Received: 4 March 2020 | Revised: 15 July 2020 | Accepted: 16 July 2020

DOI: 10.1002/dvg.23391

genesis WILEY

TECHNOLOGY REPORT

Intron with transgenic marker (InTraM) facilitates high-throughput screening of endogenous gene reporter lines

Russell T. Collins[1] | Baptiste Coxam[1,2] | Ines Fechner[1,2] | Iris Unterweger[1,3] | Anna Szymborska[1,2] | Katja Meier[1,2] | Holger Gerhardt[1,2,4]

[1]Integrative Vascular Biology Lab, Max-Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany

[2]DZHK (German Centre for Cardiovascular Research), partner site Berlin, Berlin, Germany

[3]Novo Nordisk Foundation Center for Stem Cell Biology, Copenhagen, Denmark

[4]Berlin Institute of Health (BIH), Berlin, Germany

Correspondence
Prof. Dr. Holger Gerhardt, Integrative Vascular Biology, BIH and DZHK professorship for Experimental Cardiovascular Research at the Charité, Robert-Rössle-Straße 10, 13125 Berlin, Germany.
Email: holger.gerhardt@mdc-berlin.de

Summary

The generation and maintenance of genome edited zebrafish lines is typically labor intensive due to the lack of an easy visual read-out for the modification. To facilitate this process, we have developed a novel method that relies on the inclusion of an artificial intron with a transgenic marker (InTraM) within the *knock-in* sequence of interest, which upon splicing produces a transcript with a precise and seamless modification. We have demonstrated this technology by replacing the stop codon of the zebrafish *fli1a* gene with a transcriptional activator KALTA4, using an InTraM that enables red fluorescent protein expression in the heart.

