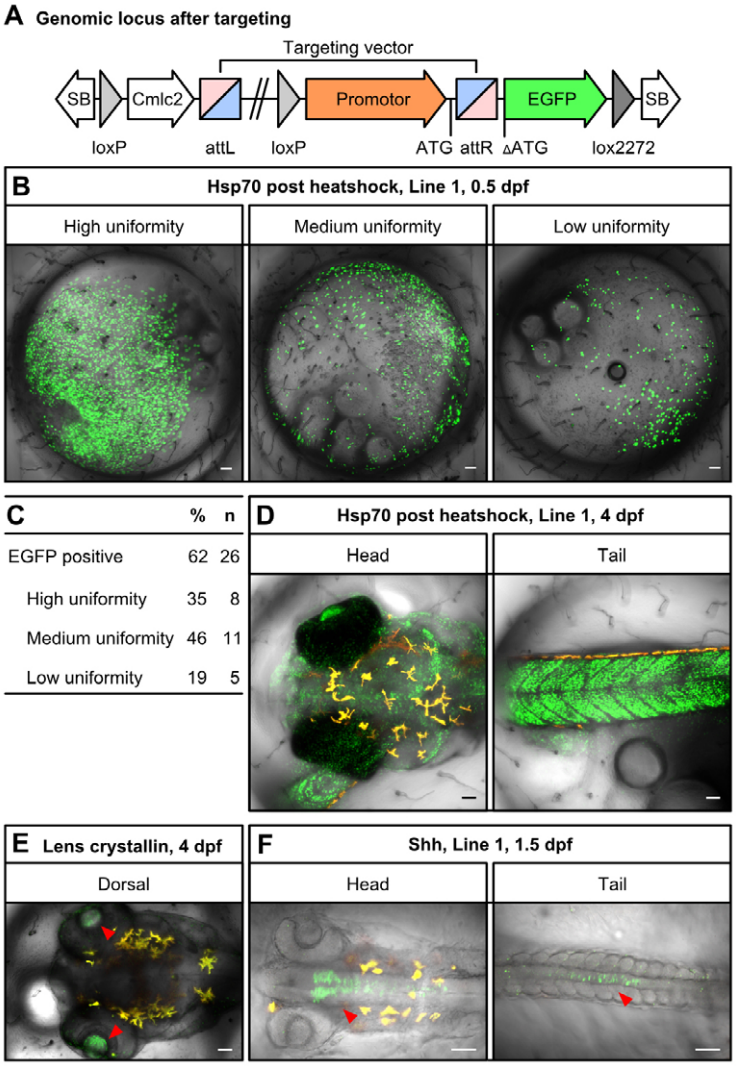

Fig. 5. Testing targeting efficiency by locus-specific analysis of promoters in the injected generation. (**A**) The genomic landing site after targeting and molecular complementation is shown. The promoter sequence provided with the targeting vector can drive the expression of an attR-EGFP open reading frame (ORF) only upon proper integration of the targeting vector. (**B-D**) For the determination of targeting efficiency after injection, the *hsp70* promoter was chosen. (B) Heatshock at ≤50% epiboly resulted in reporter gene expression patterns that fall into three groups with high, medium and low uniformity (quantification is shown in C). (D) A heatshock at 4 dpf triggers EGFP expression in the vast majority of cells, highlighting the high degree of uniformity achieved already in the injected generation. (**E**,**F**) Two other promoters, namely the lens crystallin (E, red arrowheads) (Emelyanov and Parinov, 2008) and the sonic hedgehog promoter (F, red arrowheads) (Neumann and Nuesslein-Volhard, 2000) were tested and showed reproducibly the expected expression pattern. Scale bars: 50 μm.

