Open
LETTER TO THE EDITOR

Cell Research (2015) 25:634-637.
© 2015 IBCB, SIBS, CAS All rights reserved 1001-0602/15
www.nature.com/cr

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

Cell Research (2015) **25**:634-637. doi:10.1038/cr.2015.43; published online 7 April 2015

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 pretectum (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

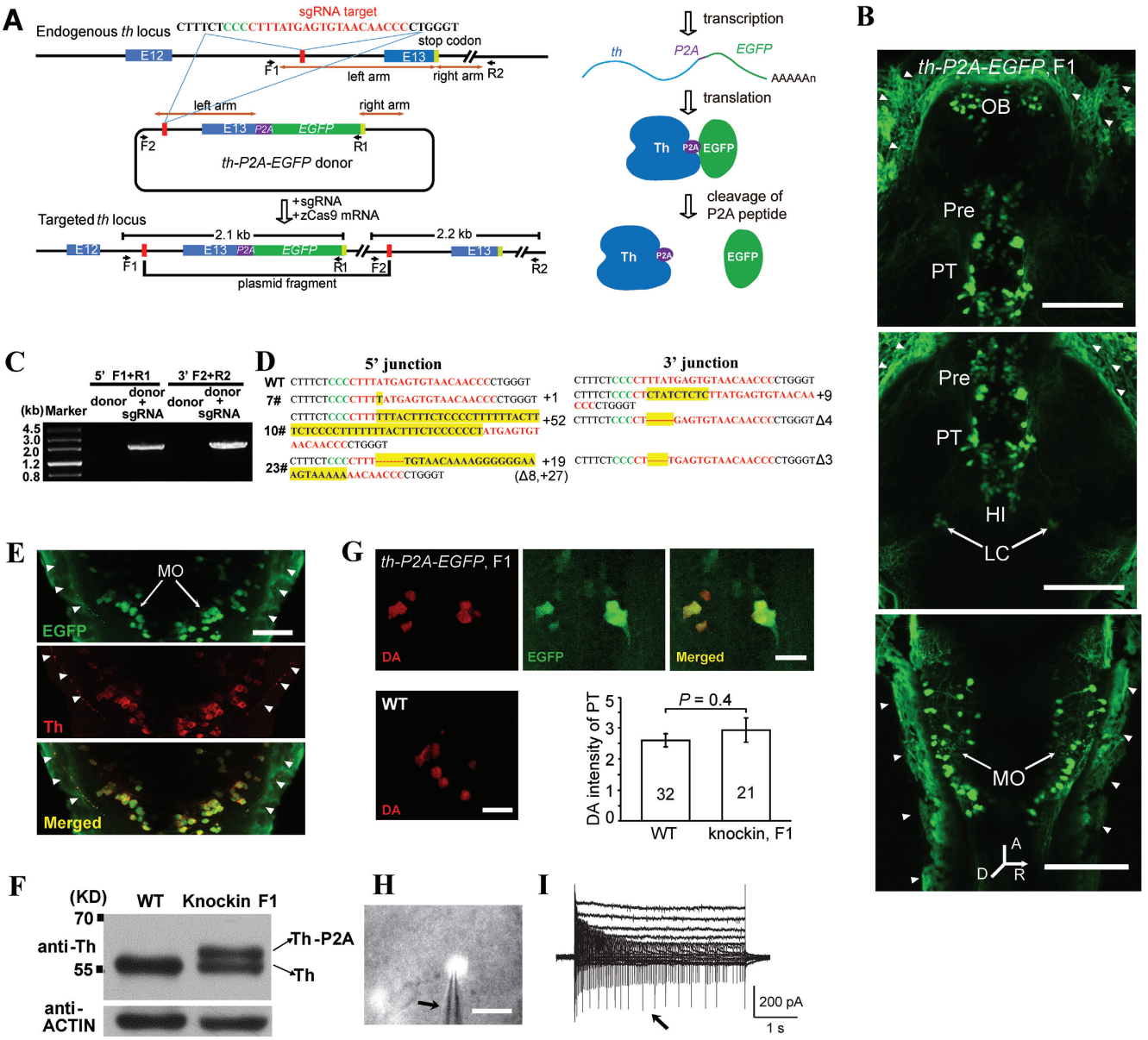


Figure 1 Intron targeting-mediated *EGFP* knockin at the zebrafish *th* locus. **(A)** Schematic of the intron targeting-mediated strategy for generating *EGFP* knockin at the zebrafish *th* locus by using the CRISPR/Cas9 system. The sgRNA target sequence is shown in red and the protospacer adjacent 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 μm. **(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 μm. **(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 μm. **(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 μm. **(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% ± 1%, mean ± 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 ~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 serotoninergic 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 Ren[1], Shan-ye Gu[1, 4], Yuan-hang Xiang[3], Cheng Huang [4], Jiu-lin Du[1, 2, 3]

[1]*Institute 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;* [2]*Graduate School, University of Chinese Academy of Sciences, 320 Yue-Yang Road, Shanghai 200031, China;* [3]*School of Life Science and Technology, ShanghaiTech University, 319 Yue-Yang Road, Shanghai 200031, China;* [4]*Drug 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