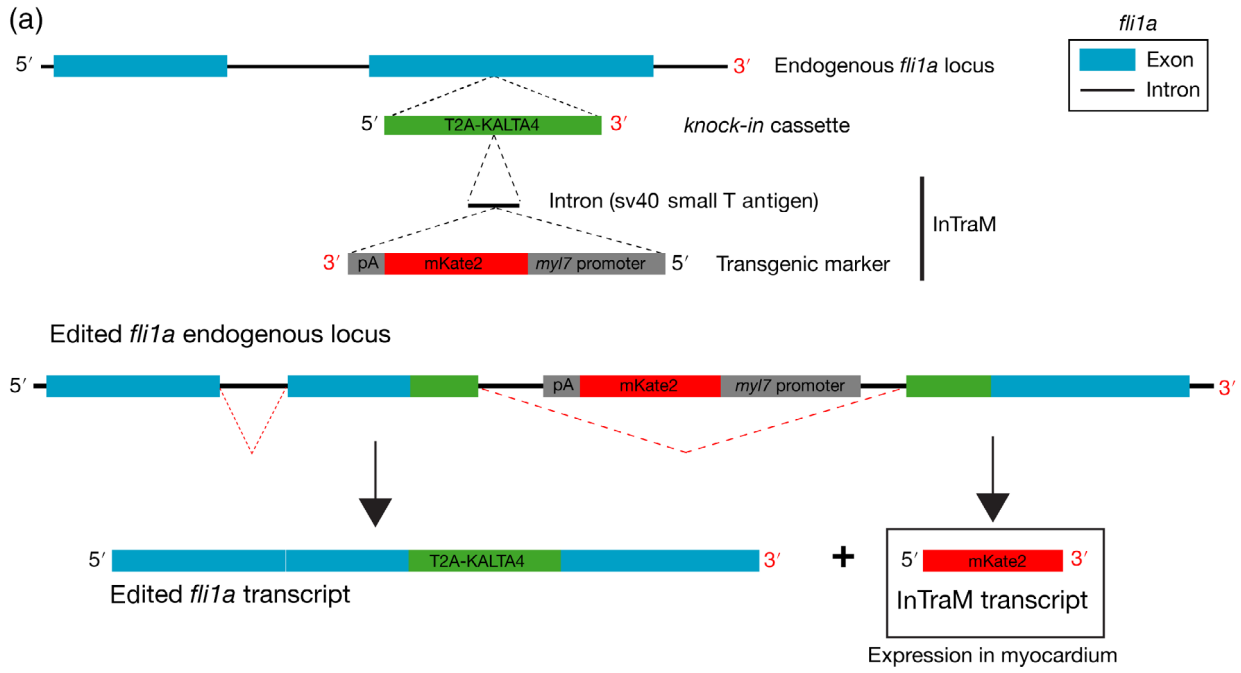
KEYWORDS

animal model, CRISPR/Cas9, genome editing, screening, transgenesis, zebrafish

1 | INTRODUCTION

The manipulation of genomes of model organisms has been a mainstay of biological research for decades, enabling the study of basic biological processes and advancing our understanding of disease mechanisms. In recent years, the advent of precise genome editing tools, such as CRISPR/Cas9 (Adli, 2018) and related systems (TALENS, ZNFs) (Bedell et al., 2012; Foley et al., 2009), has enabled researchers to mimic missense mutations associated with human diseases, generate gain- and loss-of-function gene variants, and precisely alter genomic loci with reporter tags. However, the generation and the maintenance of transgenic lines with these modifications can be difficult and time consuming. In zebrafish, for example, the process of creating a transgenic line through CRISPR/Cas9 mediated DNA cleavage and homology directed repair, involves generating mosaic P_0 individuals, of which only a subset will transmit the transgene through the germ line. This necessitates screening of hundreds of F_1 progeny to identify stable transgenics. The establishment of transgenic lines with small modifications or single nucleotide changes is particularly challenging, as it requires performing PCR on the genomic DNA of individual animals, without sacrificing them. The addition of the coding sequence of a fluorescent protein to an endogenous gene should, in principle, be easier to screen because transgenic animals can be identified visually with a fluorescence microscope. However, in our experience, the expression levels of the endogenous proteins are often so low that the fluorescent tag can only be visualized with a microscope equipped with a high numerical aperture objective and sensitive detectors, such as laser scanning or spinning disc confocal microscope. This requires access to expensive equipment and painstaking mounting procedures, both of which are impractical when screening large numbers of embryos.

Transposon-mediated transgenesis often utilizes a linked marker, such as the *mini-white* gene used in *Drosophila* P-elements (Klemenz, Weber, & Gehring, 1987) or the *myl7*:EGFP reporter found in many zebrafish Tol2 constructs. These markers are easily visually identifiable, and substantially streamline the process of transgenic line generation and maintenance. To bring a similar visible reporter system to CRISPR/Cas9 *knock-ins*, we devised a novel genetic element consisting of an artificial *i*ntron with a *tra*nsgenic *m*arker (InTraM). When inserted into the knock-in cassette, the InTraM allows for easy