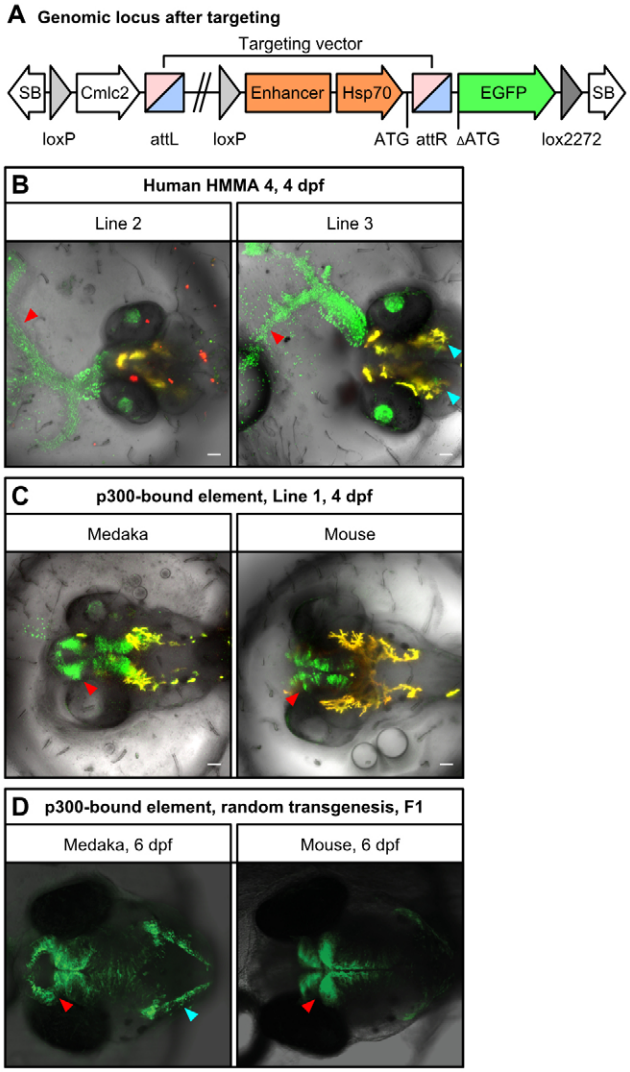

Fig. 6. Locus-specific analysis of enhancer sequences in the injected generation. (**A**) The targeting vector with the enhancer test setup after targeting and molecular complementation is depicted. Variable enhancer sequences are cloned upstream to the *hsp70* promoter and analyzed in the injected generation by imaging live embryos at 4 dpf (B,C). (**B**) The activity of the enhancer HMMA_4 is depicted (frontal view, ventral to the left). Reporter gene expression is detected in blood cells (red arrowheads). In line 3, additional activity from the characterized enhancer trap was detected in the brain (cyan arrowheads). (**C**) A mouse p300-bound element identified from mouse forebrain and the orthologous medaka sequence were tested. Both show specific reporter gene expression in the forebrain highly reminiscent to the stable lines (red arrowhead, compare with D). Dorsal view, anterior to the left. (**D**) Reporter gene expression patterns detected in a transgenic line generated with the same regulatory elements by a meganuclease-based approach (red arrowheads, compare with C). Additional expression in the hindbrain (cyan arrowhead) occurred at later stages (6 dpf, M. Eichenlaub, personal communication). Figures are taken from Eichenlaub and Ettwiller (Eichenlaub and Ettwiller, 2011). Scale bars: 50 μm.

and Nuesslein-Volhard, 2000). Upon injection of landing site lines 1 and 2, the lens crystallin promoter drives reporter gene expression solely in the lens (Fig. 5E). The shh promoter drives expression in the floorplate and the notochord (Fig. 5F). All of the tested promoters (*hsp70*, lens crystallin and *shh*) showed expression patterns consistent with the literature (Blechinger et al., 2002; Emelyanov and Parinov, 2008; Neumann and Nuesslein-Volhard, 2000). Therefore, we conclude that promoter activity can be readily identified in the injected generation.

Next, we tested whether our molecular complementation assay is applicable to enhancer analysis. Separate enhancer elements are cloned upstream of an *hsp70* promoter. In a meganuclease-based assay, the enhancer HMMA_4 (Bernstein et al., 2012) drives reporter gene expression in blood cells. Only blood cells are labeled in lines 1 and 2. Interestingly, when injected in line 3, the aforementioned enhancer trap is detectable along with reporter gene expression in blood cells (Fig. 6). Additionally, we tested an enhancer identified via a p300 pulldown assay in the mouse forebrain (Visel et al., 2009). This enhancer, as well as the orthologous enhancer from the medaka genome (Eichenlaub and Ettwiller, 2011), is shown to drive reporter gene expression in the forebrain. Importantly, the obtained pattern strongly resembles the pattern in the transgenic lines that were generated with meganuclease-based transgenesis (Eichenlaub and Ettwiller, 2011). Subtle differences in the activity of the two orthologous enhancers are readily uncovered by our approach in the injected generation (Fig. 6).

To address the range of enhancer activity during development, we tested the Fam44b enhancer (Eichenlaub and Ettwiller, 2011). As shown previously with the meganuclease-based approach, this enhancer drives reporter gene expression in a variety of neuronal structures that are formed at different developmental stages (Fig. 7B) (Eichenlaub and Ettwiller, 2011). Interestingly, this enhancer element is not sufficient to trigger reporter gene expression with PhiC31-mediated targeting. Only the *hsp70*-dependent expression in the heart and the lens was detected (Fig. 7A). As indicated by ubiquituous reporter gene expression after heatshock, strong mosaicity is not the cause for the lack of reporter gene expression (Fig. 7B). The difference between our assay and the meganuclease approach lies in the number of transgene insertions. PhiC31 targets a single locus specifically, whereas meganuclease facilitates the insertion of multiple and potentially concatemeric transgene copies (Grabher and Wittbrodt, 2007). Therefore, we hypothesized that the activity of this single copy enhancer is too low to reach detectable reporter gene expression levels. However, a multi-copy integration with meganuclease-based transgenesis overcomes this limitation. In order to amplify the expression levels from a single locus, we tested seven copies of the enhancer upstream of a single *hsp70* promoter. We detected the expected enhancer activity and we were able to follow reporter gene expression over time during development of the injected fish. The resulting pattern (Fig. 7) strongly resembles the pattern obtained with a transgenic line generated using the meganuclease-based approach (Eichenlaub and Ettwiller, 2011).

It is noteworthy that in transgenic lines obtained with the meganuclease-based assay, the final integration sites cannot be controlled (Mongin et al., 2011). Using the same enhancers in our assay, we reproducibly obtained highly similar reporter gene expression patterns from a single locus, already in the injected generation. Thus, our assay enables comparative analysis of enhancers.

Taken together, our assay allows quick and accurate assessment of the activity of enhancer and promoter elements *in vivo*. With the high efficiency of locus-specific PhiC31-mediated targeting, the injected embryos can be analyzed *in vivo* without the need to compare a large number of embryos or to establish germline transgenics. These features are particularly important for an in-depth analysis of enhancers on a large-scale.

