5′ Modifications Improve Potency and Efficacy of DNA Donors for Precision Genome Editing

Krishna S. Ghanta[†,1], Gregoriy A. Dokshin[†,1], Aamir Mir[†,1], Pranathi Meda Krishnamurthy[1], Hassan Gneid[1,2], Alireza Edraki[1], Jonathan K. Watts[*,1,3], Erik J. Sontheimer[*,1,4], Craig C. Mello[*,1,4,5]

Authors and Affiliations

1. RNA Therapeutics Institute, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605-2324, USA.

2. Chemistry, University of Southampton, Southampton, SO17 1BJ, UK

3. Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA.

4. Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA.

5. Howard Hughes Medical Institute.

† Contributed equally

*Correspondence to: jonathan.watts@umassmed.edu; erik.sontheimer@umassmed.edu; craig.mello@umassmed.edu

Nuclease-directed genome editing is a powerful tool for investigating physiology and has great promise as a therapeutic approach that directly addresses the underlying genetic basis of disease. In its most precise form, genome editing can use cellular homology-directed repair (HDR) pathways to insert information from an exogenously supplied DNA repair template (donor) directly into a targeted genomic location. Unfortunately, particularly for long insertions, toxicity and delivery considerations associated with repair template DNA can limit the number of donor molecules available to the HDR machinery, thus limiting HDR efficacy. Here, we explore modifications to both double-stranded and single-stranded repair template DNAs and describe simple 5′ end modifications that consistently and dramatically increase donor potency and HDR efficacy across cell types and species.

In the nematode worm *C. elegans*, efficient genome editing can be achieved by direct injection of editing enzyme guide-RNA ribonucleoprotein (RNP) complexes into the syncytial ovary [1]. Such injections afford simultaneous access of the editing machinery to hundreds of meiotic germ nuclei within a common cytoplasm (**Supplementary Fig. 1a**). In the worm germline, high rates of HDR are readily achieved using short (under ~200 nucleotide [nt]), single-stranded oligodeoxynucleotide (ssODN) donor templates that permit insertions of up to ~150 nt in length [2, 3, 4]. However, HDR is less efficient by 1–2 orders of magnitude when longer, double-stranded DNA (dsDNA) templates are used as donors [4].

Longer repair templates are likely at a disadvantage for multiple reasons. First, toxicity associated with high concentrations of DNA limits the safe injectable amount of a ~1kb dsDNA donor to roughly ~10-fold fewer molecules than is commonly used for a 200 nt ssODN donor [2-5]. Second, long dsDNA donor molecules may not readily transit across the nuclear envelope into the post-mitotic germ-nuclei, further reducing the effective concentration at the site of repair. We

hypothesized that the disparity in availability of ssODN and dsDNA donor molecules inside germ nuclei could account for the differences in observed HDR efficiencies. To increase potency of long dsDNA donors we set out to attach an SV40 peptide containing the core nuclear localization signal (NLS) to the donor molecule, reasoning that the modification might promote nuclear uptake and retention. Previous studies using mammalian cell cultures demonstrated that the addition of an NLS enhances nuclear uptake of plasmid DNA following transfection [6]. To attach an NLS to a long donor DNA, we first conjugated 15-nucleotide 2′-O-methyl (2′OMe) RNA adapters via a tri- or tetraethylene glycol (TEG) linkage to the 5′ ends of two target-locus specific synthetic ~20-nucleotide DNA oligonucleotides. The DNA sequences in these molecules serve as PCR primers to amplify the donor from plasmid containing the homology arms and green fluorescent protein (GFP) sequence for in-frame insertion into the target gene of interest (**Fig. 1a**). In addition, we synthesized an NLS peptide linked to a peptide nucleic acid (PNA) complementary to the 15nt 2′OMe-RNA adapter. Attachment of the NLS to the donor was then achieved by simply annealing the PNA::NLS molecules to the 2′OMe-RNA adapters on the ends of the PCR product.

To test these modified donors, we first attempted to insert GFP into the *csr-1* locus (see Methods for details). We employed a co-CRISPR assay [7, 8], to measure HDR efficacy. For this assay we chose a CRISPR RNP designed to generate indels in the easily scored *dpy-10* locus. HDR efficacy was scored as a fraction of F1 *dpy-10* mutant animals that properly express GFP::CSR-1, which forms bright peri-nuclear foci in germ cells (**Supplementary Fig. 1b**). To increase assay sensitivity, we chose an initial donor DNA concentration of 10 ng/µl (~35-fold less than previously recommended) [2]. With unmodified dsDNA donors, this donor concentration yielded only 0.92% GFP insertions (**Fig. 1b**). Strikingly, however, under otherwise identical injection conditions, this same dsDNA donor modified at both 5′ ends with 2′OMe-RNA::TEG annealed to PNA::NLS

yielded 29.5% GFP positive animals (~8 GFP+ F1 progeny per injected animal) – an increase of >30-fold (**Fig. 1b**). Interestingly, donors prepared with 2′OMe-RNA::TEG end-modification, but without annealing PNA::NLS, also substantially improved HDR efficacy to 15.7% (~4 GFP+ F1 progeny per injected animal). To our surprise, annealing a PNA without an NLS resulted in 35.1% GFP insertion efficacy (~9 GFP+ F1 progeny per injected animal), indicating that the NLS was not essential to promote HDR. Simply co-injecting PNA with plain dsDNA donor failed to boost HDR (2.2%), indicating that PNA-based HDR improvement depends on the presence of the 2′OMe-RNA::TEG moiety **(Fig. 1b**). To test the generality of these findings we targeted GFP insertions into an independent locus, *glh-1*, and found comparable increases in efficiency and dependence on 2′OMe-RNA::TEG, as well as PNA (**Fig. 1c**).

We next tested whether end modifications would increase dsDNA donor potency in mammalian cell culture systems. We first used HEK293T cells carrying a modified version of the "traffic light" reporter (TLR) [9] (Mir,A. *et al.*, manuscript in preparation). Briefly, in this system Cas9 targets a "broken" GFP followed by a frameshifted mCherry reporter. Imprecise repair restores the reading frame in a subset of indels, resulting in mCherry (red) fluorescence. Conversely, precisely templated repair of the same lesion results in GFP (green) fluorescence. Using flow cytometry, the percentage of cells expressing either GFP (precise) or mCherry (imprecise) among the total number of cells can be easily quantified. To perform the TLR assay we electroporated reporter cells with Cas9- and single guide RNA (sgRNA)- encoding plasmids, along with either an unmodified dsDNA donor, a 2′OMe-RNA::TEG donor, or the end-modified donor annealed to PNA::NLS. Strikingly, across a range of donor amounts we observed a consistent and significant increase in HDR efficacy with the end-modified donors (**Fig. 2a**). The efficacy peaked at 1.2 pmol of 2′OMe-RNA::TEG modified donor that yielded 51.8% GFP-

positive cells, compared to 22% GFP+ cells obtained with the same amount of unmodified donor (**Fig. 2a**). This gain of GFP+ cells was accompanied by a corresponding reduction in mCherry+ cells (**Fig. 2b**). As expected, with reduced donor amount the HDR efficacy declined for all donor types, and the number of GFP+ cells also declined at donor amounts over 2 pmol (**Fig. 2a, b**). The maximum HDR efficacy for unmodified donors (25% GFP+ cells) was achieved at 1.6 pmol. Notably, the 2′OMe-RNA::TEG modified donors matched this efficacy of 25% at less than 0.4 pmol, illustrating that the modified donor is approximately 4-fold more potent (**Fig. 2a**). Although less dramatic than the 30-fold efficacy increase observed in worms (which appear to have a much lower basal HDR efficacy), we consistently observed more than 2-fold increases in HDR efficacy using this mammalian cell culture system. Interestingly, however, the addition of PNA::NLS or PNA alone to the 2′OMe-RNA::TEG end-modified donor provided no additional increase in HDR efficacy in cell culture (**Fig. 2a, Supplementary Fig 2**).

