Efficient single-copy HDR by 5' modified long dsDNA donors

Jose Arturo Gutierrez-Triana[1,2,3], Thomas Thumberger[1,2], Tinatini Tavhelidse[1,2], Isabelle Thomas[2], Beate Wittbrodt[2], Tanja Kellner[2], Erika Tsingos[2] and Joachim Wittbrodt[2]*

[1]These authors contributed equally to this work.

[2]Centre for Organismal Studies (COS), Heidelberg University, Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany.

[3]Current address: Escuela de Microbiología, Facultad de Salud, Universidad Industrial de Santander, Cra. 32 #29-31, Bucaramanga, Santander, Colombia.

*Corresponding author:

Joachim Wittbrodt

jochen.wittbrodt@cos.uni-heidelberg.de

Developmental Biology and Physiology

Centre for Organismal Studies (COS), Heidelberg University

Im Neuenheimer Feld 230, D-69120 Heidelberg, Germany

Phone +49 (6221) 54-6499

Fax +49 (6221) 54-5639

ABSTRACT

CRISPR/Cas9 efficiently induces targeted mutations via non-homologous-end-joining but for genome editing, precise, homology-directed repair (HDR) of endogenous DNA stretches is prerequisite. To favour HDR, many approaches interfere with the repair machinery or manipulate Cas9 itself. Using Medaka we show that the modification of 5' ends of long dsDNA donors strongly enhances HDR, favours single-copy integration (60% efficiency) by retaining a monomeric donor conformation thus facilitating efficient gene replacement or tagging.

33 **Main**

34 The implementation of the bacterial CRISPR/Cas9 system in eukaryotes has triggered
35 a quantum leap in targeted genome editing in literally any organism with a
36 sequenced genome or targeting region (Cong et al., 2013; Jinek et al., 2012; Mali et
37 al., 2013). Site specific double strand breaks (DSBs) are catalysed by the Cas9 enzyme
38 guided by a single RNA with short complementary region to the target site. In
39 response, the endogenous non-homologous end joining (NHEJ) DNA repair
40 machinery seals the DSB. Since perfect repair will restore the CRISPR/Cas9 target site,
41 mutations introduced by imprecise NHEJ are selected for. To acquire precise genome
42 editing the initial DSB should be fixed via the homology directed repair (HDR)
43 mechanism which is preferentially active during the late S/G2 phase of the cell cycle
44 (Heyer et al., 2010). Yet, donor DNA templates with flanking regions homologous to
45 the target locus are used to introduce specific mutations or particular DNA
46 sequences. The high activity of NHEJ re-ligating CRISPR/Cas9 mediated DSBs also
47 multimerizes injected (linear) dsDNA donor templates (Burdon and Wall, 1992;
48 Winkler et al., 1991). This poses a problem, since consequently precise HDR-
49 mediated recombination of single-copy donor templates is very rare.
50 To enhance HDR, NHEJ was interfered with pharmacologically by inhibiting DNA
51 ligase IV (Maruyama et al., 2015), or HDR was favoured by fusing the HDR-mediating
52 yeast protein Rad52 to Cas9 (Wang et al., 2017), as well as by linking Cas9 to the N-
53 terminal region of the DNA replication inhibitor Geminin, targeting it for depletion in
54 the G1 phase (Gutschner et al., 2016). Those approaches improved HDR-mediated
55 integration of the homology flanks, while they did not tackle reported integration of

56 multimers (Auer et al., 2014; Kumagai et al., 2017; Stemmer et al., 2015) arising after
57 injection of dsDNA templates such as plasmids or PCR products (Burdon and Wall,
58 1992; Winkler et al., 1991).
59 To enhance HDR without interfering with the endogenous DNA repair machinery we
60 aimed at establishing DNA donor templates that escape multimerization or NHEJ
61 events. We thus blocked both 5´ends of PCR amplified long dsDNA donor cassettes
62 using "bulky" moieties like Biotin, Amino-dT (A-dT) and carbon spacers (e. g. Spacer
63 C3, SpC3). This should shield the DNA donor from multimerization and integration
64 via NHEJ, thus favouring precise and efficient single-copy integration via HDR (Fig.
65 1a).
66 We first addressed the impact of the donor 5' modification on the formation of
67 multimers *in vivo*. We injected modified and unmodified dsDNA donors into one cell
68 stage medaka (*Oryzias latipes*) embryos and analysed the conformational state of
69 the injected material during zygotic development. dsDNA donors were generated by
70 PCR employing 5' modified and non-modified primers respectively, and by
71 additionally providing traces of DIG-dUTP for labelling of the resulting PCR product.
72 The conformation of the injected dsDNA donors was assayed in the extracted total
73 DNA after 2, 4 and 6 hours post-injection, respectively. The DNA was size
74 fractionated by gel electrophoresis and donor DNA conformation was detected after
75 blotting the DNA to a nylon membrane by anti-DIG antibodies (Fig. 1b). In
76 unmodified DIG-labelled control donors we uncovered multimerization already at 2h
77 post injection as evident by a ladder of labelled donor DNA representing different
78 copy number multimers (Fig. 1b). In contrast, Biotin and SpC3 modification of DIG-
79 labelled donors prevented multimerization within six hours post-injection (Fig. 1b).

80 dsDNA donors established by A-dT modified primes, however, multimerized and
81 produced similar results as unmodified DIG-labelled dsDNA donors (Fig. 1b). Our
82 results reveal that Biotin and SpC3 5' modifications efficiently prevent donor
83 multimerization *in vivo*. While strongly blocking multimerization, the 5' modification
84 of dsDNA did not apparently enhance the stability of the resulting dsDNA (compare
85 modified and unmodified donors over time, Fig. 1b).
86 To test whether 5' modification not only reduces the degree of multimerization but
87 also impacts on single-copy HDR-mediated integration of long dsDNA donors, we
88 designed *gfp* containing donor cassettes for an immediate visual readout. We
89 generated *gfp* in-frame fusion donors for three different genes: the retinal
90 homeobox transcription factor *rx2* (Reinhardt et al., 2015; Stemmer et al., 2015), the
91 non-muscle cytoskeletal *beta-actin* (*actb*) (Stemmer et al., 2015) and the *DNA methyl*
92 *transferase 1* (*dnmt1*). Donor cassettes contained the respective 5' homology flank
93 (HF), followed by the in frame *gfp* coding sequence, a flexible linker in case of *rx2*
94 and *dnmt1*, and the corresponding 3' HF (schematic representation in Fig. 1a,
95 Supplementary Fig. 1 for detailed donor design). To amplify the long dsDNA donors
96 we employed a pair of universal primers (5' modified or unmodified as control)
97 complementary to the backbone of the cloning vectors (pDestSC-ATG (Kirchmaier et
98 al., 2013) or pCS2+ (Rupp et al., 1994)) encompassing the entire assembled donor
99 cassette (Supplementary Fig. 1).
100 Modified or unmodified long dsDNA donors were subsequently co-injected together
101 with *Cas9* mRNA and the respective locus-specific sgRNA into medaka one-cell stage
102 zygotes (Loosli et al., 1999). For all three loci (and all 5' modifications) tested we

