Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair

Thomas O. Auer,1,2,3,4 Karine Duroure,1,2,3 Anne De Cian,5,6,7 Jean-Paul Concordet,5,6,7,8 and Filippo Del Bene1,2,3,8

1Institut Curie, Centre de Recherche, Paris F-75248, France; 2CNRS UMR 3215, Paris F-75248, France; 3INSERM U934, F-75248 Paris, France; 4Centre for Organismal Studies Heidelberg, University of Heidelberg, 69120 Heidelberg, Germany; 5Muséum National d’Histoire Naturelle, Paris F-75231, France; 6CNRS UMR 7196, Paris F-75231, France; 7INSERM U565, Paris F-75231, France

Sequence-specific nucleases like TALENs and the CRISPR/Cas9 system have greatly expanded the genome editing possibilities in model organisms such as zebrafish. Both systems have recently been used to create knock-out alleles with great efficiency, and TALENs have also been successfully employed in knock-in of DNA cassettes at defined loci via homologous recombination (HR). Here we report CRISPR/Cas9-mediated knock-in of DNA cassettes into the zebrafish genome at a very high rate by homology-independent double-strand break (DSB) repair pathways. After co-injection of a donor plasmid with a short guide RNA (sgRNA) and Cas9 nuclease mRNA, concurrent cleavage of donor plasmid DNA and the selected chromosomal integration site resulted in efficient targeted integration of donor DNA. We successfully employed this approach to convert eGFP into Gal4 transgenic lines, and the same plasmids and sgRNAs can be applied in any species where eGFP lines were generated as part of enhancer and gene trap screens. In addition, we show the possibility of easily targeting DNA integration at endogenous loci, thus greatly facilitating the creation of reporter and loss-of-function alleles. Due to its simplicity, flexibility, and very high efficiency, our method greatly expands the repertoire for genome editing in zebrafish and can be readily adapted to many other organisms.

[Supplemental material is available for this article.]

Methods of genome engineering are becoming increasingly powerful owing to breakthroughs in the design of artificial nucleases that induce site-specific double-strand breaks (DSBs) in the genome (Gaj et al. 2013). These DSBs, as has been shown nearly 20 years ago using the homing endonuclease FokI, efficiently stimulate homologous recombination (HR) with a gene targeting vector in cultured cells and plants (Jasin 1996). Several types of artificial nucleases can now be designed to make the initial DSB that induces modification of a sequence of interest. Among these, zinc finger and TALE nucleases (TALENs) are fusions of artificial DNA binding domains—arrays of zinc fingers and TALE effector repeats, respectively—to the endonuclease domain of the FokI restriction enzyme. The latter is only active as a dimer and therefore needs to be recruited to the target sequence by fusion to two separate zinc finger or TALE domains binding complementary sequences separated by a short DNA spacer.

More recently, novel RNA-guided nucleases (RGNs) have been developed based on the CRISPR/Cas9 mechanism of bacterial defense against exogenous DNA (Jinek et al. 2012). A short guide RNA (sgRNA) complexed to Streptococcus pyogenes Cas9 endonuclease binds to its complementary DNA target sequence and leads to specific DNA cleavage by Cas9. By changing the 20-bp sgRNA sequence, one can redirect the Cas9 nuclease to predetermined chromosomal target sites (Cho et al. 2013; Cong et al. 2013; Hwang et al. 2013b; Mali et al. 2013).

Important pioneer studies using zinc finger nucleases (ZFNs) have demonstrated the potential of artificial sequence-specific nucleases in the genome engineering of many experimental systems. TALE and CRISPR/Cas9 nucleases have emerged as powerful alternatives that are much easier to engineer. While sequence-specific TALE nucleases can be readily assembled from TALE repeats specific to each nucleotide (Cermak et al. 2011; Huang et al. 2011; Sander et al. 2011), sgRNAs for the CRISPR/Cas9 system can be easily generated by cloning of target-specific oligonucleotides into sRNA expression vectors. The constraints of the sequences that can be targeted are minimal since TALE nucleases can be assembled to target TN(4,4)-NGG sequences (Miller et al. 2011) and sgRNA (G/A)(G/A)N18-NGG sequences (Hwang et al. 2013b). Importantly, both systems have been shown to be active in a very high proportion of cases, although efficiencies may vary considerably (Reyon et al. 2012; Hwang et al. 2013b).

The use of sequence-specific TALENs or RGNs based on the CRISPR/Cas9 system allows specific gene disruption in many organisms not previously amenable to forward genetic analyses, for instance, in common experimental models such as the rat or the zebrafish (Huang et al. 2011; Sander et al. 2011; Tesson et al. 2011; Hwang et al. 2013b). Gene inactivation results from small insertions or deletions (indels) introduced during the repair of cleaved DNA by nonhomologous end joining (NHEJ), causing frameshifts and premature stop codons.

However, a broader range of DNA sequence modifications is highly desirable for many purposes such as locus-specific insertion of reporter genes or tagging of open reading frames. Since their first application, both systems have been used for the targeted insertion
of short DNA sequences. By co-injection of single-stranded oligonucleotides bearing sequences flanking the cleaved target, site-specific DNA integration was recently demonstrated in mouse and zebrafish (Bedell et al. 2012; Chang et al. 2013; Hwang et al. 2013a; Wang et al. 2013; Wefers et al. 2013). Inducing DSBs with TALEns or RGNs at two sites on a chromosome can be used to trigger chromosomal deletions and inversions in cultured cells and zebrafish (Carlson et al. 2012; Gupta et al. 2013; Lim et al. 2013; Xiao et al. 2013). Artificial nucleases can also stimulate highly precise sequence modification by HR, but the efficiency is generally low. For example, using extremely active TALEN pairs that were able to induce indel mutations at rates up to 98%, Zu et al. could show gene targeting by HR in zebrafish with efficiencies at ~1.5% (Zu et al. 2013). Linearized donors with >800-bp perfect homology flanking the TALEN target site served as a template for gene targeting by HR and allowed integration of inserts up to 1 kb. In living organisms, low efficiency limits the widespread application of gene targeting by HR because screening a large number of animals may be required to isolate founders carrying the mutation of interest. Here we report highly efficient CRISPR/Cas9-mediated knock-in of >5.7-kb-long DNA cassettes into the zebrafish genome based on homology-independent DSB repair. We show that, due to its flexibility and high efficiency, our method considerably expands the practical possibilities of genome engineering in model organisms.

Results

It was recently shown that zinc finger nucleases and TALEns can drive targeted integration of DNA cassettes in cultured cells (Cristea et al. 2013; Maresca et al. 2013) via homology-independent DSB repair. Although the design strategy slightly differed between the two studies, they both showed that if a donor plasmid is cleaved in transfected cells, it is frequently integrated at a site concomitantly targeted by zinc finger or TALE nucleases. We were interested in testing this approach in a model organism—the zebrafish—as a potential alternative to gene targeting by homologous recombination. Due to its easier design compared to ZFNs and TALE nucleases, we decided to first utilize the CRISPR/Cas9 system to introduce targeted DSBs.