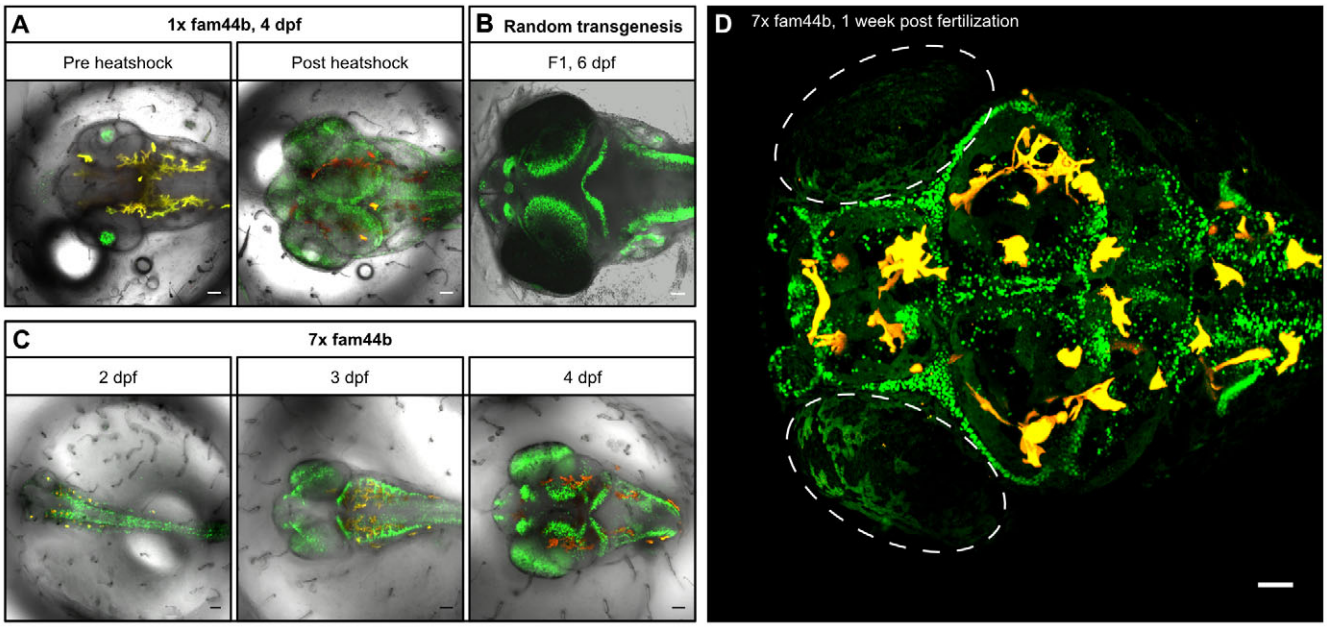

Fig. 7. Concatemeric assembly of the Fam44b enhancer increases reporter gene expression levels. (**A**) A single copy of the Fam44b enhancer element (Eichenlaub and Ettwiller, 2011) shows no detectable reporter gene expression except for *hsp70*-dependent expression in lens and heart. A heatshock confirms uniform targeting and integration. (**B**) Expected expression pattern from a transgenic line established using the meganuclease-based approach at 4 dpf. Figure is taken from Eichenlaub and Ettwiller (Eichenlaub and Ettwiller, 2011). (**C**,**D**) Seven copies of the Fam44b enhancer element cloned upstream of a single *hsp70* promoter lead to prominent reporter gene expression. Time course to show expression at different stages of development in injected fish. Note the high degree of similarity of the expression pattern at 4 dpf between the injected fish and the transgenic line (B). Scale bars: 50 μm.

DISCUSSION

We developed a novel site-specific transgenesis system for medaka using PhiC31 integrase with three independent landing site lines. PhiC31-targeted transgenic embryos are efficiently pre-selected via targeting sensors. We used our flexible system design to establish a locus-specific enhancer test. It is based on the molecular complementation of any enhancer/promoter fragment in the targeting vector with the EGFP reporter gene in the landing site. The high uniformity of reporter gene expression after injection allows immediate analysis of promoter activity. Interestingly, a defined enhancer, Fam44b (Eichenlaub and Ettwiller, 2011), is not active as a single copy in a single-locus configuration but shows enhancer activity in a multicopy configuration.

Our system design, especially the inclusion of the attP site within a selection cassette is adaptable to other model organisms. Consequently, animals or cells that contain properly integrated targeting vectors can be efficiently selected. For example, antibiotic resistance genes can be exchanged using PhiC31-mediated targeting, thereby facilitating the selection of properly targeted cells.

We successfully used SB100x for the generation of transgenic lines harboring one single-copy landing site. Obtaining single insertions may require the segregation of the initial set of integrations by at least three outcross generations. However, as SB100x is a cut and paste transposon (Ivics et al., 1997; Mátés et al., 2009), new single copy landing sites can be readily established by remobilization of single-copy landing sites resulting in local or distant transpositions (Ruf et al., 2011; Yergeau et al., 2011). Alternatively, landing sites could be introduced at specifically chosen positions by targeted homology-directed repair. Initial reports suggest that integration of DNA elements using TALENs is achievable (Bedell et al., 2012) but rather inefficient. The combination of homology-directed repair for a targeted insertion of the PhiC31 landing site appears to be a promising approach.