We next used the TLR assay to define the features of the 2′OMe-RNA::TEG adapter that promote HDR at the optimal 1.2 pmol donor amount (**Fig. 2c, d**). Interestingly, we found that donors modified with either the 2′OMe-RNA alone or with TEG alone consistently boosted HDR efficiencies (**Fig. 2c**). Moreover, even donors with TEG modification at the 5′ end of only one of the two strands provided a significant boost in HDR efficacy (**Supplementary Fig. 3a).** Finally, different lengths of PEG (4, 6, 9 or 12 ethylene glycol repeats) showed similar efficacy (**Supplementary Fig. 4a**). In all cases we observed a corresponding decline in mCherry+, imprecisely edited, cells (**Fig. 2d, Supplementary Fig. 3b, 4b**).

To explore the utility of end-modified donors for repair at other genomic locations, we generated donors and guides to integrate full-length *eGFP* at the endogenous *GAPDH* and *TOMM20* loci in HEK293T cells. The *GAPDH* donor was designed to integrate IRES-eGFP in the

3′-UTR of the *GAPDH* locus, whereas the *TOMM20* donor was designed to tag the C-terminus of the mitochondrial protein TOM20 [10, 11]. By measuring the fraction of cells expressing eGFP by flow cytometry, we found that the TEG, 2′OMe-RNA or 2′OMe-RNA::TEG modifications consistently increased the fraction of eGFP cells (by up to 4-fold) when compared to unmodified dsDNA donor (**Fig. 2e, f**). Again, in these loci, as in the TLR assay (**Fig. 2c**), we noted that the presence of TEG was necessary for maximal HDR and that TEG alone performed better than 2′OMe-RNA alone. As expected, in all cases, precise insertion of eGFP was Cas9-dependent.

We next tested modified donors in cell types that are typically more resistant to HDR than are the HEK293T cells used in the above studies. To do this we generated 2′OMe-RNA::TEG donors and Cas9 RNPs designed to target the insertion of eGFP at the *TOMM20* locus in hTERT-immortalized human foreskin fibroblasts (HFFs) and at the *Gapdh* locus in Chinese hamster ovary (CHO) cells. Although the overall efficiencies of HDR were lower than those observed in HEK293T cells, the use of modified donors led to a 2.3-fold and 6-fold increase in HDR in HFF and CHO cells respectively (**Fig. 2g, h**).

The experiments described thus far employed dsDNA donors that are easy to generate by PCR; however, single stranded DNA (ssDNA) donors have become widely used in many HDR editing protocols. We therefore tested end modifications to ~800 nt long ssDNA donors generated by transcription and reverse transcription of PCR products. As observed for dsDNA donors, we found that the addition of a 2′OMe-RNA::TEG tail to a long ssDNA donor elicited a consistent and significant boost in HDR efficacy over unmodified ssDNA donors (with corresponding reductions in imprecise repair) across a range of concentrations in the TLR assay (**Fig. 2i, Supplementary Fig. 5a**). Efficacy of modified ssDNA donor was 2.3 times higher than unmodified donor at 1 pmol and peaked to 22.5% at 6 pmol (versus 15.9% for unmodified).

Furthermore, the maximal efficacy of the plain donor (16% at 8 pmol) could be achieved at less than 2 pmol of 2′OMe-RNA::TEG donor, again highlighting the improved potency of the modified donors.

The highest reported yields of HDR in both cultured mammalian cells and *C. elegans* have been achieved using short (≤ 200 nt), synthetic ssODN donors delivered at high concentrations [2-4, 12]. To test 5′ modified ssODNs for HDR efficacy we used the sensitive GFP-to-BFP conversion assay in K562 cells (**Fig. 2j**). At first, we were surprised to find that 5′ modified ssODNs gave no increase in HDR efficacy, ~30%, when tested at 10pmol, an amount at which unmodified ssODNs gave a peak efficacy of >45% HDR (**Fig. 2j**). In order to fully investigate the relative potency of modified ssODNs we decided to explore a broader range of concentrations. Our studies revealed that both modified and unmodified donors exhibited similar maximal levels of BFP positive cells but differed greatly in their donor amount dependence. Strikingly, the end modified ssODNs achieved maximal efficacy at amounts 10-fold lower than unmodified donor (**Fig.2j**). Interestingly TEG-modified ssODN donors showed dose-limiting toxicity that scaled with their increased HDR potency (**Fig. 2j**) – suggesting that the mechanisms of toxicity and improved HDR efficacy may be related (see further discussion below). The increased potency of HDR was observed with concurrent reductions in imprecise editing as measured by the frequency of GFP(-) and BFP(-) cells (**Supplementary Fig. 5b**).

To explore how 3′ modifications affect HDR efficacy we added either internal phosphorothioate (PS) or terminal 2′OMe-RNA and/or TEG modifications to the 3′ ends of ssODNs. Interestingly, 2′OMe-RNA and/or TEG modifications on the 3′ terminal hydroxyl group – modifications that would prevent these molecules from priming DNA synthesis or undergoing ligation – led to a significant decrease in HDR efficacy relative to 5′ modification alone (**Fig. 2k,**

Supplementary Fig. **6**). In contrast, modification of the 3′ internal phosphate linkages as PS modifications, which do not block the terminal hydroxyl group, did not impede HDR. The differences between the effects of terminal (2′OMe-RNA and/or TEG) and non-terminal (PS) 3′ modifications on HDR efficacy of the 5′-terminally modified donors suggest that the mechanism of HDR improvement requires the availability of a 3′-OH group. Unlike the modifications conjugated to the 5′ terminus, at 0.5 pmol of donor, internal PS modifications at the 5′-end did not improve HDR efficacy compared to the unmodified donor (**Supplementary Fig. 6**).

Here we have employed the simple approach of chemically modifying the repair template to promote HDR. We initially hypothesized that exploiting active nuclear import would increase donor potency. In *C. elegans,* the addition of complementary PNA and PNA::NLS to 2′OMe-RNA::TEG donors led to improved HDR efficacy. In contrast, these adducts did not improve efficacy in mammalian cells over use of the 2′OMe-RNA::TEG donor itself. One potential explanation for this interesting difference is that the site of injection in *C. elegans* is a post-mitotic ovary, as compared with rapidly dividing mammalian cells. Thus, the putative role of PNA::NLS in getting donor past the nuclear envelope should be further explored in other systems involving post-mitotic or slowly-dividing cell types. The finding that annealing PNA and PNA-peptide fusions to the 2′OMe-RNA stimulates HDR in *C. elegans* raises the question of what other nucleic acid or nucleic acid-like adapters might also stimulate HDR. We have yet to scratch the surface of potential chemistries to explore with this modular system.

We were surprised to find that the 5′ addition of either TEG or 2′OMe-RNA alone was sufficient to dramatically improve HDR potency. Oligonucleotides bearing these simple, easy to synthesize chemistries have already proven to be safe and well tolerated in other clinical applications [13]. These modifications increased the potency of dsDNA, ssDNA and ssODN donors,

allowing efficient editing at significantly lower amounts. We find it intriguing that long DNA donors exhibited both increased potency and maximal efficacy when modified, while short ssODNs exhibited increased potency over a broader range of donor amounts without an increase in the maximal levels of templated repair. This difference requires further study but could be explained if ssODN and longer DNA donors experience different dose-limiting barriers. Perhaps the number of molecules of DNA (or DNA free ends inside cells) represents a dose limiting barrier for small ssODNs that longer dsDNA molecules never encounter due to other limitations. Consistent with this idea, unmodified long dsDNA donors begin to plateau in efficacy at nearly 4-fold more mass, but ~10-fold lower molar amounts than ssODNs. When end modified, both types of donor exhibit similar maximal efficacy in the 1-2pmol range. It is possible that 5′ end modifications boost HDR by preventing end joining reactions that would otherwise decrease the available number of donor molecules in cells. However, this mechanism would also increase any toxicity associated with DNA free ends. For long DNA where free-end related toxicity may not be limiting, both potency and maximal efficacy increase. In contrast, for shorter ssODNs, where free ends may be driving the dose limiting toxicity, potency is increased but not maximal efficacy. Nevertheless, this dramatic potency increase suggests that terminally modified ssODNs may be particularly useful under conditions where donor concentration is limiting (such as delivery *in vivo*).