Targeted knock-in of KaflA4 into the Tg(neurode:eGFP) locus

We chose a neurode:eGFP transgene (Obholzer et al. 2008) that is broadly expressed in the central nervous system during embryonic development as the target integration site. The eGFP transgene allows the direct visualization of target gene disruption and should not compromise survival upon loss of gene function.

In our donor plasmid, we inserted the target sequences for two sgRNAs specific to eGFP (hereafter referred to as “bait” sequence) followed by the coding sequence of an improved version of the transcriptional transactivator Gal4 (KaflA4) (Distel et al. 2009). This reading frame was preceded by an E2A peptide linker for multicistronic expression (Fig. 1A; Szyczmak et al. 2004). When the donor plasmid was co-injected into an eGFP transgenic line with sgRNAs/Cas9 mRNA, concurrent cleavage of the genomic eGFP locus and bait plasmid sequence occurred. As NHEJ was shown to be highly active in early zebrafish development (Hagmann et al. 1998; Dai et al. 2010; Liu et al. 2012), we speculated that it would trigger integration of the donor plasmid into the opened chromosomal locus through nonspecific ligation of cleaved DNA ends.

After integration of the donor plasmid resulting in in-frame insertions of the E2A-KaflA4 DNA (Fig. 1A), former eGFP-positive cells were expected to express KaflA4. The simple loss of eGFP expression demonstrates gene disruption by the CRISPR/Cas9 system. In order to visualize integration events of the donor plasmid, we performed injections in embryos also carrying an UAS-RFP transgene [Tg(neurode:eGFP) × Tg(UAS:RFP, cry1:eGFP)] (Fig. 1B,C). If KaflA4 is inserted in-frame at the neurode:eGFP locus (which happens theoretically in 16.6% of integration events given three different frames and two insertion directions of the donor plasmid), the expressed KaflA4 will transactivate RFP expression by binding to the UAS sequence and triggering RFP transcription.

We designed two different sgRNAs targeting the eGFP bait sequence and estimated their efficiency at inducing indel mutations. For this purpose we pooled ten eGFP transgenic embryos after injection of sgRNAs and Cas9 mRNA, isolated genomic DNA, performed locus-specific PCR amplification on the eGFP locus, and estimated the rate of mutations by sequencing individual PCR clones. While sgRNA eGFP 1 was able to induce indel mutations at a rate of 66% (10/15 clones carrying mutations) (Table 1; Supplemental Table 1), the rate for sgRNA eGFP 2 was significantly lower (20%, 3/15 clones carrying mutations).

Using sgRNA eGFP 1 and co-injecting it with our eGFP bait-E2A-KaflA4 donor and Cas9 mRNA into Tg(neurode:eGFP) × Tg(UAS:RFP, cry1:eGFP) embryos, we observed RFP-positive cells within the neurod pattern in >75% (293/388) of injected embryos (Table 1). In about 22% (85/388) of injected embryos, RFP-positive cells were largely recapitulating neurode:eGFP expression (Supplemental Fig. 1; Table 1). In such embryos, RFP expression could be simultaneously detected in the brain and caudal neural tube, indicating integration events had likely occurred during the earliest stages of development.

In all confocal images acquired, we never observed co-expression of eGFP and RFP in the same cell. In about 80% of embryos (303/388), eGFP expression was strongly reduced compared to uninjected controls (Supplemental Fig. 1), indicating disruption of the eGFP open reading frame. RFP expression was more often observed in embryos that lost large parts of their neurode:eGFP expression, arguing for higher activity of the CRISPR/Cas9 system in these embryos. Within the group of RFP-positive embryos, <3% (9/388) showed RFP-expressing cells outside the neurode:eGFP expression domain (in muscle or skin cells). To further check for potential off-target integration of the donor plasmid, we performed injections in Tg(UAS:RFP, cry1:eGFP) embryos without the neurode:eGFP target locus. Within these embryos, we could only rarely observe some red muscle or skin cells in 1/300 (0.3%) embryos, arguing for a very low frequency of off-target integration events leading to expression of a functional KaflA4. In Tg(neurode:eGFP) × Tg(UAS:RFP, cry1:eGFP) embryos injected with eGFP bait-E2A-KaflA4 donor DNA and Cas9 mRNA but no sgRNA, we did not detect any RFP-expressing cells (0/243) (Fig. 1C). This indicates that the sgRNA is necessary to trigger integration of the donor plasmid.

After injection of the donor plasmid with the RGNs, successful targeted knock-in events were verified by PCR amplification (Fig. 1D) using integration site- and donor-specific primers (Fig. 1A). Subsequent analysis of the junction sequences revealed indel events typical for DSB repair by classical NHEJ and alternative end-joining mechanisms (Fig. 1E; Supplemental Table 1; Dai et al. 2010; Liu et al. 2012). Analyzing all junction sequences (between target locus and knocked-in donors) obtained in the course of this study, 50% exhibited small deletions (24/48 sequences) and 33% small
insertions (16/48), while 17% (8/48) corresponded to ligation of nonmodified DNA sequences from the targeted locus and plasmid (perfect repair).

In a further set of experiments, we made use of the second sgRNA specific for eGFP, sgRNA eGFP 2, and again found phenotypic and molecular evidence for targeted DNA integration (Supplemental Fig. 2). The number of successfully converted embryos (22/149), however, was much lower (15% vs. 76% with sgRNA eGFP 1), consistent with a reduced efficiency of this sgRNA (20%) at directing site-specific indel mutations in the eGFP ORF compared to sgRNA eGFP 1 (66%).

Comparison to co-injection of linearized donor plasmid

We wanted to test whether co-injected linearized donor plasmids would be integrated at the genomic locus cleaved by the CRISPR/Cas9 system. We therefore linearized our donor plasmid prior to injection in vitro with a restriction enzyme, cutting just upstream of the E2A-KalTA4 sequence (close to the sgRNA eGFP 1 binding site). When we co-injected linearized eGFP bait-E2A-KalTA4 donor DNA with sgRNA eGFP 1 and Cas9 mRNA into one-cell stage embryos of the Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) cross, we observed an increased death rate compared to when co-injecting circular plasmid (35% vs. 15%, respectively) (Supplemental Fig. 3C). Frequency of in-frame integration events as scored by RFP expression was much lower (11% vs. 76% with circular plasmid) and observed in a sparse manner (Supplemental Fig. 3A,E). Altogether, this experiment demonstrates that co-injection of a circular plasmid that is cleaved concurrently with the endogenous target locus is less toxic and more efficient in triggering plasmid integration at the desired locus.

Targeted knock-in of KalTA4 into the Tg(vsx2:eGFP) transgenic line

We next sought to confirm the efficiency of our approach using a second eGFP transgenic line [Tg(vsx2:eGFP)] (Kimura et al. 2006) integrated at a different genomic locus and with a more restricted expression pattern. Vsx2:eGFP drives eGFP expression in the zebrafish embryonic retina and hindbrain cells in 2-dpf-old embryos (Fig. 2A). The eGFP bait-E2A-KalTA4 donor plasmid was co-injected with sgRNA eGFP 1 and fish embryos
examined at 2 dpf. As shown in Figure 2B, conversion of the eGFP to the KalTA4 transgene could be directly visualized by the appearance of red fluorescent cells in the retina in the Tg(eye2:eGFP) × Tg(UAS:RFP, cry1:eGFP) genetic background. Cells in the hindbrain also switched from eGFP to RFP expression (Fig. 2C). Efficiency of targeted DNA integration was estimated to range around 60% (83/144 embryos) (Table 1), based on the green to red fluorescence conversion. Eleven percent of embryos (16/144) thereby showed a broad expression pattern, with red cells spread over the whole retina (~5% of retinal cells) (Fig. 2B) and the hindbrain (Fig. 2C). PCR and sequence analysis further confirmed that targeted DNA integration had taken place, and indel mutations typical of homology-independent repair pathways such as NHEJ were detected at junction sequences (Fig. 2D).

