Knock-in tagging in zebrafish facilitated by insertion into non-coding regions

Daniel S Levic, Naoya Yamaguchi, Siyao Wang, Holger Knaut and Michel Bagnat
DOI: 10.1242/dev.199994

Editor: Steve Wilson

Review timeline
Original submission: 8 July 2021
Editorial decision: 13 August 2021
First revision received: 25 August 2021
Accepted: 27 August 2021

Original submission

First decision letter

MS ID#: DEVELOP/2021/199994

MS TITLE: Knock-in tagging in zebrafish facilitated by insertion into non-coding regions

AUTHORS: Daniel S Levic, Naoya Yamaguchi, Siyao Wang, Holger Knaut, and Michel Bagnat

I have now received all the referees reports on the above manuscript, and have reached a decision. The referees' comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

You will be pleased to see that the referees like you manuscript and only have relatively minor comments for improvements. Please attend to all of the reviewers' comments in your revised manuscript and detail them in your point-by-point response. If you do not agree with any of their criticisms or suggestions explain clearly why this is so.

Reviewer 1

Advance summary and potential significance to field

Although molecular genetic tools for zebrafish have been advancing rapidly over the past 10 years, there have remained barriers to generating knock-in lines. Most published methods rely on homology-based approaches which yield precise DNA integration, however, this is an inefficient process. In this manuscript, the authors outline a clever strategy to improve the efficiency of tagging endogenous loci, which exploits targeting and repair of untranslated regions for imprecise integration. The authors describe strategies for N- and C-terminal targeting and present multiple lines in which successful targeting has been achieved. There are many beautiful images, and the authors also present quantitative image analysis of GFP-Rab11a on apical vesicles, without concern about overexpression.
Overall, this manuscript presents a very exciting and useful approach to facilitate creation of knock-in alleles: this is currently a goal for many researchers in the community. The questions below are minor.

Comments for the author

• General comment: Many important data related to the method and its efficiency are in Figures S2 and S3. For this type of Technique Report, it seems important to include this in the main portion of the figures, perhaps as part of Figure 1.

• Figure 1, lines 88-89: do the authors have any thoughts or comments regarding intron lengths to target? For a C-terminal knock-in, it is noted that cutting should be targeted 100 bp upstream of the last exon. Should there also be some distance downstream from the upstream exon?

• Figure 1, lines 98-99: for an N-terminal fusion, how far upstream should the dsDNA break be induced? Is it the same as for the C-terminal fusion (~100 bp upstream)?

• Figure S3: It is really remarkable that so few F0 adults needed to be screened to find F1 carriers. Therefore, it seems that screening injected larvae for fluorescence is a crucial factor.
How many larvae were screened at the injected F0 stage? How did the mosaicism seen in injected embryos compare to other published methods? Did the authors have any trouble with survival when raising F0 injected larvae to adulthood?

• Figure S3: Were independent F0 founders of the same line sequenced and shown to have different NHEJ-repaired sites? Was there any evidence of insertion of the donor in the wrong direction?

• This is a minor clarification, but for the Rab11 knock-in, was there precise integration on the 5' end? Supplementary file 5 does not have any indel annotated on the 5' end. Also, do the authors have confirmation via cDNA sequencing that the splicing proceeds as predicted?

• The authors refer to having done antibody staining for at least one of the endogenous proteins targeted (Rab11). Can they include images for this (in a wild type non-targeted context), and any others they might have antibodies for, in this manuscript for comparison?

Minor Comments:
• Paragraph lines 177-182 should refer to Fig. 2I-K
• Typo: in line 276, "revers" should be "reverse"

Reviewer 2

Advance summary and potential significance to field

Determining the subcellular localization or abundance of proteins is crucial to understanding their activity how they are regulated, and their contribution to a process. Visualizing or detecting proteins in tissues and measuring their abundance generally requires robust and specific antibodies, or when unavailable tagged fusion proteins that are often expressed at higher than endogenous levels or without their endogenous regulatory mechanisms intact – under the control of another gene's promoter or by injecting RNAs lacking 5' or 3' regulatory elements. This overexpression or misexpression can lead to inappropriate levels of protein or result in expression of a protein in a cell type where it is not normally expressed. To circumvent these issues, knock-ins to the endogenous locus are generated such that the gene is expressed at the correct time and place and with gene-specific regulatory elements intact. Knock-in technology in zebrafish have been limited due to technical issues. Even with the discovery of CRISPR and Talens, insertion of nondisruptive tags remains challenging, particularly due to disruptive indels, insertions or deletions flanking the inserted tag. This techniques paper reports a method to introduce N-terminal and C-terminal fluorescent tags in the endogenous loci. The approach uses amplified donor sequences flanked with splice donor or acceptor sites that are targeted to noncoding regions to ameliorate effects due to insertion – splicing removes any deleterious changes to the surrounding sequence. Insertions were made at several loci to tag genes required for epithelial development. Multiple founders were generated for each line and for some both N and C terminal tags were generated. Evidence is provided that the patterns observed recapitulate previously established patterns and without showing the disruptive phenotypes associated with prior overexpression studies using tagged proteins. Proof of principle data are presented for measuring membrane trafficking of eGFP-Rab11a