Potential benefits of end modifications may include prolonged intracellular donor molecule stability, improved nuclear localization, or an increased ability of the donor molecule to engage the HDR machinery. For example, as suggested above, the 5′ end modifications employed here may block end joining reactions that would otherwise reduce the available molar quantities of repair template molecules. However, it is unlikely that merely blocking end joining explains the

improved efficacy, as modifications to the 3′ ends, which should also block end-to-end ligation, were unhelpful. Taken together our findings suggest that a combination of blocking the 5′ end while leaving a 3′ OH end, perhaps to prime repair synthesis, are both important.

We initiated these investigations to explore why long DNA donors were so much less efficacious than ssODN donors in *C. elegans*. The chemical modifications we have made to the 5′ ends of donor molecules have very significantly closed that gap in efficacy. The findings that these same modifications also increase the potency of short ssODN donors raise the possibility that long modified dsDNA donors are approaching a level of efficacy where donor availability is no longer limiting for HDR. Thus, methods that block alternative imprecise repair pathways [14-18] or reduce damage responses may synergize with 5′ end modified donors to stimulate HDR to even higher levels.

228 **Acknowledgements**. This work was funded by: American Cancer Society Fellowship (G.A.D.),
229 Howard Hughes Medical Institute (C.C.M.), NIH PO1# HD078253 (C.C.M), and institutional
230 funds to J.K.W and E.J.S from University of Massachusetts Medical School. We are grateful to
231 Scot Wolfe for sharing the Cas9 expression plasmid and to Ahmet Ozturk for help with statistical
232 analyses.

233

Fig. 1. 5′ dsDNA donor modifications promote HDR in *C. elegans* germ cells

(**a**). Schematic representation of the end-modified dsDNA donor design. 2′OMe-RNA::TEG moieties are covalently attached to 5′ ends of each DNA strand and PNA::NLS is annealed to the RNA overhangs. HDR efficacy at the (**b**). *csr-1* and (**c**). *glh-1* loci using unmodified or 2′OMe-RNA::TEG-modified donors alone, or with addition of PNA or PNA::NLS, is plotted as a fraction of GFP+ *dpy-10* animals. Each point represents the average GFP+ fraction in a brood from a single injected P0 animal. Horizontal lines mark the mean. The number of GFP+ over total number of *dpy-10* edited animals scored for each condition is shown in parentheses. P-values were calculated using One-way ANOVA and in all cases end-modified donors were compared to the unmodified donor (Tukey's multiple comparisons test (****P < 0.0001; ***P < 0.001; **P < 0.01; ns, not significant).

Fig. 2. TEG and 2′OMe-RNA at 5′ ends of donors promote HDR in mammalian cells.

Editing efficacy plotted as percentage of (**a**) GFP+ (HDR) and (**b**) mCherry+ (NHEJ) HEK293T TLR cells at different amounts of unmodified, 2′OMe-RNA::TEG-modified, or modified and PNA::NLS-annealed dsDNA donors. Editing efficacy plotted as percentage of (**c**) GFP+ (precise) and (**d**) mCherry+ (imprecise) HEK293T TLR cells at 1.2 pmol of donor. Efficacy of eGFP integration at **e**. *TOMM20* and (**f**) *GAPDH* loci in HEK293T cells with and without Cas9, using modified donors and plotted as percentage of GFP+ cells. Efficacy of eGFP integration at the (**g**). *TOMM20* locus in HFF and (**h**) *Gapdh* locus in CHO cells using dsDNA donors with and without 2′OMe-RNA::TEG modifications at the 5′ ends. (**i**) Editing efficacy plotted as percentage of GFP+ (precise) HEK293T TLR cells at different amounts of unmodified and 2′OMe-RNA::TEG-modified long (800 nt) ssDNA donors. (**j**) Editing efficacy of GFP-to-BFP reporter conversion in

K562 cells using different amounts of unmodified and 2′OMe-RNA::TEG-modified 66 nt ssODN donors plotted as percentage of BFP+ cells (precise). (**k**). Editing efficacy of GFP-to-BFP conversion in K562 cells using 0.5 pmol of ssODN donors modified at the 5′ end alone, the 3′ end alone, or at both the 5′ and 3′ ends, with phosphorothioate (PS), TEG, 2′OMe-RNA, or 2′OMe-RNA::TEG, plotted as percentage of BFP+ cells (precise). Complete figure of panel **k** is shown, along with other modifications, in **Supplementary Fig**. **6**. All data points represent a mean of at least three independent replicates and all error bars represent standard deviation. P-values were calculated using either one-way or two-way ANOVA and in all cases end-modified donors were compared to the unmodified donor unless indicated otherwise (Tukey's multiple comparisons test; ****P < 0.0001; ***P < 0.001; **P < 0.01; *P< 0.05; ns- not significant).

References

1. Cho, S.W., Lee, J., Carroll, D., Kim, J.S. & Lee, J. Heritable gene knockout in Caenorhabditis elegans by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* **195**, 1177-1180 (2013).
2. Paix, A., Folkmann, A., Rasoloson, D. & Seydoux, G. High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201**, 47-54 (2015).
3. Prior, H., Jawad, A.K., MacConnachie, L. & Beg, A.A. Highly Efficient, Rapid and Co-CRISPR-Independent Genome Editing in Caenorhabditis elegans. *G3 (Bethesda)* **7**, 3693-3698 (2017).
4. Dokshin, G.A., Ghanta, K.S., Piscopo, K.M. & Mello, C.C. Robust genome editing with short single-stranded and long, partially single-stranded DNA donors in C. elegans. *bioRxiv* (2018).
5. Mello, C.C., Kramer, J.M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *EMBO J* **10**, 3959-3970 (1991).
6. Branden, L.J., Mohamed, A.J. & Smith, C.I. A peptide nucleic acid-nuclear localization signal fusion that mediates nuclear transport of DNA. *Nat Biotechnol* **17**, 784-787 (1999).
7. Kim, H. et al. A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. *Genetics* **197**, 1069-1080 (2014).
8. Arribere, J.A. et al. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics* **198**, 837-846 (2014).
9. Certo, M.T. et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods* **8**, 671-676 (2011).
10. He, X. et al. Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res* **44**, e85 (2016).
11. Roberts, B. et al. Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol Biol Cell* **28**, 2854-2874 (2017).
12. Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. & Corn, J.E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34**, 339-344 (2016).
13. Fowles, J., Banton, M., Klapacz, J. & Shen, H. A toxicological review of the ethylene glycol series: Commonalities and differences in toxicity and modes of action. *Toxicol Lett* **278**, 66-83 (2017).
14. Chu, V.T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-548 (2015).
15. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538-542 (2015).
16. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G.F. & Chin, L. Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep* **14**, 1555-1566 (2016).
17. Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).

18. Yang, D. et al. Enrichment of G2/M cell cycle phase in human pluripotent stem cells enhances HDR-mediated gene repair with customizable endonucleases. *Sci Rep* **6**, 21264 (2016).

