Efficient genome editing in zebrafish using a CRISPR-Cas system

Woong Y Hwang[1,7], Yanfang Fu[2,3,7], Deepak Reyon[2,3], Morgan L Maeder[2,4], Shengdar Q Tsai[2,3], Jeffry D Sander[2,3], Randall T Peterson[1,5,6], J-R Joanna Yeh[1,5] & J Keith Joung[2–4]

In bacteria, foreign nucleic acids are silenced by clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated (Cas) systems. Bacterial type II CRISPR systems have been adapted to create guide RNAs that direct site-specific DNA cleavage by the Cas9 endonuclease in cultured cells. Here we show that the CRISPR-Cas system functions *in vivo* to induce targeted genetic modifications in zebrafish embryos with efficiencies similar to those obtained using zinc finger nucleases and transcription activator–like effector nucleases.

Bacteria and archaea have evolved an adaptive defense mechanism that uses CRISPR, together with Cas proteins, to provide themselves with acquired resistance to invading viruses and plasmids[1–3]. The type II CRISPR-Cas system relies on uptake of foreign DNA fragments into CRISPR loci[4] and subsequent transcription and processing of these RNA transcripts into short CRISPR RNAs (crRNAs)[5], which in turn anneal to a trans-activating crRNA (tracrRNA) and direct sequence-specific silencing of foreign nucleic acids by Cas proteins[5–7] (**Fig. 1a**). Recent *in vitro* work showed that a synthetic single guide RNA (sgRNA) consisting of a fusion of crRNA and tracrRNA can direct Cas9 endonuclease–mediated cleavage of target DNA[6] (**Fig. 1b**). In addition, Cas9 can function with either crRNA and tracrRNA together or sgRNA to efficiently induce targeted alterations in cultured human cells[8,9]. However, whether CRISPR-Cas–based RNA-guided endonucleases (RGENs) can be used like zinc finger nucleases (ZFNs)[10] or transcription activator–like effector nucleases (TALENs)[11] for genome editing in whole organisms is not known.

Here we show that customizable sgRNAs can direct Cas9 endonuclease–mediated alteration of endogenous genes in zebrafish embryos. First, we constructed expression vectors that enable T7 RNA polymerase–mediated production of a capped, polyadenylated mRNA, encoding the monomeric Cas9 endonuclease, and of a customizable sgRNA bearing 20 nucleotides (nts) of sequence complementary to a target site (**Fig. 1c**). The sequence of our sgRNA, like that of another recently described[8], differs from a sgRNA used *in vitro*[6] in that our sgRNA contains additional tracrRNA-derived sequences at its 3′ end (**Fig. 1b,c** and **Supplementary Table 1**). For initial experiments, we designed a sgRNA with a targeting region complementary to a sequence in the *fh* gene (site no. 1) (**Supplementary Table 2**).

To determine the optimal quantity of each RNA species to use for genome editing, we microinjected varying amounts of *fh*-targeted sgRNA and Cas9-encoding mRNA into one-cell-stage zebrafish embryos; we then assessed the frequency of altered alleles in single embryos using a T7 endonuclease I (T7EI) assay (**Supplementary Methods**). We observed targeted insertion/deletion mutations (indels) at all concentrations of RNAs examined and in nearly all embryos tested (**Supplementary Table 3**). However, the highest mean frequency of mutations was obtained with a solution containing 12.5 ng/µl sgRNA and 300 ng/µl Cas9-encoding mRNA (**Supplementary Table 3**), so we used these concentrations for all subsequent experiments. Sequencing of mutated *fh* alleles revealed indels that begin within or encompass the 5′ end of the DNA sequence complementary to the sgRNA (**Supplementary Fig. 1**). This pattern of mutations is consistent with the expected induction of a Cas9-induced double-stranded break at this position[6] within the genomic *fh* target site followed by error-prone nonhomologous end joining–mediated repair.