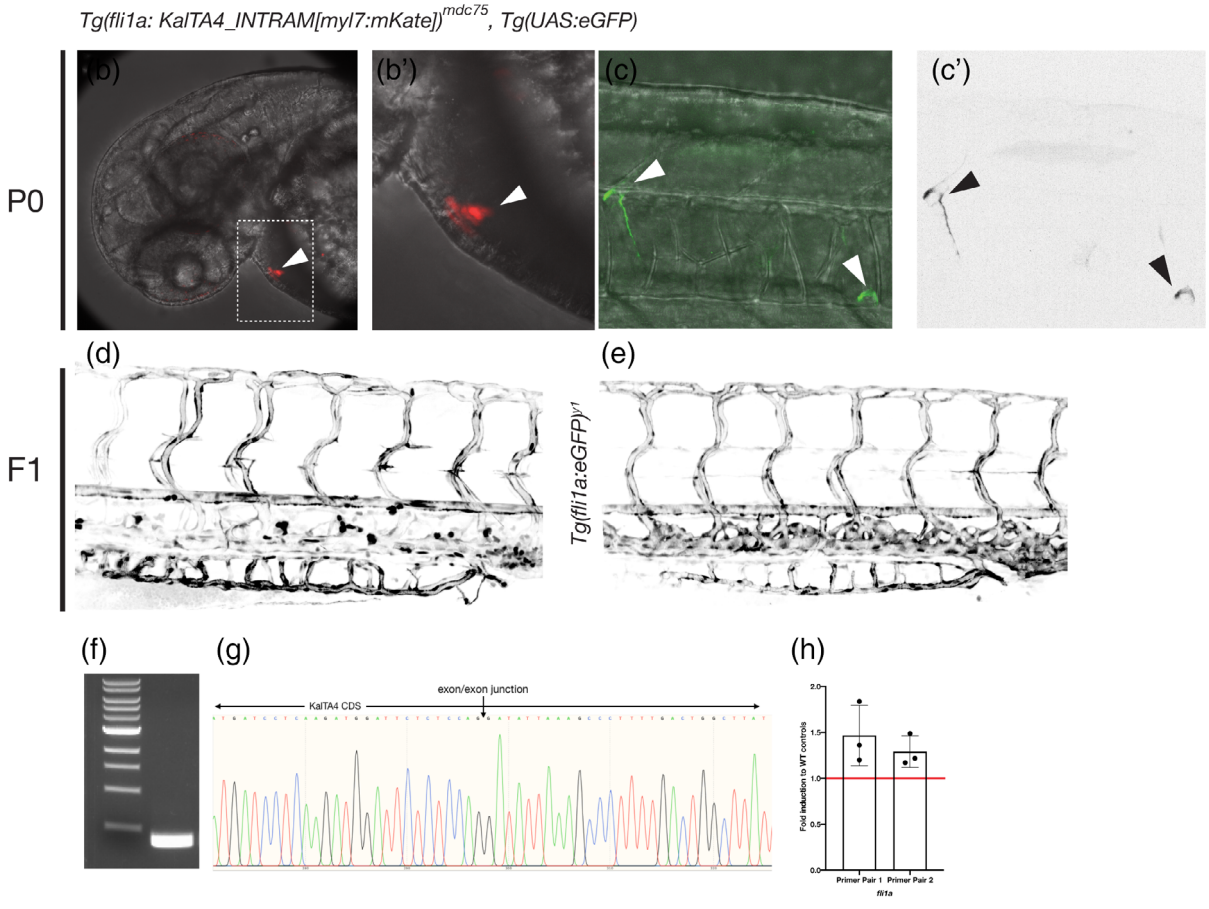


FIGURE 1 Legend on next page.

identification of the gene edited animals, but it is spliced out from mRNAs transcribed from the targeted locus, resulting in the expression of proteins with precise and seamless modifications.

2 | RESULTS

2.1 | Design of the *knock-in* cassette with InTraM

As proof-of-principle, we replaced the stop codon of the zebrafish *fli1a* gene with the ribosome skipping T2A sequence from *Thosea asigna* (Liu et al., 2017) and the coding sequence for KALTA4 (a zebrafish optimized GAL4 transcriptional activator) (Distel, Wullimann, & Koster, 2009). To make the InTraM we adapted the intron from the *Simian virus 40* small T antigen (previously demonstrated to function in zebrafish embryos [Collins, Linker, & Lewis, 2010]), by inserting a reporter gene consisting of the *myl7* (also known as *cmlc2*) promoter, the coding DNA sequence (CDS) for the mKate2 fluorescent protein and a herpes simplex virus thymidine kinase polyA signal (Figure 1a, Supplementary Figure 1a). By 24 hr postfertilization (hpf), the *myl7* promoter is strongly and exclusively active in the myocardium (Burns et al., 2005). The InTraM sequence was inserted in the antisense orientation to the T2A-KALTA4 CDS to prevent interference with the transcription from the targeted locus. We placed the InTraM between the guanines of the sequence AGG at position 260 of the KalTA4 CDS to generate consensus splice donor and acceptor sites (Collins et al., 2010; Roca, Krainer, & Eperon, 2013; Stephens & Schneider, 1992).