Fig. 1

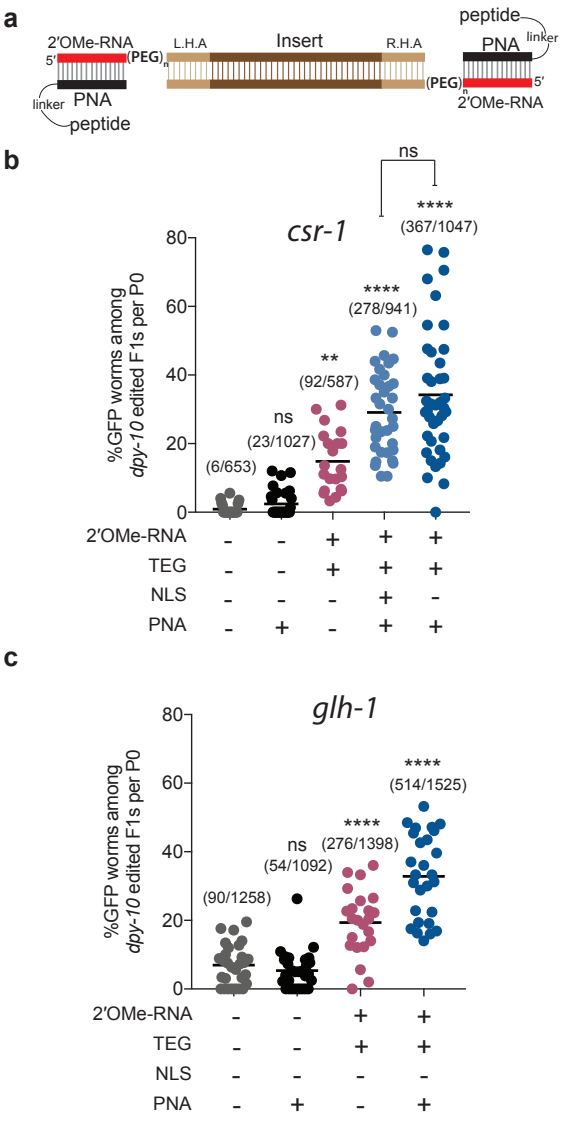
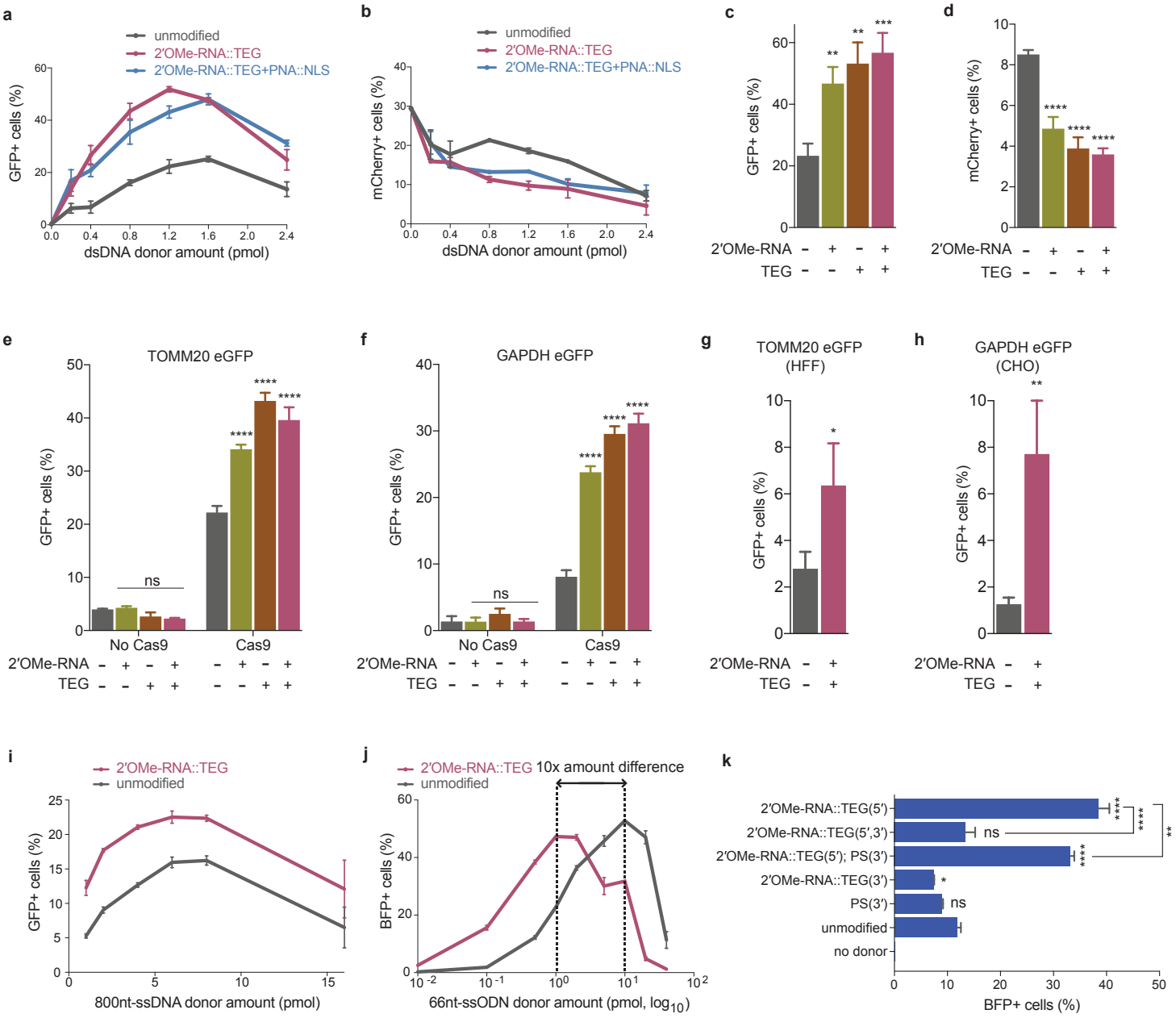


Fig. 2



Supplementary Materials for

5′ Modifications Improve Potency and Efficacy of DNA Donors for Precision Genome Editing

This PDF file includes:

Materials and Methods
Supplementary Figs. 1 to 7

Materials and Methods

Synthesis of PNA-NLS peptide. PNA oligomers were synthesized at 2μmol scale on Fmoc-PAL-PEG-PS solid support (Applied Biosystems) using an Expedite 8909 synthesizer. Fmoc/Bhoc-protected PNA monomers (Link Technologies) were dissolved to 0.2M in anhydrous *N*-methylpyrrolidinone. Amino acid monomers (Sigma Aldrich) and AEEA linker (Link Technologies) were dissolved to 0.2 M in anhydrous dimethylformamide. Coupling time was 8.5 min using HATU (Alfa Aesar) as activator; double coupling was performed on all PNA monomers and amino acids. PNAs were cleaved and deprotected by treating the resin with 400 μL of 19:1 TFA:*m*-Cresol for 90 min at room temperature. The resin was then removed with a PTFE centrifugal filter and PNAs were precipitated from cold diethyl ether and resuspended in deionized water. PNAs were purified by HPLC on a Waters XSelect CSH C18 5μm column at 60 °C, using gradients of acetonitrile in water containing 0.1% TFA, and were characterized on an Agilent 6530 Q-TOF LC/MS system with electrospray ionization. The PNA::NLS sequence used was GCGCTCGGCCCTTCC-[AEEA linker]-PKKKRK.

