

Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system

Cell Research advance online publication 7 April 2015; doi:10.1038/cr.2015.43

Dear Editor,

Animals carrying exogenous genes integrated at specific genomic loci are versatile tools for biological research [1]. Zebrafish (*Danio rerio*), an emerging vertebrate animal model, is widely used in studies on genetics, developmental biology and neurobiology. Although loss-of-function genomic editing for zebrafish has been well developed [2-5], lack of feasible methods for inserting a large exogenous DNA sequence into the zebrafish genome is becoming a bottleneck for zebrafish-relevant research. It was reported that the coding sequence of enhanced green fluorescent protein (EGFP) can be integrated at the zebrafish tyrosine hydroxylase (*th*) locus through TALEN-mediated double-stranded breaks and homologous recombination (HR) with a low efficiency [6]. However, the targeted gene was destroyed and EGFP failed to express [6]. Recently, by using the type II bacterial clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) 9 system (CRISPR/Cas9), two non-HR-based knockin approaches were developed to insert *Gal4* (a transcriptional transactivator) and EGFP into zebrafish genomic loci with a relatively high efficiency [7, 8]. However, the coding sequence of targeted endogenous genes was disrupted or the expression pattern of inserted exogenous genes could not well recapitulate the endogenous ones as insertion occurred within either the exon [7] or *cis*-regulatory elements of targeted genes [8]. These disadvantages will limit their application in neuroscience research. Here, using the CRISPR/Cas9 system, we developed an intron targeting-mediated and HR-independent efficient knockin approach for zebrafish, with which the intactness of the coding sequence and regulatory elements of targeted endogenous genes are maintained.

The Th protein is a rate-limiting enzyme for synthesizing two important neuromodulators, dopamine and noradrenline. As dopamine and noradrenline are synthesized and released by dopaminergic and noradrenergic

neurons, respectively, Th is a specific marker for these cells. We designed a short guide RNA (sgRNA) targeting the last intron of the zebrafish *th* and performed co-injection of sgRNA and mRNA of zebrafish codon-optimized Cas9 (zCas9) into one-cell-stage zebrafish embryo, which yielded a cleavage efficiency of ~83% (Supplementary information, Table S1A and S1B). Next, we designed a donor plasmid *th-P2A-EGFP* consisting of three parts: a left arm, a P2A-EGFP coding sequence, and a right arm (Figure 1A). To retain the full coding sequence of *th*, the left arm begins from the upstream of the 5' side of the sgRNA target site in the last intron, spans the whole last exon E13, and ends at the last base just before the stop codon of *th*. To keep the normal control of Th expression, the right arm includes the stop codon and 3' regulatory elements of *th*. The P2A peptide is a linker for multicistronic expression [9].

We co-injected the donor plasmid, sgRNA and the zCas9 mRNA into one-cell-stage fertilized zebrafish egg. As both the donor plasmid DNA and the last intron of *th* contain sgRNA target site, concurrent cleavage by sgRNA/Cas9 would result in efficient and specific integration of the donor DNA into the *th* locus via non-HR. Indeed, we observed EGFP expression in the brain of injected larvae 3 days post fertilization (dpf) (33/139; Supplementary information, Figure S1A1 and Table S1B). Based on their location and morphology [10], EGFP-expressing cells included dopaminergic neurons in the posterior tubercular (PT), intermediate hypothalamus (HI) and preteectum (Pre), and noradrenergic neurons in the locus coeruleus (LC) and medulla oblongata (MO) (Supplementary information, Figure S1A1). Successful non-HR-mediated insertion of the *th-P2A-EGFP* donor was then verified by PCR using target site- and donor-specific primers and junction sequencing analysis (Supplementary information, Figures S1A2 and S1A3). The specificity of knockin events was further confirmed by *in situ* immunohistochemistry staining, which revealed that EGFP was co-localized with Th in 46 out of

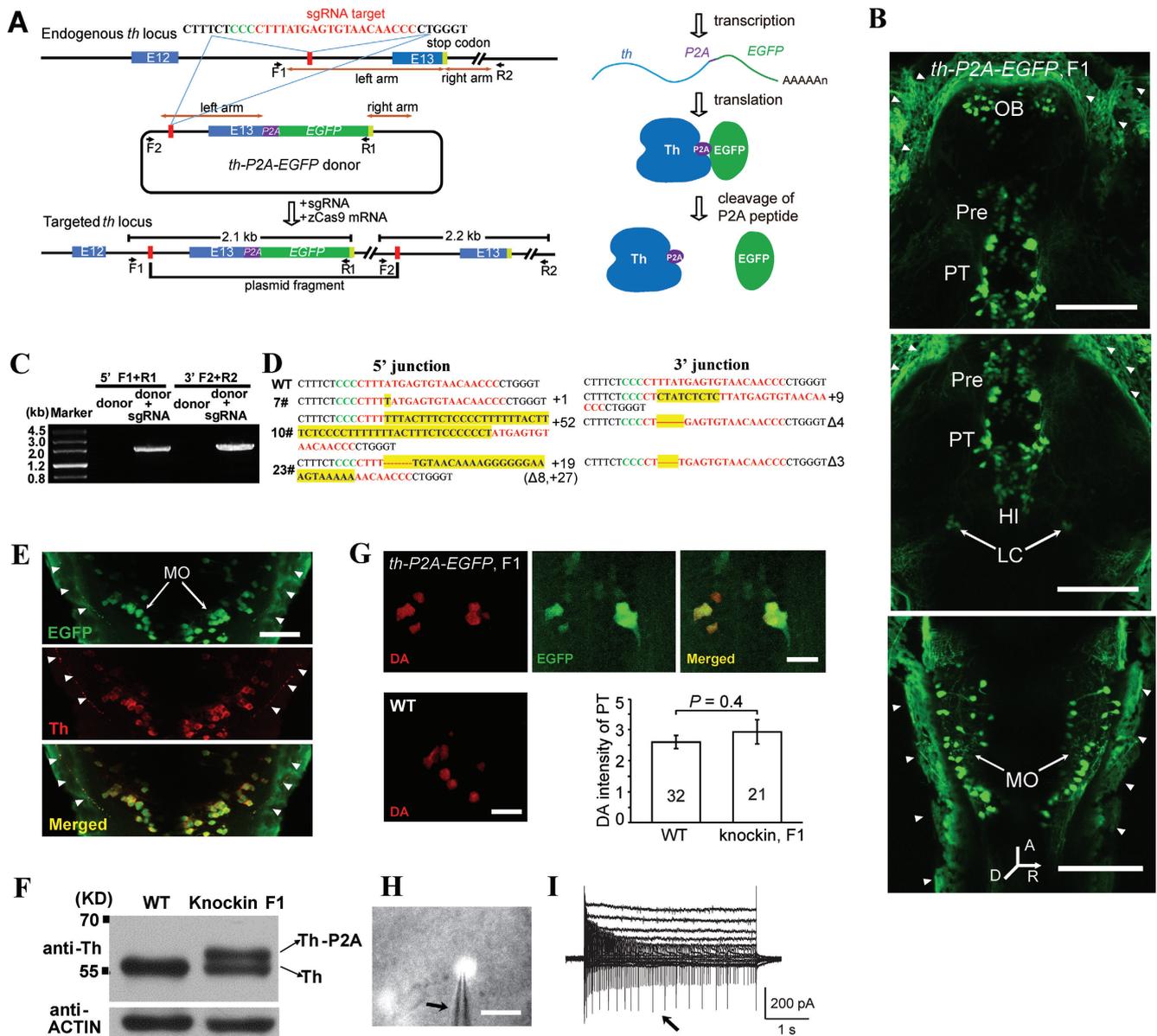


Figure 1 Intron targeting-mediated *EGFP* knockin at the zebrafish *th* locus. **(A)** Schematic of the intron targeting-mediated strategy for generating *EGFP* knockin at the zebrafish *th* locus by using the CRISPR/Cas9 system. The sgRNA target sequence is shown in red and the protospacer adjacent motif (PAM) sequence in green. The left and right arm sequences of the donor plasmid are indicated by the brown lines with double arrows. The left arm is 1 298 bp, and the right arm is 671 bp. The *th*-P2A-*EGFP* cassette was integrated into the *th* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The zebrafish *th* has 13 exons, and E12 and E13 represent the 12th and 13th exons, respectively. Right: the schematic of the mRNA and protein of the targeted *th* gene. **(B)** Representative projected *in vivo* confocal images (dorsal view) of *th*-P2A-*EGFP* knockin F1 larvae at 3 dpf, showing specific *EGFP* expression in various dopaminergic (OB, Pre, PT, and HI) and noradrenergic neurons (LC and MO). The white arrowheads mark non-specific signaling on the skin. A, anterior; D, dorsal; R, right. Scale bar, 50 μ m. **(C)** PCR analysis of the 5' and 3' junctions of F1 progenies from the 7# founder. The F1, R1, F2 and R2 primers are shown in **A**. **(D)** 5' and 3' junction sequences of F1 progenies of three *th*-P2A-*EGFP* knockin F0 founders. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are shown in green and red, respectively. **(E)** Whole-mount *in situ* double immunostaining of *th*-P2A-*EGFP* knockin F1 larvae, showing that *EGFP* signaling (green) co-localizes with Th signals (red). The white arrowheads mark non-specific signaling on the skin. Scale bar, 20 μ m. **(F)** Western blot of the Th expression in WT and heterozygous *th*-P2A-*EGFP* knockin F1 embryos. **(G)** Dopamine (DA) immunostaining of *th*-P2A-*EGFP* knockin F1 (top) and WT larvae (bottom left). Bottom right: DA intensity of DA-positive neurons in WT and knockin F1 larvae. The numbers on the bars represent the numbers of cells examined. Scale bar, 10 μ m. **(H)** Bright-field image showing *in vivo* whole-cell recording of *EGFP*-expressing LC neurons in a homozygous *th*-P2A-*EGFP* knockin F2 larvae. The black arrow indicates the recording microelectrode. Scale bar, 10 μ m. **(I)** Whole-cell currents of an *EGFP*-expressing LC neuron. Under voltage-clamp mode, the neuron was held at -60 mV and voltage pulses from -100 to 30 mV with an interval of 10 mV were applied. Action potential currents (arrow) appear near -30 mV.