As we used cry1:eGFP (resulting in eGFP expression in the lens) as a transgenesis marker for the UAS:RFP transgene in the Tg(UAS:RFP, cry1:eGFP) line, we offered a further potential target site for eGFP-specific sgRNAs. In a few cases, we could observe RFP expression in the lens of the Tg(UAS:RFP, cry1:eGFP) transgenic fish (Supplemental Fig. 4). This event likely reflects the insertion of the KalTA4 DNA cassette into the cry1:eGFP transgene and was rarely detected (8/388 [2%] of injected embryos), owing to the extremely restricted expression pattern of the cry1 promoter.

Targeted knock-in at the Tg(pou4f3:mGFP) locus

Subsequently, to test our method with a different target gene while still benefiting from the visual read-out of the eGFP-to-KalTA4 switch, we targeted a transgene encoding an older, noncodon optimized version of GFP present in the Tg(pou4f3:mGFP) transgenic line (Xiao et al. 2005). We designed a sgRNA specific to the noncodon optimized GFP coding sequence and generated a new matching bait sequence for our E2A-KalTA4 donor plasmid. Co-injection with Cas9 mRNA into the Tg(pou4f3:mGFP) × Tg(UAS:RFP, cry1:eGFP) cross led to the GFP-to-KalTA4 switch (Supplemental Fig. 5A,B), and targeted DNA integration was confirmed at the DNA level by PCR and DNA sequence analysis of the junctions at the integration site (Supplemental Fig. 5C,D). The previous experiments show that we can successfully target eGFP and GFP transgenes and convert them to KalTA4 expression.

Targeted knock-in at the zebrafish kif5aa locus

To further extend the validity of CRISPR/Cas9-mediated knock-in on an endogenous target gene, we chose to target integration of KalTA4 CDNA to the kinesin family member Saa (kif5aa, ENSEMBL ID: ENSDARG0000005470.9) locus. Using in situ hybridization, we detected mRNA expression of this gene from 24 h postfertilization onward in the spinal cord (Fig. 3A), consistent with a recently published expression pattern (Campbell and Marlow 2013). At 3 dpf, kif5aa is broadly expressed in the brain, while BAC transgenesis using the medaka (Oryzias latipes) ortholog showed additional kif5aa transcription in the spinal cord and motoneurons at later stages of development (Kawasaki et al. 2012). We first designed a sgRNA specific to kif5aa, whose efficiency at inducing indel mutations was determined to range around 22% (4/18) (Supplemental Table 1). Furthermore, we replaced the eGFP bait sequence in the previously described KalTA4 targeting vector with a bait sequence for kif5aa (Fig. 3D). Successful integration of KalTA4 was revealed by RFP expression after co-injection of the kif5aa-bait-E2A-KalTA4 donor vector, sgRNA kif5aa 1, and Cas9 mRNA into Tg(UAS:RFP, cry1:eGFP) embryos. RFP-positive cells could be detected in 4% (6/150) of injected embryos within the endogenous kif5aa expression domain (Fig. 3B,C; Table 1), while the remaining 96% of embryos did not show any RFP expression. We observed RFP-expressing cells in the spinal cord, hindbrain, cerebellum, and motoneurons. Insertion in the kif5aa locus was confirmed by PCR and subsequent sequence analysis (Fig. 3E). In contrast to experiments on the two eGFP transgenes, however, we did not observe embryos with extensive red fluorescent labeling, indicating that knock-in efficiency was lower (76% of RFP-positive cells when using the eGFP knock-in set

Table 1. Knock-in efficiencies at the eGFP and the kif5aa locus

<table>
<thead>
<tr>
<th>Transgenic line used</th>
<th>Targeting system</th>
<th>Indel mutation efficiency</th>
<th>Donor plasmid used for knock in</th>
<th>Embryos with RFP expression</th>
<th>Broad RFP pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP)</td>
<td>sgRNA eGFP 1</td>
<td>66.6%</td>
<td>10/15</td>
<td>eGFP-bait-E2A-KalTA4</td>
<td>293/388 (75.6%)</td>
</tr>
<tr>
<td>Tg(xys2:eGFP) × Tg(UAS:RFP, cry1:eGFP)</td>
<td>sgRNA eGFP 1</td>
<td>66.6%</td>
<td>10/15</td>
<td>eGFP-bait-E2A-KalTA4</td>
<td>83/144 (57.6%)</td>
</tr>
<tr>
<td>Tg(UAS:RFP, cry1:eGFP)</td>
<td>sgRNA kif5aa 1</td>
<td>22.2%</td>
<td>4/18</td>
<td>kif5aa-bait-E2A-KalTA4</td>
<td>6/150 (4.0%)</td>
</tr>
<tr>
<td>Tg(UAS:RFP, cry1:eGFP)</td>
<td>sgRNA eGFP 1</td>
<td>66.6%</td>
<td>10/15</td>
<td>eGFP-bait-E2A-KalTA4</td>
<td>58/604 (9.6%)</td>
</tr>
</tbody>
</table>

Figure 1. CRISPR/Cas9-mediated knock-in of KalTA4 into the Tg(neurod:eGFP) transgenic line. (A) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (B) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (C) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (D) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (E) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (F) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (G) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (H) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line. (I) A schematic of the donor plasmid consisting of an N-terminal eGFP-bait with two sgRNA target sites (in orange, PAM sequence in blue). After co-injection of the donor with Cas9 mRNA and one eGFP sgRNA, insertion at the eGFP locus occurs. In-frame fusion of the E2A-KalTA4 transgenic line.
Improved by the co-injection of a more efficient sgRNA for in vivo cleavage of the donor vector. In addition, this experiment demonstrated that our knock-in strategy is independent from any sequence homology between the target locus and the bait sequence in the donor plasmid.