Apart from position-dependent enhancer activity assays, there are a number of additional applications that will benefit from PhiC31-mediated site-specific transgenesis in fish. It will strongly improve the reliability of transcriptional signaling pathway sensors (Barolo, 2006), inducible transactivation systems (Leneuve et al., 2003), RNA interference systems (Ni et al., 2009) as well as recombination systems (Vooijs et al., 2001). Furthermore, our system allows the integration of two different transgenes at the same locus in heterozygous chromosomes. This offers the unique opportunity to test regulatory elements not only in cis, but also in trans (Bateman et al., 2012). Additionally, this is the first step to mosaic analysis with a repressible cell marker (MARCM)/mosaic analysis with double markers (MADM)-like studies in fish model systems, which so far have been restricted to species amenable to homologous recombination (Lee and Luo, 2001; Luo, 2007). Finally, it is noteworthy that PhiC31 has been employed for BAC transgenesis in *Drosophila* (up to the size of 146 kb) (Venken et al., 2006), which suggests that it is applicable for this purpose in fish.

Acknowledgements

We thank the entire Wittbrodt department, especially S. Albadri, L. Centanin, M. Eichenlaub, D. Inoue, S. Lemke and R. Sinn, for fruitful discussions throughout the project. We thank O. Symmons for providing the various plasmid constructs for the Sleeping Beauty system. We are grateful to A. Saraceno, E. Leist and M. Majewski for excellent fish husbandry and to I. Kirchmaier for comments on the manuscript.

Funding

This project was supported by funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) [grant agreement number 223210 CISSTEM to J.W. and F.S.]; S.K. and B.H. received fellowships from The

Hartmut Hoffmann-Berling International Graduate School (HBIGS). Deposited in PMC for immediate release.

Competing interests statement
The authors declare no competing financial interests.

Author contributions
S.K., B.H., F.S. and J.W. designed the vectors of the transgenesis system. S.K. and E.K.M. established the transgenic lines. S.K., B.H. and E.K.M. characterized the transgenic lines. S.K., B.H. and D.B. performed the enhancer tests. S.K., B.H. and J.W. wrote the manuscript.

Supplementary material
Supplementary material available online at
http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.096081/-/DC1