2.2 | Functional testing of an InTraM containing transgene

We injected the *fli1a* targeting construct, along with Cas9 mRNA and target-specific single guide RNA (sgRNA), into *Tg[5xUAS:EGFP]*^{zf82} embryos. By 24 hpf mKate2 expression in the myocardium was clearly detectable with a fluorescence stereomicroscope in a subset of the injected embryos (Figure 1b,b′). The mKate2-positive embryos also showed EGFP expression in individual endothelial cells (Figure 1c,c′) consistent with the expected endogenous expression of the targeted *fli1a* gene in mosaic P_0 embryos. Repeating the injections in wild-type embryos resulted in 18% (of more than 800) embryos with detectable mKate2 expression in the myocardium. We extracted genomic DNA from 12 mKate2 positive embryos and performed PCR over the insertion site using a forward primer up-stream of the 5′ homology arm and a reverse primer in the KALTA4 sequence. Eight of the 12 reactions resulted in a PCR product of the size expected for precise insertion in the *fli1a* locus, suggesting that approximately two-thirds of the mKate2 positive embryos had integrations from homologous recombination. We raised the remaining mKate2-positive P_0 embryos and from the 34 surviving P_0 adults we successfully identified a single *Tg[fli1a-T2A-KalTA4_InTraM(myl7:mKate2)]*^{md75} founder fish using only the mKate2 myocardial expression in F_1 embryos as a readout. PCR and sequencing of the genomic DNA from F_1 embryos demonstrated that T2A-KalTA4 CDS was inserted in frame into the *fli1a* CDS. When F_1 fish were crossed to homozygous *Tg[5xUAS:EGFP]*^{zf82} fish, all of the mKate2-positive progeny displayed robust expression of EGFP in endothelial cells (Figure 1d). This expression pattern was similar to that of *Tg(fli1a: eGFP)*^{y1} (Figure 1e), further supporting that the expression of the T2A-KaltA4 CDS is consistent with an integration within the *fli1a* locus. Reverse transcription PCR on total RNA from mKate2-positive F_2 embryos, using primers upstream and downstream of the InTraM element, produced a single PCR product of the size expected for a spliced transcript (Figure 1f). Sequencing of this PCR product confirmed that splicing occurred as expected, demonstrating that InTraM was efficiently and precisely excised from the modified transcript (Figure 1g). The level of *fli1a* expression was assessed by qPCR using cDNA generated from pools of mKate2 positive and mKate2 negative embryos from out-crossing the *Tg[fli1a-T2A-KalTA4_InTraM(myl7:mKate2)]*^{md75} line to wild-type fish. The level of *fli1a* transcripts in mKate2 positive embryos was 1.3- to 1.4-fold higher than their mKate2 negative siblings (Figure 1h), suggesting that integration of the transgene into the *fli1a* locus resulted in a moderately increased expression relative to that of the endogenous *fli1a* gene.

3 | DISCUSSION

The generation of transgenic lines with targeted insertions has been, and continues to be, challenging in large part due to inefficiencies in the process. In zebrafish, only a limited number of the injected embryos will contain cells with targeted insertions, of which only a small fraction will survive to adulthood and transmit the transgene

FIGURE 1 (a) Schematic representation of the InTRaM *knock-in* strategy targeting the *fli1a* endogenous locus. (b,b′) Representative image of *Tg[fli1a: KalTA4_INTRAM(myl7:mKate)]*, *Tg[5xUAS:EGFP]*^{zf82} expression in positive P_0 embryos at 52 hpf (transmitted light and mKate2 signal). (c,c′) In the same embryo, positive endothelial cells can be detected in the trunk at 52 hpf (transmitted light and eGFP). eGFP only signal is shown in panel c′. (d) Representative image of *Tg[fli1a: KalTA4_INTRAM(myl7:mKate)]*, *Tg[5xUAS:EGFP]*^{zf82} expression in F_1 embryos at 4 days postfertilization (dpf) (green channel only). (e) Representative image of *Tg[fli1a: eGFP]*^{y1} expression in embryos at 4 dpf. (f) Reverse transcription PCR on total RNA from mKate2-positive F_2 embryos produced a single band of the expected size (416 bp) for spliced transcripts, ladder: New England Biolab 1 kb ladder. (g) Sequencing of this PCR product demonstrated precise excision of the InTraM from the modified transcript. (h) qPCR quantification of *fli1a* expression levels in 72 hpf *Tg[fli1a: KalTA4_INTRAM(myl7:mKate)]*, *Tg[5xUAS:EGFP]*^{zf82} mKate2-positive embryos, compared to mKate2-negative embryos, using two different *fli1a* primer pairs