48 Th-positive cells in three larvae examined (Supplementary information, Figures S1A4 and S1A5).

To examine the germline transmission of knockin events, 25 embryos showing mosaic expression of EGFP were raised to adulthood. Each of them was then outcrossed to wild-type (WT) zebrafish, and their F1 progenies were screened for EGFP signal. Three F0 founders were identified, and EGFP-positive F1 progenies were produced at rates ranging from 15.5% to 21.1% (Supplementary information, Table S1C). As expected, in comparison with F0 (Supplementary information, Figure S1A1), more EGFP-expressing dopaminergic and noradrenergic neurons were observed in F1 progenies (Figure 1B), including neurons in the olfactory bulb (OB), Pre, PT, HI, LC and MO. PCR and junction sequencing analysis of F1 progenies confirmed the inheritance of the genomic integration of their corresponding F0 founders (Figure 1C and 1D). Immunostaining was also performed in the F1 embryos of *th-P2A-EGFP* knockin fish, and EGFP signal was found to be well co-localized with the Th protein ($98\% \pm 1\%$, mean \pm SEM, in 5 larvae; Figure 1E), suggesting the high specificity of EGFP expression in the stable knockin lines.

As the full reading frame and regulatory elements of *th* were maintained by using this knockin strategy, both the integrity and expression pattern of the gene product should be normal. To examine these points, we extracted the total protein from F1 embryos carrying EGFP knockin at *th* and performed western blot analysis. F1 embryos were heterozygous because they were generated by crossing knockin F0 founders with WT fish. By using a Th antibody, two bands for knockin embryos were detected (Figure 1F). The lower band at around 56 kDa represents the WT Th protein derived from a WT *th* allele. The P2A peptide is about 2 kDa and cleaved between the last two amino acids. If knockin events did not affect the integrity of the Th protein, the cleavage of the Th-P2A-EGFP protein will result in two products: Th-P2A fusion protein (58 kDa) and EGFP protein (Figure 1A). Therefore, the upper band at around 58 kDa indicates the integrity of the Th protein produced from a knockin *th* allele. Furthermore, the expression levels of the WT Th protein and the knockin Th-P2A fusion protein were almost equal (Figure 1F), further suggesting that our knockin strategy does not impair the expression level of the targeted endogenous gene. To examine whether knockin events affect Th functions, we then performed immunostaining of dopamine, the level of which can reflect the activity of Th. The intensities of dopamine signals in dopaminergic neurons were not significantly different between knockin F1 and WT embryos ($P = 0.4$; Figure 1G), suggesting that Th function is not affected by knockin events.

To examine the physiological normality of neurons carrying targeted integration, *in vivo* whole-cell recording was subsequently performed in homozygous *th-P2A-EGFP* knockin F2 larvae (Figure 1H). EGFP-expressing neurons exhibited a normal intrinsic membrane property, as reflected by outwardly rectifying whole-cell currents (Figure 1I).

To extend the application of our knockin strategy to other exogenous genes, we generated knockin fish carrying the transactivator protein *Gal4* at the *th* locus by using the same strategy, in which only the *EGFP* coding sequence was replaced with the *Gal4* sequence (*th-P2A-Gal4*; Supplementary information, Figure S1B1). After injection of *Gal4* knockin-relevant elements into fertilized eggs of Tg(UAS:GCaMP5) transgenic zebrafish, integration events were visualized by the expression of GCaMP5 in dopaminergic or noradrenergic neurons (Supplementary information, Figure S1B2). As GCaMP5 is a genetically encoded calcium indicator, we could observe mechanical stimulus-induced calcium responses in neurons by puffing water to the fish tail through a micropipette (Supplementary information, Figure S1B3). We also injected the *Gal4* knockin-relevant elements into WT fish to screen for *th-P2A-Gal4* knockin founders. As the Gal4 protein has no fluorescence, we raised the injected knockin embryos to adulthood without prior selection and crossed these adults with Tg(UAS:Kaede) transgenic fish. Two founders were identified among the total of 28 injected fish (Supplementary information, Table S1D), as evidenced by the fact that dopaminergic neurons were labeled by Kaede in their progenies, which were produced at a mean rate of $\sim 7\%$ (Supplementary information, Figure S1B4 and Table S1D). Successful insertion of the *th-P2A-Gal4* donor was then verified by PCR and junction sequencing analysis in F1 progenies (Supplementary information, Figures S1B5 and S1B6).

It was reported that the CRISPR/Cas9 system shows a high frequency of off-target (OT) cleavage in human cell lines, and the specificity of Cas9 targeting can tolerate up to three base pair (bp) mismatches between a sgRNA and its target DNA [11]. We therefore searched all zebrafish genomic loci containing up to 3-bp mismatches in comparison with the coding sequence of the *th* sgRNA, and found three potential OT sites. PCR and sequencing analysis of those potential OT sites in the genome of injected WT embryos or *th-P2A-EGFP* knockin F1 embryos did not reveal indels (Supplementary information, Table S1E), suggesting a low OT rate associated with our knockin strategy.

The applicability of our knockin strategy was further validated by targeting other endogenous genes specifically expressed in different types of cells, as exemplified

by the integration of *EGFP* into the zebrafish *tryptophan hydroxylase 2 (tph2)*, *glial fibrillary acidic protein (gfap)*, and *flk1* loci. These *EGFP* insertions resulted in the specific labeling of serotonergic neurons, glia and vascular endothelial cells, respectively (Supplementary information, Figures S1C-S1E, and Table S1A and S1B). It is worth noticing that, in the case of the *tph2* knockin, the second last intron was selected for targeting, indicating that the last intron is the first but not the only choice for targeting. Furthermore, by replacing the *P2A* in the *gfap-P2A-EGFP* plasmid with a flexible serine-serine linker sequence, we succeeded in fusing an EGFP tag to endogenous Gfap (Supplementary information, Figure S1F), demonstrating that our knockin strategy can also be used to tag endogenous proteins.

Taking advantage of both the HR for donor design and the non-HR for donor integration, we developed a novel CRISPR/Cas9-mediated intron-targeting knockin strategy, by which knockin zebrafish can be efficiently generated without disruption of targeted endogenous genes. Compared with HR, error-prone non-homologous end joining (NHEJ)-involved non-HR knockin for zebrafish has two advantages. First, NHEJ is at least 10-fold more active than HR during early zebrafish development [12]. Second, unlike HR, NHEJ does not need the precise homology between the parent zebrafish and the targeting donor, avoiding time-consuming screening and genotyping of parent animals. More importantly, to maintain the integrity of targeted endogenous genes, we designed sgRNAs targeting introns, so that NHEJ-mediated indel mutations do not change the reading frame of targeted genes. In addition, intron targeting also theoretically increases the rate of in-frame insertion up to 3-fold in comparison with exon-based targeting. Furthermore, we artificially added the endogenous genome sequence spanning from the sgRNA target site to the 3' intergenic region into donor plasmids. Therefore, the predicted forward ligation of the donor into the targeted locus retains the original reading frame and both 5' and 3' regulatory elements of targeted genes. Taken together, this strategy has two advantages: (1) inserted exogenous genes can faithfully recapitulate the expression pattern of targeted endogenous genes; (2) the expression and function of targeted endogenous genes are maintained. Thus, the readiness, high efficiency and targeted gene integrity maintenance make our strategy an applicable knockin approach for zebrafish and even other organisms.

Acknowledgments

We thank Drs Bo Zhang, Hui Yang and Filippo Del Bene for comments on the manuscript and Dr Bo Zhang for providing zCas9. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02040003), the National Basic Research Program of China (973 Program; 2011CBA00400, 2012CB945101), the National Outstanding Young Scientist Program of the National Natural Science Foundation of China (31325011), Shanghai Subject Chief Scientist Program of the Science and Technology Commission of Shanghai (14XD1404100), and the Postdoctoral Research Program of Shanghai Institutes for Biological Sciences (2014KIP307).

Jia Li^{1,*}, Bai-bing Zhang^{1,2,*}, Yong-gang Ren¹, Shan-ye Gu^{1,4}, Yuan-hang Xiang³, Cheng Huang⁴, Jiu-lin Du^{1,2,3}

¹Institute of Neuroscience, State Key Laboratory of Neuroscience, Center for Excellence in Brain Science, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China; ²Graduate School, University of Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China; ³School of Life Science and Technology, ShanghaiTech University, 319 Yue-Yang Road, Shanghai 200031, China; ⁴Drug Discovery Laboratory, School of Pharmacy, Shanghai University of Traditional Chinese Medicine, 1200 Cai-Lun Road, Shanghai 201203, China

*These two authors contributed equally to this work.