References

Allen, B. G. and Weeks, D. L. (2005). Transgenic Xenopus laevis embryos can be generated using phiC31 integrase. *Nat. Methods* **2**, 975-979.
Auman, H. J., Coleman, H., Riley, H. E., Olale, F., Tsai, H.-J. and Yelon, D. (2007). Functional modulation of cardiac form through regionally confined cell shape changes. *PLoS Biol.* **5**, e53.
Barolo, S. (2006). Transgenic Wnt/TCF pathway reporters: all you need is Lef? *Oncogene* **25**, 7505-7511.
Bateman, J. R., Johnson, J. E. and Locke, M. N. (2012). Comparing enhancer action in cis and in trans. *Genetics* **191**, 1143-1155.
Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., 2nd, Tan, W., Penheiter, S. G., Ma, A. C., Leung, A. Y. H. et al. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-118.
Belteki, G., Gertsenstein, M., Ow, D. W. and Nagy, A. (2003). Site-specific cassette exchange and germline transmission with mouse ES cells expressing phiC31 integrase. *Nat. Biotechnol.* **21**, 321-324.
Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., Snyder, M. et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.
Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312-3317.
Blechinger, S. R., Evans, T. G., Tang, P. T., Kuwada, J. Y., Warren, J. T., Jr and Krone, P. H. (2002). The heat-inducible zebrafish hsp70 gene is expressed during normal lens development under non-stress conditions. *Mech. Dev.* **112**, 213-215.
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.
Eichenlaub, M. P. and Ettwiller, L. (2011). De novo genesis of enhancers in vertebrates. *PLoS Biol.* **9**, e1001188.
Ellingsen, S., Laplante, M. A., König, M., Kikuta, H., Furmanek, T., Hoivik, E. A. and Becker, T. S. (2005). Large-scale enhancer detection in the zebrafish genome. *Development* **132**, 3799-3811.
Emelyanov, A. and Parinov, S. (2008). Mifepristone-inducible LexPR system to drive and control gene expression in transgenic zebrafish. *Dev. Biol.* **320**, 113-121.
Grabher, C. and Wittbrodt, J. (2007). Meganuclease and transposon mediated transgenesis in medaka. *Genome Biol.* **8** Suppl., S10.
Groth, A. C., Olivares, E. C., Thyagarajan, B. and Calos, M. P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proc. Natl. Acad. Sci. USA* **97**, 5995-6000.
Groth, A. C., Fish, M., Nusse, R. and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782.
Hu, G., Goll, M. G. and Fisher, S. (2011). ΦC31 integrase mediates efficient cassette exchange in the zebrafish germline. *Dev. Dyn.* **240**, 2101-2107.
Huang, C.-J., Tu, C.-T., Hsiao, C.-D., Hsieh, F.-J. and Tsai, H.-J. (2003). Germ-line transmission of a myocardium-specific GFP transgene reveals critical regulatory elements in the cardiac myosin light chain 2 promoter of zebrafish. *Dev. Dyn.* **228**, 30-40.
Ivics, Z., Hackett, P. B., Plasterk, R. H. and Izsvák, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**, 501-510.
Kapusi, E., Kempe, K., Rubtsova, M., Kumlehn, J. and Gils, M. (2012). phiC31 integrase-mediated site-specific recombination in barley. *PLoS ONE* **7**, e45353.
Koster, R., Stick, R., Loosli, F. and Wittbrodt, J. (1997). Medaka spalt acts as a target gene of hedgehog signaling. *Development* **124**, 3147-3156.
Lee, T. and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development. *Trends Neurosci.* **24**, 251-254.
Leneuve, P., Colnot, S., Hamard, G., Francis, F., Niwa-Kawakita, M., Giovannini, M. and Holzenberger, M. (2003). Cre-mediated germline mosaicism: a new transgenic mouse for the selective removal of residual markers from tri-lox conditional alleles. *Nucleic Acids Res.* **31**, e21.
Lister, J. A. (2010). Transgene excision in zebrafish using the phiC31 integrase. *Genesis* **48**, 137-143.
Lu, J., Maddison, L. A. and Chen, W. (2011). PhiC31 integrase induces efficient site-specific excision in zebrafish. *Transgenic Res.* **20**, 183-189.
Luo, L. (2007). Fly MARCM and mouse MADM: genetic methods of labeling and manipulating single neurons. *Brain Res. Rev.* **55**, 220-227.
Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J. et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41**, 753-761.
Maurano, M. T., Wang, H., Kutyavin, T. and Stamatoyannopoulos, J. A. (2012). Widespread site-dependent buffering of human regulatory polymorphism. *PLoS Genet.* **8**, e1002599.
Mongin, E., Auer, T. O., Bourrat, F., Gruhl, F., Dewar, K., Blanchette, M., Wittbrodt, J. and Ettwiller, L. (2011). Combining computational prediction of cis-regulatory elements with a new enhancer assay to efficiently label neuronal structures in the medaka fish. *PLoS ONE* **6**, e19747.
Mosimann, C., Puller, A. C., Lawson, K. L., Tschopp, P., Amsterdam, A. and Zon, L. I. (2013). Site-directed zebrafish transgenesis into single landing sites with the phiC31 integrase system. *Dev. Dyn.* **242**, 949-963.
Neumann, C. J. and Nuesslein-Volhard, C. (2000). Patterning of the zebrafish retina by a wave of sonic hedgehog activity. *Science* **289**, 2137-2139.
Ni, J.-Q., Liu, L.-P., Binari, R., Hardy, R., Shim, H.-S., Cavallaro, A., Booker, M., Pfeiffer, B. D., Markstein, M., Wang, H. et al. (2009). A Drosophila resource of transgenic RNAi lines for neurogenetics. *Genetics* **182**, 1089-1100.
Pastinen, T. and Hudson, T. J. (2004). Cis-acting regulatory variation in the human genome. *Science* **306**, 647-650.
Preibisch, S., Saalfeld, S. and Tomancak, P. (2009). Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463-1465.
Rausch, H. and Lehmann, M. (1991). Structural analysis of the actinophage phi C31 attachment site. *Nucleic Acids Res.* **19**, 5187-5189.
Raymond, C. S. and Soriano, P. (2007). High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. *PLoS ONE* **2**, e162.
Rembold, M., Lahiri, K., Foulkes, N. S. and Wittbrodt, J. (2006). Transgenesis in fish: efficient selection of transgenic fish by co-injection with a fluorescent reporter construct. *Nat. Protoc.* **1**, 1133-1139.
Ritter, D. I., Dong, Z., Guo, S. and Chuang, J. H. (2012). Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. *PLoS ONE* **7**, e35202.
Ruf, S., Symmons, O., Uslu, V. V., Dolle, D., Hot, C., Ettwiller, L. and Spitz, F. (2011). Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat. Genet.* **43**, 379-386.
Shakes, L. A., Du, H., Wolf, H. M., Hatcher, C., Norford, D. C., Precht, P., Sen, R. and Chatterjee, P. K. (2012). Using BAC transgenesis in zebrafish to identify regulatory sequences of the amyloid precursor protein gene in humans. *BMC Genomics* **13**, 451.
Thiel, C., Schwarz, M., Peng, J., Grzmil, M., Hasilik, M., Braulke, T., Kohlschütter, A., von Figura, K., Lehle, L. and Körner, C. (2003). A new type of congenital disorders of glycosylation (CDG-li) provides new insights into the early steps of dolichol-linked oligosaccharide biosynthesis. *J. Biol. Chem.* **278**, 22498-22505.
Thomson, J. G., Chan, R., Thilmony, R., Yau, Y.-Y. and Ow, D. W. (2010). PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the Arabidopsis genome. *BMC Biotechnol.* **10**, 17.
Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. and Calos, M. P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol. Cell. Biol.* **21**, 3926-3934.
Venken, K. J., He, Y., Hoskins, R. A. and Bellen, H. J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* **314**, 1747-1751.
Venken, K. J. T., Schulze, K. L., Haelterman, N. A., Pan, H., He, Y., Evans-Holm, M., Carlson, J. W., Levis, R. W., Spradling, A. C., Hoskins, R. A. et al. (2011). MiMIC: a highly versatile transposon insertion resource for engineering Drosophila melanogaster genes. *Nat. Methods* **8**, 737-743.
Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. et al. (2009a). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858.
Vooijs, M., Jonkers, J. and Berns, A. (2001). A highly efficient ligand-regulated Cre recombinase mouse line shows that LoxP recombination is position dependent. *EMBO Rep.* **2**, 292-297.
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.
Yergeau, D. A., Kelley, C. M., Kuliyev, E., Zhu, H., Johnson Hamlet, M. R., Sater, A. K., Wells, D. E. and Mead, P. E. (2011). Remobilization of Sleeping Beauty transposons in the germline of Xenopus tropicalis. *Mobile DNA* **2**, 15.
Yonemura, N., Tamura, T., Uchino, K., Kobayashi, I., Tatematsu, K., Iizuka, T., Sezutsu, H., Muthulakshmi, M., Nagaraju, J. and Kusakabe, T. (2012). PhiC31 integrase-mediated cassette exchange in silkworm embryos. *Mol. Genet. Genomics* **287**, 731-739.