To test the robustness of the sgRNA:Cas9 system in zebrafish, we constructed ten additional sgRNAs, one targeted to another sequence in the *fh* gene (site no. 2) and the remaining nine targeted to sites in nine other endogenous genes (**Supplementary Table 2**). We detected high frequencies (means of 24.1–59.4%) of targeted indels at eight of these ten sites in all individual embryos tested (**Table 1**). Frequencies of mutagenesis did not appear to depend upon which DNA strand (sense or anti-sense) was targeted by the sgRNA. Notably, we detected highly efficiencient mutagenesis at sites in the *gsk3b* and *drd3* genes, which we were not able to alter using TALENs (**Supplementary Table 4**). Mutation rates in the other six successfully targeted sites were similar to those seen in these same genes using ZFNs and/or TALENs (**Table 1** and **Supplementary Table 4**). Sequencing of all eight of these target sites confirmed the efficient introduction of targeted indels at the expected genomic locations (**Fig. 1d**, **Supplementary Table 5** and **Supplementary Fig. 2**). The lengths of indel mutations induced by these sgRNA:Cas9 constructs are similar to those of mutations induced by ZFNs and TALENs previously made by our groups (**Supplementary Fig. 3a**). Furthermore, the nature of the mutations (that is, the relative abundance of insertions and deletions) also appears to be similar among all three systems (**Supplementary Fig. 3b**). Thus, the sgRNA:Cas9 system successfully targeted >80% of the sites we tested in zebrafish.
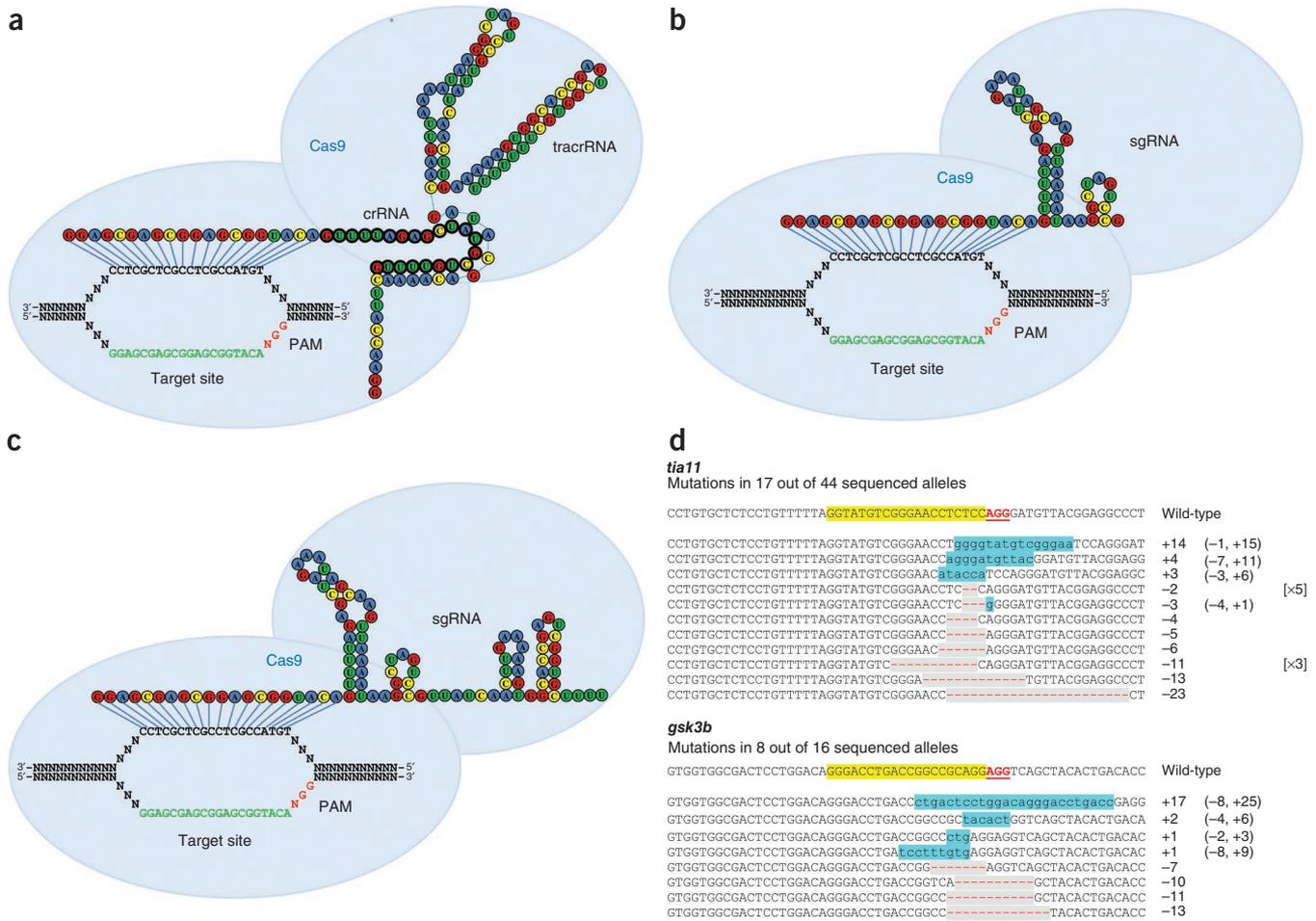