Synthesis of PEGylated oligos. PEG-modified oligonucleotides were synthesized using standard phosphoramidite methods on an ABI 394 synthesizer. Phosphoramidites were purchased from ChemGenes. Coupling times for 2′OMe-RNA and spacer phosphoramidites were extended to 5 min. Oligonucleotides were deprotected in concentrated aqueous ammonia at 55 °C for 16 h. Oligonucleotides were desalted using either Nap-10 (Sephadex) columns or Amicon ultrafiltration. All the PEG-modified oligonucleotides were characterized on an Agilent 6530 Q-TOF LC/MS system with electrospray ionization. The 2′-OMe RNA sequence appended to the 5′-end of donor DNAs was GGAAGGGCCGAGCGC.

dsDNA Donor generation. Donor template sequences with the homology arms and the desired insert for knock-in (eg: gfp), were generated by PCR. PCR products were cloned into ZeroBlunt TOPO vector (Invitrogen, #450245) and plasmids were purified using Macherey-Nagel midi-prep kits (cat# 740412.50). Using the purified plasmids as templates and PEGylated oligos as primers, donor sequences were PCR amplified with iProof (Bio-Rad,1725302, *C. elegans*) or Phusion polymerase (NEB, #M0530S, mammalian). Before use in *C. elegans* microinjections, the resulting PEGylated PCR products were excised from 0.8-1% TAE agarose gel and purified using spin-columns (Omega, #D2501-02). For use in mammalian cells, the PEGylated PCR products were purified using spin columns (Qiagen, # 28104). PCR conditions were optimized for each primer set with a gradient for the annealing temperature [1) 98°C for 1:00 min,2) 98°C for 15 sec, 3) 50°C to 64°C for 30 sec (choose optimal), 4) 72°C for 1:00 min (34 cycles), 5) 72°C for 5:00 min, 6) 4°C forever].

Single Strand DNA donor generation. Long single stranded DNA donors were prepared using the protocol described by Li et al [1]. Briefly, the donor template containing the T7 promoter was amplified using standard PCR and purified using SPRI magnetic beads (Core Genomics). T7 *in vitro* transcription was performed using the HiScribe T7 High Yield RNA Synthesis kit (NEB) and the RNA was purified using the SPRI magnetic beads. Finally, the ssDNA donor was synthesized by TGIRT™-III (InGex) based reverse transcription using the synthesized RNA as a template and a TEG-modified or unmodified DNA primer. We then performed base-treatment to remove RNA. The donor was again purified using SPRI beads.

***C. elegans* microinjection and HDR screening.** Microinjections were performed using Cas9-RNPs at final concentrations of: 0.25 µg/µl of SpyCas9 protein (IDT, # 1074181), 0.04 µg/µl of crRNA (against the target sequence), 0.016 µg/µl of crRNA (*dpy-10*) (IDT, Alt-R CRISPR-Cas9

crRNA) and 0.1µg/µl of tracrRNA (IDT, # 1072533) along with a series of modified or unmodified DNA donors (10µg/ul) [2]. To prepare a 20µl injection mix, Cas9, tracrRNA, crRNAs and nuclease free water were incubated at 37° C for 12 min; then, DNA donors were added to the RNPs. This injection mixture was spun down at 14000 RPM for 2 min and 17 µl of the mixture was carefully taken from the top and transferred to a fresh 1.5-ml tube. Prior to adding the donors to the injection mixture, we mixed them either with PNA-peptide ligands or water, heated the mixture to 95°C and cooled to 4°C using the thermal cycler (95°C-2:00 min; 85°C-10 sec, 75°C-10 sec, 65°C-10 sec, 55°C-1:00 min, 45°C-30 sec, 35°C-10 sec, 25°C- 10 sec, 4°C-forever). As controls, we performed microinjections with the DNA donors that lack chemical modifications. After the microinjections, injected animals (P0) were singly picked onto Normal Growth Media (NGM) plates and cultured at 20°-22°C for about 3.5 days. F1 animals exhibiting evidence of CRISPR induced lesions at a marker locus (*dpy-10*) [3, 4] were separated onto fresh NGM plates (5 animals/plate) and screened under a fluorescent dissection microscope for the presence or absence of GFP expression in the gonad. We then plotted the percentage of GFP positive animals among the total number of dumpys/rollers produced by each P0 animal. We have considered only those P0s that produced at least 15 F1 dumpy and/or roller animals to account for the injection quality.

***C. elegans* GFP tagging strategy.** In *C. elegans*, we aimed to insert the entire coding sequence of Green Fluorescent Protein (GFP) immediately after the ATG start codon of *csr*-1 and *glh-1* genes. For *csr-1*, we used a two-step CRISPR protocol, such that the same exact reagents could be used to fuse any gene to a large DNA sequence, without requiring synthesis of modified primers for every locus (such as GFP or mCherry ~900 bp). Briefly, we employed Cas9-RNP complexes and single stranded DNA (ssODN) donor oligo (as described in [2]to knock-in FLAG (x3)::Glycine(x3) linker::TEV tag. Starting strain that is homozygous for 3XFLAG::GlyGlyGly::TEV::CSR-1 allele

was used to knock-in gfp sequence between flag and glycine-linker. We used a crRNA (CTATAAAGACGATGACGATA NGG) with PAM site in the glycine-linker and donor DNA with arms homologous to 35 bp of 3xflag and 30 bp of 3xglycine-linker::tev flanking the gfp sequence. Once conditions were optimized with universal 2-step system, endogenous *glh-1* locus was targeted for gfp insertion directly using *glh-1* guide and 35bp homology arms.

Expression and purification of SpyCas9. The pMCSG7 vector containing the 6xHis-tagged 3xNLS SpyCas9 was a gift from Scot Wolfe at UMass Medical School. This construct was transformed into the Rosetta 2 DE3 strain of *E. coli* for protein production. Expression and purification of SpyCas9 was performed as described previously [5]. Briefly, cells were grown at 37°C to OD600 of 0.6, at which point 1 mM IPTG (Sigma) was added and the temperature was lowered to 18°C. Cells were grown overnight and harvested by centrifugation at 4,000 g. The protein was purified first by Ni^{2+} affinity chromatography, then by cation exchange and finally by size-exclusion chromatography.

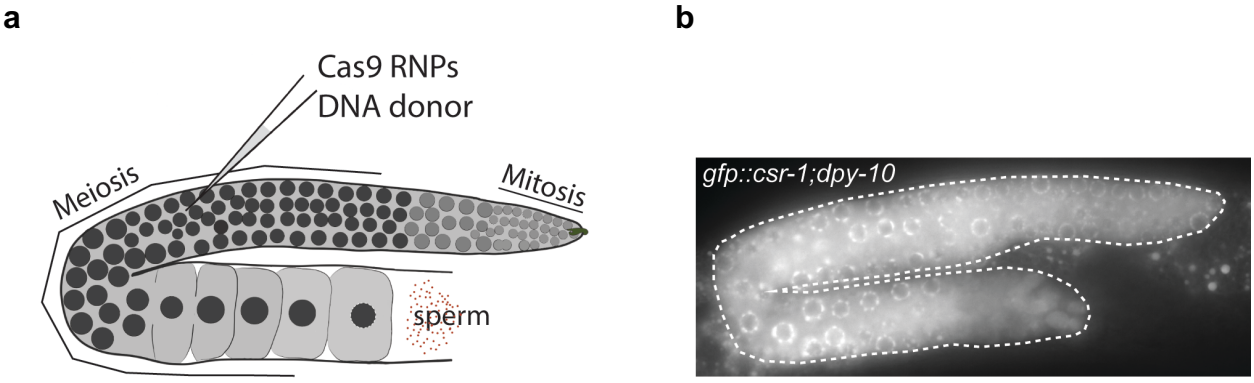
Cell culture and transfections. HEK293T cells were obtained from ATCC and were cultured in standard DMEM medium (Gibco, #11995) supplemented with 10% fetal bovine serum (FBS) (Sigma, #F0392). Human foreskin fibroblasts (HFF) were maintained in DMEM medium supplemented with 20% FBS. Chinese hamster ovary (CHO) cells (obtained from ATCC) were cultured in F-12K medium (Gibco 21127022) supplemented with 10% FBS, and K562 cells were cultured in IMDM medium (Gibco 12440053) supplemented with 10% FBS. Electroporations were performed using the Neon transfection system (ThermoFisher). SpyCas9 was delivered either as a plasmid or as protein. For plasmid delivery of Cas9 and sgRNA, appropriate amounts of plasmids were mixed in ~10 µl Neon buffer-R (ThermoFisher) followed by the addition of 100,000 cells. For RNP delivery of Cas9, 20 pmol of 3xNLS-SpyCas9 and 25 pmol of crRNA-tracrRNA

were mixed in 10 µl of buffer R. This mixture was incubated at room temperature for 30 minutes followed by the addition of 100,000 cells that were already resuspended in buffer R. This mixture was then electroporated using the 10 µl Neon tips. Electroporation parameters (pulse voltage, pulse width, number of pulses) were 1150 v, 20 ms, 2 pulses for HEK293T cells, 1650 v, 10 ms, 3 pulses for CHO cells, 1400 v, 30 ms, 1 pulse for HFF cells and 1600 v, 10 ms, 3 pulses for K562 cells. Electroporated cells were harvested for FACS analysis 48-72 hr post electroporation unless mentioned otherwise in results.