through their germline. We have demonstrated that incorporation of an artificial intron carrying a transgenic marker into CRISPR/Cas9 mediated *knock-in* cassette is a powerful strategy for mitigating these inefficiencies by providing a simple tool to screen out many, if not most, of the embryos that do not contain targeted insertions. We found that the expression of *fli1a* from the targeted locus was higher than from the endogenous *fli1a* gene. It is not known if the altered expression is due to the inclusion of the InTraM or the other components of the targeting construct. Any modification of an endogenous locus carries the risk of altering expression from that locus, and any alterations in expression is likely to be specific to the targeted gene and the nature of the modification.

In our proof-of-concept, we inserted additional coding sequences into the target gene, but the same approach can also be used to edit genes with small or single nucleotide modifications. By including the desired modification in the homology arms flanking the InTraM, transgenic lines with genes coding for dominant-negative or constitutively active proteins, or genes with disease associated missense mutations, can be generated efficiently (Supplementary Figure 1b).

One disadvantage of the CRISPR/Cas9 mediated HDR is the necessity to clone long homology arms. There are other targeting methods that rely on nonhomologous end joining repair, such as Homology Independent Targeted Insertion (HITI) (Suzuki et al., 2016). However, these methods do not permit seamless modification of the target CDS, and often introduce small insertions and deletions at the insertion site (Hruscha et al., 2013; Hwang et al., 2013a). To make use of the quick and easy cloning of HITI targeting constructs while retaining the ability to generate precise and seamless protein modifications, we are currently testing a variant of InTraM that is inserted into an intron, rather than an exon, of the target gene. In this artificial Exon with Transgenic Marker (ExTraM), the CDS for the desired modification is flanked by splice acceptor and donor sites. When the targeted gene is transcribed, the CDS within the ExTraM is spliced in-frame with the CDS from the endogenous locus, resulting in the seamless modification of the translated protein (Supplementary Figure 1c). We expect that together, the InTraM and ExTraM techniques will enable researchers to generate novel transgenic lines with precise genome and/or protein modifications in a more efficient manner.

4 | METHODS

4.1 | Construct sequences

SV40 small T antigen intron has the following sequence:

5′ GTAAATATAAAATTTTTAAGTGTATAATGTGTT[**reporter gene**]AAACTACTGATTCTAATTGTTTGTGTATTTTAGGA 3′.

Tg(fli1a: KalTA4_INTRAM[myl7:mKate]) construct for CRISPR/CAS9-homology independent targeted integration is available upon request.

In the final T2A-KALTA4 targeting construct the *knock-in* cassette was flanked with 958 bp 5′ and 1,146 bp 3′ homology arms to facilitate the CRISPR/Cas9 mediated HDR [4]. Target sites of the sgRNA were included at the ends of homology arms to allow for plasmid linearization upon injection as described in Irion, Krauss, and Nusslein-Volhard (2014). The construct was assembled using NEB Builder HiFi DNA Assembly kit (New England Biolabs).

sgRNA was designed using ZiFiT (http://zifit.partners.org/ZiFiT/references.aspx) and the plasmid for production of sgRNA was generated as previously described (Hwang et al., 2013b). sgRNA was transcribed using the MEGAscript T7 Transcription Kit (Invitrogen) and purified using the MEGAclear Transcription Clean-Up Kit (Invitrogen) following the manufacturer's instructions.

Cas9 mRNA was transcribed from the pT7-Cas9 plasmid (Hwang et al., 2013b) using the mMESSAGE mMACHINE T7 ULTRA Transcription Kit (Invitrogen).