Correspondence: Jiu-lin Du

Tel: 86-21-5492-1825; Fax: 86-21-5492-1735

E-mail: forestdu@ion.ac.cn

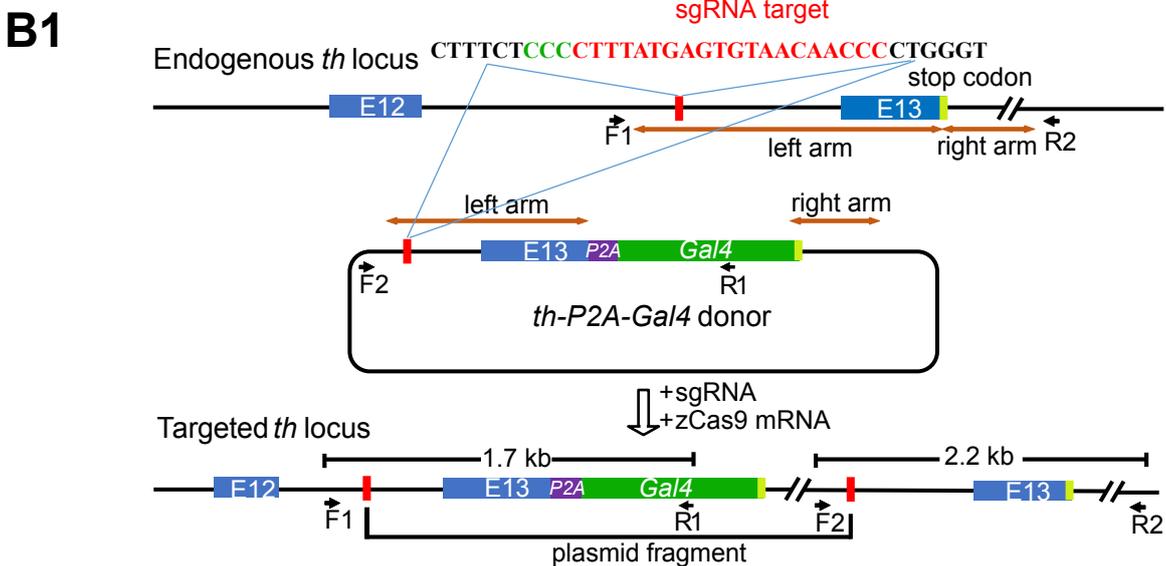
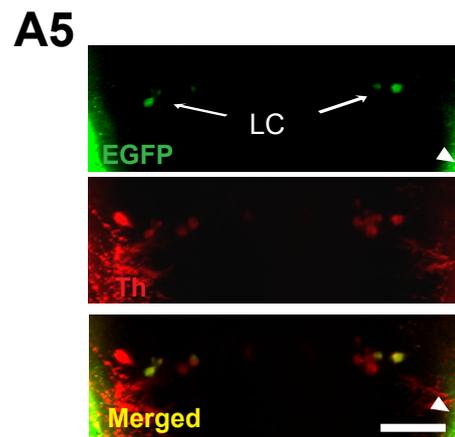
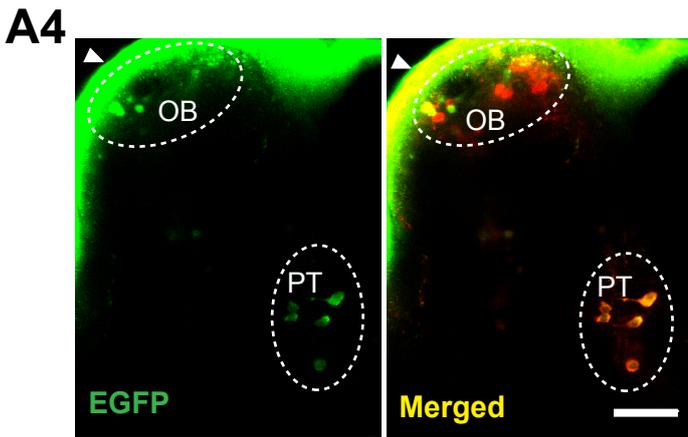
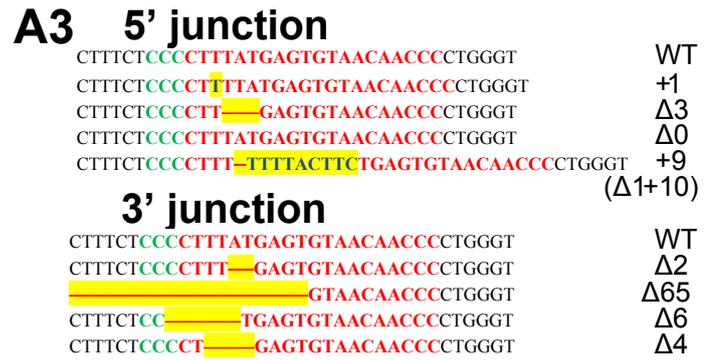
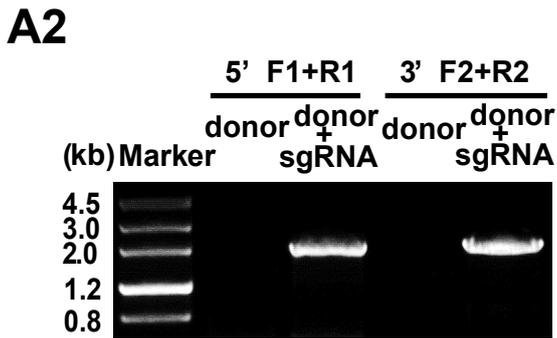
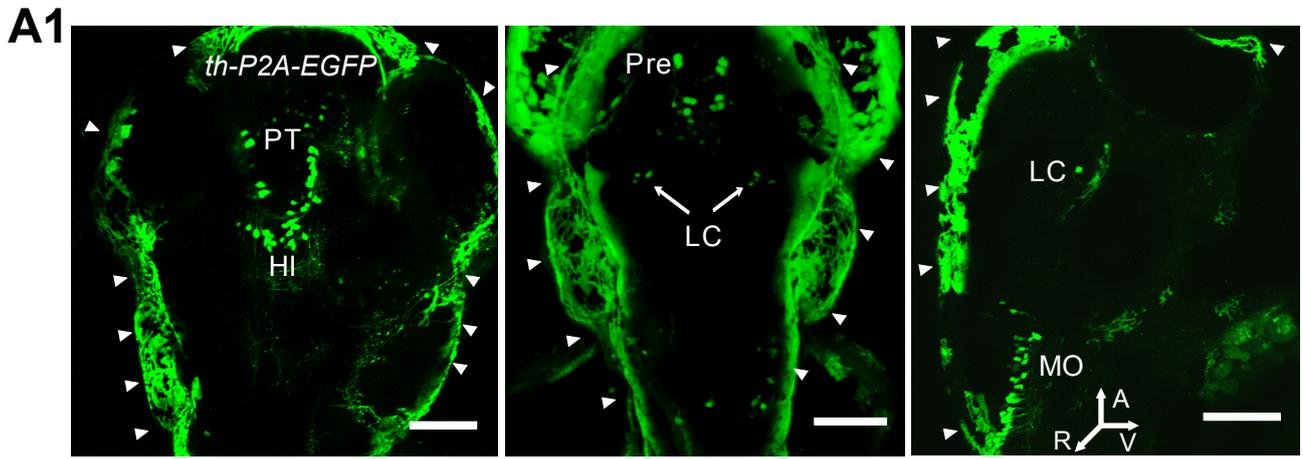
References

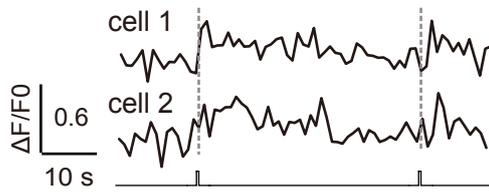
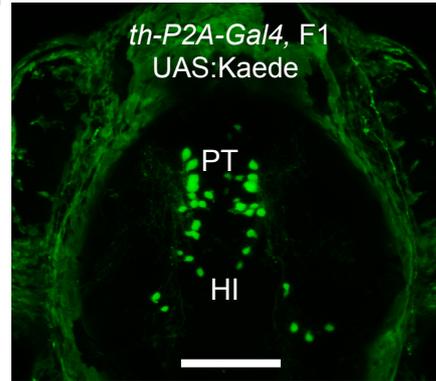
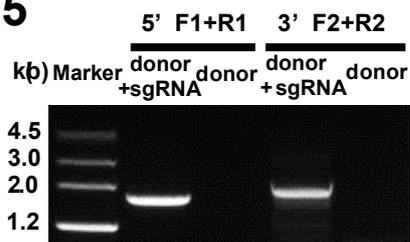
- 1 Capecchi MR. *Nat Rev Genet* 2005; **6**:507-512.
- 2 Huang P, Xiao A, Zhou M, et al. *Nat Biotechnol* 2011; **29**:699-700.
- 3 Bedell VM, Wang Y, Campbell JM, et al. *Nature* 2012; **491**:114-118.
- 4 Chang N, Sun C, Gao L, et al. *Cell Res* 2013; **23**:465-472.
- 5 Hwang WY, Fu Y, Reyon D, et al. *Nat Biotechnol* 2013; **31**:227-229.
- 6 Zu Y, Tong X, Wang Z, et al. *Nat Methods* 2013; **10**:329-331.
- 7 Auer TO, Duroure K, De Cian A, et al. *Gen Res* 2014; **24**:142-153.
- 8 Kimura Y, Hisano Y, Kawahara A, et al. *Scientific Rep* 2014; **4**:6545.
- 9 Kim JH, Lee SR, Li LH, et al. *PLoS One* 2011; **6**:e18556.
- 10 Wen L, Wei W, Gu W, et al. *Dev Biol* 2008; **314**:84-92.
- 11 Fu Y, Foden JA, Khayter C, et al. *Nat Biotechnol* 2013; **31**:822-826.
- 12 Dai J, Cui X, Zhu Z, et al. *Int J Biol Sci* 2010; **6**:756-768.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>