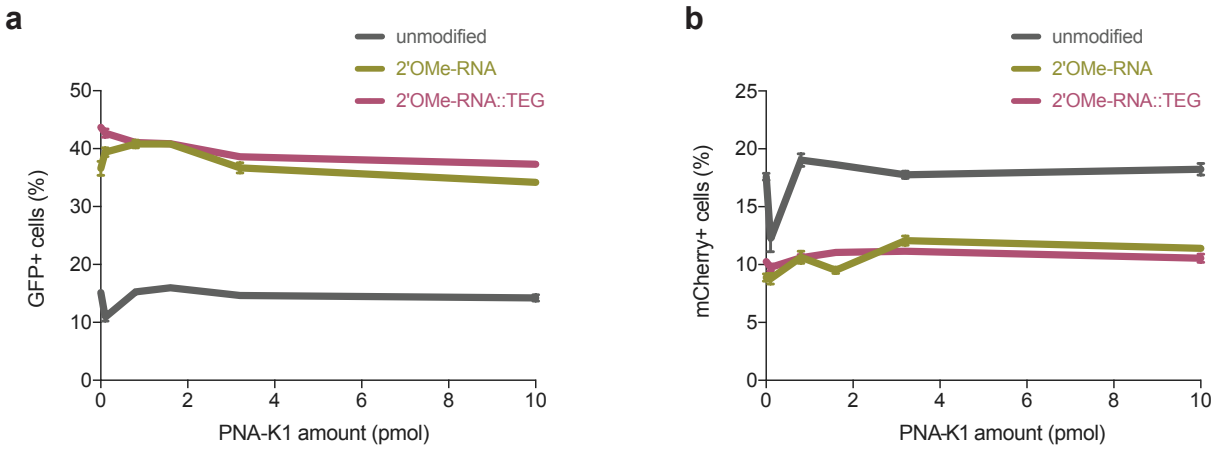
K562 GFP+ stable cell line generation. Lentiviral vector expressing EGFP was cloned using the Addgene plasmid #31482. The EGFP sequence was cloned downstream of the SFFV promoter using Gibson assembly. For lentivirus production, the lentiviral vector was co-transfected into HEK293T cells along with the packaging plasmids (Addgene 12260 & 12259) in 6-well plates using TransIT-LT1 transfection reagent (Mirus Bio) as recommended by the manufacturer. After 24 hours, the medium was aspirated from the transfected cells and replaced with fresh 1 ml of fresh DMEM media. The next day, the supernatant containing the virus from the transfected cells was collected and filtered through a 0.45 µm filter. 10 µl of the undiluted supernatant along with 2.5 µg of Polybrene was used to transduce ~1 million K562 cells in 6-well plates. The transduced cells were selected using media containing 2.5 µg/ml of puromycin. Less than 20% of the transduced cells survived, and these were then diluted into 96-well plates to select single clones. One of the K562 GFP+ clones was used for the analysis shown in this study. Cas9 was electroporated into the K562 GFP+ cells as RNP (20 pmol) with a crRNA targeting the *GFP* sequence. ssODN (66 nt) with or without end modifications was provided as donor template to convert the *GFP* coding sequence to the *BFP* coding sequence. % BFP+ (precise) and % GFP (-); BFP (-) (imprecise) cells were quantified using flow cytometry.

Flow cytometry. The electroporated cells were analyzed on a MACSQuant VYB from Miltenyi Biotec. Cells were gated first based on forward and side scattering to select "live" cells and then for single cells. GFP-positive cells were identified using the blue laser (488 nm) and 525/50 nm filter whereas for the detection of mCherry positive cells, yellow laser (561 nm) and 615/20 nm filter were used. BFP-positive cells were identified using the violet laser (405 nm) and 450±50 nm filter. The gating strategy is shown in **Supplementary Fig.7**.

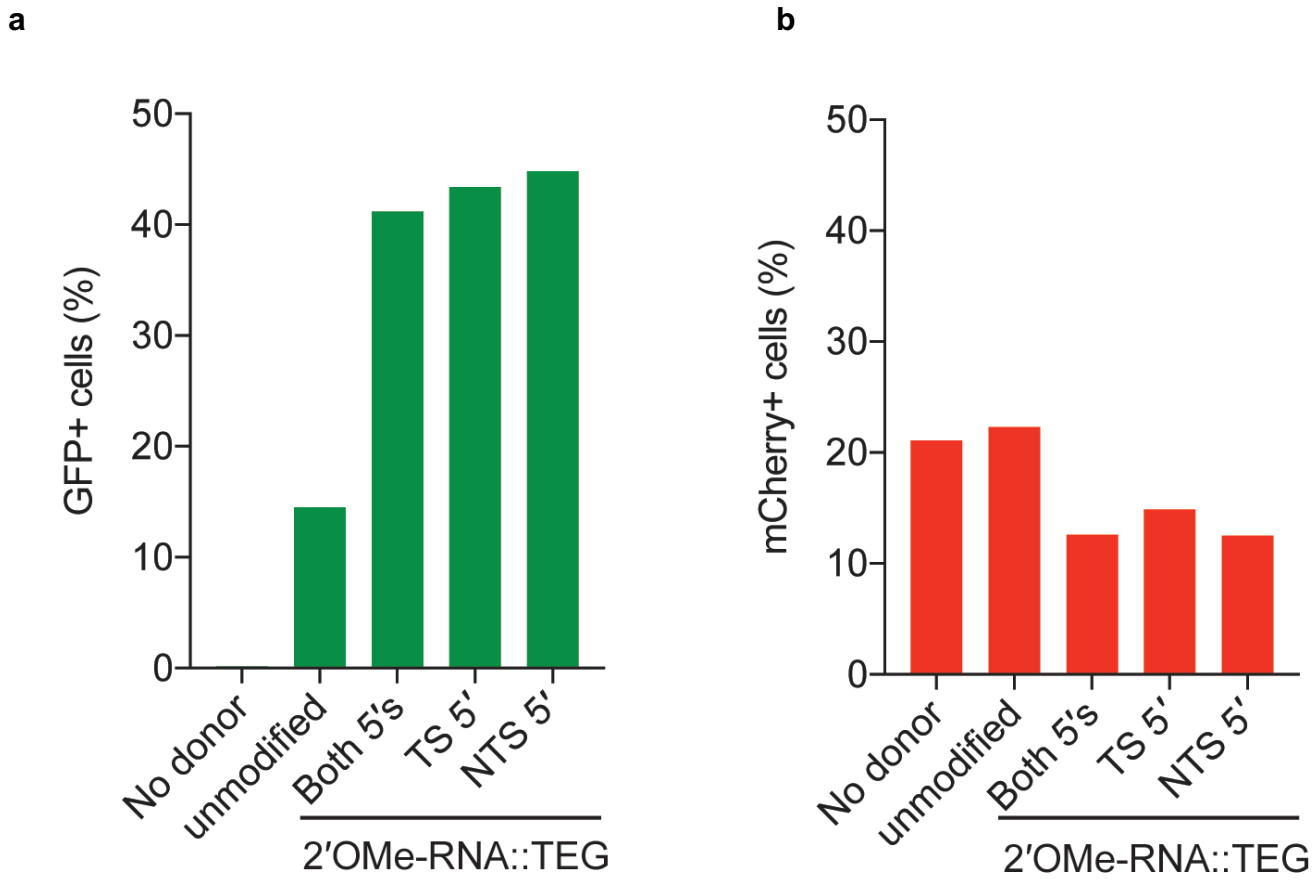
Statistics. ANOVA tests were performed, and adjusted P values were obtained using Tukey's multiple comparisons test. All mammalian experiments were performed with at least 3 replicates and the n values for number of *dpy-10* edited worms scored for GFP expression are shown in the figures. All the statistical analyses were performed using GraphPad Prism (v7.0).