Figure 1 Schematic illustrating naturally occurring and engineered CRISPR-Cas systems. (**a**) Naturally occurring dual-RNA:Cas9 endonuclease. crRNA interacts with the complementary strand of the DNA target site harboring an adjacent PAM sequence (green and red text, respectively), tracrRNA base pairs with the crRNA, and the overall complex is recognized and cleaved by Cas9 nuclease (light blue shape). Folding of the crRNA and tracrRNA molecules is depicted as predicted by Mfold[20] and the association of the crRNA to the tracrRNA is depicted partially based on the model previously proposed[6]. (**b**) Engineered sgRNA:Cas9 system previously used *in vitro*. sgRNA composed of portions of the crRNA and tracrRNA from **a** is illustrated interacting with the DNA target site. Folding of sgRNA is as predicted by Mfold[20]. (**c**) Modified engineered sgRNA:Cas9 system used *in vivo* in this study. Components are illustrated the same way as in **b**, except the sgRNA contains additional sequence from the 3′ end of the tracrRNA. Folding of sgRNA is as predicted by Mfold[20]. The sgRNA depicted is essentially identical to that previously described[8]. (**d**) Targeted indel mutations induced by engineered sgRNA:Cas9 at the *tia1l* and *gsk3b* genes. For each gene, the wild-type sequence is shown at the top with the target sites highlighted in yellow and the PAM sequence highlighted as red underlined text. Deletions are shown as red dashes highlighted in gray and insertions as lower case letters highlighted in blue. The net change in length caused by each indel mutation is to the right of each sequence (+, insertion; −, deletion). Note that some alterations have both insertions and deletions of sequence and in these instances the alterations are enumerated in the parentheses. The number of times each mutant allele was isolated is shown in brackets. For each gene locus, DNA fragments were amplified by PCR from genomic DNA of ten pooled modified embryos and then sequenced as described in **Supplementary Methods**. (See also **Supplementary Fig. 2**.)