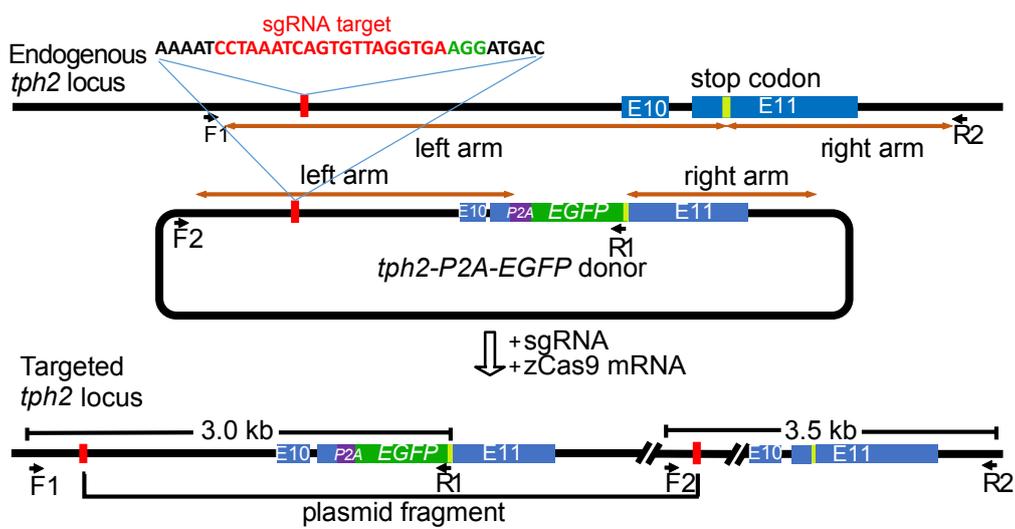
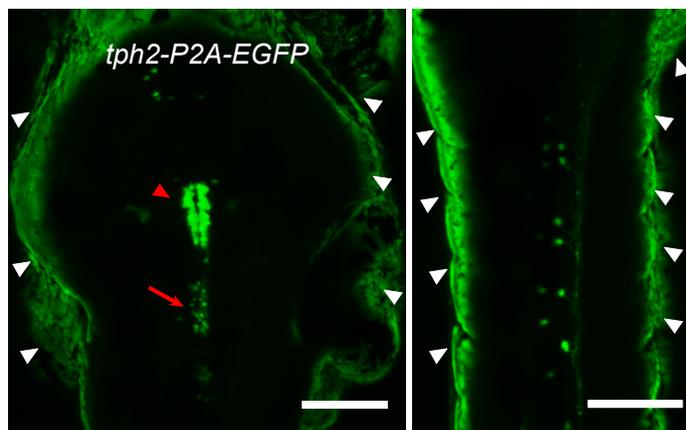
B2**B3****B4****B5****B6**

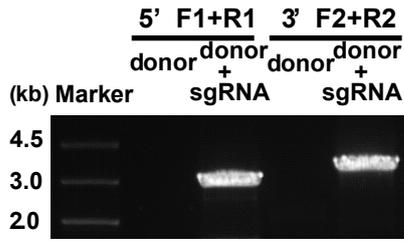
5' junction

WT CTTTCTCCCTTTATGAGTGTAAACAACCCCTGGGT
 5# CTTTCTCCCTTATGAGTGTAAACAACCCCTGGGT Δ2
 8# CTTTCTCCCTGGGTGTAAACAACCCCTGGGT Δ4
 (Δ6, +2)

3' junction

WT CTTTCTCCCTTTATGAGTGTAAACAACCCCTGGGT
 5# CTTTCTCCCTTGTAACAACCCCTGGGT Δ8
 8# CCCCCTTGTAGTGTAAACAACCCCTGGGT Δ7

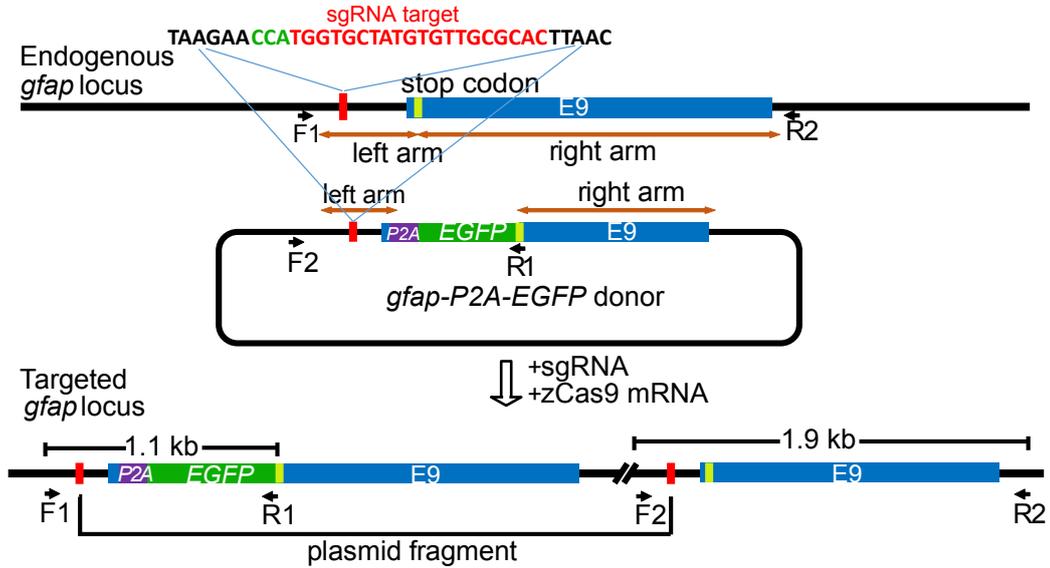
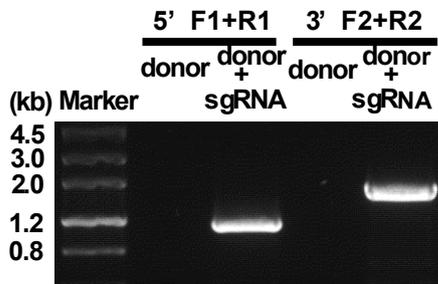
C1**C2**

C3**C4** 5' junction

CAGAAAACCTGAGCATACTGACTGAGAGGTTGTTCAA WT
 CAGAAAACCTGAGCATACTGAGTTTCTCAGCATACTGAGAGGTTGTTCAA +13
 TTTCCAA Δ38
 CAGAAAACCTGAGCATACTGA—GCATACAGAGGTTGTTCAA +3(Δ3,+6)

3' junction

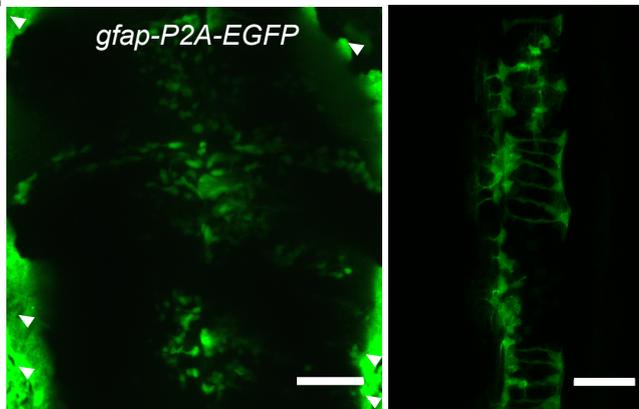
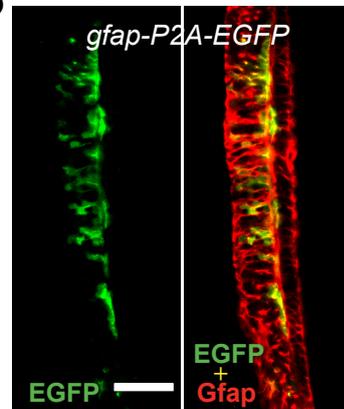
CAGAAAACCTGAGCATACTGACTGAGAGGTTGTTCAA WT
 CAGAAAACCTGAGCATACTGAC—AGAGGTTGTTCAA Δ2
 CAGAAAACCTGAGCATACT—GAGAGGTTGTTCAA Δ5
 GTGTTCAA Δ63