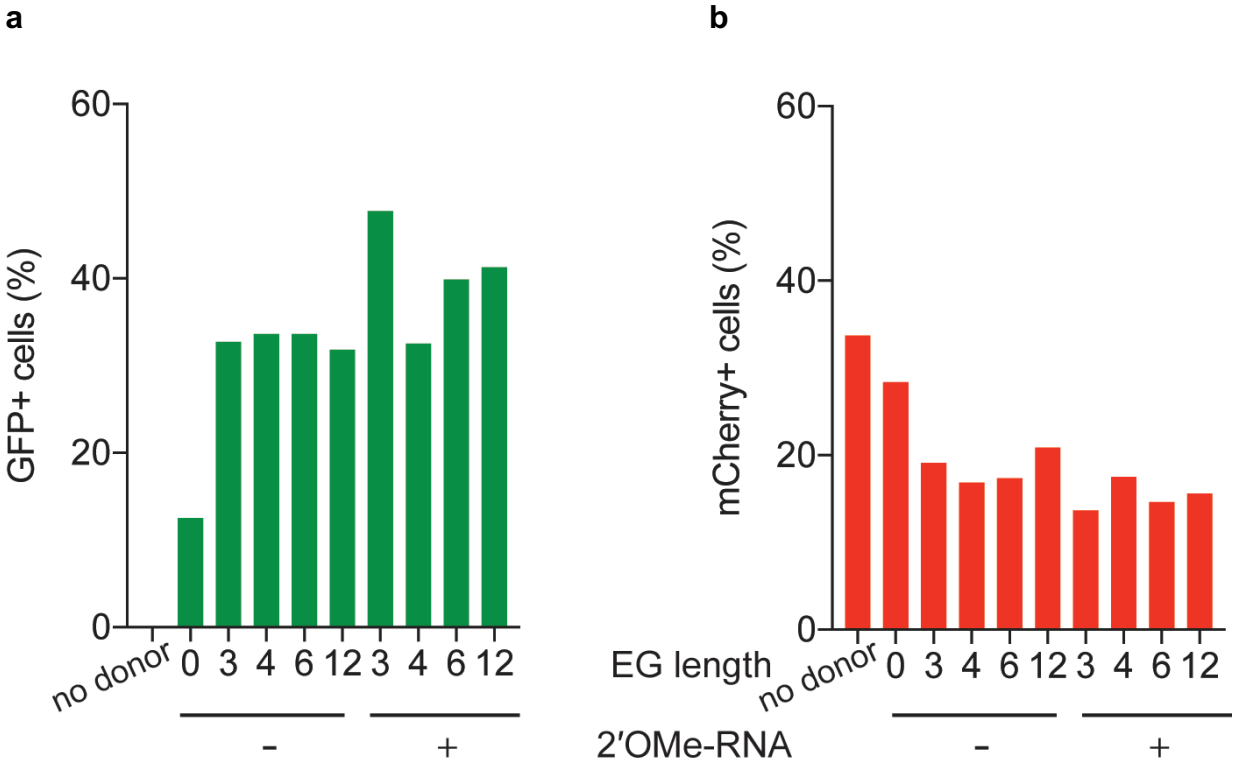
Supplementary Fig. 1. (**a**) Illustration of microinjection into the *C. elegans* hermaphrodite gonad arm. Most of the germline is in the meiotic stage with non-dividing nuclei sharing the common cytoplasm (syncytium). The total number of fertilized eggs is limited by the number of available sperm. Cas9 RNPs and the donor DNA are injected into the syncytium. (**b**) After the injection, F1 *dpy-10* edited animals were screened for GFP::CSR-1 expression in the gonad (P-granules).



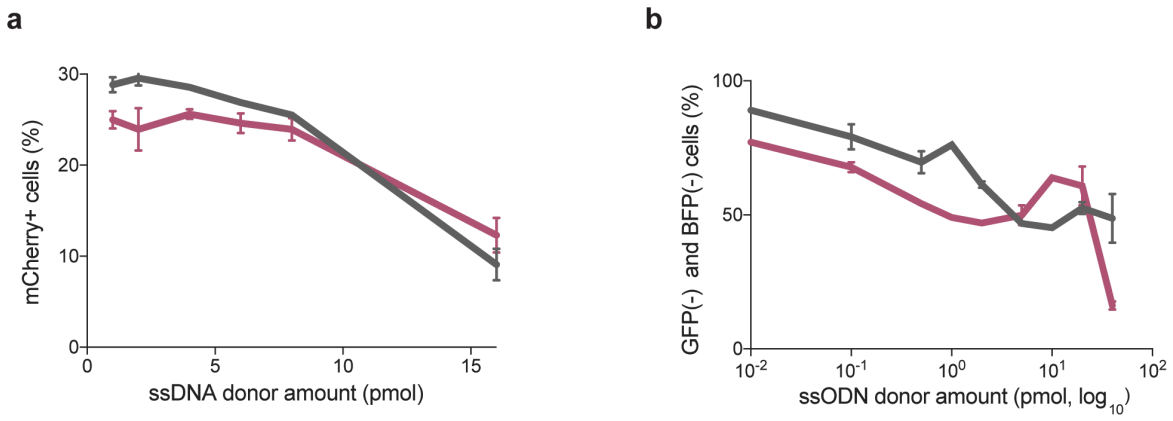
Supplementary Fig. 2. Addition of PNA-K1 (without NLS) to unmodified or end-modified donors does not further improve HDR efficiency in mammalian cells. 0.8 pmol of each type of donor was annealed to PNA-K1 (0.1 to 10 pmol). Editing efficiency was plotted as percentage of (**a**) GFP+ (precise) cells and (**b**) mCherry+ (imprecise) cells. Percentages were calculated by sorting the cells through flow cytometry (see Methods).



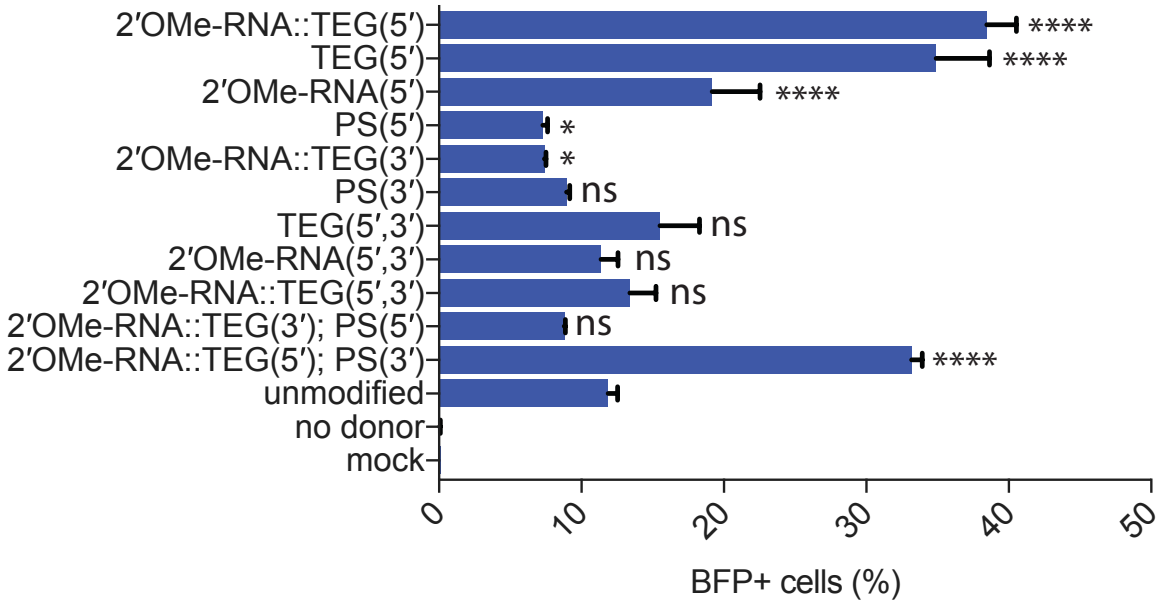
Supplementary Fig. 3. 5′ end modifications on either strand of dsDNA donor promote HDR in TLR-HEK293 cells. Editing efficiency is plotted as percentage of (**a**) GFP+ (precise) and (**b**) mCherry+ (imprecise) with 2′OMe-RNA::TEG modifications at 5′ ends of both the strands, Target Strand (TS), or Non-target Strand (NTS). 1.2 pmol of donors were used; percentages of GFP+ and mCherry+ cells were calculated using flow cytometry (see Methods). TS is defined as crRNA-complementary strand, and NTS is defined as crRNA-noncomplementary strand.