For reverse transcription PCR and qPCR, 30 mKate positive and 30 mKate negative F_2 embryos were flash frozen in liquid nitrogen and homogenized in 300 μl TRI Reagent Solution (Ambion). Total RNA was purified using the Direct-zol RNA microprep kit (Zymo Research). To assess splicing of the InTram cassette, PCR was performed using a forward primer [CCGGGCTGCAGGAATTCGATAGTCAGCTGTCTCTGTCCCT] 5′ to the splice donor site, and a reverse primer [ACGGTAT CGATAAGCTTGATAGGGACAGAGACAGCTGACT] 3′ to the splice acceptor site. For sequencing, the resulting PCR product was cloned into EcoRV digested pISceI (Thermes et al., 2002) using the NEB Builder HiFi DNA Assembly kit. qPCR was performed on a QuantStudio 6 Real Time PCR system (Applied Biosystems) using two sets of primers: *fli1a* pair 1 [GCAATAAAATGGTGGGCGGG and AGAGACGGGTCTGCGGG] and *fli1a* pair 2: TGAGAAATGTGCTCCGCCTC and CAGAGACGGG TCTGCGGG]. Each reaction was performed in triplicate, each qPCR mixture contained 0.45 μl ROX (EURx), 0.25 μl uracil-N-glycosylase (UNG, EURx), 10 μl 2X SYBR green qPCR Master Mix (EURx), 2 μl Primers (0.5uM each mix), 2 μl of 1:10 dilution of cDNA and 5.3 μl of dH2O. The efficiency of PCR amplification was determined using LinReg PCR (Ruijter et al., 2009). Gene expression was normalized to that of the housekeeping genes *rpl13* and *rps29*. Calculations of fold induction changes were performed following published models (Pfaffl, 2001).

4.2 | Injections

For CRISPR/Cas9 mediated homologous recombination one-cell stage embryos were injected with 2 nl of 20 ng/μl targeting construct plasmid, 125 ng/μl Cas9 mRNA and 25 ng/μl of target specific sgRNA. The sg target sequence for *fli1a* was GGTTTAGTAGTAACTACCA (minus strand), PAM sequence: AGG.

4.3 | Zebrafish husbandry

Zebrafish (*Danio rerio*) were raised and staged as previously described (Kimmel, Ballard, Kimmel, Ullmann, & Schilling, 1995). Transgenic lines used were: *Tg[fli1a-T2A-KalTA4_InTraM(myl7:mKate2)]*^{md75}, *Tg[5xUAS:EGFP]*^{zf82} and *Tg(fli1:EGFP)*^{y1}. The wild-type fish used were from a

cross of the AB and TL strains. For growing and breeding of transgenic lines, we comply with regulations of the ethical commission animal science of MDC Berlin and with FELASA guidelines (Alestrom et al., 2019).

4.4 | Imaging

Embryos were anesthetized in 0.014% tricaine (Tricaine-Pharmaq 1,000 mg/g, PHARMAQ Limited), mounted in plastic petri dishes (94 × 16 mm - Sarstedt Ref#82.1473) containing 0.014% tricaine, and bathed in E3 media containing 0.007 (0.5×) to 0.014% (1×) tricaine and 0.003% Phenylthiourea (as indicated). Imaging was performed on an upright 3i spinning-disc confocal microscope using Zeiss Plan-Apochromat 20× or 40× /1.0 NA water-dipping objectives. Screening of embryos was performed using a Leica M205 FA stereomicroscope with filter sets ET GFP M205FA/M165FC and ET mCherry M205FA/M165FC.

ACKNOWLEDGMENT

Open access funding enabled and organized by Projekt DEAL.