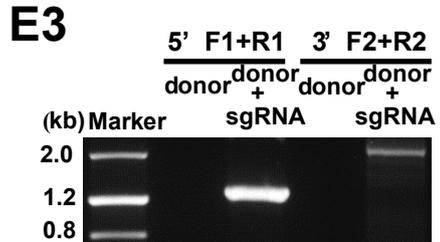
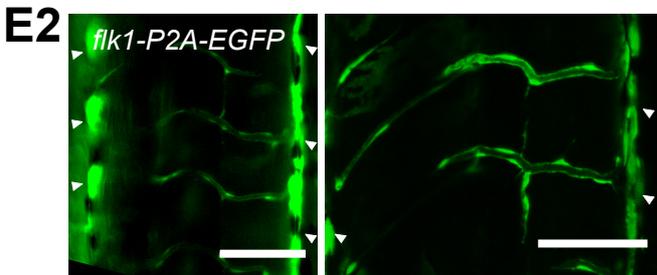
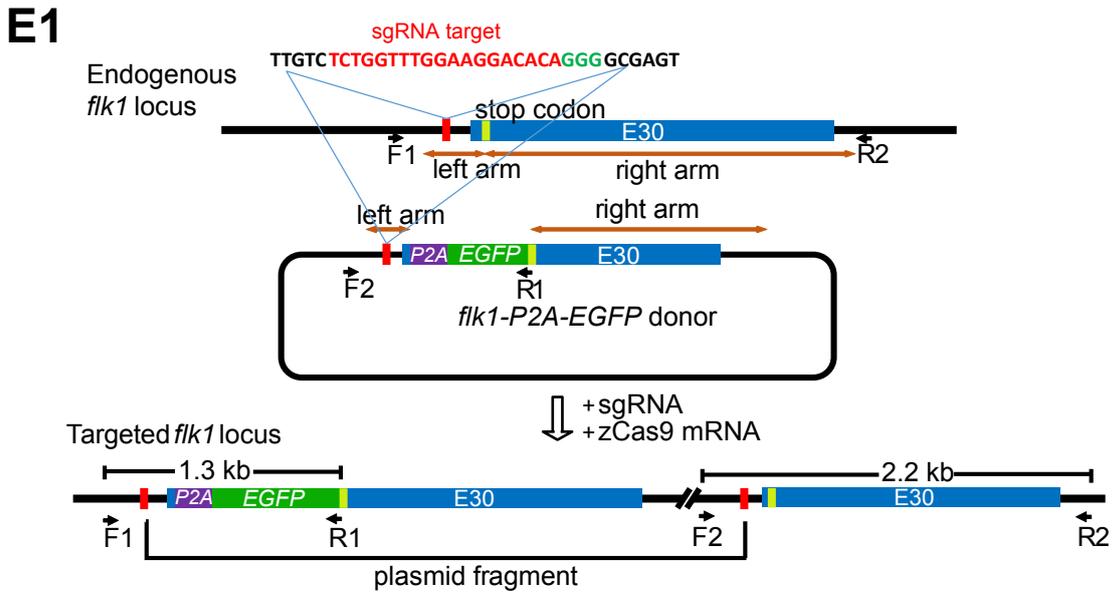
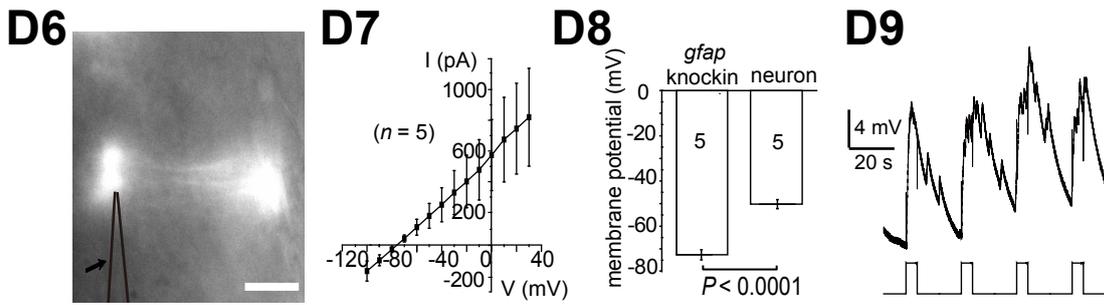
D1**D2****D3** 5' junction

AAGAACCATGGTGCTATGTGTTGCGCACTTAAC WT
 AAGAACCATG—TATGTGTTGCGCACTTAAC Δ4
 AAGAACCATGGGACTGCTATGTGTTGCGCACTTAAC +3
 AAGAACCAT—GTTGCGCACTTAAC Δ10
 AAGAACCATG—GGGTTGCGCACTTAAC Δ8(Δ11,+3)

3' junction

CATAAGAACCATGGTGCTATGTGTTGCGCACTTAAC WT
 CATAAGAACCA—TGCTATGTGTTGCGCACTTAAC Δ3
 CATAAGAA—CTATGTGTTGCGCACTTAAC Δ8
 CTATGTGTTGCGCACTTAAC Δ21
 CATAAGAACCAT—TGCTATGTGTTGCGCACTTAAC +1(Δ1,+2)

D4**D5**



E4

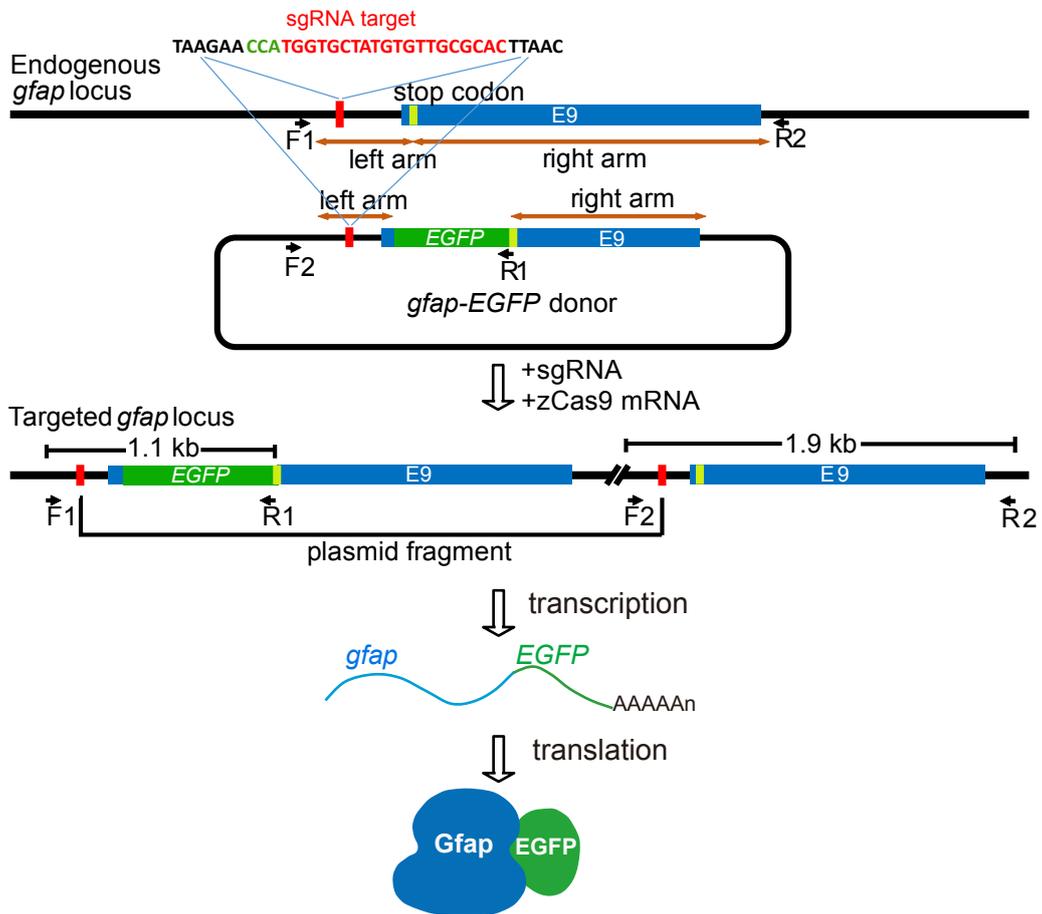
5' junction

CTTGTCCTGGTTTGAAGGACACAGGGCGAGT	WT
CTTGTCCTGGTTTGAAGGAC-CAGGGCGAGT	Δ1
CTTGTCCTGGTTTGAAGGACA-GGGCGAGT	Δ2
CTTGTCCTGGTTTGAAGGACACAGGGCGAGT	Δ38
CTTGTCCTGGTTTGAAGGACACAGGGCGAGT	Δ0

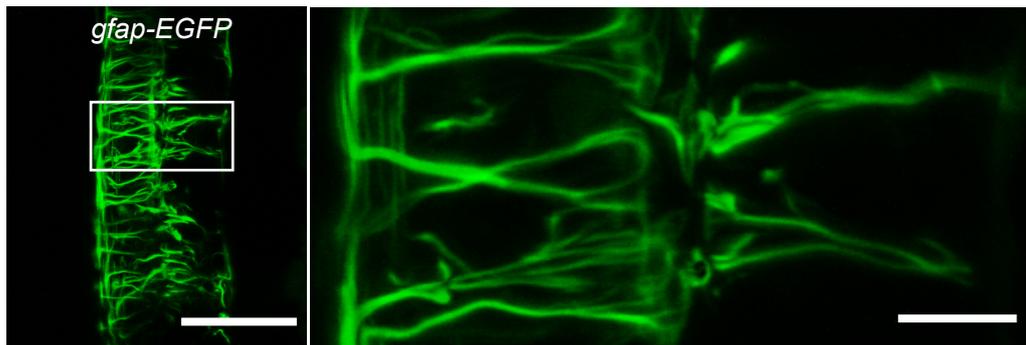
3' junction

TTGTCCTGGTTTGAAGGACACAGGGCGAGT	WT
TTGTCCTGGTTTGAAGGACAAACAGGGCGAGT	+3
TTGTCCTGGTTTGAAGGAC-AGGGCGAGT	Δ14
TTGTCCTGGTTTGAAGGAC-GAGT	Δ8
TTGTCCTGGTTTGAAGGACAGTCCCCACGCTTCAGGGCGAGT	+13