With the CRISPR-Cas system used in this study, only one customized sgRNA is required to target a specific sequence and the same Cas enzyme is suitable for all sequences. In contrast, TALENs or ZFNs require the design and assembly of two nucleases for each target site. In addition, sgRNAs are encoded by short ~100-bp sequences and are therefore much simpler and easier to construct than TALENs or ZFNs. The short length of sgRNA sequences also avoids complications associated with longer (typically 3 kb or more) and highly repetitive TALEN-encoding vectors (e.g., delivery using viral vectors[12], challenges with DNA sequencing, potential for recombination). It will be of interest to determine why our sgRNA:Cas9 reagents efficiently mutagenized sites in endogenous zebrafish genes that we could not alter using TALENs.

The sgRNA:Cas9 system described here can in principle target any sequence of the form 5′-GG-N$_{18}$-NGG-3′. Such sites occur once in every 128 bp of random DNA sequence. Constraints on the range

of targetable sequences are due to sequence requirements imposed by the T7 promoter used to make sgRNAs (GG at the 5′ end of the transcript) (**Supplementary Fig. 5**) and by the requirement for a protospacer adjacent motif (PAM) sequence (NGG) in genomic DNA just 3′ to the target site[6]. Previous studies suggest that the T7 promoter requirement for a pair of guanines at the 5′ end of the transcript could be relaxed to allow for an adenine at either position[13]. Loosening this constraint would enable targeting of sequences of the form 5′-(G/A)(G/A)-N$_{18}$-NGG-3′, which occur once in every 32 bp of random DNA sequence. In addition, it is also possible that mismatches between the two sgRNA bases nearest to the 5′ end and the target site might be tolerated; although as this may enable targeting of additional sites, it may have implications for off-target effects. Future studies should perform larger-scale tests of the targeting range of the sgRNA:Cas9 system for both matched and mismatched target sites.

Table 1 Mutation frequencies induced by customized sgRNA:Cas9 nucleases at ten endogenous gene target sites in the zebrafish genome

Gene	Indel mutation frequency										
	Embryo no. 1	Embryo no. 2	Embryo no. 3	Embryo no. 4	Embryo no. 5	Embryo no. 6	Embryo no. 7	Embryo no. 8	Embryo no. 9	Embryo no. 10	Mean ± s.e.m.
fh1 (site no. 2)	62.05%	52.16%	50.82%	60.18%	64.73%	66.13%	61.18%	55.45%	64.96%	55.95%	59.4 ± 1.7%
apoea	6.28%	15.91%	2.70%	3.72%	40.78%	55.62%	6.56%	59.61%	11.67%	37.87%	24.1 ± 7.0%
gria3a	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
th1	15.08%	29.26%	51.11%	41.21%	46.37%	45.85%	53.43%	19.99%	15.92%	37.53%	35.6 ± 4.6%
rgs4	53.45%	42.14%	50.85%	44.44%	48.67%	26.43%	27.35%	15.12%	19.13%	30.13%	35.8 ± 4.4%
tia1l	59.64%	59.88%	63.12%	57.33%	67.91%	58.02%	61.54%	56.39%	14.36%	72.18%	57.0 ± 5.0%
tph1a	7.32%	37.81%	28.59%	49.83%	40.92%	41.24%	37.49%	44.56%	41.84%	30.75%	36.0 ± 3.7%
slc6a3	0.00%	N/A	N/A	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
gsk3b	39.02%	4.94%	55.62%	4.31%	44.09%	19.12%	24.31%	3.07%	43.11%	33.31%	27.1 ± 6.0%
drd3	32.46%	20.34%	13.80%	34.20%	44.24%	33.13%	20.14%	23.17%	30.75%	31.35%	28.4 ± 2.8%

For each target site, up to ten individual embryos were assessed for indel mutation frequency using the PCR-based T7EI assay (**Supplementary Methods**). Mean frequencies of mutation for the ten embryos assayed for each target site are also shown with s.e.m. N/A, PCR reaction failed.

Data from such experiments may also yield insights into why some sgRNAs fail to mediate efficient sequence alterations and whether such failures can be predicted in advance.

The modified CRISPR-Cas system described here may be adapted to modify the genome of any organism into which RNA can be introduced. The plasmids expressing short ~100-nt sgRNAs with customized targeting regions can be easily and rapidly assembled simply by ligating pairs of short annealed oligonucleotides into our T7 promoter–based sgRNA vector (**Supplementary Methods** and **Supplementary Figure 5**). This process is considerably simpler than other methods for assembling TALEN- or ZFN-encoding plasmids[11] and therefore should be readily amenable to automation and high-throughput use. We have updated our web-based ZiFiT Targeter program[14] to enable identification of sgRNA:Cas9–targetable sites and to generate the sequences of oligonucleotides required to construct customized sgRNAs (http://zifit.partners.org/). All plasmids described in this report are available through the nonprofit reagent distribution service Addgene (http://www.addgene.org/crispr/jounglab). Any updates to reagents, protocols and software will be made available on the website (http://www.crispr-cas.org/).