103 observed efficient targeting as apparent by the GFP expression within the expected
104 expression domain (Fig. 1c, Supplementary Table 1, Supplementary Fig. 2).
105 The survival rates of embryos injected with Biotin and SpC3 5' modified dsDNA
106 donors did not differ significantly from embryos injected with the unmodified dsDNA
107 control donors (Supplementary Table 1, Supplementary Fig. 2). In contrast, the
108 injection of the A-dT 5' modified dsDNA donors resulted in high embryonic lethality
109 (Supplementary Table 1, Supplementary Fig. 2).
110 We next analysed the frequency of single-copy HDR events by PCR on genomic DNA
111 of GFP expressing embryos injected with unmodified, Biotin, SpC3 5' modified
112 dsDNA donors. Our approach allowed distinguishing alleles without *gfp* integration
113 (i.e. size of wild type locus) from those generated by HDR and NHEJ respectively and
114 addressed the size of the integration by a locus spanning PCR with a reduced number
115 of PCR cycles (<=30) to omit *in vitro* fusion-PCR artefacts (own data and (Won and
116 Dawid, 2017)). We genotyped GFP expressing embryos using locus primers (Lf / Lr)
117 located distal to the utilized HFs (Supplementary Fig. 1) and addressed the fusion of
118 the *gfp* donor sequence to the target genes (Fig. 1d, Supplementary Figs. 3, 4).
119 In unmodified donors, the rate of HDR was very low as evidenced by the
120 predominant amplification of the alleles without *gfp* integration (Fig. 1d,
121 Supplementary Fig. 3). In strong contrast, the 5' modified long dsDNA donors
122 resulted in efficient HDR already detectable in the injected generation. This was
123 evident for all targeted loci and prominent in the *gfp* tagging of *rx2*. Here 6/10 GFP-
124 expressing embryos injected with a 5'Biotin dsDNA donor showed precise HDR-
125 mediated single-copy integration (Supplementary Fig. 3) as sequence confirmed by
126 the analysis of the locus (Supplementary Fig. 4).

127 We observed the highest efficiency of HDR targeting by 5' modified dsDNA for all
128 three loci tested for 5'Biotin modified donors (Fig. 1d). Already in the injected
129 generation we prominently detected and validated the HDR-mediated fusion of the
130 long dsDNA donors with the respective locus (Supplementary Fig. 4b). While still
131 giving rise to a high percentage of HDR events, SpC3 5' modified dsDNA donors also
132 resulted in elevated levels of additional bands indicative for the integration of higher
133 order multimers and NHEJ events (Supplementary Fig. 3).
134 The precise integration detected in the injected generation was successfully
135 transmitted to the next generation (Fig. 2, Supplementary Fig. 4). We addressed the
136 nature of the insertion predicted to be single-copy by a combination of PCR and
137 expression studies. We validated the genomic organization of the *gfp-rx2* knock-in in
138 homozygous F2 animals (Fig. 2a) by genomic sequencing (Supplementary Fig. 4c) and
139 Southern Blot (Southern, 2006) analysis (Fig. 2b, 2b'). Here we detected a single
140 band indicative for single-copy HDR-mediated integration, when probing digested
141 genomic DNA of F2 *gfp-rx2*$^{+/+}$ knock-in fish by enzyme combinations releasing the
142 target region including the HFs (BglII/HindIII, Fig. 2b-b') or cutting in the 5' HF
143 (ScaI/HindIII, Fig. 2b, b'), validating the PCR based prediction and sequencing results.
144 Furthermore, transcript analysis in homozygous F3 embryos exclusively uncovered a
145 single fusion *gfp-rx2* transcript (Fig. 2c, 2c'). This molecular analysis confirmed that
146 the 5'Biotin modification of the long dsDNA donor promoted precise single-copy
147 HDR-mediated integration with high efficiency.
148 Taken together the simple 5'Biotin modification at both ends of long dsDNA donors
149 by conventional PCR amplification presented here provides the means to favour HDR
150 without interfering with the cellular DNA repair machinery. Our approach facilitates

151 the highly efficient detection of HDR events by GFP tagging even in non-essential
152 genes. Other than in non-modified dsDNA donors, the locus specific expression of
153 GFP is an excellent predictor for precise, single-copy HDR-mediated integration
154 already in the injected generation. This allows an easy selection and results in HDR
155 rates of up to 60%. This rate can even be higher in essential loci, where the full
156 functionality of the (modified) gene under investigation is required for the survival of
157 the affected cells, tissues, organs or organisms, thus providing the means for
158 effective positive selection for single, fully functional integrates. It is interesting to
159 note that approximately 400 bp flanking the insertion site in 5' and 3' direction are
160 sufficient for HDR. The use of modified universal primers for the two donor template
161 vectors employed allowed a flexible application of the procedure and ensured a high
162 quality of the PCR generated 5' modified long dsDNA donors. Strikingly, the primer-
163 introduced non-homology regions in the very periphery of the long dsDNA donors
164 did not negatively impact on the HDR efficiency.
165 Taken together, the simplicity and high reproducibility of our proof-of-concept
166 analysis highlight that the presented protection of both 5' ends of long dsDNA
167 donors prevent multimerization and promote precise insertion/replacement of DNA
168 elements, thus facilitating functional studies in basic research as well as therapeutic
169 interventions.
170

171 **Methods**

172 A step-by-step protocol is available as a Supplementary Protocol and at Protocol
173 Exchange.

174

175 **Fish maintenance**

176 All fish are maintained in closed stocks at Heidelberg University. Medaka (*Oryzias*
177 *latipes*) husbandry (permit number 35–9185.64/BH Wittbrodt) and experiments
178 (permit number 35-9185.81/G-145/15 Wittbrodt) were performed according to local
179 animal welfare standards (Tierschutzgesetz §11, Abs. 1, Nr. 1) and in accordance
180 with European Union animal welfare guidelines (Bert et al., 2016). The fish facility is
181 under the supervision of the local representative of the animal welfare agency.
182 Embryos of medaka of the wildtype Cab strain were used at stages prior to hatching.
183 Medaka were raised and maintained as described previously (Köster et al., 1997).

184

185 **Donor plasmids**

186 *Rx2* and *actb* template plasmids for *gfp* donor cassette amplification are described in
187 Stemmer et al. (Stemmer et al., 2015) and were generated by GoldenGATE assembly
188 into the pGGDestSC-ATG destination vector (addgene #49322) according to
189 Kirchmaier et al. (Kirchmaier et al., 2013). See Supplementary Table 2 for primers
190 used to amplify respective homology flanks. The *dnmt1 gfp* plasmid was cloned with
191 homology flanks (5' HF 402 bp, primers dnmt1 5'HF f / dnmt1 5'HF r; 3' HF 405 bp,
192 dnmt1 3'HF f/ dnmt1 3'HF r) that were PCR amplified with Q5 polymerase (New
193 England Biolabs, 30 cycles) from wildtype medaka genomic DNA. *mgfp-flexible linker*
194 was amplified with primers mgfpf / mgfpr. The respective restriction enzyme was