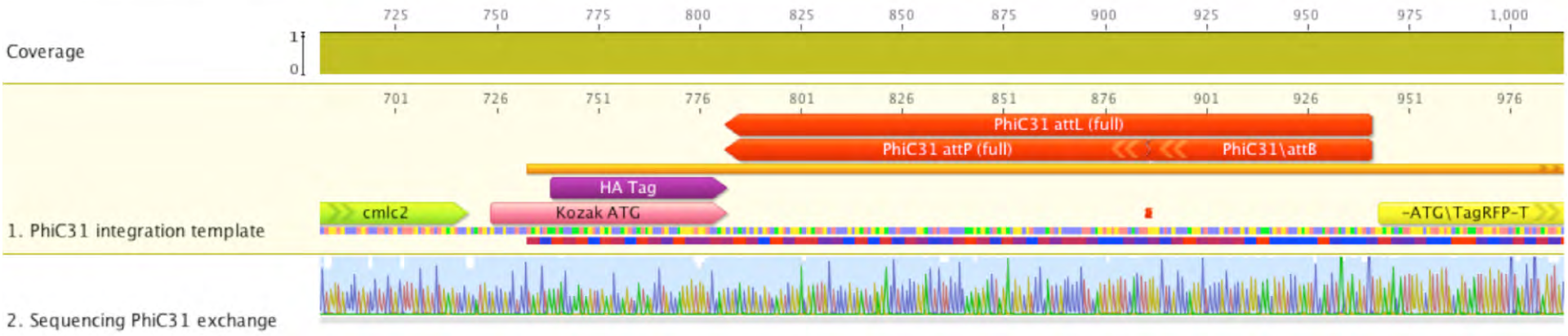

Figure S1 – Sequence verification of PhiC31 mediated vector targeting

The cmlc2 promoter (green arrow) drives the expression of an ORF (orange bar) that includes an HA tag (violet arrow), a full attL site (red arrows along with partial attP and attB sites) and a TagRFP-T fragment. Sequencing of injected docking site line fishes as well as germline transgenics verified PhiC31-mediated vector targeting and the maintenance of the ORF. Sequences were analyzed using Geneious software (Biomatters).

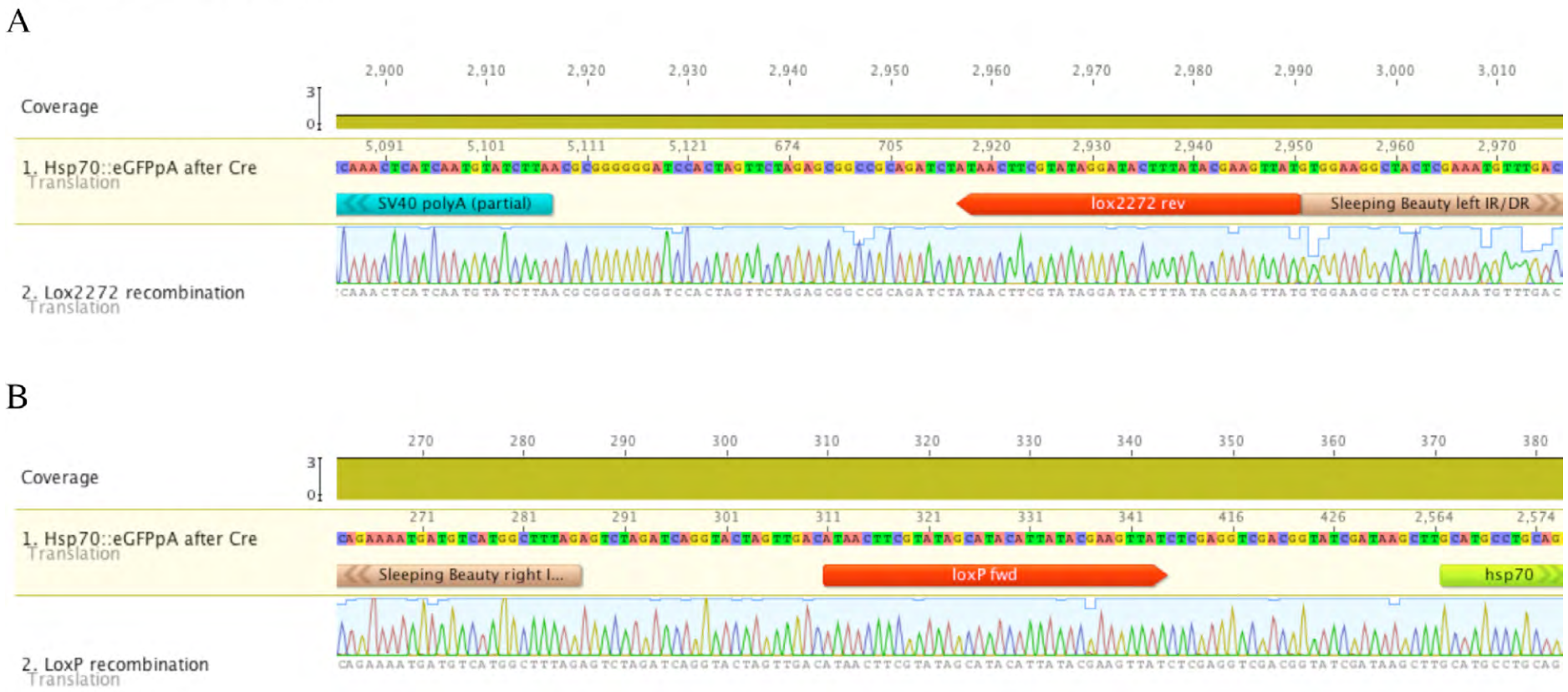

Figure S2 – Sequence verification of Cre-mediated locus cleanup

Genomic DNA of targeted embryos injected with CreNLS was analyzed by a genotyping PCR and subsequent sequencing. Sequencing verifies Cre-mediated locus cleanup over the Lox2272 (A) as well as LoxP (B) sites.