Previous studies showed efficient germline transmission of all ZFN- and TALEN-engineered somatic mutations present at rates of 2% or greater in zebrafish embryos that develop normally[15–17]. Because all of the active sgRNA:Cas9 endonuclease combinations described here induced somatic mutation rates well above 10%, and these mutations were detected in normally developing embryos, we expect that germline transmission of sgRNA:Cas9-induced mutations will be as efficient as those induced by ZFNs or TALENs. Based on the rates of mutagenesis we observed, we calculate that the mean frequency of cells bearing bi-allelic alterations in these embryos would range from ~6% to ~36% (**Supplementary Discussion**). Thus, a proportion of the cells in many founders is likely to experience complete loss of target gene function. Although these rates might allow screening for certain phenotypes in the founders themselves, germline transmission will still be desirable in most instances to create nonmosaic knockout animals.

Additional work is needed to determine the off-target effects of the sgRNA:Cas9 system used here. Previous work *in vitro* with purified components and in cultured human cells has suggested that the 3′ end of the sgRNA target recognition sequence may be the most critical with regard to specificity[6,9], but whether this will also be true *in vivo* remains to be determined. The toxicity induced by sgRNA:Cas9–encoding mRNA in zebrafish (as judged by the numbers of deformed and dead embryos; **Supplementary Fig. 4**) appeared to vary among the different sgRNAs but did not correlate with their abilities to induce indels at the intended target sites, a phenomenon we

have also observed with ZFNs and TALENs[15,17–19]. The frequencies of deformed or dead embryos are similar to what we have observed in previous experiments using ZFNs[15,19] and TALENs[17,18].

This demonstration that a customized CRISPR-Cas system can be used to efficiently induce site-specific modifications in zebrafish should encourage wider use of this robust and easy-to-use technology, potentially in a variety of other organisms.

Note: Supplementary information is available in the online version of the paper.

AUTHOR CONTRIBUTIONS
W.Y.H., Y.F., M.L.M., D.R., S.Q.T., J.D.S., R.T.P., J.-R.J.Y. and J.K.J. designed experiments. W.Y.H., Y.F., M.L.M., D.R., S.Q.T. and J.D.S. performed experiments. W.Y.H., Y.F., M.L.M., D.R., S.Q.T., J.D.S., R.T.P., J.-R.J.Y. and J.K.J. wrote the paper.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Published online at http://www.nature.com/doifinder/10.1038/nbt.2501.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. *Nature* **482**, 331–338 (2012).
2. Horvath, P. & Barrangou, R. *Science* **327**, 167–170 (2010).
3. Terns, M.P. & Terns, R.M. *Curr. Opin. Microbiol.* **14**, 321–327 (2011).
4. Barrangou, R. *et al. Science* **315**, 1709–1712 (2007).
5. Brouns, S.J. *et al. Science* **321**, 960–964 (2008).
6. Jinek, M. *et al. Science* **337**, 816–821 (2012).
7. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
8. Mali, P. *et al. Science* doi:10.1126/science.1232033 (3 January 2013).
9. Cong, L. *et al. Science* doi:10.1126/science.1231143 (3 January 2013).
10. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. *Nat. Rev. Genet.* **11**, 636–646 (2010).
11. Joung, J.K. & Sander, J.D. *Nat. Rev. Mol. Cell Biol.* **14**, 49–55 (2012).
12. Holkers, M. *et al. Nucleic Acids Res.* doi:10.1093/nar/gks1446 (28 December 2012).
13. Schneider, T.D. & Stormo, G.D. *Nucleic Acids Res.* **17**, 659–674 (1989).
14. Sander, J.D. *et al. Nucleic Acids Res.* **38**, W462–W468 (2010).
15. Foley, J.E. *et al. PLoS ONE* **4**, e4348 (2009).
16. Huang, P. *et al. Nat. Biotechnol.* **29**, 699–700 (2011).
17. Cade, L. *et al. Nucleic Acids Res.* **40**, 8001–8010 (2012).
18. Sander, J.D. *et al. Nat. Biotechnol.* **29**, 697–698 (2011).
19. Sander, J.D. *et al. Nat. Methods* **8**, 67–69 (2011).
20. Zuker, M. *Nucleic Acids Res.* **31**, 3406–3415 (2003).