Homology-independent knock-in with TALE nucleases

Because the current design of the CRISPR/Cas9 system allows one to target statistically one sequence every 32 bp, in specific cases it may be necessary to use TALENs to target DSBs at specific loci (Hwang et al. 2013b). Therefore, we wanted to test the compatibility of our knock-in method with TALE nucleases in zebrafish. We designed a TALEN pair targeting the kif5aa locus. As previously described for our sgRNAs, we estimated the CRISPR efficiency at inducing indel mutations by PCR amplification on genomic DNA from a pool of ten injected embryos and subsequent sequence analysis of individual PCR clones. Thereby, this TALEN pair showed an efficiency of 60% (6/10 clones carrying mutations) at inducing indel mutations (Supplemental Table 1). For the visualization of integration events, we designed a plasmid donor with a kif5aa bait sequence followed by an UAS:eGFP cassette (Supplemental Fig. 6A). This DNA reporter construct shows eGFP expression independently from the direction and the frame of its insertion, allowing an easy assessment of integration events. Injections of the donor plasmid together with the kif5aa TALEN mRNAs were performed into the double transgenic line Tg(UAS:mcherry) × Et(1);Shp70;Gal4-VP16×101B (Scott et al. 2007) that expresses Gal4 and mcherry in the central nervous system and the notochord. This approach can be used without any prior knowledge of the target gene expression pattern and allows an efficient preselection of potential founders with targeted integration. More than 30% of injected embryos showed correct eGFP expression in the notochord compared to controls (injection without TALEN mRNAs or injection of TALEN mRNAs plus donor with scrambled bait sequence) (Supplemental Fig. 6B) showing no eGFP signal. Integration events were verified by PCR and sequence analysis (Supplemental Fig. 6C,D). In a few cases, eGFP fluorescence could also be detected in muscle cells in control embryos, which may correspond to rare random DNA integration or persistence of plasmid DNA at later developmental stages.

Germline transmission of knocked-in transgenes

To investigate the transmission of knocked-in donor plasmids through the germline to the next generation, we raised embryos of the Tg(neurod:eGFP) transgenic line that were injected with the eGFPbait-E2A-KalTA4 donor plasmid together with sgRNA eGFP 1 and Cas9 mRNA. This allowed an unbiased determination of the

Combination of multiple sgRNAs to increase knock-in efficiency

To overcome this reduced efficiency and demonstrate the flexibility of the CRISPR/Cas9 system for targeted knock-in, we co-injected Cas9 mRNA, the eGFPbait-E2A-KalTA4 donor plasmid, and the more efficient sgRNA eGFP 1 together with sgRNA kif5aa 1 (Fig. 4A). While sgRNA eGFP 1 guides Cas9 nuclease activity to cut the donor plasmid in the eGFP bait sequence, sgRNA kif5aa 1 is used to target the endogenous target locus. By more efficient cutting of the donor plasmid in the transgenic embryos, scale bar, 50 μm. (C) eGFP to KalTA4 conversion was assessed with the developing hindbrain. (D) Upon integration of the 5’ junction and the 3’ junction. (E) F1 embryo (from founder A) with stable expression of the Tg(vsx2:eGFPbait-E2A-KalTA4) transgene activating RFP expression from UAS:RFP in the retina. Scale bar, 300 μm. (F) List of 5’ junctions of alleles identified in stable transgenic founders. Within 12 screened potential founder fish, six alleles could be detected, whereas four founders showed in-frame integration of the transgene. (Orange) sgRNA binding site; (blue) binding sequence homology between the target locus and donor vector. The integration frequency can be significantly improved by the co-injection of a more efficient sgRNA for in vivo cleavage of the donor vector. In addition, this experiment demonstrated that our knock-in strategy is independent from any sequence homology between the target locus and the bait sequence in the donor plasmid.

Homology-independent knock-in with TALE nucleases

Because the current design of the CRISPR/Cas9 system allows one to target statistically one sequence every 32 bp, in specific cases it may be necessary to use TALENs to target DSBs at specific loci (Hwang et al. 2013b). Therefore, we wanted to test the compatibility of our knock-in method with TALE nucleases in zebrafish. We designed a TALEN pair targeting the kif5aa locus. As previously described for our sgRNAs, we estimated the CRISPR efficiency at inducing indel mutations by PCR amplification on genomic DNA from a pool of ten injected embryos and subsequent sequence analysis of individual PCR clones. Thereby, this TALEN pair showed an efficiency of 60% (6/10 clones carrying mutations) at inducing indel mutations (Supplemental Table 1). For the visualization of integration events, we designed a plasmid donor with a kif5aa bait sequence followed by an UAS:eGFP cassette (Supplemental Fig. 6A). This DNA reporter construct shows eGFP expression independently from the direction and the frame of its insertion, allowing an easy assessment of integration events. Injections of the donor plasmid together with the kif5aa TALEN mRNAs were performed into the double transgenic line Tg(UAS:mcherry) × Et(1);Shp70;Gal4-VP16×101B (Scott et al. 2007) that expresses Gal4 and mcherry in the central nervous system and the notochord. This approach can be used without any prior knowledge of the target gene expression pattern and allows an efficient preselection of potential founders with targeted integration. More than 30% of injected embryos showed correct eGFP expression in the notochord compared to controls (injection without TALEN mRNAs or injection of TALEN mRNAs plus donor with scrambled bait sequence) (Supplemental Fig. 6B) showing no eGFP signal. Integration events were verified by PCR and sequence analysis (Supplemental Fig. 6C,D). In a few cases, eGFP fluorescence could also be detected in muscle cells in control embryos, which may correspond to rare random DNA integration or persistence of plasmid DNA at later developmental stages.
CRISPR/Cas9 mediated knock-in in zebrafish

Figure 3. CRISPR/Cas-mediated knock-in of KalTA4 into the kif5aa locus. (A) Kif5aa expression in zebrafish embryos revealed by in situ hybridization. Dorsal (A') and lateral (A'') views of 24-hpf embryos and dorsal view of 3-dpf embryo head and trunk region (A''') showing kif5aa expression in various brain regions and the spinal cord. (B,C) Representative confocal pictures of a Tg(UAS:RFP, cry1:eGFP) embryo injected with the eGFP-bait-E2A-KalTA4 donor plasmid together with sgRNA eGFP 1 that showed expression of RFP in parts of the neurod:eGFP expression domain. Within the pool of RFP-selected embryos, we found germline transmission of two in-frame knock-in events in five founders screened (40%, 2/5) (Table 2). This argues for an enrichment of in-frame integration by selection for RFP expression in F0 fish as expected.

We used five different primer combinations to detect head-to-tail concatemer formation as shown in Supplemental Figure 8B. These results were further confirmed by Southern blot analysis (Supplemental Fig. 8C,D).

To further confirm locus-specific knock-in events, we performed Southern blot analysis. With a probe hybridizing to the neurod locus flanking sequence (Fig. 5A), we could detect a specific band of the expected size (6.6 kb) for insertion of the donor plasmid into the neurod:eGFP locus in the progeny of founder C (Fig. 5E, black arrow). Furthermore, the probe detected a 2.7-kb band in the wild-type zebrafish embryos corresponding to the endogenous neurod locus (white arrowhead), present also in all other samples as expected. In the transgenic animals used for our knock-in experiments, this band was accompanied by a smaller 2.6-kb band corresponding to a partially digested fragment (see Fig. 5A for a graphic explanation). The signals corresponding to the transgenic locus were much stronger and 3' junctions of integrations at the target locus and potential head-to-tail concatemers were no longer detected and were replaced by the 6.6-kb band resulting from the KalTA4 integration.

germline transmission rate without prior selection for positive integration events. Potential founder fish were out-crossed to Tg(UAS:eGFP, cry1:eGFP) embryos and screened for RFP expression. We could detect germline transmission of in-frame knock-in events in three out of 29 (10.3%) F0 fish (Fig. 5B; Table 2). The degree of transmission of the knocked-in transgene to the next generation thereby ranged from 1.2% (3/244) to 34.2% (93/272) in F1 progeny (Supplemental Table 2). If no RFP expression was observed in at least 50 embryos, these were pooled and analyzed by PCR for out-of-frame insertion of the targeting vector not resulting in expression of a functional KalTA4. In six further founders, we detected forward insertion of the KalTA4 transgene by PCR, and sequence analysis confirmed out-of-frame insertion into the eGFP locus (Fig. 5C; see Fig. 5D for a list of sequenced 5' junctions). This argues for a germline transmission rate of forward integrated donors of 31% (9/29 tested founders) (Table 2).