195 used to digest the amplicons (5'HF: SalI HF (New England Biolabs), AgeI HF (New
196 England Biolabs); *mgfp-flexible linker*: AgeI HF (New England Biolabs), SpeI HF (New
197 England Biolabs); 3'HF: SpeI HF (New England Biolabs), NotI HF (New England
198 Biolabs)) followed by gel purification (AnalyticJena) and ligation into pCS2+ (Rupp et
199 al., 1994) (digested with SalI HF (New England Biolabs), NotI HF (New England
200 Biolabs)). All primers were obtained from Eurofins Genomics.
201
202 **Donor amplification**
203 We designed universal primers that match the pGGDestSC-ATG (Kirchmaier et al.,
204 2013) (addgene #49322) or pCS2+ (Rupp et al., 1994) backbone encompassing the
205 assembled inserts (i.e. the *gfp* donor cassette). Unmodified control primers (pDest f,
206 pDest r, pCS2 f, pCS2 r) were ordered from Eurofins Genomics. Modified primers
207 obtained from Sigma-Aldrich (pDest f mod, pDest r mod, pCS2 f mod, pCS2 r mod)
208 consist of the same sequences with phosphorothioate bonds in the first five
209 nucleotides and 5'moiety extension: 5'Biotin, 5'Amino-dT or 5'Spacer C3.
210 The dsDNA donor cassettes were amplified by PCR using 1x Q5 reaction buffer, 200
211 µM dNTPs, 200 µM primer forward and reverse and 0.6 U/µl Q5 polymerase (New
212 England Biolabs). Conditions used: initial denaturation at 98°C 30 sec, followed by 35
213 cycles of: denaturation at 98°C 10 sec, annealing at 62°C 20 sec and extension at
214 72°C 30 sec per kb and a final extension step of 2 min at 72°C. The PCR reaction was
215 treated with 20 units of DpnI (New England Biolabs) to remove any plasmid template
216 following gel purified using the QIAquick Gel Extraction Kit (Qiagen, 28706) and
217 elution with 20µl nuclease free water.
218

219 The *LacZ* cassette of the pGGDestSC-ATG (Kirchmaier et al., 2013) (addgene #49322)
220 which served as DIG labelled dsDNA fragment to test *in vivo* multimerization was
221 amplified via Q5-PCR as above using a mixture of 200 µM dATP, dCTP, dGTP, 170 µM
222 dTTP and 30 µM DIG-dUTP and purified as detailed.
223
224 **sgRNA target site selection**
225 *Dnmt1* sgRNAs were designed with CCTop as described in Stemmer et al. (Stemmer
226 et al., 2015). sgRNAs for *rx2* and *actb* were the same as in Stemmer et al. (Stemmer
227 et al., 2015). The following target sites close to the translational start codons were
228 used (PAM in brackets): *rx2* (GCATTTGTCAATGGATACCC[TGG]), *actb*
229 (GGATGATGACATTGCCGCAC[TGG]) and *dnmt1* (TGACATCGTCTGGCAAAGAC[AGG]).
230 Cloning of sgRNA templates and *in vitro* transcription was performed as detailed in
231 Stemmer, et al. (Stemmer et al., 2015).
232
233 ***In vitro* transcription of mRNA**
234 The pCS2+ *Cas9* plasmid was linearized using NotI and the mRNA was transcribed *in*
235 *vitro* using the mMessage_mMachine SP6 kit (ThermoFisher Scientific, AM1340).
236
237 **Microinjection and screening**
238 Medaka zygotes were injected with 10 ng of DIG labelled donors and were allowed
239 to develop until 2, 4 and 6 hours post injection. For the CRISPR/Cas9 experiments,
240 medaka zygotes were injected with 5 ng/µl of either unmodified and modified long
241 dsDNA donors together with 150 ng/µl of *Cas9* mRNA and 15-30 ng/µl of the gene
242 specific sgRNAs. Injected embryos were maintained at 28°C in embryo rearing

243 medium (ERM, 17 mM NaCl, 40 mM KCl, 0.27 mM $CaCl_2$, 0.66 mM $MgSO_4$, 17 mM
244 Hepes). One day post injection (dpi) embryos were screened for survival, GFP
245 expression was scored at two dpi.
246
247 **Southern Blot**
248 In order to check for multimerization of unmodified and modified donors, we used a
249 modified Southern Blot approach. In brief, embryos were injected with DIG-labelled
250 donors which were PCR-amplified from pGGDestSC-ATG (addgene #49322) using
251 primers pDest f/pDest r (*LacZ* cassette) harbouring either no 5' moiety or one of the
252 following: 5'Biotin, Amino-dT or Spacer C3. 2, 4 and 6 hours post injection, 30
253 embryos were lysed in TEN buffer plus proteinase K (10 mM Tris pH 8, 1 mM EDTA,
254 100 mM NaCl, 1 mg/ml proteinase K) at 60°C over night. DNA was ethanol
255 precipitated after removal of lipids and proteins by phenol-chloroform extraction.
256 Total DNA was resuspended in TE buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA pH 8.0).
257 200 ng of each sample was run on a 0.8% agarose gel. As control, 100 pg of
258 uninjected donor PCR product were loaded. The agarose gel was transferred to a
259 nylon membrane over night using 10x SSC (1.5 M NaCl, 0.15 M $C_6H_5Na_3O_7$) as
260 transfer solution. The crosslinked membrane was directly blocked in 1% w/v blocking
261 reagent (Roche) in 1x DIG1 solution (0.1 M maleic acid, 0.15 M NaCl, pH 7.5) and the
262 labelled DNA was detected using CDP star (Roche) following the manufacturer's
263 instructions.
264 In order to check for copy number insertions in the *gfp-rx2* transgenic line, genomic
265 DNA was isolated as described above from F2 embryos expressing GFP. 10 µg
266 digested genomic DNA were loaded per lane on a 0.8% agarose gel and size

267 fractionated by electrophoresis. The gel was depurinated in 0.25 N HCl for 30 min at
268 room temperature, rinsed with H_2O, denatured in 0.5 N NaOH, 1.5 M NaCl solution
269 for 30 min at room temperature and neutralised in 0.5 M Tris HCl, 1.5 M NaCl, pH 7.2
270 before it was transferred over night at room temperature onto a Hybond membrane
271 (Amersham). The membrane was washed with 50 mM NaPi for 5 min at room
272 temperature, then crosslinked and pre-hybridized in Church hybridization buffer
273 (0.5 M NaPi, 7% SDS, 1 mM EDTA pH 8.0) at 65°C for at least 30 min. The probe was
274 synthesised from the donor plasmid with primers gfp probe f and gfp probe r using
275 the PCR DIG Probe Synthesis Kit (Roche, 11636090910) and the following PCR
276 protocol: initial denaturation at 95°C for 2 min, 35 cycles of 95°C 30 sec, 60°C 30 sec,
277 72°C 40 sec and final extension at 72°C 7 min. The probe was boiled in hybridization
278 buffer for 10 min at 95°C and the membrane was hybridized over night at 65°C. The
279 membrane was washed with pre-heated (65°C) Church washing buffer (40 mM NaPi,
280 1% SDS) at 65°C for 10 min, then at room temperature for 10 min and with 1x DIG1
281 and 0.3% Tween for 5 min at room temperature. The membrane was blocked in 1%
282 w/v blocking reagent (Roche) in 1x DIG1 solution at room temperature for at least
283 30 min. The membrane was incubated with 1:10,000 anti-digoxigenin-AP Fab
284 fragments (Roche) for 30 min at room temperature in 1% w/v blocking reagent
285 (Roche) in 1x DIG1 solution. Two washing steps with 1x DIG1 and 0.3% Tween were
286 performed for 20 min at room temperature, followed by a 5 min washing step in 1x
287 DIG3 (0.1 M Tris pH 9.5, 0.1 M NaCl) at room temperature. Detection was performed
288 using 6 µl/ml CDP star (Roche).
289
290