in lysosome-rich-enterocytes and pronephros. Data are also provided demonstrating the extent of mosaicism and transmission efficiencies for three genes. The manuscript is clearly written, the images are beautiful, the approach is clearly described, the schematics are helpful, and the quantitative data and data overall are clearly presented and compelling. This will be a powerful tool for the field. I have only one very minor suggestion below.

Comments for the author

The manuscript is clearly written, the images are beautiful, the approach is clearly described, the schematics are helpful, and the quantitative data and data overall are clearly presented and compelling. This will be a powerful tool for the field. I have only one very minor suggestion below.

Very Minor

In Figure 2 and in Figure S2, different pseudo colors are used for the same gene. For Figure 2, please consider making A-D either all magenta or cyan, or if there was a resolution issue - fine details were easier to visualize with cyan than magenta consider using greyscale for panel D otherwise one thinks this is a different marker at first glance. Same comment for Figure S2 panels A and B. Please consider making both either green or cyan.

Reviewer 3

Advance summary and potential significance to field

In this manuscript “Knock-in tagging in zebrafish facilitated by insertion into non-coding regions”, Levic et al. describe a method to generate knock-in fusions in zebrafish. Although gene knock-ins by CRISPR-Cas9 gene editing have been previously described, the authors modify the approach by using intronic sequences as homology-arms, to reduce the deleterious consequences of small insertion/deletions induced by error-prone homology directed repair.
The authors demonstrate the efficacy and versatility of this approach by generating several knock-in lines including N-terminal as wells as C-terminal fusions. In the second part of the manuscript the authors further characterize their knock-in technique by testing different parameters, such as efficacy of different donor substrates (mutated or single-stranded DNA). Furthermore, they evaluate germline transmission of different donor substrates and the fidelity of protein localization. Lastly. the authors demonstrate that knock-in reporters are amenable to in vivo protein quantification.

A lot of effort in the zebrafish community has gone into the establishment of knock-in models in recent years.
Although CRISPR/Cas9 technology has been used in generating zebrafish knock-ins, success has been limited due to low efficacy and error-prone homology directed repair. The technology presented in this manuscript is not entirely novel but instead aims at specific modifications in experimental design, which allow to generate protein fusions with high efficacy and precision. Therefore, the work presented here is of great interest to zebrafish community and will encourage the generation of many additional knock-in lines.

Comments for the author

In my opinion, the data presented are of high quality and should be published without major revision. I only have some minor comments:
1) line 174: What do the authors mean by “well tolerated”. Is there no increase in lethality?
2) Many of the figures are poorly labeled. e.g. Figure 2 – what are the tissues labeled by ZO1-EGFP?

First revision

Author response to reviewers' comments

We would like to thank the reviewers for their positive assessment of our manuscript. In response to the insightful comments made by the reviewers, we have made changes to the text and minor changes to figures as requested. Additionally, during the review period we obtained a stable KI line for aPKC (*prkci*), which was initially presented only in F0s. We have added data panels for this line in Figure 3 and a sequence file of the genomic integration site in supplement. Changes have been kept to a minimum and did not alter any of the findings reported in the initial manuscript. However, we feel that these minor edits and additions have improved the manuscript. The changes are shown in the manuscript text in blue font. Please note that we were asked to reduce the word count of the manuscript text and figure legends. Minimal changes were made to accommodate the word limit. Below we respond to specific points made by the reviewers in blue font.