Taking advantage of the visual readout of integration, we also selectively raised Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) embryos injected with the eGFP-bait-E2A-KalTA4 donor plasmid together with sgRNA eGFP 1 that showed expression of RFP in parts of the neurod:eGFP expression domain. Within the pool of RFP-selected embryos, we found germline transmission of two in-frame knock-in events in five founders screened (40%, 2/5) (Table 2). This argues for an enrichment of in-frame integration by selection for RFP expression in F0 fish as expected.

Similarly, for the Tg(vsx2:eGFP) transgene, we could identify transmission through the germline at a comparable rate of 50% (6/12) in RFP-selected Tg(vsx2:eGFP) × Tg(UAS:RFP, cry1:eGFP) embryos (Table 2), with 33% (4/12) showing in-frame integration (Fig. 2E; see Fig. 2F for a list of sequenced 5’ junctions).

The identical expression pattern of RFP and eGFP clearly argues for the insertion of the eGFP-bait-E2A-KalTA4 transgene into the eGFP locus as confirmed by PCR analysis.

To further confirm locus-specific knock-in events, we performed Southern blot analysis. With a probe hybridizing to the neurod locus flanking sequence (Fig. 5A), we could detect a specific band of the expected size (6.6 kb) for insertion of the donor plasmid into the neurod:eGFP locus in the progeny of founder C (Fig. 5E, black arrow). Furthermore, the probe detected a 2.7-kb band in the wild-type zebrafish embryos corresponding to the endogenous neurod locus (white arrowhead), present also in all other samples as expected. In the transgenic animals used for our knock-in experiments, this band was accompanied by a smaller 2.6-kb band corresponding to the neurod:eGFP BAC transgene (black arrowhead), as well as an additional weaker band of 4.4 kb (asterisk), that likely corresponded to a partially digested fragment (see Fig. 5A for a graphic explanation). The signals corresponding to the transgenic locus were much more intense than the wild-type one, consistent with the presence of multiple transgene copies in the Tg(neurod:eGFP) line (Fig. 5E, inset), which is frequently observed in classical BAC transgenesis used to generate this line (Obholzer et al. 2008). In the knock-in animals derived from founder C, the bands corresponding to the neurod:eGFP transgene were no longer detected and were replaced by the 6.6-kb band resulting from the KalTA4 integration.

To examine if multiple copies of donor plasmid were integrated, we performed PCR analysis on DNA of founder progeny. We used five different primer combinations to detect 5' and 3' junctions of representative targeted integration events after PCR-based amplification. Binding sites of primers used for amplification are shown in D. (Orange) sgRNA binding site; (blue) PAM sequence NGG; (red) integrated additional base pairs. Note that the sgRNA is targeting the minus strand.
mutations by T7 endonuclease I digestion in pools of Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) embryos with and without injection of sgRNA eGFP 1, Cas9 mRNA, and the eGFPbait-E2A-KalTA4 donor plasmid. As expected, we could detect T7E1-mediated cleavage at the neurod:eGFP locus in the pool of injected embryos (Supplemental Fig. 9). In contrast, no mutations could be detected at eight of the 11 potential off-target loci tested. For off#7 we saw the same T7E1 activity in controls as in injected embryos, and we determined by sequencing of PCR products that this was caused by a polymorphism in the Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) genetic background (16:43707701–43707722: TGGTTATTTTTGTTTTTTA — TG— - - - - - - - - - G). At two loci (off#1, off#8), we detected T7E1-mediated cleavage that was more prominent in injected embryos compared to controls (Supplemental Fig. 9). By direct sequencing of PCR clones from these loci, we did not detect any indel mutations at the potential off-target site off#8 in 33 clones (0/33 clones carrying mutations), arguing for a cleavage frequency <3% at this locus. For off#1, we sequenced 34 clones. Thus we detected the presence of a polymorphic microsatellite region with various alleles within our amplicon (1:40240770–40240783: GTGTGTGTGTGT) that would lead to fragment sizes, after T7E1 cleavage, of around 140 bp + 250 bp. Furthermore, we did not detect any indel mutations at the potential off-target site (0/34 clones carrying mutations). Also at this locus, the cleavage frequency of sgRNA eGFP 1/ Cas9 must be <3%.

To check for knock-in of our donor plasmid at the two off-target sites, off#1 and off#8, we looked for plasmid insertion by PCR at these two locations in injected embryos and could not detect any evidence for off-target insertion (Supplemental Fig. 10). Similarly, when analyzing the progeny of one founder fish [Tg(neurod:eGFPbait-E2A-KalTA4)—founder H], no integration of our donor plasmid at these two potential off-target genomic locations could be observed, consistent with our Southern blot data.

**Discussion**

In the experiments described here, we showed for the first time in an in vivo model that CRISPR/Cas9-mediated DSBs can be used to efficiently knock-in donor plasmids at predetermined target sites. We were able to knock-in donors as large as 5.7 kb compared to up to 1 kb when gene targeting was performed by HR in zebrafish (Zu et al. 2013).