291 **Genotyping**

292 Single injected GFP positive embryos were lysed in DNA extraction buffer (0.4 M
293 Tris/HCl pH 8.0, 0.15 M NaCl, 0.1% SDS, 5 mM EDTA pH 8.0, 1 mg/ml proteinase K) at
294 60°C overnight. Proteinase K was inactivated at 95°C for 10 min and the solution was
295 diluted 1:2 with H_2O. Genotyping was performed in 1x Q5 reaction buffer, 200 µM
296 dNTPs, 200 µM primer forward and reverse and 0.012 U/µl Q5 polymerase and 2 µl
297 of diluted DNA sample and the respective locus primers. The conditions were: 98°C
298 30 sec, 30 cycles of 98°C 10 sec, annealing for 20 sec and 72°C 30 sec per kb
299 (extension time used would allow for detecting potential NHEJ events on both ends
300 of the donor) (rx2 Lf/rx2 Lr: 68°C annealing, 90 sec extension time; actb Lf/actb Lr:
301 66°C annealing, 84 sec extension time; dnmt1 Lf/dnmt1 Lr: 65°C annealing, 90 sec
302 extension time) and a final extension of 2 min at 72°C. PCR products were analysed
303 on a 1% agarose gel.

304

305 **RT-PCR**

306 Total RNA was isolated from 60 homozygous embryos (stage 32) by lysis in TRIzol
307 (Ambion) and chloroform extraction according to the manufacturer's protocol. RNA
308 was precipitated using isopropanol and resuspended in H_2O. cDNA was reverse
309 transcribed with Revert Aid Kit (Thermo Fisher Scientific) after DNAse digestion and
310 inactivation following the manufacturer's instructions. PCR was performed using
311 5'UTRf, 3'UTRr and Q5 polymerase (New England Biolabs): 98°C 30 sec, 35 cycles of
312 98°C 10 sec, annealing 65°C for 20 sec and 72°C 210 sec and a final extension of 2
313 min at 72°C. PCR products were analysed on a 1.5% agarose gel.

314

315 **Sequencing**

316 Plasmids and PCR fragments were sequenced with the indicated primers by a

317 commercial service (Eurofins Genomics).

318

Acknowledgements:

This research was funded through the German Science funding agency (DFG, CRC 873, Project A3 to J.W.). T.Ta. and E.T. are members of HBIGS, the Heidelberg Biosciences International Graduate School. We are grateful to M. Majewsky, E. Leist and A. Saraceno for fish husbandry. We thank all members of the Wittbrodt lab for their critical, constructive feedback on the procedure and the manuscript.

Author contributions:

A.G., T.T., T.Ta. and J.W. designed the study, analysed the data and wrote the manuscript. A.G., T.T. and T.Ta. performed injection experiments and acquired the data. I.T. contributed to the rx2 experiments, B.W. conducted genotyping PCR experiments and bred fish lines, T.K. cloned the donor plasmids and sgRNA vectors and generated *Cas9* mRNA and sgRNAs, E.T. performed the Southern blot.

Competing interest statement:

The authors state that there are no competing financial interests

337 **References**

338 Auer, T.O., Duroure, K., De Cian, A., Concordet, J.P., and Del Bene, F. (2014). Highly
339 efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA
340 repair. Genome Research *24*, 142–153.

341 Bert, B., Chmielewska, J., Bergmann, S., Busch, M., Driever, W., Finger-Baier, K.,
342 Hößler, J., Köhler, A., Leich, N., Misgeld, T., et al. (2016). Considerations for a
343 European animal welfare standard to evaluate adverse phenotypes in teleost fish.
344 The EMBO Journal *35*, 1151–1154.

345 Burdon, T.G., and Wall, R.J. (1992). Fate of microinjected genes in preimplantation
346 mouse embryos. Mol. Reprod. Dev. *33*, 436–442.

347 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W.,
348 Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas
349 Systems. Science *339*, 819–823.

350 Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F., and Chin, L. (2016). Post-
351 translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair.
352 CellReports *14*, 1555–1566.

353 Heyer, W.-D., Ehmsen, K.T., and Liu, J. (2010). Regulation of homologous
354 recombination in eukaryotes. Annu. Rev. Genet. *44*, 113–139.

355 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E.
356 (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial
357 immunity. Science *337*, 816–821.

358 Kirchmaier, S., Lust, K., and Wittbrodt, J. (2013). Golden GATEway cloning--a
359 combinatorial approach to generate fusion and recombination constructs. PLoS ONE
360 *8*, e76117.

361 Köster, R., Stick, R., Loosli, F., and Wittbrodt, J. (1997). Medaka spalt acts as a target
362 gene of hedgehog signaling. Development *124*, 3147–3156.

363 Kumagai, H., Nakanishi, T., Matsuura, T., Kato, Y., and Watanabe, H. (2017).
364 CRISPR/Cas-mediated knock-in via non-homologous end-joining in the crustacean
365 Daphnia magna. PLoS ONE *12*, e0186112.

366 Loosli, F., Winkler, S., and Wittbrodt, J. (1999). Six3 overexpression initiates the
367 formation of ectopic retina. Genes Dev. *13*, 649–654.

368 Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for
369 engineering biology. Nat Meth *10*, 957–963.

370 Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R., and Ploegh,
371 H.L. (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by
372 inhibition of nonhomologous end joining. Nat. Biotechnol. *33*, 538–542.

373 Reinhardt, R., Centanin, L., Tavhelidse, T., Inoue, D., Wittbrodt, B., Concordet, J.-P.,
374 Martinez-Morales, J.R., and Wittbrodt, J. (2015). Sox2, Tlx, Gli3, and Her9 converge
375 on Rx2 to define retinal stem cells in vivo. The EMBO Journal *34*, 1572–1588.

376 Rupp, R.A., Snider, L., and Weintraub, H. (1994). Xenopus embryos regulate the
377 nuclear localization of XMyoD. Genes Dev. *8*, 1311–1323.

378 Southern, E. (2006). Southern blotting. Nature Protocols *1*, 518–525.

379 Stemmer, M., Thumberger, T., del Sol Keyer, M., Wittbrodt, J., and Mateo, J.L.
380 (2015). CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool.
381 PLoS ONE *10*, e0124633.

382 Wang, L., Yang, L., Guo, Y., Du, W., Yin, Y., Zhang, T., and Lu, H. (2017). Enhancing
383 Targeted Genomic DNA Editing in Chicken Cells Using the CRISPR/Cas9 System. PLoS
384 ONE *12*, e0169768.

385 Winkler, C., Vielkind, J.R., and Schartl, M. (1991). Transient expression of foreign
386 DNA during embryonic and larval development of the medaka fish (Oryzias latipes).
387 Mol. Gen. Genet. *226*, 129–140.

388 Won, M., and Dawid, I.B. (2017). PCR artifact in testing for homologous
389 recombination in genomic editing in zebrafish. PLoS ONE *12*, e0172802.

390

391 **Figure legends**

392

393 **Figure 1. Modification of 5' ends of long dsDNA fragments prevents *in vivo***

394 **multimerization and promotes HDR-mediated single-copy integration**.