Reviewer 1 Advance Summary and Potential Significance to Field...
Although molecular genetic tools for zebrafish have been advancing rapidly over the past 10 years, there have remained barriers to generating knock-in lines. Most published methods rely on homology-based approaches which yield precise DNA integration, however, this is an inefficient process. In this manuscript, the authors outline a clever strategy to improve the efficiency of tagging endogenous loci, which exploits targeting and repair of untranslated regions for imprecise integration. The authors describe strategies for N- and C-terminal targeting and present multiple lines in which successful targeting has been achieved. There are many beautiful images, and the authors also present quantitative image analysis of GFP-Rab11a on apical vesicles, without concern about overexpression.
Overall, this manuscript presents a very exciting and useful approach to facilitate creation of knock-in alleles: this is currently a goal for many researchers in the community. The questions below are minor.

Reviewer 1 Comments for the Author...
- General comment: Many important data related to the method and its efficiency are in Figures S2 and S3. For this type of Technique Report, it seems important to include this in the main portion of the figures, perhaps as part of Figure 1.

We thank the reviewer for the positive comments and helpful suggestions. We agree that including these data in the main figure would be helpful for readers. We have moved the F0/F1 efficiency data from Fig. S2-3 to the revised Fig. 1.

- Figure 1, lines 88-89: do the authors have any thoughts or comments regarding intron lengths to target? For a C-terminal knock-in, it is noted that cutting should be targeted 100 bp upstream of the last exon. Should there also be some distance downstream from the upstream exon?

3’ splice acceptors utilize both the splice site at the intron-exon boundary as well as the branch point sequence, which is typically located about 50 bp upstream from the intron-exon boundary. By targeting at least 100 bp upstream from the last exon, it is highly unlikely that functional splicing elements will be affected. On the other hand, 5’ splice donors typically rely only on the splice site at the exon-intron boundary. Therefore, specifying a recommended distance downstream from the upstream exon is not an important design factor.

- Figure 1, lines 98-99: for an N-terminal fusion, how far upstream should the dsDNA break be induced? Is it the same as for the C-terminal fusion (~100 bp upstream)?

We have not systematically tested KI efficiencies resulting from dsDNA break points at varying distances from the insertion site. On the one hand, it is generally accepted that minimizing the distance between a genomic dsDNA break and the integration site enhances efficiency of KI. On

the other hand, KI in zebrafish is highly error-prone and is often accompanied by INDELs. Therefore, a dsDNA break close to the coding sequence integration site may be more likely to have functional consequences. Because we found that stable KI lines can be readily recovered when the dsDNA break distance is 100-300 bp away from the integration site, we recommend this for N-terminal tagging as well as C-terminal tagging.

- Figure S3: It is really remarkable that so few F0 adults needed to be screened to find F1 carriers. Therefore, it seems that screening injected larvae for fluorescence is a crucial factor. How many larvae were screened at the injected F0 stage? How did the mosaicism seen in injected embryos compare to other published methods? Did the authors have any trouble with survival when raising F0 injected larvae to adulthood?

The reviewer is correct; visual screening of injected embryos is crucial for obtaining stable KI lines. The numbers of animals reported in Figure 1I-J (previously in Fig. S2-S3 in the original manuscript) are representative for all the stable KI lines we show. For example, for cldn15la we performed 1 experiment where we injected 800 embryos. 720 embryos survived and were screened visually, and we raised 5 larvae to adulthood. We outcrossed 3 of the 5 adult F0s to obtain a founder line. Similarly, for ZO1-GFP we performed 1 experiment where we injected 1,120 embryos. 960 embryos survived and were screened visually, and we raised 8 larvae to adulthood. We outcrossed 3 of the 8 adult F0s to obtain a founder line. We only attempted to raise larvae that appeared healthy and morphologically normal (as indicated by formation of the swim bladder), so most of the F0s survived to adulthood.

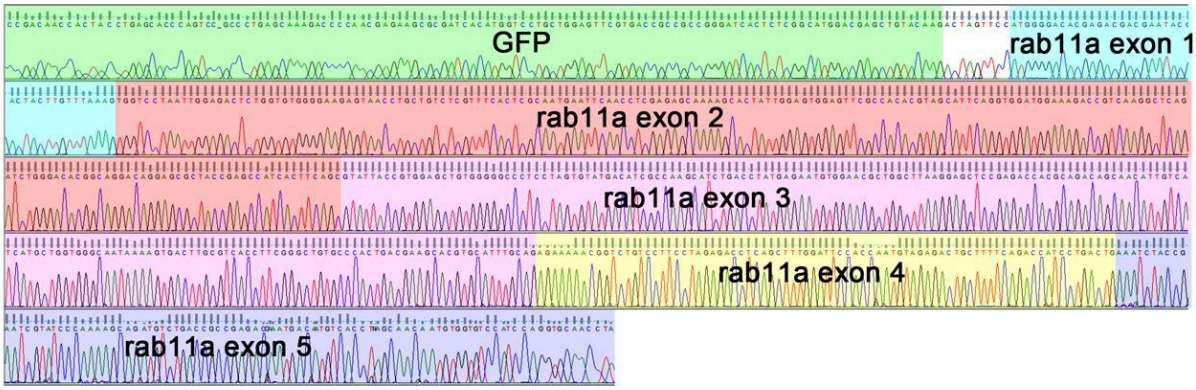
- Figure S3: Were independent F0 founders of the same line sequenced and shown to have different NHEJ-repaired sites? Was there any evidence of insertion of the donor in the wrong direction?