In previous cell culture studies, Cristea et al. (2013) showed that including a short DNA sequence bearing the nuclease target site (5'-NGG-PAM-3') in the donor plasmid was essential for efficient knock-in. However, the success rate was relatively low (<3%). In our study, we found that the use of the correct PAM sequence (NGG) and the selection of appropriate binding sites were critical for achieving high efficiency and specificity in the in vivo system. The results suggest that CRISPR/Cas9-mediated knock-in has great potential for genetic manipulation in zebrafish and other live model organisms, offering a powerful tool for studying gene function and disease mechanisms.
Figure 5. Analysis of stable germline transmission of the Tg(neurod:eGFPbait-E2A-KalTA4) transgene. (A) Schematic depicting the Southern blot design to detect KalTA4 transgene integration. The neurod locus-specific probe 1 detects a 2.7-kb fragment after HindIII digest in the wild-type allele. The transgenic BAC neurod:eGFP locus is digested into a 2.6-kb fragment and, in the case of a partial digest in the BAC backbone, into a 4.4-kb fragment. After insertion of the KalTA4 cassette, a 6.6-kb fragment is detected. (B) Brightfield and fluorescent images of a transgenic Tg(neurod:eGFPbait-E2A-KalTA4) embryo at 2 dpf. (C) Screening for transgene integration by PCR in eight potential founders. Two show the expected fragment size (478 bp) (cf. Fig. 1A for primer positions and amplicon size). Note that the amplicon of founder B is slightly larger, as confirmed by sequencing and shown in D. (D) Sequences of 5’ junction sites of alleles identified in stable transgenic founders. Out of 11 founders showing stable transgene integration and transmission, five had an in-frame integration of the transgene. (Orange) sgRNA binding site; (blue) PAM sequence NGG; (red) integrated additional base pairs. (E) Analysis of the stable founder C for site-specific transgene integration by Southern blot analysis. As controls, wild-type and Tg(neurod:eGFP) embryos were used. Compare the schematic shown in A for expected fragment sizes. The 2.7-kb wild-type neurod fragment can be seen in all three samples (white arrow). The Tg(neurod:eGFP) sample shows a further fragment at 2.6 kb with greater intensity (black arrow) consistent with multiple insertions of the BAC construct. A shorter exposure is shown below to better distinguish the two separate bands. A further fragment at 4.4 kb is visible (asterisk), probably arising from incomplete digest of the neurod:eGFP BAC transgene. In founder C, the neurod:eGFP band is no longer visible—instead, a fragment at 6.6 kb corresponding to the integration of KalTA4 into the eGFP sequence is detected.
sequence onto a plasmid (that we call bait sequence) was sufficient for targeted integration at the nuclease chromosomal target sequence by homology-independent pathways of DSB repair upon cotransfection of plasmid and nuclease expression vectors. In contrast, Maresca et al. (2013) reported a different design where further cleavage of the integrated plasmid was prevented due to the specific utilization of nucleases with FokI mutants that only heterodimerize. In our case, when using CRISPR/Cas9, we showed that both designs were efficient, since reclereaving of the integrated sequence is possible in the case of the KalTA4 insertion into the eGFP locus but impossible after the insertion of the same DNA cassette into the kif5aa locus, as shown in Figure 4. In the first case, upon integration and end-joining in the absence of indels, we expect to re-create a complete sgRNA target sequence necessary for Cas9 activity, while in the second case, a hybrid sequence between the endogenous gene and the GFP bait sequence will be generated and no longer be recognized by the sgRNAs. In our study, we have not examined which homology-independent mechanisms are mediating DNA integration. Further studies would be necessary to determine if classical NHEJ or alternative end-joining pathways are involved. Nevertheless, in agreement with previous studies in cell culture systems (Maresca et al., 2013), classical NHEJ is the most likely mechanism involved.

In order to test our knock-in method we chose to target eGFP transgenes, and we have shown that our eGFP bait-E2A-KalTA4 construct can be directly applied to efficiently convert any eGFP into a KalTA4 transgenic line. Given the wealth of eGFP enhancer and gene trap lines previously generated in zebrafish (Kawakami et al. 2004; Parinov et al. 2004; Ellingsen et al. 2005), this offers new possibilities for deeper analysis of the marked cell types by tissue-specific expression of various UAS-driven constructs. The same approach, using the same target plasmid and sgRNA, can also be used in other species, such as Drosophila, where large collections of eGFP transgenic lines exist and CRISPR/Cas9 has been shown to work (Gratz et al. 2013).

Previously, when performing a knock-in by HR, Zu et al. (2013) showed germline transmission in zebrafish at rates of 1.5%, using highly efficient TALEN pairs (up to 98% indel rates). In our case, the most efficient nucleae, sgRNA eGFP 1/Cas9, had an indel mutation rate of 66%. Nevertheless, we observed germline transmission rates for the neurod:eGFP locus up to 31%. Even just taking in-frame integrations into account, with 10.3%, the rate of functional targeting of the locus was still higher. Taking advantage of positive selection, as done when screening for RFP-positive founders, we could increase this rate up to 40%. This high rate of in-frame founders after selection held true for a second locus, Tg(vsx2:eGFP), with four in-frame insertion events in 12 screened founder fish (4/12, 33%). Therefore, it seems that knock-in events by homology-independent DSB repair mechanisms are more frequent and lead to higher rates of germline transmission than HR-mediated events. This is in line with previous studies that showed that NHEJ, the major homology-independent mechanism of DSB repair, is at least 10-fold more active than HR during early zebrafish development (Hagmann et al. 1998; Dai et al. 2010; Liu et al. 2012).

Importantly, when targeting the kif5aa locus, we found that integration efficiency was considerably increased by using a combination of the kif5aa-specific sgRNA kif5aa 1 and sgRNA eGFP 1, with its corresponding eGFP DNA donor (Fig. 4). This strategy can be easily applied to any gene of interest without designing locus-specific donor plasmids. Our efficient sgRNA 1 for eGFP seems to direct only a very low degree of off-target nuclease activity, and no integration of the donor vector at predicted off-target sites could be detected. Therefore, sgRNA eGFP 1 together with its donor plasmid can be used to efficiently insert KalTA4 at any genomic locus targeted by a site-specific sgRNA, even of modest efficacy. Furthermore, KalTA4 can be easily replaced with reporter genes such as GFP to generate fluorescent fusion proteins, or other heterologous transcription factors such as TetR or LexA.

Our strategy, due to its simplicity and high efficiency, may become a new standard to generate mutant alleles that can be readily visualized and screened for in different transgenic backgrounds. This has the advantage of creating reporter lines at the same time (as compared to BAC recombineering), as we demonstrated for the kif5aa locus. The possibility to select for integration events already in the F0 will greatly reduce the number of animals to raise and screen to obtain mutants, so far blindly selected by PCR. In addition, the simplicity of the DNA target vector preparation will offer an easier alternative to BAC transgenesis. In fact, as bait sequences are of small size, they can be generated easily by PCR or oligonucleotide cloning, and no long homology stretches between donor and target site are required.

However, in contrast to gene targeting by HR, which allows for precise, predetermined transgene insertion sites, knock-in events mediated by homology-independent mechanisms have to be selected for appropriate in-frame insertions. In our case, this did not seem to be a major limitation due to the high knock-in rate. In many cases, choosing target sequences within introns and employing splice acceptor sites in the donor plasmid will avoid problems due to imprecise end-joining, and it could even further increase the number of functional insertions. As a great advantage, CRISPR/Cas9 allows the simultaneous targeting of several sequences (Cong et al. 2013; Wang et al. 2013) and may also be used for gene replacement by targeting sequences upstream of and downstream from a given locus at the same time.

### Methods

**Fish lines and husbandry**

For this study, the Tg(neurod:eGFP) (Obholzer et al. 2008), Tg(vsx2:eGFP) (Kimura et al. 2006), Tg(pou4f3:smGFP) (Xiao et al. 2005), Tg(UAS:Cherry) × Et(1.5shp70:Gα14-VP16)s1013t16 (Scott et al. 2008) and Tg(UAS:RFP, cry1:eGFP) (Auer et al., 1999) transgenic lines were used.

### Table 2. Rate of germline transmission of KalTA4 knock-in into the eGFP locus

<table>
<thead>
<tr>
<th>Pool of F0 fish screened</th>
<th>Number of fish screened</th>
<th>Founder with forward integration</th>
<th>Founder with forward integration in-frame</th>
<th>Rate of germline transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tg(neurod:eGFP) Injected embryos without selection</td>
<td>29</td>
<td>9</td>
<td>3</td>
<td>31% (9/29)</td>
</tr>
<tr>
<td>Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) Injected embryos screened for RFP expression</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>40% (2/5)</td>
</tr>
<tr>
<td>Tg(vsx2:eGFP) × Tg(UAS:RFP, cry1:eGFP) Injected embryos screened for RFP expression</td>
<td>12</td>
<td>6</td>
<td>4</td>
<td>50% (6/12)</td>
</tr>
</tbody>
</table>
et al. 2007), and Tg(UAS:RFP, cry1:eGFP) transgenic lines were used. Breeding and raising of zebrafish followed standard protocols.