395 (a) Schematic representation of long dsDNA donor cassette PCR amplification with

396 universal primers (black arrows) complementary to the cloning vector backbone

397 outside of the assembled donor cassette (e. g. *gfp* with homology flanks). Bulky

398 moieties like Biotin at the 5' ends of both modified primers (red octagon) prevent

399 multimerization/NHEJ of dsDNA, providing optimal conditions for HDR-mediated

400 single-copy integration following CRISPR/Cas9 introduced DSB at the target locus

401 (grey scissors). Representation of locus (Lf/Lr) and internal *gfp* (Gf/Gr) primers for

402 PCR genotyping of putative HDR-mediated *gfp* integration events. (b) Southern Blot
403 analysis reveals the monomeric state of injected dsDNA fragments *in vivo* for 5'
404 modification with Biotin or Spacer C3. Long dsDNAs generated with control
405 unmodified primers or Amino-dT attached primers multimerize as indicated by high
406 molecular weight ladder apparent already within two hours post injection (hpi).
407 Note: 5' moieties did not enhance the stability of injected DNA. (c) GFP expression in
408 the respective expression domain after HDR-mediated integration of modified
409 dsDNA *gfp* donor cassettes into *rx2*, *actb* and *dnmt1* ORFs in the injected generation.
410 (d) Individual embryo PCR genotyping highlights efficient HDR-mediated single-copy
411 integration of 5'Biotin modified long dsDNA donors, but not unmodified donor
412 cassettes. Locus PCR reveals band size indicative of single-copy *gfp* integration
413 (asterisk) besides alleles without *gfp* integration (open arrowhead). Amplification of
414 *gfp* donor (white arrow) for control.
415
416 **Figure 2. Single-copy integration of long dsDNA donor establishes stably**
417 **transmitted *gfp-rx2* fusion gene.** (a) F2 homozygous embryos exhibit GFP-Rx2 fusion
418 protein expression in the pattern of the endogenous gene in the retina. (b) Southern
419 Blot analysis of individual F2 homozygous fish reveals a single band for a digestion
420 scheme cutting outside the donor cassette (BglII/HindIII) or within the 5' donor
421 cassette and in intron 2 (ScaI/HindIII) indicating precise single-copy donor integration.
422 (b') Schematic representation of the modified locus indicating the restriction sites
423 and the domain complementary to the probe used in (b).

424 (c) RT-PCR analysis on mRNA isolated from individual homozygous F3 embryos
425 indicates the transcription of a single *gfp-rx2* fusion mRNA in comparison to the
426 shorter wild type *rx2* mRNA as schematically represented in (c').
427
428 **Supplementary Figure 1. Schematic representation of the donor plasmids**
429 (a-c) Schematic to-scale representation of respective target locus (a, *rx2*; b, *actb*; c,
430 *dnmt1*) with UTR (white boxes with red outlines) and exons (red boxes) highlighted.
431 Homology flanks (HF, grey), sgRNA target sites (white scissors) and respective locus
432 primers (black arrows, Lf, Lr, see Supplementary Table 2) are indicated. Respective
433 assembled donor plasmid (a, b, Golden GATEway (Kirchmaier et al., 2013; Stemmer
434 et al., 2015) or c, conventional cloning) that served as template for PCR amplification
435 of unmodified/modified long dsDNA *gfp* donor cassette is depicted below (FL,
436 flexible linker; backbone sequence in blue). Entry vectors (EV) or restriction enzyme
437 sites for cloning are indicated. Position of primers flanking the donor cassette is
438 indicated, modification highlighted by red octagon.
439
440 **Supplementary Figure 2. Quantification of survival and GFP expression of injected**
441 **embryos**
442 Embryos injected with unmodified or 5'Biotin, Amino-dT or Spacer C3 modified long
443 dsDNA *gfp-rx2* donor cassettes were scored for survival at one dpi, and for GFP
444 expression at two dpi. n, total number of injected embryos.
445
446

447 **Supplementary Figure 3. 5'Biotin modification of long dsDNA donors strongly**

448 **enhances HDR-mediated integration.**

449 (a) *rx2* locus PCR genotyping of individual GFP-Rx2 positive embryos injected with

450 unmodified or 5'Biotin or Spacer C3 modified long dsDNA *gfp* donor cassettes (green

451 asterisk, single-copy HDR-mediated integration of *gfp*, 2547 bp, open arrowhead, *rx2*

452 allele without *gfp* integration, 1719 bp;). Horizontal bar, individual embryo; L, *rx2*

453 locus PCR with rx2 Lf / rx2 Lr; G, *gfp* internal PCR for control with Gf / Gr. (b)

454 Qualitative summary of band spectrum (single-copy HDR-mediated *gfp* integration,

455 *rx2* allele without *gfp* integration, other) resulting from PCR genotyping in (a).

456 n, number of genotyped GFP-Rx2 expressing embryos.

457

458

459 **Supplementary Figure 4. Stable germ line transmission of the single-copy HDR-**

460 **mediated precise *gfp* integration.**

461 (a) Schematic to-scale representation of the *gfp-rx2* locus with UTR (white boxes

462 with red outlines), exons (red boxes) and homology flanks (HF, grey) highlighted.

463 (b, c) Individual GFP-Rx2 expressing embryos (F0, b; F1, c) were genotyped using

464 primers rx2 Lf / rx2 Lr (black arrows) and sequenced with gfpf and gfpr primers

465 (green arrows).

466

467

468 **Supplementary Table 1. Analysis of injected embryos.**

469 Embryos injected with unmodified or modified (5'Biotin, Amino-dT, Spacer C3) long

470 dsDNA *gfp* donor cassettes matching the *rx2*, *actb* or *dnmt1* locus, were scored for

471 GFP expression and survival. Injections without *Cas9* mRNA for control.

472

locus	*rx2*						*actb*		*dnmt1*	
modification	-	-	Biotin	Biotin	A-dT	SpC3	-	Biotin	-	Biotin
Cas9	-	+	-	+	+	+	+	+	+	+
Total injected	93	640	133	474	379	395	65	105	94	105
GFP positive	0	41	0	69	29	65	17	33	11	9
dead	3	54	1	38	192	87	15	34	41	4
%GFP/injected	0.00	6.41	0.00	14.56	7.65	16.46	26.15	31.43	11.70	8.57
% death rate	3.23	8.44	0.75	8.02	50.66	22.03	23.08	32.38	43.62	3.81