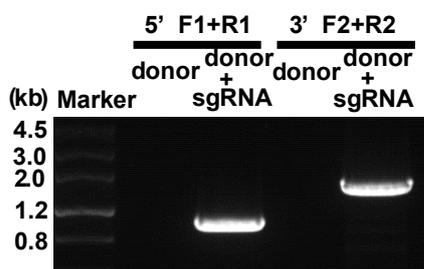
F1



F2



F3



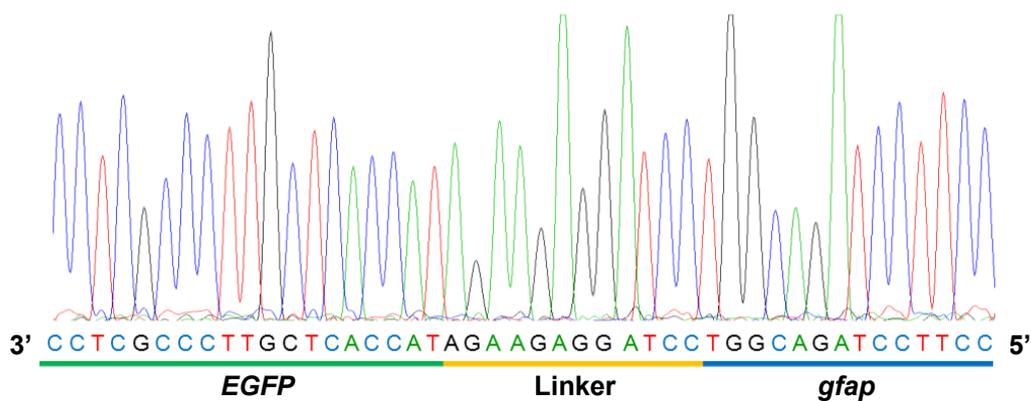
F4 5' junction

AAGAACCATGGTGCTATGTGTTGCGCACTTAAC WT
 AAGAACCATG GCTATGTGTTGCGCACTTAAC Δ2
 AAGAACCATG GAACCACTATGTGTTGCGCACTTAAC +3(Δ3,+6)
 AAGAACCATG TGTGGCTAAGAGGCTATGTGTTGCGCACTTAAC +11(Δ2,+13)

3' junction

AAGAACCATGGTGCTATGTGTTGCGCACTTAAC WT
 AAGAACCAT AAGAACTATGTGTTGCGCACTTAAC +1(Δ4,+5)
 AAGAACCATGGTGCTATGTGTTGCGCACTTAAC Δ0
 AAGAACCATGGT ACCGTACCGTACCTATGTGTTGCGCACTTAAC +11(Δ1,+12)

F5



Supplementary information, Figure S1 (A) Transient *EGFP* knockin at the zebrafish *th* locus. **(A1)** Projected in vivo confocal images of three transient *th-P2A-EGFP* knockin larvae at 3 dpf, showing specific EGFP expression in various dopaminergic (PT, HI and Pre) and noradrenergic neurons (LC and MO). The white arrowheads mark non-specific signaling on the skin. Left and Middle: dorsal view; Right: lateral view. HI, intermediate hypothalamus; LC, locus coeruleus; MO, medulla oblongata; Pre, pretectum; PT, posterior tuberculum. A, anterior; R, right; V, ventral. Scales, 50 μ m. **(A2)** PCR analysis of the 5' and 3' junctions of the targeted *th* locus. The primers of F1, R1, F2, and R2 are showed in **Figure 1A**. A 2.1-kb band was amplified by using the 5' junction primers F1 and R1, and a 2.2-kb band was amplified by using the 3' junction primers F2 and R2. **(A3)** Sequences of the 5' and 3' junctions in four representative targeted knockin events. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(A4, A5)** Whole-mount *in situ* double immunostaining of transient *th-P2A-EGFP* knockin larvae, showing that EGFP signaling (green) co-localizes with Th signaling (red) in the olfactory bulb (OB) **(A4)** and LC **(A5)**. The white arrowheads mark non-specific signaling on the skin. Scales, 20 μ m. **(B)** Intron targeting-mediated *Gal4* knockin at the zebrafish *th* locus. **(B1)** Schematic of the intron targeting-mediated strategy for generating *Gal4* knockin at the zebrafish *th* locus by using the CRISPR/Cas9 system. The sgRNA target sequence is showed in red and the protospacer adjacent motif (PAM) sequence in green. The left and right arm sequences of the donor plasmid are indicated by the brown lines with double arrows. The left arm is 1298 bp, and the right arm is 671 bp in length. The *th-P2A-Gal4* cassette was integrated into the *th* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The zebrafish *th* has 13 exons, and E12 and E13 represent the 12th and 13th exons, respectively. **(B2)** Confocal image of a 3-dpf Tg(UAS:GCaMP5) larva carrying transient *th-P2A-Gal4* knockin, showing GCaMP5 expression in 4 neurons of bilateral LC. Scale, 20 μ m. **(B3)** In vivo two-photon calcium imaging showing Ca²⁺ responses of the two GCaMP5-expressing LC neurons (shown in **B2**) evoked by repeated mechanical stimuli (bottom) with an interval of 40 s at the fish tail. **(B4)** Projected in vivo confocal image of a 3-dpf F1 progeny of *th-P2A-Gal4* founders crossed with Tg(UAS:Kaede) fish, showing Kaede expression in dopaminergic neurons. Scale, 50 μ m. **(B5)** PCR analysis of the 5' and 3' junctions of *th-P2A-Gal4* founders. The F1, R1, F2, and R2 primers are showed in **(B1)**. **(B6)** Sequences of the 5' and 3' junctions in two *th-P2A-Gal4* founders. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(C)** Intron targeting-mediated *EGFP* knockin at the zebrafish *tph2* locus. **(C1)** Schematic of the intron targeting-based strategy for generating *EGFP* knockin at the zebrafish *tph2* locus. In the *tph2-P2A-EGFP* donor plasmid, the left arm is 2187 bp, and the right arm is 1081 bp in length. The *tph2-P2A-EGFP* cassette was integrated into the *tph2* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The second last intron was selected for targeting. The zebrafish *tph2* has 11 exons, and E10 and E11 represent the 10th and 11th exons, respectively. **(C2)** Representative projected confocal images (dorsal view) of a 3-dpf *tph2-P2A-EGFP* knockin larva, showing EGFP expression in the Raphe nucleus (RN, left) and the spinal cord (right). The red arrowhead and arrow indicate neurons in the rostral and caudal Raphe nuclei, respectively. The white arrowheads mark non-specific signaling on the skin. Scales, 50 μ m. **(C3)** PCR analysis of the 5' and 3' junctions of the targeted *tph2* locus. The F1, R1, F2, and R2 primers are shown in **(C1)**. A 3.0-kb band was amplified by using the 5' junction primers F1 and R1, and a 3.5-kb band was amplified by using the 3' junction primers F2 and R2. **(C4)** Sequences of the 5' and 3' junctions in three representative knockin events. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(D)** Intron targeting-mediated *EGFP* knockin at the zebrafish *gfap* locus. **(D1)** Schematic of the intron targeting-based strategy for generating *EGFP* knockin at the zebrafish *gfap* locus. In the *gfap-P2A-EGFP* donor plasmid, the left arm is 243 bp, and the right arm is

1329 bp in length. The *gfap-P2A-EGFP* cassette was integrated into the *gfap* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The zebrafish *gfap* has 9 exons and E9 represents the 9th exon of *gfap*. **(D2)** PCR analysis of the 5' and 3' junctions of the targeted *gfap* locus. The F1, R1, F2, and R2 primers are shown in **(D1)**. A 1.1-kb band was amplified by using the 5' junction primers F1 and R1, and a 1.9-kb band was amplified by using the 3' junction primers F2 and R2. **(D3)** Sequences of the 5' and 3' junctions in four representative targeted knockin events. The indel is highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(D4)** Representative projected confocal images (dorsal view) of a 3-dpf *gfap-P2A-EGFP* knockin larva, showing EGFP expression in the brain (left) and the spinal cord (right). The white arrowheads mark non-specific signaling on the skin. Scales: 50 μ m (left), 30 μ m (right). **(D5)** Whole-mount *in situ* double immunostaining of a *gfap-P2A-EGFP* knockin larva, showing that EGFP (green) co-localizes with Gfap (red). Scale, 50 μ m. **(D6)** Bright-field image showing *in vivo* whole-cell recording of EGFP-expressing glia in the spinal cord of a 3-dpf *gfap-P2A-EGFP* knockin larva. The arrow indicates the recording microelectrode. Scale, 10 μ m. **(D7)** Mean I-V curve of five EGFP-expressing glial cells. **(D8)** Resting membrane potentials of EGFP-expressing glia and nearby neurons. The numbers on the bars represent the numbers of cells examined. **(D9)** Sluggish depolarization responses of an EGFP-expressing glia evoked by 5-s light flashes (bottom) with an interval of 25 s. **(E)** Intron targeting-mediated *EGFP* knockin at the zebrafish *flk1* locus. **(E1)** Schematic of the intron targeting-based strategy for generating EGFP knockin at the zebrafish *flk1* locus. In the *flk1-P2A-EGFP* donor plasmid, the left arm is 333 bp, and the right arm is 1812 bp in length. The *flk1-P2A-EGFP* cassette was integrated into the *flk1* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The zebrafish *flk1* has 30 exons, and E30 represent the 30th exon. **(E2)** Representative projected confocal images (lateral view) of two 3-dpf *flk1-P2A-EGFP* knockin larvae, showing EGFP expression in trunk segmental vessels. The arrowheads indicate non-specific signaling on the skin. Scales, 50 μ m. **(E3)** PCR analysis of the 5' and 3' junctions of the targeted *flk1* locus. The F1, R1, F2, and R2 primers are shown in **(E1)**. A 1.3-kb band was amplified by using the 5' junction primers F1 and R1, and a 2.2-kb band was amplified by using the 3' junction primers F2 and R2. **(E4)** Sequences of the 5' and 3' junctions in four representative targeted knockin events. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(F)** Intron targeting-mediated knockin of an *EGFP* tag at the zebrafish *gfap* locus. **(F1)** Schematic of the intron targeting-based strategy for generating fused *gfap-EGFP* knockin alleles. In comparison with the P2A strategy (see Supplementary information, Figure S1D), a flexible serine-serine linker sequence was used to replace the P2A sequence in the donor plasmid (*gfap-EGFP*). The left arm is 243 bp, and the right arm is 1329 bp in length. The *gfap-EGFP* cassette was integrated into the *gfap* locus after co-injection of the donor with the sgRNA and zCas9 mRNA. The zebrafish *gfap* has 9 exons and E9 represents the 9th exon of *gfap*. **(F2)** Representative projected confocal images (dorsal view) of a 3-dpf *gfap-EGFP* fused knockin larva, showing EGFP expression in the spinal cord. Right, enlarged view of the area outlined by the rectangle in the left. Scales: 50 μ m (left), 10 μ m (right). **(F3)** PCR analysis of the 5' and 3' junctions of the targeted *gfap* locus. The F1, R1, F2, and R2 primers are shown in **(F1)**. A 1.1-kb band was amplified by using the 5' junction primers F1 and R1, and a 1.9-kb band was amplified by using the 3' junction primers F2 and R2. **(F4)** Sequences of the 5' and 3' junctions in three representative knockin events. The indel mutations are highlighted in yellow, and the PAM and sgRNA target sequences are showed in green and red, respectively. **(F5)** Sequence of the fusion region of *gfap* with *EGFP*.