We did analyze this for one of the target genes. We obtained 3 independent lines for eGFP-rab11a KI generated either by 1) dsDNA donor with 5' UTR gRNA, 2) dsDNA donor with both 5' UTR gRNA and intron 1 gRNA, or 3) ssDNA donor with both 5' UTR gRNA and intron 1 gRNA. The 3 lines seem to be identical by expression, and the 5' end of the integration site is seamless and appears to have been incorporated by homology-directed repair for all 3 lines. However, the 3' end of the integration site is shorter in the ssDNA line compared to that of the 2 dsDNA lines. In the dsDNA lines the 3' intron fragment is 491 bp, whereas in the ssDNA line the 3' intron fragment is 102 bp. We speculate that this is due to the instability of ssDNA, which is prone to degradation during handling. Additionally, there are small sequence differences between the 2 dsDNA lines. There is a small INDEL at the 3' integration boundary in the intron that is present in one of the lines. These sequence discrepancies, because they all reside in non-coding regions, appear to have no effect on eGFP-Rab11a expression at the protein level. These data are described in the revised text and are presented in the revised Fig. S2.

Regarding the second point, it is likely that some C-terminal KI integrations proceeded in the wrong direction but were not observed. The reason for this is that we rely on visual screening in injected embryos, thus selecting for correct integrations. The fluorescent protein coding sequence in the donors we prepared for C- terminal KI lack a start codon, so it is unlikely that improper or random genomic integrations would result in fluorescent protein expression.

- This is a minor clarification, but for the Rab11 knock-in, was there precise integration on the 5' end? Supplementary file 5 does not have any indel annotated on the 5' end. Also, do the authors have confirmation via cDNA sequencing that the splicing proceeds as predicted?

As noted above, our sequencing data do indeed indicate that the 5' ends for 3 independent eGFP-rab11a alleles proceeded seamlessly as intended. The 3' integration boundaries within intron 1, on the other hand, differ in each of the 3 lines. We did sequence the eGFP-Rab11a transcript by RT-PCR to verify proper mRNA splicing.



- The authors refer to having done antibody staining for at least one of the endogenous proteins targeted (Rab11). Can they include images for this (in a wild type non-targeted context), and any others they might have antibodies for, in this manuscript for comparison?

Thank you for the suggestion. In the revised Fig. S4, we have included antibody staining comparisons for Rab11a and aPKC. For Cldn15la, we have added a comparison with a BAC transgenic line.

Minor Comments:
- Paragraph lines 177-182 should refer to Fig. 2I-K
- Typo: in line 276, “revers” should be “reverse”

Thank you for pointing out these errors.

Reviewer 2 Advance Summary and Potential Significance to Field...
Determining the subcellular localization or abundance of proteins is crucial to understanding their activity, how they are regulated, and their contribution to a process. Visualizing or detecting proteins in tissues and measuring their abundance generally requires robust and specific antibodies, or when unavailable tagged fusion proteins that are often expressed at higher than endogenous levels or without their endogenous regulatory mechanisms intact – under the control of another gene’s promoter or by injecting RNAs lacking 5’ or 3’ regulatory elements. This overexpression or misexpression can lead to inappropriate levels of protein or result in expression of a protein in a cell type where it is not normally expressed. To circumvent these issues, knock-ins to the endogenous locus are generated such that the gene is expressed at the correct time and place and with gene-specific regulatory elements intact. Knock-in technology in zebrafish have been limited due to technical issues. Even with the discovery of CRISPR and Talens, insertion of nondisruptive tags remains challenging, particularly due to disruptive indels, insertions or deletions flanking the inserted tag. This techniques paper reports a method to introduce N-terminal and C-terminal fluorescent tags in the endogenous loci. The approach uses amplified donor sequences flanked with splice donor or acceptor sites that are targeted to noncoding regions to ameliorate effects due to insertion – splicing removes any deleterious changes to the surrounding sequence. Insertions were made at several loci to tag genes required for epithelial development. Multiple founders were generated for each line and for some both N and C terminal tags were generated. Evidence is provided that the patterns observed recapitulate previously established patterns and without showing the disruptive phenotypes associated with prior overexpression studies using tagged proteins. Proof of principle data are presented for measuring membrane trafficking of eGFP-Rab11a in lysosome-rich-enterocytes and pronephros. Data are also provided demonstrating the extent of mosaicism and transmission efficiencies for three genes. The manuscript is clearly written, the images are beautiful, the approach is clearly described, the schematics are helpful, and the quantitative data and data overall are clearly presented and compelling. This will be a powerful tool for the field. I have only one very minor suggestion below.