473

474

475 **Supplementary Table 2. Oligonucleotides used in this work.**

476 Restriction enzyme sites used for cloning of PCR amplicons are indicated in italics.
477 Substituted nucleotides to facilitate T7 *in vitro* transcription of the sgRNA
478 oligonucleotides are shown in small letters (Stemmer et al., 2015). Locus primers
479 forward (Lf) and reverse (Lr) of respective gene loci, *gfp* primers forward (Gf) and
480 reverse (Gr), *gfp* sequencing primers gfpf and gfpr and primers to amplify the *mgfp-*
481 *flexible linker* as well as the *gfp* probe for Southern Blot analysis are given. Asterisks
482 indicate phosphorothioate bonds, "5'moiety" was either 5'Biotin, Amino-dT or
483 Spacer C3 in the pDest f mod, pDest r mod, pCS2 f mod and pCS2 r mod primers.
484

primer name	**primer sequence 5'-3'**
rx2	
rx2 5'HF f (BamHI)	GCC*GGATCC*AAGCATGTCAAAACGTAGAAGCG
rx2 5'HF r (KpnI)	GCC*GGTACC*CATTTGGCTGTGGACTTGCC
rx2 3'HF f (BamHI)	GCC*GGATCC*CATTTGTCAATGGACACGCTTGGGATGGTGGACG AT
rx2 3'HF r (KpnI)	GCC*GGTACC*TGGACTGGACTGGAAGTTATTT
rx2 sgRNA f	TAgGCATTTGTCAATGGATACCC
rx2 sgRNA r	AAACGGGTATCCATTGACAAATG
rx2 Lf / 5'UTRf	TGCATGTTCTGGTTGCAACG

rx2 Lr	AGGGACCATACCTGACCCTC
3'UTRr	GCGACAGCTATTCCACATAAAAAC
actb	
actb 5'HF f (BamHI)	GG*GGATCC*CAGCAACGACTTCGCACAAA
actb 5'HF r (KpnI)	GG*GGTACC*GGCAATGTCATCATCCATGGC
actb 3'HF f (BamHI)	GG*GGATCC*GACGACGATATAGCTGCACTGGTTGTTGACAACG GATCTG
actb 3'HF r (KpnI)	GG*GGTACC*CAGGGGCAATTCTCAGCTCA
actb sgRNA f	TAGGATGATGACATTGCCGCAC
actb sgRNA r	AAACGTGCGGCAATGTCATCAT
actb Lf	GTCCGAGTTGAGGGTGTCTG
actb Lr	CATGTGCTCCACTGTGAGGT
dnmt1	
dnmt1 5'HF f (SalI)	AATTT*GTCGAC*GCTTTGACAGTTAACCTACACG
dnmt1 5'HF r (AgeI)	AATTT*ACCGGT*CGTAACTGCAAACTAAAAAATAAAAC
dnmt1 3'HF f (SpeI)	AATTT*ACTAGT*ATGCCATCCAGAACGTCCTTATCTCTACCAGAC GATGTCAGAAAAAGGTAC
dnmt1 3'HF r	AATTT*GCGGCCGC*CTACACATATTGTCTGTGATAC

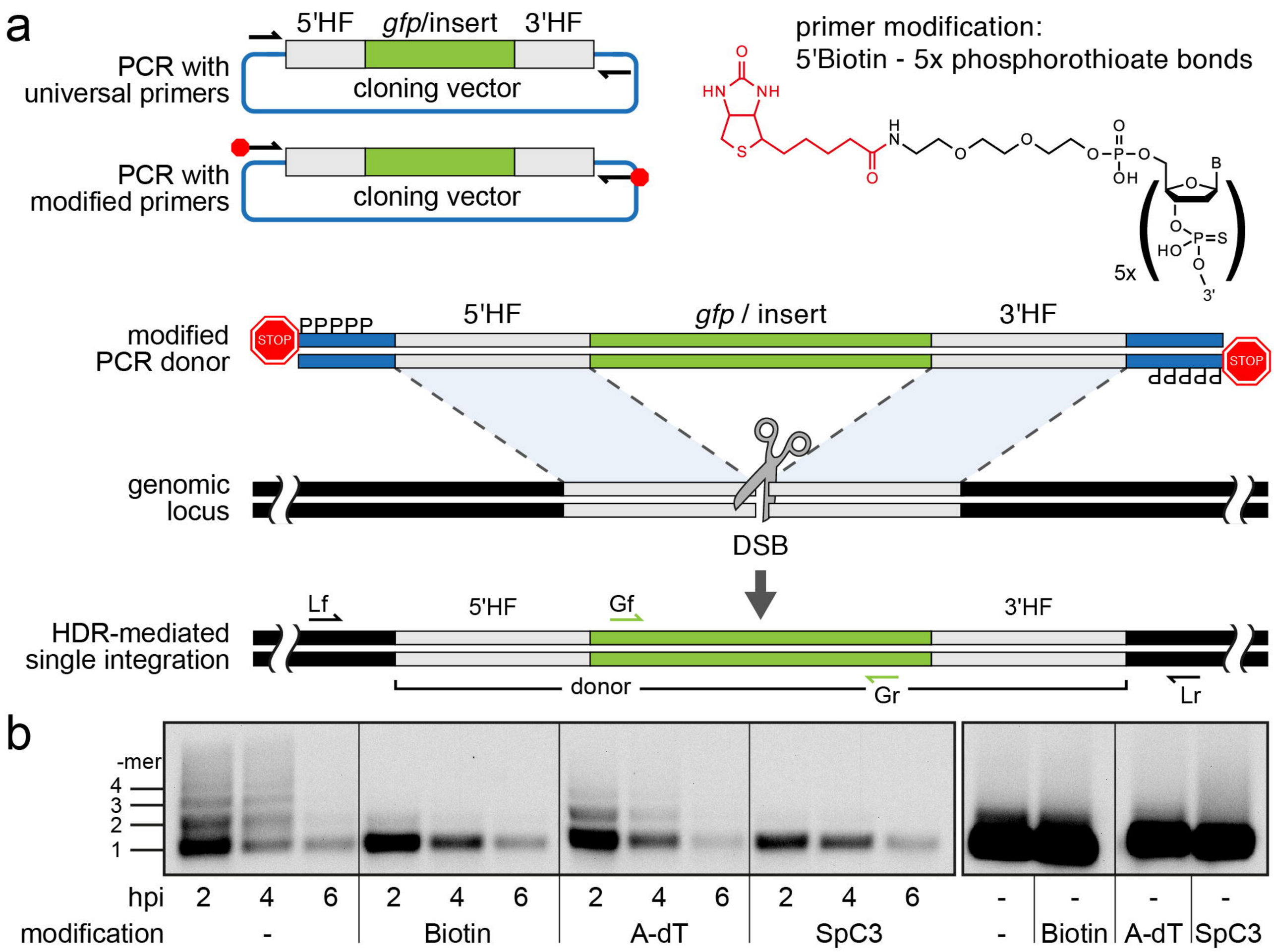
(NotI)	
mgfpf (*AgeI*)	AATTT*ACCGGT*ACTAGTACCATGAGTAAAGGAGAAGAACTTTT CAC
mgfpr (*SpeI*)	AATTT*ACTAGT*CGCGGCTGCACTTCCACCGCCTCCCGATCCGCC ACCGCCAGAGCCACCTCCGCCTGAACCGCCTCCACCGCTCAGG CTAGCTTTGTATAGTTCATCCATGCCATG
dnmt1 sgRNA f	TAgGACATCGTCTGGCAAAGAC
dnmt1 sgRNA r	AAACGTCTTTGCCAGACGATGT
dnmt1 Lf	CTCAATGTAAACACTTCGTGTCGCTTC
dnmt1 Lr	TTGCATGCATATTCAAAGTTGTCAAAG
Backbone primers for *gfp* cassette amplification	
pDest f (control)	CGAGCGCAGCGAGTCAGTGAG
pDest r (control)	CATGTAATACGACTCACTATAG
pDest f mod	5'moiety-C*G*A*G*C*GCAGCGAGTCAGTGAG
pDest r mod	5'moiety-C*A*T*G*T*AATACGACTCACTATAG
pCS2 f	CCATTCAGGCTGCGCAACTG
pCS2 r	CACACAGGAAACAGCTATGAC
pCS2 f mod	5'moiety-C*C*A*T*T*CAGGCTGCGCAACTG
pCS2 r mod	5'moiety-C*A*C*A*C*AGGAAACAGCTATGAC
diagnostic *gfp* primer	
Gf	ATGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC
Gr	CTCAGGTAGTGGTTGTCG
gfpf	GCTCGACCAGGATGGGCA

gfpr	CTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG
gfp probe f	GTGAGCAAGGGCGAGGAGCT
gfp probe r	CTTGTACAGCTCGTCCATG