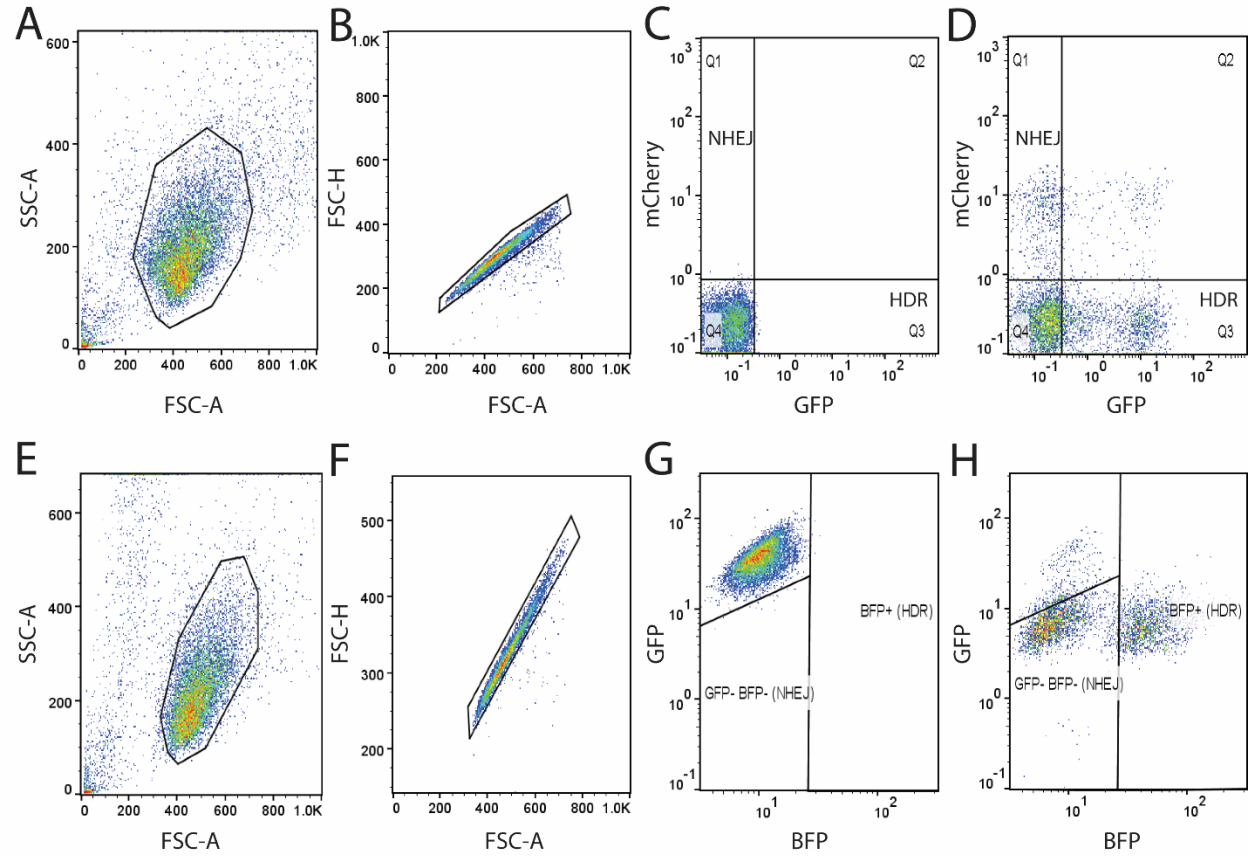
Supplementary Fig. 4. Different lengths of ethylene glycol units promote HDR in TLR-HEK293 cells. Editing efficiency is plotted as percentage of (**a**) GFP+ (precise) and (**b**) mCherry+ (imprecise). Numbers on the x-axes represent the number of ethylene glycols used on the 5′ ends of the donor DNA, without (-) or with (+) 2′OMe-RNA. 1.2 pmol of donors were used; percentages of GFP+ and mCherry+ cells were calculated using flow cytometry (see Methods).



Supplementary Fig. 5. 2′OMe-RNA::TEG modification of long ssDNA and ssODN donors results in reduced imprecise editing. (**a**) Imprecise editing efficiency plotted as percentage of mCherry+ HEK293T TLR cells at different amounts of unmodified or TEG::2′OMe-RNA-modified ssDNA donor. The efficiency of precise editing is plotted in **Fig. 2i.** (**b**) Imprecise editing efficiency plotted as percentage of GFP- and BFP- cells in GFP-to-BFP reporter K562 cells using different amounts of unmodified and TEG::2′OMe-RNA-modifed ssODN donors. The efficacy of precise editing is plotted in **Fig. 2j**.



Supplementary Fig. 6. Effects of terminal and non-terminal modifications of ssODN donors on HDR efficacy. Editing efficacy of GFP to BFP conversion in K562 cells using 0.5 pmol of ssODN donors modified at the 5′ end alone, the 3′ end alone, or at both the 5′ and 3′ ends, with phosphorothioate (PS) TEG, 2′OMe-RNA, or 2′OMe-RNA::TEG, plotted as percentage of BFP+ cells (precise). This figure consists of the data shown in **Fig. 2k** along with other controls and modifications of the donors. Note that the PS modification is at the 5′ or 3′ *internal* linkages while TEG modifications are appended to the 5′ or 3′ terminus.



Supplementary Fig. 7. Flow cytometry analysis to determine the percentage of precise and imprecise genome editing events. The gating strategies used for HEK293T TLR cells (**Top**) and K562 GFP+ cells (**Bottom**) are shown. Cells were first gated based on forward and side scattering to select "live" cells (**a**, **e**), and then gated to select singlets (**b, f**). **c, d.** Quadrant gates were drawn to isolate GFP+ mCherry- cells (indicating successful HDR) and mCherry+ GFP- cells (indicating imprecise repair) for mock-transfected sample (**c**) and treated sample (**d**). **g, h.** K562 GFP+ cells were gated for BFP+ events (precise) and double-negative events (imprecise). Representative mock-transfected sample is shown in **g** and treated sample in **h**.

Supplementary References

1. Li, H. et al. Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv* (2017).
2. Dokshin, G.A., Ghanta, K.S., Piscopo, K.M. & Mello, C.C. Robust genome editing with short single-stranded and long, partially single-stranded DNA donors in C. elegans. *bioRxiv* (2018).
3. Kim, H. et al. A co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. *Genetics* **197**, 1069-1080 (2014).
4. Arribere, J.A. et al. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans. *Genetics* **198**, 837-846 (2014).
5. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).