Molecular cloning

The UAS:RFP/cry1:eGFP construct was cloned combining the cry1:eGFP fragment (Balcıunas et al. 2004) with an 14×UAS sequence upstream of RFP (Koster and Fraser 2001) in a vector containing Tol2 sites (Kawakami et al. 2000). The eGFP-bait-E2A-KalTA4 donor plasmid was generated by forward insertion of a PCR-amplified eGFP fragment into the pCRII-TOPO (TOPO TA Cloning Kit Dual Promoter, Invitrogen) vector. Primers were used (5’ to 3’) eGFP_fwd: ATAGTTGATCCCATCGTGAAGCAAGGC GAGGAGC, eGFP_rev: GTAGCGGCTGAAGCACTGCACGC. The E2A-KalTA4-pA fragment was generated by fusion of individual PCR products using Phusion High-Fidelity DNA Polymerase (Thermo Scientific); E2A was amplified with the primers (5’ to 3’) E2A_fwd: TGCAGATATCCAGGAGGAGGACAGTGTACTAATTAT and XhoI. The amplified, and inserted into pCRII-TOPO–(Distel et al. 2009). Subsequently, both fragments were fused, and inserted into pCRII-TOPO–eGFP-bait with EcoRV and Xhol. The GFP-bait-E2A-KalTA4-pA donor plasmid was generated by forward insertion of a PCR-amplified GFP fragment into the pCRII-TOPO vector. Primers used (5’ to 3’) E2A_fwd: TGCAGATATCCAGGAGGAGGACAGTGTACTAATTAT and XhoI/XbaI-digested DR274 (Addgene ref 42250) plasmid vector for synthesis of sgRNA proteins are listed in Supplemental Table 4. sgRNAs guide sequence upstream of UAS:RFP/cry1:eGFP

TALEN and sgRNA generation

TALENs were assembled by a method derived from Huang et al. (2011). For each TALEN subunit, the fragment containing the 16 RVD segment was obtained from single-unit plasmids kindly provided by Bo Zhang (Peking University, China). The assembled TALE repeats were subcloned into a pCS2 vector containing appropriate Δ152 Nter TALE, +63 Cter TALE, and FokI cDNA sequences with the appropriate half-TALE repeat (derived from the original D. rerio TALEN-2 sequence (Balciunas et al. 2004) with an 14×UAS sequence upstream of RFP (Koster and Fraser 2001) in a vector containing Tol2 sites (Kawakami et al. 2000). The eGFP-bait-E2A-KalTA4 donor plasmid was generated by forward insertion of a PCR-amplified eGFP fragment into the pCRII-TOPO (TOPO TA Cloning Kit Dual Promoter, Invitrogen) vector. Primers used were eGFP_fwd: ATAGTTGATCCCATCGTGAAGCAAGGC GAGGAGC, eGFP_rev: GTAGCGGCTGAAGCACTGCACGC. The E2A-KalTA4-pA fragment was generated by fusion of individual PCR products using Phusion High-Fidelity DNA Polymerase (Thermo Scientific); E2A was amplified with the primers (5’ to 3’) E2A_fwd: TGCAGATATCCAGGAGGAGGACAGTGTACTAATTAT and XhoI/XbaI-digested DR274 (Addgene ref 42250) plasmid vector for synthesis of sgRNA proteins are listed in Supplemental Table 4. sgRNAs guide sequence upstream of UAS:RFP/cry1:eGFP

Injection of zebrafish embryos

TALEN mRNAs or sgRNA/Cas9 mRNA were co-injected into one-cell stage zebrafish embryos with fresh Qiagen midiprep (Qiagen) purified donor DNA. Each embryo was injected with 1 nl of solution containing ~75 ng/μl of each TALEN mRNA or ~7 ng/μl of sgRNA and ~150 ng/μl Cas9 mRNA together with ~7 ng/μl of donor plasmid. When two sgRNAs were co-injected, 7 ng/μl of each sgRNA were used. On the next day, injected embryos were inspected under a stereomicroscope. Only embryos that developed normally were assayed. Fluorescent protein expression was monitored over consecutive days. Genomic DNA was extracted from either single embryos or pools of embryos (as indicated) and then used for PCR, mapping, and DNA sequencing experiments as described below.

Insertion mapping

For insertion mapping, the primers used are listed in Supplemental Table 5. Genomic DNA was extracted following standard protocols. PCR was performed using Phusion High-Fidelity DNA Polymerase (Thermo Scientific). For sequence analysis of PCR products, PCR amplicons were tagged using Taq Polymerase (Life Technologies), cloned into the pCRII-TOPO (TOPO TA Cloning Kit Dual Promoter, Life Technologies) vector, and sent for sequencing. Mutant alleles were identified by sequencing the wild-type unmodified sequence. Mapping products were compared to the theoretical fusion products of cutting sites.

Detection of germline transmission

Potential founder fish were out-crossed to the Tg(UAS:RFP, cry1:eGFP) transgenic line. Fluorescent protein expression was monitored over the following days of development and the rate of mosaicism of germline transmission determined for RFP-positive in-frame founders. If no RFP signal was detected in at least 50 embryos, embryos were pooled, and genomic DNA was extracted and screened for locus-specific transgene integration by PCR. Subsequently PCR amplicons were sequenced.

Immunohistochemistry

Zebrafish larvae were processed for immunohistochemistry using standard protocols. Briefly, 4-dpf larvae were fixed in 4% paraformaldehyde (PFA; w/v, pH 7.4) overnight at 4°C, equilibrated in 30% sucrose (w/v) in phosphate-buffered saline (PBS) overnight at 4°C, and embedded in Tissue-Tek O.C.T. Compound (Sakura Fine-Tech). Blocks were then frozen at ~80°C on dry ice. Embedded larvae were sectioned horizontally on a cryostat (Leica Instruments). The 12-μm sections were collected on Superfrost Plus slides (Fisher Scientific), air dried for 30 min–2 h, and rehydrated in PBS. Sections were incubated with blocking reagent containing 10% (v/v) normal goat serum (Jackson ImmunoResearch Laboratories) and 0.1% Tween-20 (v/v; Sigma) in PBS (pH 7.4) for 1 h at room temperature. Slides were left overnight in primary antibody diluted in blocking solution at 4°C in a humidified chamber. The following day, sections were washed three times in PBS/0.1% Tween-20 and then incubated for 2 h in a blocking solution containing Alexa fluorophore-conjugated secondary antibody diluted 1:500 (Invitrogen Molecular Probes) with DAPI nuclear marker (Sigma), washed three times in PBS/0.1% Tween-20, and mounted in Fluoromount (Sigma). Slides were air-dried in the dark from 4 h to overnight. Images were acquired using a Zeiss LSM 710 confocal microscope (Zeiss). Primary antibody used and concentrations: anti-GFP antibody (GeneTex), 1:100; anti-RFP antibody (Evrogen), 1:400.
In situ hybridization
In situ hybridization was performed on 24-hpf and 3-dpf-old embryos (TL) as described (Di Donato et al. 2013). For generation of a kif5aa specific antisense-probe, the following primers were used (5’ to 3’): Kif5aa-is-fwd: AGCATGCTCTAATGACGGGTT TTT, Kif5aa-is-rv: GCTGCTCCCCTCATCCTGACCTCTT.