485

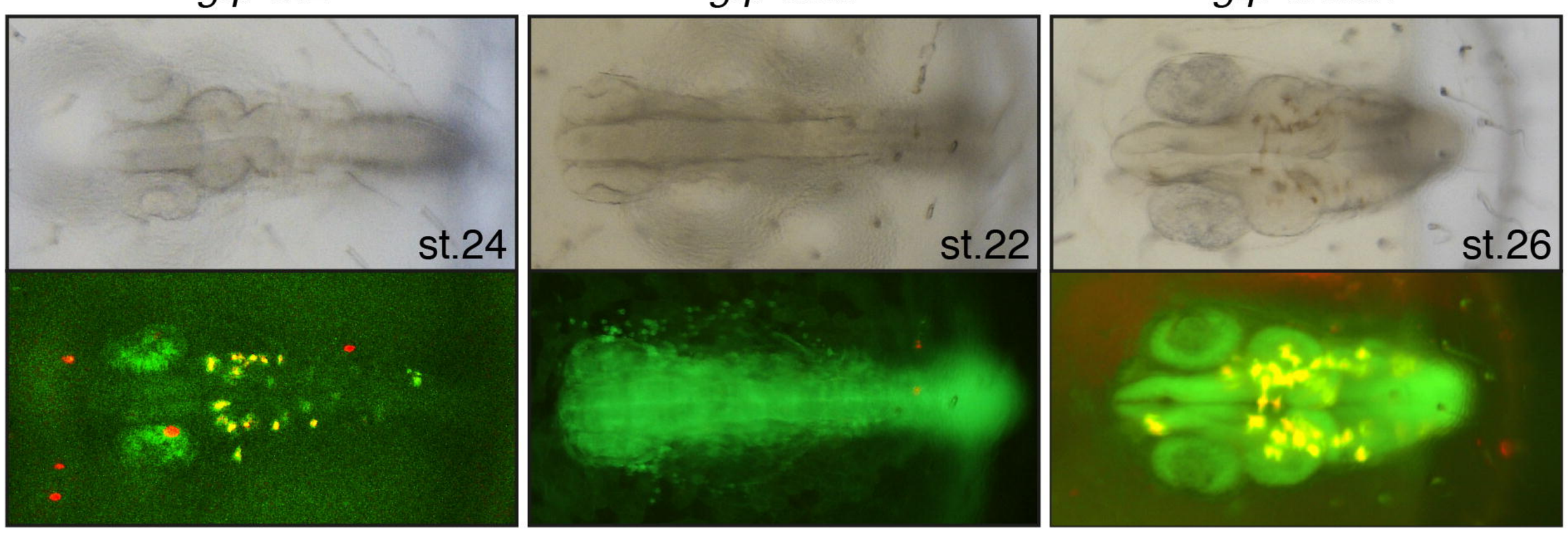
486

a


5'HF
gfp/insert
3'HF
PCR with universal primers
cloning vector
PCR with modified primers
cloning vector
primer modification:
5'Biotin - 5x phosphorothioate bonds
5x
modified PCR donor
PPPPP
STOP
5'HF
gfp / insert
3'HF
genomic locus
DSB
HDR-mediated single integration
Lf
5'HF
Gf
3'HF
donor
Gr
Lr
-mer
4
3
2
1
hpi 2 4 6 2 4 6 2 4 6 2 4 6 - - - -
modification - Biotin A-dT SpC3 - Biotin A-dT SpC3