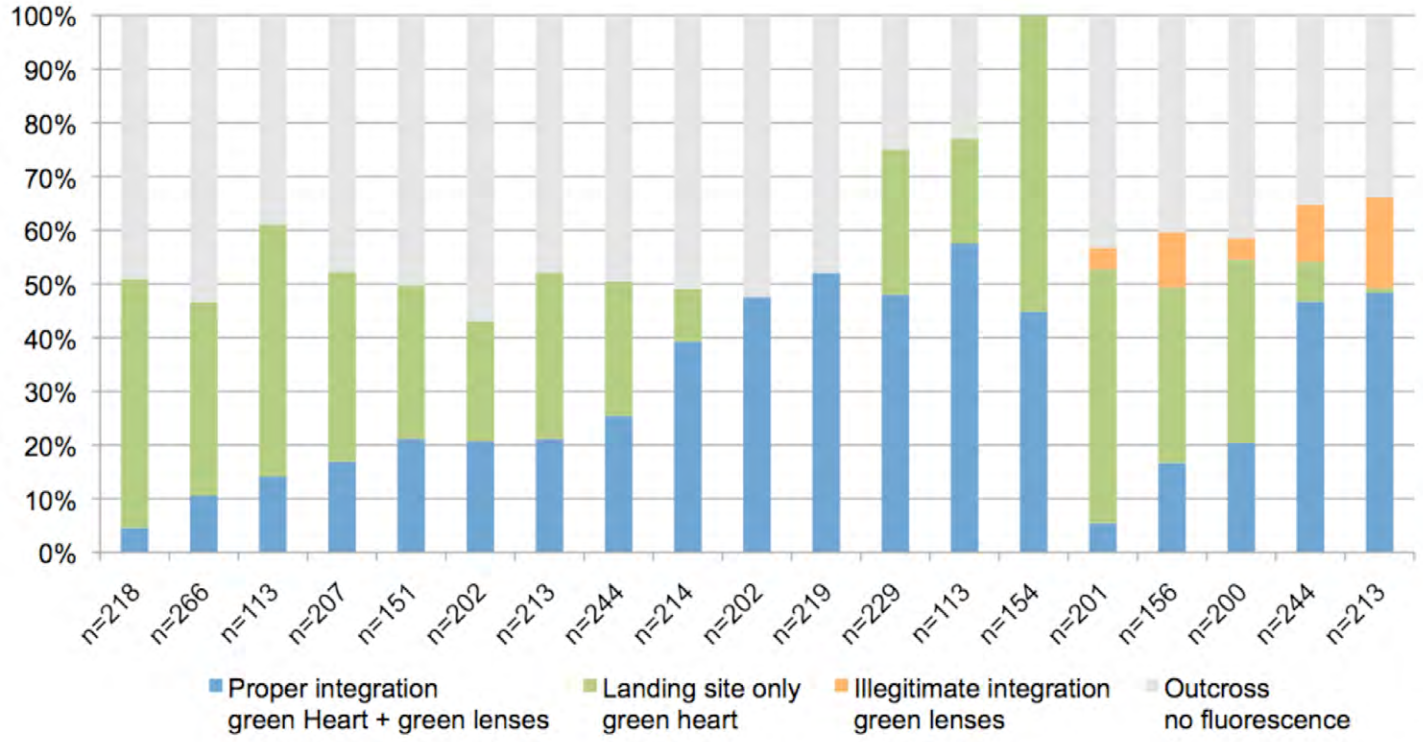

Figure S3 - Germline transmission rates for all positive founders

From a pool of founders carrying unmapped landing sites, 19 out of 26 (65%) of preselected fish transmitted the targeting vector through the germline. The transmission was scored in an outcross against wildtype (Cab), hence the maximum achievable transmission rate of a single genomic landing site is 50%. 3 fish yielded higher germline transmission, indicating the presence of multiple non-linked genomic landing sites. Illegitimate integrations of targeting vector outside of the landing site (orange) was identified by the presence of transgene-specific expression features (EGFP in lens) without co-occurring targeting sensor and/or landing site features (EGFP/TagRFP-T in heart muscle). Illegitimate integrations were restricted to 5 fish and in every case occurred only in a subset of targeted offspring.

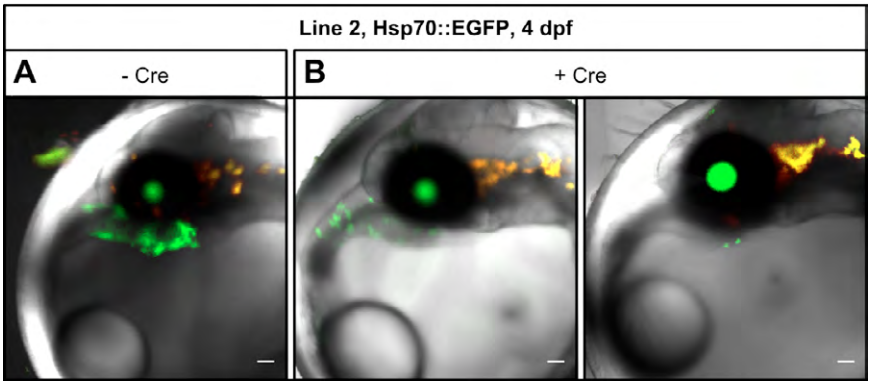

Figure S4 – Cre-mediated locus cleanup in line 2

(A) A lateral view of a fish transgenic for the hsp70::EGFP insert. Strong EGFP expression is detectable in the lens and the heart. As described in the main text, the heart-specific expression originates from an interaction between the cmlc2 promotor of the landing site and the insert-specific hsp70 promotor.
(B) Embryos from (A) were injected with CreNLS mRNA at the one-cell stage. Cre-mediated locus cleanup in the injected generation can be visualized by highly mosaic EGFP reporter gene expression in the heart. These fish submit the "cleaned" locus via the germline at high frequencies (100%, n=2)