Supplementary information, Table S1A sgRNA-induced cleavages at zebrafish *th*, *tph2*, *gfap* and *flk1* loci.

sgRNA cleavage efficiency at the <i>th</i> locus (10/12)		
CTTTCTCCCTTTATGAGTGTAACAACCCCTGGGT	WT	× 2
CTTTCTCCCTTCTCCCCGGC_____GAGTGTAATGACCCCTGGGT	+6 (Δ8, +14)	
CTTTCTCTCC_____TGAGTGTAACAACCCCTGGGT	Δ3 (Δ5, +2)	
CTTTCTCCCTTTTTTTTTT_____GAGTGTAACAACCCCTGGGT	+6 (Δ2, +8)	
CTTTCTCCCTTGAGAAATTATGAGTGTAACAACCCCTGGGT	+4 (Δ1,+5)	× 2
CTTTCTCCCT_____TGAGTGTAACAACCCCTGGGT	Δ3	× 2
CTTTCTCCCT_____TAACAACCCCTGGGT	Δ9	× 3
sgRNA cleavage efficiency at the <i>tph2</i> locus (7/11)		
AGAAAACCTGAGCATACTGACTGAGAGGTGTTTCAA	WT	× 4
AGAAAACCTGAGCATACTGACTTGAGAGGTGTTTCAA	+1	× 2
_____CTGAGAGGTGTTTCAA	Δ81	
AGAAAACCTGAGCATA_____GAGAGGTGTTTCAA	Δ4 (Δ6, +1)	
AGAAAACCTGAGCATACTGAC_____AGAGGTGTTTCAA	Δ2	× 2
AGAAAACCTGAGCAGACAGGAAAACCT_____GAGGTGTTTCAA	(Δ10, +13)	
sgRNA cleavage efficiency at the <i>gfap</i> locus (6/15)		
AAGAACCATGGTGCTATGTGTTGCGCACTTAAC	WT	× 9
AAGAACCATGGTTCACCATGGCTATGTGTTGCGCACTTAAC	+8	
AAGAACCATGGTGAACCATGCTATGTGTTGCGCACTTAAC	+7	
AAGAA_____CACGTGGCTATGTGTTGCGCACTTAAC	Δ1 (Δ7, +6)	
AAGAACCATGG_____TATGTGTTGCGCACTTAAC	Δ3	
AAGAACCATGG_____CTATGTGTTGCGCACTTAAC	Δ2	
AAGAAC_____ATGTGTTGCGCACTTAAC	Δ8	
sgRNA cleavage efficiency at the <i>flk1</i> locus (3/8)		
TTGTCTCTGTTTGAAGGACACAGGGCGAGTAA	WT	× 5
TTGTCTCTGTTTGAAGGA_____GGCGAGTAA	Δ6	
TTGTCTCTGTTTGAAGGACACAGTAAGGGCGAGTAA	+4	
TTGTCTCTG_____	Δ40	

Supplementary information, Table S1B Knockin efficiency at zebrafish *th*, *tph2*, *gfap* and *flk1* loci.

Targeted locus	Cas9 efficiency	Donors	Rate of EGFP-positive embryos
<i>th</i>	83.3% (10/12)	<i>th</i> -P2A-EGFP	23.7% (33/139)
<i>tph2</i>	63.6% (7/11)	<i>tph2</i> -P2A-EGFP	12.0% (6/50)
<i>gfap</i>	40.0% (6/15)	<i>gfap</i> -P2A-EGFP	35.6% (78/219)
		<i>gfap</i> -EGFP	35.2% (31/88)
<i>flk1</i>	37.5% (3/8)	<i>flk1</i> -P2A-EGFP	16.9% (14/83)

Supplementary information, Table S1C Ratio of EGFP-positive F1 progenies of three *th-P2A-EGFP* knockin founders.

Founder	Number of positive F1 embryos	Number of F1 progenies screened	Rate of positive F1 embryos
7#	36	171	21.1%
10#	53	342	15.5%
23#	11	69	15.9%

Supplementary information, Table S1D Ratio of Gal4-positive F1 progenies of two *th-P2A-Gal4* knockin founders.

Founder	Number of positive F1 embryos	Number of F1 progenies screened	Rate of positive F1 embryos
5#	1	36	2.8%
8#	6	50	12.0%

Supplementary information, Table S1E Off-target analysis of the Cas9 for the *th* locus.

Site name	Sequence	Indel mutation frequency	Coordinate
Target <i>th</i>	GGGTTGTTACTCATAAAGGGG	/	Chr:25 24779734-24779756
OT1	GGGTgGTTACTCATAttGAGG	0/10	Chr:23 5872409-5872431
OT2	GtGTTtTTgCACTCATAAAGGGG	0/10	Chr:17 50862552-50862574
OT3	GtGTTGTgACACTaATAAAGTGG	0/10	Chr:24 14631093-14631115

Supplementary information, Table S1F Sequences of the primers for amplifying the left and right arms of donor plasmids.

<i>th</i>-knockin	
left arm F	cga ggtacc GCGATTACATGGCGTAGTACGCT
left arm R	cga ggatcc AGCCAACACATTCAGGGCAT
right arm F	cga accggt TTCTGCTTAATCATTCTCTATGC
right arm R	cga gtcgac ATGGGAACCTCCTAGTACCTTAA
<i>tph2</i>-knockin	
left arm F	cga ggtacc GGCCGATTATTCTGATAAGTAGCCC
left arm R	cga ggatcc GATACCGAGGTATTTGTTTC
right arm F	cga accggt TAAACAAACACAGTCCCTCAAC
right arm R	cga gtcgac GCTGCTTTGACGCAATCTGC
<i>gfap</i>-kncokin	
left arm F	cga ggtacc TAAGCCGGGAATCGAACTCGGA
left arm R	cga ggatcc TGGCAGATCCTTCCTCTCCGT
right arm F	cga accggt TAATGAGGCTCAGACACTGGC
right arm R	cga gtcgac ATTCACAAATACAGGCAAATGCAG
<i>flk1</i>-knockin	
left arm F	cga ggtacc TCTTAATTTGGGTTATCTTGTC
left arm R	cga ggatcc GACGGGTGGTGTGGAGTAACG
right arm F	cga accggt TGACATGCTCTGCGCTGGAAAAGC
right arm R	cga gtcgac GATGGAGGAGGCAAGAAGGCTTTC

Supplementary information, Table S1G Sequences of the primers for PCR analysis of the 5' and 3' junctions of the targeted genomic loci.

<i>th</i>-EGFP knockin	
F1	ATTCCGCCGAAACGCAGTAAAGA
R1	CTTGTACAGCTCGTCCATGCC
F2	TGTAAAACGACGGCCAGT
R2	AGTGTACCGTAAAGGTGTTA
<i>th</i>-Gal4 knockin	
F1	ATTCCGCCGAAACGCAGTAAAGA
R1	ATATCAGTCTCCACTGAAGCCAA
F2	TGTAAAACGACGGCCAGT
R2	AGTGTACCGTAAAGGTGTTA
<i>tph2</i>-EGFP knockin	
F1	AGTCTGGGTGAGGCATAAGAC
R1	CTTGTACAGCTCGTCCATGCC

F2	TGTAAAACGACGGCCAGT
R2	TGCATCGCAACAATGCATCA
<i>gfap</i>-EGFP knockin	
F1	TGGTGTAGGGCAGTGGAGGTTACA
R1	CTTGTACAGCTCGTCCATGCC
F2	CAACTGTTGGGAAGGGCGA
R2	ACCAACTTGTTGCTTGAAGGC
<i>flk1</i>-EGFP knockin	
F1	CCTGGTAGCTGACCATATGCC
R1	CTTGTACAGCTCGTCCATGCC
F2	TGTAAAACGACGGCC AGT
R2	CATATCTGGGGAAAGATGGAG

Supplementary information, Table S1H Sequences of the primers for off-target analysis of *th* locus.