Microscopy
For low magnification imaging, a Leica MZ FLII stereomicroscope (Leica) equipped with a Leica DFC310FX digital camera (Leica) was used. Confocal microscopy was performed using a Zeiss LSM 710 confocal microscope (Zeiss) and a 40× or 20× water immersion or 10× objective. Z volumes were acquired with a 1- to 3-μm resolution and images processed using Adobe Photoshop and Adobe Illustrator software. Three-dimensional reconstructions of Z-volumes were done using Imaris.

Genomic DNA extraction for Southern blot analysis
Genomic DNA was isolated from pools of 20–50 out-crossed embryos harvested 5 dpf. Samples were digested for 1 h at 55°C in 0.5 ml lysis buffer (10 mM Tris, pH 8.0, 10 mM NaCl, 10 mM EDTA, and 2% SDS) with protease K (0.17 mg/ml, Roche Diagnostics) and centrifuged for 10 min at 14,000 rpm. The supernatant was transferred to a phase lock gel tube (Dutscher) and centrifuged for 10 min at 14,000 rpm. One milliliter of 100% ethanol and 10% of 3 M sodium acetate, pH 6.0 were added to the supernatant and centrifuged for 30 min at 14,000 rpm at 4°C. The pellet was washed with 70% ethanol, dried, and resuspended in 100 μL H2O.

Southern blot analysis
Genomic DNA (3–5 μg) was digested overnight with 50 units of HindIII (New England Biolabs, High Fidelity) restriction enzyme. The digested genomic DNA was separated by standard gel electrophoresis on a 1% agarose gel in 1× TAE buffer. Transfer of DNA was done overnight by upward capillarity transfer in 10× SSC to a Hybond N+ membrane (Amersham Biosciences). The membrane was UV cross-linked using a UV cross-linker (Fisher Biotech). A neurod locus-specific probe (565 bp, probe 1) and an E2A-KalTA4-specific probe (491 bp, probe 2) were amplified using the PCR DIG Probe Synthesis Kit (Roche), according to the manufacturer’s protocol. Probes were amplified starting from genomic wild-type DNA. The digested genomic DNA was separated by standard gel electrophoresis and images processed using Adobe Photoshop and Adobe Illustrator software. Three-dimensional reconstructions of Z-volumes were done using Imaris.

Identification of off-target sites and T7EI assay
Potential off-targets of sgRNA eGFP 1 (GGCGAGGGCCATGCA CCTACCG) in the Danio rerio Zv9 assembly were identified using fuzznuc from the EMBOS suite, and no off-targets bearing up to three mismatches were detected. Out of 21 sequences with up to five mismatches, 14 were annotated as part of genes in the UCSC database (Supplemental Table 3). For amplification of these loci and the neurod:eGFP locus, primers listed in Supplemental Table 6 were used.

Genomic DNA was isolated from pools of 25 5-dpf embryos of the Tg(neurod:eGFP) × Tg(UAS:RFP, cry1:eGFP) cross with and without injection of the eGFP-bait-E2A-KalTA4 donor plasmid together with sgRNA eGFP 1 and Cas9. PCR was performed using Phusion Polymerase (New England Biolabs) following the manufacturer's protocol. Five microliters of unpurified PCR product + 5 μL of NEBuffer 2 (2×) (New England Biolabs) were melted and annealed (95°C for 5 min, 95°C to 25°C at −0.5°C/30 sec, and 4°C for 15 min) to form heteroduplex DNA. The annealed DNA was treated (or untreated) with 0.75 units of T7 endonuclease 1 (New England Biolabs) for 20 min at 37°C and run on a 2.4% agarose gel after stopping the reaction by adding 10 μL of Proteinase K (0.4 mg/μL) in 50% sucrose. To check for frequency of indel mutations at the off-target sites off#1 and off#8, PCR amplicons were tailed using Taq Polymerase (Life Technologies), cloned into the pCRII-TOPO (TOPO TA Cloning Kit Dual Promoter, Life Technologies) vector and sent for sequencing. Mutant alleles were identified by comparison to the wild-type unmodified sequence. For detection of the polymorphism at off#7, multiple PCR clones were sent for sequencing and alleles compared. For insertion mapping at the two off-target sites, the primers listed in Supplemental Table 6 were used.

Data access
Sequences of the primers are listed in the Methods and Supplemental Tables 5 and 6. The target sites of the sgRNAs and TALENs are listed in Supplemental Table 4. The TALEN RVD sequences are provided in Supplemental Table 4.

Acknowledgments
We thank A. Shkumatava, H. Baier, and L. Poggi for the Tg(neurod: eGFP), Tg(pou4f3:mGFP), and Tg(vsx2:eGFP) lines, respectively. We thank I. Rentero Rebollo and A. Shkumatava for critical reading of the manuscript. We also thank D. Balciunas for the UAS:RFP plasmid, C. Giovannangeli and members of the Del Bene laboratory for general discussion and comments, and M. Charpentier for the generation of the kif5aa TALEN plasmids and excellent technical help. We thank the members of the Developmental Biology Curie imaging facility (PCT-IBiSA@BDD, UMR3215/U934) for their help and advice with confocal microscopy. The Del Bene laboratory “Neural Circuits Development” is part of the Laboratoire d’Excellence (LABEX) entitled DEEP (ANR-11-LABX-0044), and of the École des Neurosciences de Paris Ile-de-France network. T.O.A. was supported by a Boehringer Ingelheim Fonds Ph.D. fellowship. This work has been supported by an ATIP/AVENIR program starting grant (F.D.B.), ERC-SIG #311159 (F.D.B.), ANR TEFOR (J.P.C.), CNRS, INSERM, Institut Curie and Museum National d’Histoire Naturelle.

Author contributions: T.O.A., J.P.C., and F.D.B. conceived and designed experiments; J.P.C. prepared sgRNA constructs. A.D.C. performed the T7E1 assays. T.O.A. and K.D. built donor constructs, carried out the PCR diagnosis, Southern blotting, and in situ hybridization. T.O.A. performed microinjections and microscopy and...
CRISPR/Cas9 mediated knock-in in zebrafish


Received June 5, 2013; accepted in revised form October 28, 2013.
Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair

Thomas O. Auer, Karine Duroure, Anne De Cian, et al.

*Genome Res.* 2014 24: 142-153 originally published online October 31, 2013
Access the most recent version at doi:10.1101/gr.161638.113

**Supplemental Material**
http://genome.cshlp.org/content/suppl/2013/11/04/gr.161638.113.DC1

**References**
This article cites 47 articles, 14 of which can be accessed free at:
http://genome.cshlp.org/content/24/1/142.full.html#ref-list-1

**Creative Commons License**
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at http://creativecommons.org/licenses/by-nc/3.0/.

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

---

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

© 2014 Auer et al.; Published by Cold Spring Harbor Laboratory Press