b

c


gfp-rx2
gfp-actb
gfp-dnmt1
st.24
st.22
st.26

d

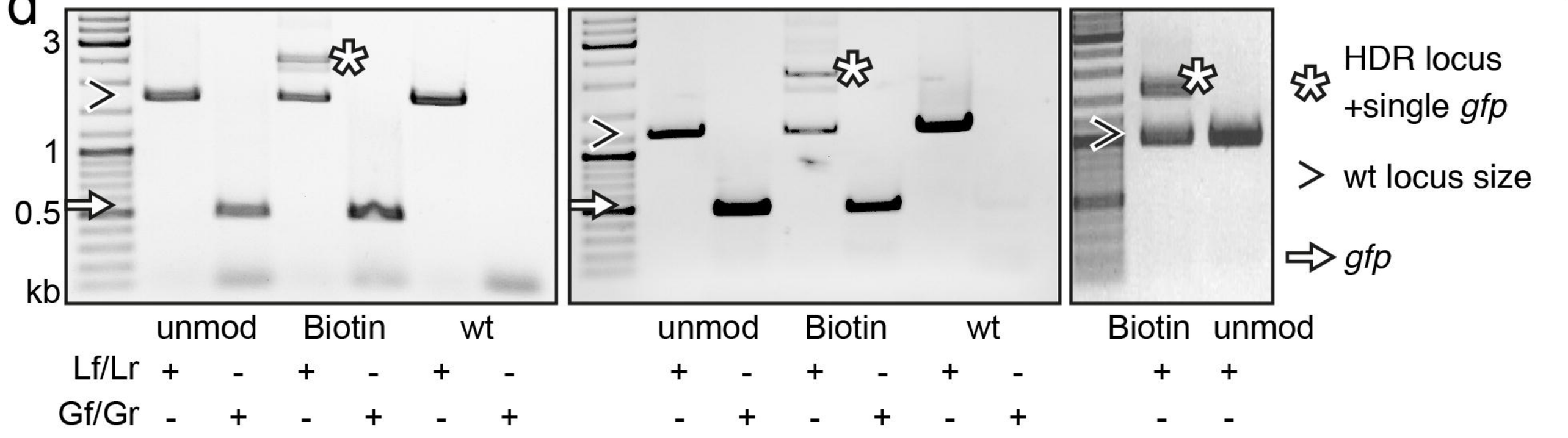
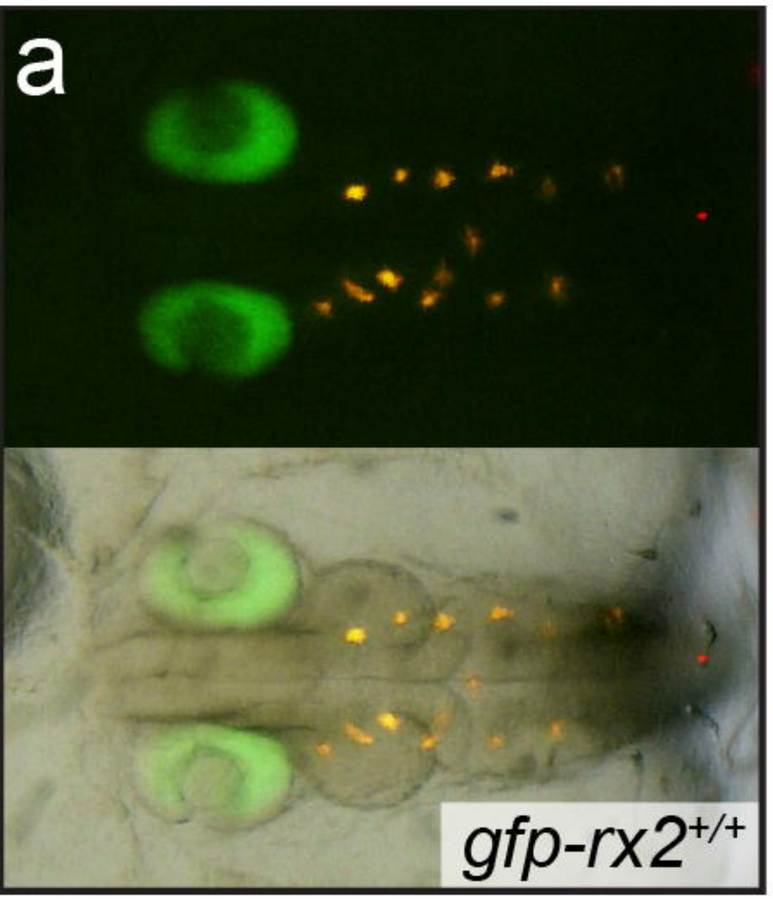
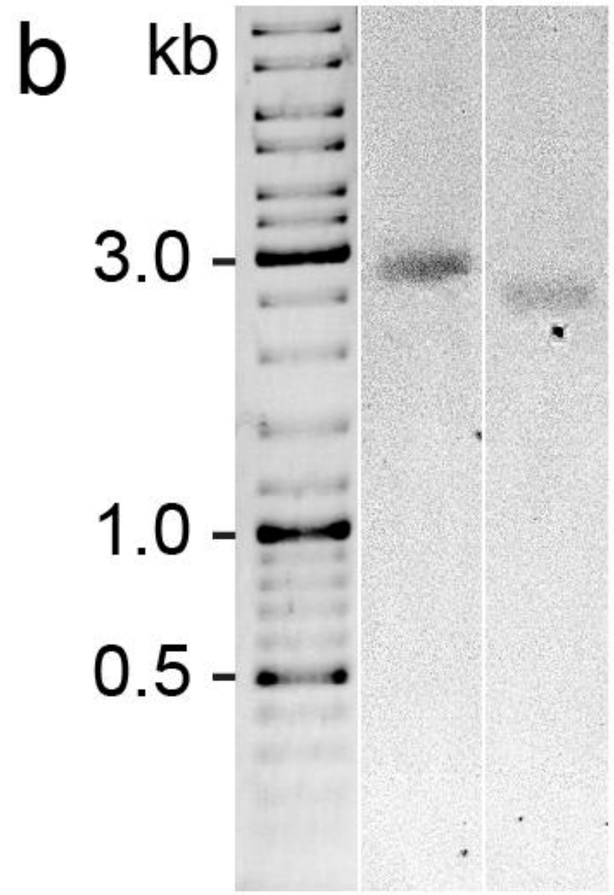
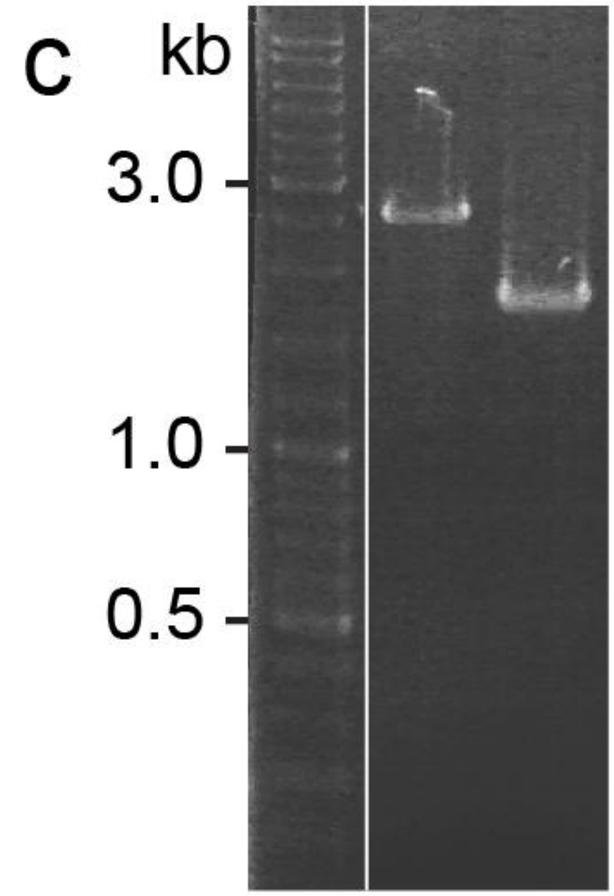
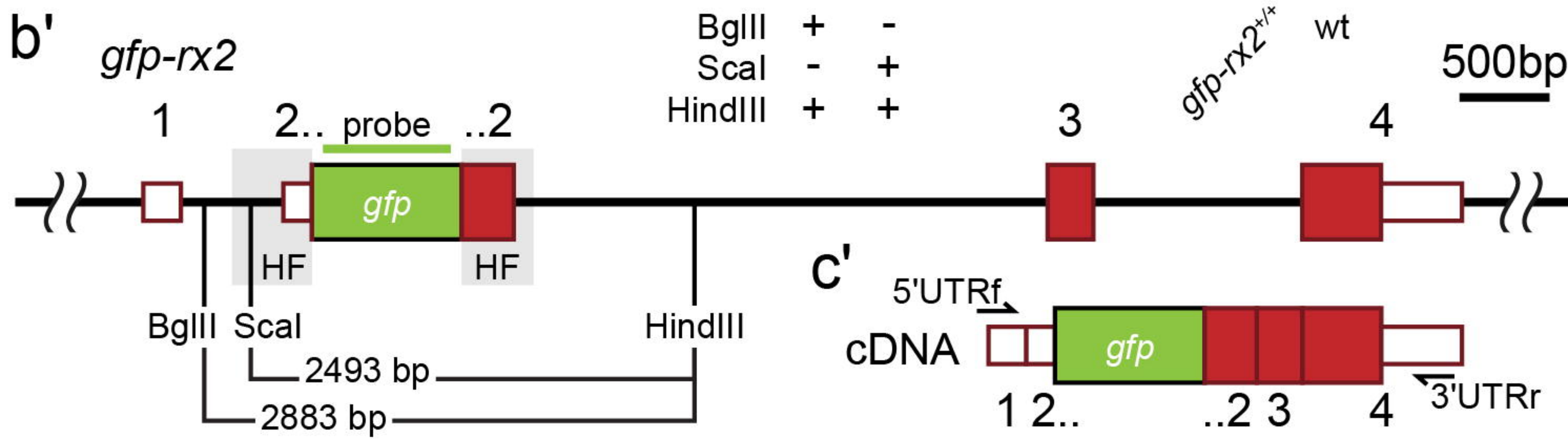

3
1
0.5
kb
HDR locus +single gfp
wt locus size
gfp
unmod Biotin wt
unmod Biotin wt
Biotin unmod
Lf/Lr + - + - + - + - + - + - + +
Gf/Gr - + - + - + - + - + - + - -

Figure 2

Gutierrez, Thumberger, Tavhelidse, et al.







	gfp-rx2+/+	wt
BglII	+	-
ScaI	-	+
HindIII	+	+