ORCID

Russell T. Collins https://orcid.org/0000-0002-1847-4517
Baptiste Coxam https://orcid.org/0000-0003-1606-6373
Ines Fechner https://orcid.org/0000-0002-0184-8324
Anna Szymborska https://orcid.org/0000-0003-1741-4106
Holger Gerhardt https://orcid.org/0000-0002-3030-0384

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, *9*(1), 1911. https://doi.org/10.1038/s41467-018-04252-2

Alestrom, P., D'Angelo, L., Midtlyng, P. J., Schorderet, D. F., Schulte-Merker, S., Sohm, F., & Warner, S. (2019). Zebrafish: Housing and husbandry recommendations. *Laboratory Animals*, *54*, 213–224. https://doi.org/10.1177/0023677219869037

Bedell, V. M., Wang, Y., Campbell, J. M., Poshusta, T. L., Starker, C. G., Krug, R. G., 2nd, … Ekker, S. C. (2012). In vivo genome editing using a high-efficiency TALEN system. *Nature*, *491*(7422), 114–118. https://doi.org/10.1038/nature11537

Burns, C. G., Milan, D. J., Grande, E. J., Rottbauer, W., MacRae, C. A., & Fishman, M. C. (2005). High-throughput assay for small molecules that modulate zebrafish embryonic heart rate. *Nature Chemical Biology*, *1*(5), 263–264. https://doi.org/10.1038/nchembio732

Collins, R. T., Linker, C., & Lewis, J. (2010). MAZe: A tool for mosaic analysis of gene function in zebrafish. *Nature Methods*, *7*(3), 219–223. https://doi.org/10.1038/nmeth.1423

Distel, M., Wullimann, M. F., & Koster, R. W. (2009). Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(32), 13365–13370. https://doi.org/10.1073/pnas.0903060106

Foley, J. E., Maeder, M. L., Pearlberg, J., Joung, J. K., Peterson, R. T., & Yeh, J. R. (2009). Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nature Protocols*, *4*(12), 1855–1867. https://doi.org/10.1038/nprot.2009.209

Hruscha, A., Krawitz, P., Rechenberg, A., Heinrich, V., Hecht, J., Haass, C., & Schmid, B. (2013). Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, *140*(24), 4982–4987. https://doi.org/10.1242/dev.099085

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Kaini, P., Sander, J. D., … Yeh, J. R. (2013a). Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One*, *8*(7), e68708. https://doi.org/10.1371/journal.pone.0068708

Hwang, W. Y., Fu, Y., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., … Joung, J. K. (2013b). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*, *31*(3), 227–229. https://doi.org/10.1038/nbt.2501

Irion, U., Krauss, J., & Nusslein-Volhard, C. (2014). Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. *Development*, *141*(24), 4827–4830. https://doi.org/10.1242/dev.115584

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, *203*(3), 253–310. https://doi.org/10.1002/aja.1002030302

Klemenz, R., Weber, U., & Gehring, W. J. (1987). The white gene as a marker in a new P-element vector for gene transfer in drosophila. *Nucleic Acids Research*, *15*(10), 3947–3959. https://doi.org/10.1093/nar/15.10.3947

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., … Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, *7*(1), 2193. https://doi.org/10.1038/s41598-017-02460-2

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, *29*(9), e45. https://doi.org/10.1093/nar/29.9.e45, 445.

Roca, X., Krainer, A. R., & Eperon, I. C. (2013). Pick one, but be quick: 5′ splice sites and the problems of too many choices. *Genes & Development*, *27*(2), 129–144. https://doi.org/10.1101/gad.209759.112

Ruijter, J. M., Ramakers, C., Hoogaars, W. M., Karlen, Y., Bakker, O., van den Hoff, M. J., & Moorman, A. F. (2009). Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*, *37*(6), e45. https://doi.org/10.1093/nar/gkp045

Stephens, R. M., & Schneider, T. D. (1992). Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *Journal of Molecular Biology*, *228*(4), 1124–1136. https://doi.org/10.1016/0022-2836(92)90320-j

Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E. J., … Belmonte, J. C. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, *540*(7631), 144–149. https://doi.org/10.1038/nature20565

Thermes, V., Grabher, C., Ristoratore, F., Bourrat, F., Choulika, A., Wittbrodt, J., & Joly, J. S. (2002). I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mechanisms of Development*, *118*(1–2), 91–98. https://doi.org/10.1016/s0925-4773(02)00218-6

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Collins RT, Coxam B, Fechner I, et al. Intron with transgenic marker (InTraM) facilitates high-throughput screening of endogenous gene reporter lines. *genesis*. 2020;e23391. https://doi.org/10.1002/dvg.23391