Reviewer 2 Comments for the Author...
The manuscript is clearly written, the images are beautiful, the approach is clearly described, the schematics are helpful, and the quantitative data and data overall are clearly presented and compelling. This will be a powerful tool for the field. I have only one very minor suggestion below.

Very Minor
In Figure 2 and in Figure S2, different pseudo colors are used for the same gene. For Figure 2, please consider making A-D either all magenta or cyan, or if there was a resolution issue - fine details were easier to visualize with cyan than magenta consider using greyscale for panel D otherwise one thinks this is a different marker at first glance. Same comment for Figure S2 panels A and B. Please consider making both either green or cyan.

Thanks for the positive comments and the suggestion. We have changed panel 2D from cyan to greyscale because magenta made it difficult to visualize fine details. Please note that the eGFP-Rab11a data formerly in Fig. S2A-B is now moved to Fig. S1D, but we only show one color now as suggested.

Reviewer 3 Advance Summary and Potential Significance to Field...
In this manuscript „Knock-in tagging in zebrafish facilitated by insertion into non-coding regions", Levic et al. describe a method to generate knock-in fusions in zebrafish. Although gene knock-ins by CRISPR-Cas9 gene editing have been previously described, the authors modify the approach by using intronic sequences as homology-arms, to reduce the deleterious consequences of small insertion/deletions induced by error-prone homology directed repair.
The authors demonstrate the efficacy and versatility of this approach by generating several knock-in lines including N-terminal as wells as C-terminal fusions. In the second part of the manuscript the authors further characterize their knock-in technique by testing different parameters, such as efficacy of different donor substrates (mutated or single-stranded DNA). Furthermore, they evaluate germline transmission of different donor substrates and the fidelity of protein localization. Lastly. the authors demonstrate that knock-in reporters are amenable to in vivo protein quantification.

A lot of effort in the zebrafish community has gone into the establishment of knock-in models in recent years. Although CRISPR/Cas9 technology has been used in generating zebrafish knock-ins, success has been limited due to low efficacy and error-prone homology directed repair. The technology presented in this manuscript is not entirely novel but instead aims at specific modifications in experimental design, which allow to generate protein fusions with high efficacy and precision. Therefore, the work presented here is of great interest to zebrafish community and will encourage the generation of many additional knock-in lines.

Reviewer 3 Comments for the Author...
In my opinion, the data presented are of high quality and should be published without major revision. I only have some minor comments:

1) line 174: What do the authors mean by "well tolerated". Is there no increase in lethality?

The *TgBAC(cldn15la-GFP)*^{pd1034} allele is homozygous lethal at embryonic stages, presumably due to a gene disruption at the genomic integration site of the BAC transgene. In contrast to the BAC transgenic, the *TgKI(cldn15la-tdTomato)*^{pd1249} line that we report here shows no obvious phenotypes in homozygous KI larvae. We have modified the text to clarify this point.

2) Many of the figures are poorly labeled. e.g. Figure 2 – what are the tissues labeled by ZO1-EGFP?

Thanks for pointing this out. We have labeled the figures and included better descriptions in the figure legends.

Second decision letter

MS ID#: DEVELOP/2021/199994

MS TITLE: Knock-in tagging in zebrafish facilitated by insertion into non-coding regions

AUTHORS: Daniel S Levic, Naoya Yamaguchi, Siyao Wang, Holger Knaut, and Michel Bagnat
ARTICLE TYPE: Techniques and Resources Report

I am happy to tell you that your manuscript has been accepted for publication in Development, pending our standard ethics checks.