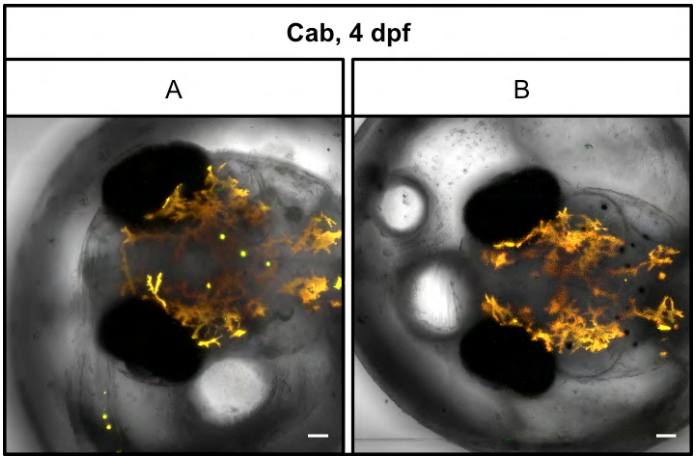

Figure S5 – Wildtype Cab strain imaged with Nikon AZ100

The wildtype Cab strain develops autofluorescent pigment cells. Autofluorescence of these cells can be excited with the 488nm laser and the emission is detected with the green as well as the red channel. Thereby, proper EGFP expression can be distinguished from the pigment cells.

Table S1.

	LM-PCR primer
5'-us-Hsp92II	GGATTTGCTGGTGCAGTACAGGCCTTAAGAGGGACTACATG
5’-ls	PO4-TAGTCCCTCTTAAGGCCT-NH2
3'-us-Hsp92II	GTAATACGACTCACTATAGGGCTCCGCTTAAGGGACTACATG
3’-ls	PO4-TAGTCCCTTAAGCGGAG-NH2
LM_SB-left2	GATGTCCTAACTGACTTGCCAAA
LM_SB-left1	TGAAAAACGAGTTTTAATGACTC
LM_SB-right1	CTCGAAATGTTTGACCCAAGT
LM_SB-right2	AAAGGCAATGCTACCAAATACT
oLM-5rev	GGATTTGCTGGTGCAGTACAG
oLM-5rev-nest	AGTACAGGCCTTAAGAGGGA
oLM-3rev	GTAATACGACTCACTATAGGGC
oLM-3rev-nest	AGGGCTCCGCTTAAGGGAC
	Verification of Cre cleanup
Cretest_SB_F	GGACATCTACTTTGTGCATGACACA
Cretest_Hsp70_R	ACCAAGCGACACCCCTGAAGGA
Cretest_SV40pA_F	GTGGTTTGTCCAAACTCATC
Cretest_SB_R	TCACATTCCCAGTGGGTCAGAAGT
	Genotyping
PhiC31_cmlc2-F2	CCAGTGACCCAGGAGCCCAGACC
PhiC31_tagRFPt-R1	TGCCCTCGTAGGGCTTGCCT
PhiC31_eGFP-R3	AGCTTGCCGGTGGTGCAGATG
Dock_Cmlc2_inner	TCGGGGTTTGCCTGGATTGTGT
Chr16_R	AAGCGTTCAGTTGCTTTAACGGTCA
Chr18_R	ACGCTGTGAGCAGGGAGGAGT
Scaffold_R	TGGACAGATGATGGTGTTGCATGGT

Table S1. Important oligos used in this study

Table S2

Site	Sequence
attP	GCTTCACGTTTTCCCAGGTCAGAAGCGGTTTTCGGGAGTAGT GCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGGCG TAGGGTCGCCGACATGACACAAGGGGTTGTGACCGGGGTGG ACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAGCG
attB	GTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGT GCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACC TCACCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCG GTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCT CGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCAC GGGACGTGCGACGGCGTCGGCGGGTGCGGATACGCGGGGC AGCGTCAGCGGGTTCTCGACGGTCACGGCGGG
attB mut	GTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGT GCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACC TC**T**CCCATCTGGTCCATCATGATGAACGGGTCGAGGTGGCG GTAGTTGATCCCGGCGAACGCGCGGCGCACCGGGAAGCCCT CGCCCTCGAAACCGCTGGGCGCGGTGGTCACGGTGAGCAC GGGACGTGCGACGGCGTCGGCGGGTGCGGATACGCGGGGC AGCGTCAGCGGGTTCTCGACGGTCACGGCGGG

Table S2 – Sequences of used PhiC31 att sites
The attB mut site is used when complementing the EGFP reporter gene with regulatory DNA. The mutation is underlined (A->T) and deletes an in-frame stop codon after proper recombination. The mutation is outside of the core attB sequence and we observed the same efficiencies with attB mut as with wildtype attB.