<i>th</i> OT1	
F	TTCGTCACCAATGAGGCCAT
R	TACAGTCTGCCCCATTCTGC
<i>th</i> OT2	
F	GGGAAACACCAACACATGGC
R	GGAGGAAGGGTTTGAGAGCC
<i>th</i> OT3	
F	TGCATTAAGCAGGCCCTTGA
R	TTCATGCTTTGTGTGTGCCG

Supplementary information, Data S1 Materials and Methods

Zebrafish husbandry

Adult zebrafish were maintained in the National Zebrafish Resources of China (Shanghai, China) with automatic fish housing system (ESEN, China) at 28°C. Embryos were raised under a 14h-10h light-dark cycle in 10% Hank's solution that consisted of (in mM): 140 NaCl, 5.4 KCl, 0.25 Na₂HPO₄, 0.44 KH₂PO₄, 1.3 CaCl₂, 1.0 MgSO₄ and 4.2 NaHCO₃ (pH 7.2). Zebrafish-handling procedures were approved by Institute of Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

Production of zCas9 mRNA and sgRNAs

The zCas9 expression plasmid pGH-T7-zCas9 [1] was linearized by *Xba* I and used as a template for Cas9 mRNA *in vitro* synthesis with the mMACHINE T7 Ultra kit (Ambion). The zCas9 mRNA was then purified by the same kit. The sequence of sgRNAs was designed according to previously reported criteria [2]. We used the CRISPR/Cas9 design tool (<http://zifit.partners.org>) to select specific targets to minimize off-target effects. The sequences of designed sgRNAs are as follows.

th: GGGTTGTTACTCATAAAG (reverse strand);

gfap: GTGCGCAACACATAGCACCA (reverse strand);

tph2: ACCTGAGCATACTGACTGAG (forward strand);

flk1: TCTGGTTTGGGAAGGACACAG (forward strand).

A pair of oligonucleotides containing the targeting sequence of sgRNAs were annealed and cloned downstream of the T7 promoter in the PT7-sgRNA vector. The sgRNAs were synthesized by the MAXIscript T7 Kit (Ambion), and were purified by using the mirVana™ miRNA Isolation Kit (Ambion).

Molecular cloning of donor plasmids

P2A-EGFP fragment was ligated into the PMD-19-T by T-A cloning to form the *T-P2A-EGFP* vector. The left and right arms for each targeted gene were amplified by the KOD-PLUS Neo

DNA polymerase from WT zebrafish genomic DNA using relevant primers (Supplementary information, Table S1F). Then, the arms were ligated into the 5' and 3' regions of the *P2A-EGFP* fragment in the *T-P2A-EGFP* vector, respectively. For the *gfap-EGFP* fusion knockin, the P2A sequence was replaced with a flexible serine-serine linker sequence (TCTTCT).

Micro-injection of one-cell stage embryos

The zCas9 mRNA, sgRNAs, and donor plasmids were co-injected into one-cell stage fertilized zebrafish eggs. Each embryo was injected with 1 nl of solution containing 800 ng/μl zCas9 mRNA, 80 ng/μl sgRNA, and 15 ng/μl donor plasmid.

Junction sequencing

Genome pool of 3-dpf injected embryos was used as template to amplify the 5' and 3' junction fragments of targeted genes by the primers (Supplementary information, Table S1G). The fragments were cloned by using the TA Cloning Kit (TAKARA) and then were sequenced.

Western blot

Twenty zebrafish embryos at 7 dpf were lysed in 40 μl of a protein extraction buffer (10 mM Tris HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 0.5% NP-40 supplemented with protease inhibitor cocktail and PMSF) on ice for 30 minutes. The lysate was then centrifuged at 13000 g for 15 minutes. The supernatant was boiled with a loading buffer and was run on SDS-PAGE gels. A Th antibody (Millipore) and actin antibody (Abmart) were then used for western analysis.

Whole-mount immunostaining of zebrafish embryos

Whole-mount immunostaining was carried out as previously reported [3]. Antibodies against Th, Gfap and dopamine [4] were obtained from Millipore, ZFIN, and Dr. H. Steinbusch's lab, respectively. For Th and Gfap immunostaining, 4-dpf larvae were fixed in 4% PFA in phosphate-buffered saline (PBS) overnight at 4°C, and larval head was cut down with ophthalmic scissors. Samples were then blocked in 10% Normal Goat Serum (NGS, Gibco) containing PBST (1% Tween-20) for two hours at RT and incubated in 1% NGS-containing

PBST with either mouse anti-Th (1:200) for 4 - 5 days or mouse anti-Gfap (1:200) for 2 days at 4°C. Subsequent wash was done with 1% NGS PBST for 12 hours at RT. Alexa 568 goat anti-mouse secondary antibodies (1:1000 in 1% NGS PBST, Molecular Probes) were then used at 4°C for 2 days.

For dopamine immunostaining, larvae were fixed in 5% glutaraldehyde in a Tris Buffer Saline (TBS, containing 50 mM Tris and 150 mM NaCl) containing 1% sodium metabisulfite (SMB) (referred as TBS/SMB, pH 7.2) overnight at 4°C. After cutting down the head, we embedded the head in 6% low melting-point agarose (Sigma). Samples were sectioned for 70 µm/slice by using a vibration microtome (Leica VT 1200s, speed 0.30 mm/s, amplitude 1.30 mm), then blocked with 10% goat-serum at RT, and subsequently incubated with rabbit anti-dopamine (1:500) at 4°C overnight. After washing for 5 hours at RT, dopamine signals were visualized with Alexa 568 goat anti-rabbit secondary antibodies (1:1000 in 1% NGS TBST/SMB, Molecular Probes). Slices were mounted in 1% low melting-point agarose and imaged under a FV1000 confocal microscope (Olympus, Japan). The intensity was measured by ImageJ and the relative fluoresce value was calculated as $(F-F_0)/F_0$ (F, fluorescent value of cells; F_0 , background fluorescent value).

Confocal imaging

Z-stack images were taken at RT, under a 20X or 40X water-immersion objective (N.A., 0.80) by using a FV1000 confocal microscope (Olympus, Japan). The resolution of all the images was either 1024×1024 or 800×600 .

In vivo two-photon calcium imaging

For in vivo two-photon calcium imaging, larvae at 3 - 5 dpf were immobilized in 1.5 % low melting-point agarose without paralysis or anesthetics. Time-lapse imaging was performed under a 40X water-immersion objective by using a Fluoview 1000 two-photon microscope (Olympus, Japan). Images were collected at 2 Hz. To analyze changes in calcium activity, images were firstly aligned by using Image J (NIH). For each experiment, individual regions of interest (ROIs) were manually drawn on the average image calculated from the entire series. The change in fluorescence intensity of each ROI was calculated as $(F-F_0)/F_0$, in which F_0 was

the average intensity of the ROI through the first 20 frames without sensory stimulation. The mechanical stimulus was applied by gas pressure-driven water ejection through a micropipettes towards fish tail (0.5-s duration, 40-s interval).

In vivo whole-cell recording

Zebrafish larvae at 3 - 5 dpf were paralyzed with α -bungarotoxin (100 μ g/ml, Sigma) for 10 – 15 minutes, and then immobilized in 1.0% low melting-point agarose. The extracellular solution consisted of (in mM): 134 NaCl, 2.9 KCl, 2.1 CaCl₂, 1.2 MgCl₂, 10 HEPES, and 10 glucose (pH 7.8). Recording micropipettes were made from borosilicate glass capillaries (BF120-69-15, WPI). The internal solution consisted of (in mM): 110 K-gluconate, 5 KCl, 0.3 Na₄-GTP, 2 Mg₂-ATP, 2 CaCl₂, 5 Phosphocreatine, 10 HEPES, and 10 EGTA (pH 7.3). The equilibrium potential of chloride ion (E_{Cl^-}) was about -60 mV according to the Nernst equation. Recording was performed at RT with the skin removed from the dorsal part about 50 μ m caudally to the location of EGFP-expressing cells. A recording micropipette (~ 20 M Ω , tip diameter < 1.5 μ m) filled with the internal solution was inserted into the brain through this cut, and rostra-ventrally approached to EGFP-expressing cells with a persistent positive pressure for keeping tip clean. After the contact of the micropipette tip with cell membrane, giga-ohm seal was formed by removing the positive pressure and applying a slight negative pressure. Whole-cell recording was achieved by delivering a few brief electrical zaps (duration, 25 μ s - 1 ms) to break the cell membrane beneath the micropipette tip. Data acquisition and signal processing were made with a patch-clamp amplifier (MultiClamp 700B, Axon Instruments), and all signals were filtered at 2 kHz and sampled at 10 kHz by using Clampex 10.2 (Molecular Devices). More details are described in a previous study [3].

Prediction and sequencing of potential off-targets

Potential off-targets of the *th* sgRNA were predicted by searching the zebrafish genome (Zv9) for potential off-target sequences, which contain up to three mismatches to the *th* sgRNA target and are followed by a NGG PAM sequence. Genomic DNA isolated from pools of 10 5-dpf injected F0 or *th-P2A-EGFP* knockin F1 embryos was collected. For sequencing, PCR products were cloned by using the TA Cloning Kit (TAKARA), and potential indel mutations were

identified by sequencing (Invitrogen). The PCR primers are showed in Supplementary information, Table S1H.

Statistics

Jarque-Bera test was first performed to examine the normality assumption of data. Two-tailed unpaired Student's *t*-test was then used for significance analysis for normal data. The *P* value less than 0.05 was considered to be statistically significant. All results are represented as mean \pm SEM.

References

- 1 Liu D, Wang Z, Xiao A *et al.* *Journal of Genetics and Genomics* 2014; **41**:43-46.
- 2 Chang N, Sun C, Gao L *et al.* *Cell Research* 2013; **23**:465-472.
- 3 Mu Y, Li XQ, Zhang B, Du JL. *Neuron* 2012; **75**:688-699.
- 4 Yamamoto K, Ruuskanen JO, Wullimann MF *et al.* *Molecular and Cellular Neuroscience* 2010; **43**:394-402.