



RESEARCH

Open Access



ssDNA is not superior to dsDNA as long HDR donors for CRISPR-mediated endogenous gene tagging in human diploid RPE1 and HCT116 cells

Akira Mabuchi^{1†}, Shoji Hata^{1,2*†} , Mariya Genova³, Chiharu Tei¹, Kei K. Ito¹, Masayasu Hirota¹, Takuma Komori¹, Masamitsu Fukuyama¹, Takumi Chinen¹, Atsushi Toyoda⁴ and Daiju Kitagawa^{1*} 

Abstract

Background Recent advances in CRISPR technology have enabled us to perform gene knock-in in various species and cell lines. CRISPR-mediated knock-in requires donor DNA which serves as a template for homology-directed repair (HDR). For knock-in of short sequences or base substitutions, ssDNA donors are frequently used among various other forms of HDR donors, such as linear dsDNA. However, partly due to the complexity of long ssDNA preparation, it remains unclear whether ssDNA is the optimal type of HDR donors for insertion of long transgenes such as fluorescent reporters in human cells.

Results In this study, we established a nuclease-based simple method for the preparation of long ssDNA with high yield and purity, and comprehensively compared the performance of ssDNA and dsDNA donors with 90 bases of homology arms for endogenous gene tagging with long transgenes in human diploid RPE1 and HCT116 cells. Quantification using flow cytometry revealed lower efficiency of endogenous fluorescent tagging with ssDNA donors than with dsDNA. By analyzing knock-in outcomes using long-read amplicon sequencing and a classification framework, a variety of mis-integration events were detected regardless of the donor type. Importantly, the ratio of precise insertion was lower with ssDNA donors than with dsDNA. Moreover, in off-target integration analyses using donors without homology arms, ssDNA and dsDNA were comparably prone to non-homologous integration.

Conclusions These results indicate that ssDNA is not superior to dsDNA as long HDR donors with relatively short homology arms for gene knock-in in human RPE1 and HCT116 cells.

Keywords CRISPR-Cas, Knock-in, Repair template, Long ssDNA

[†]Akira Mabuchi and Shoji Hata contributed equally to this work.

*Correspondence:

Shoji Hata

s.hata@mol.f.u-tokyo.ac.jp

Daiju Kitagawa

dkitagawa@mol.f.u-tokyo.ac.jp

Full list of author information is available at the end of the article



Background

Gene knock-in is a crucial technique for studying gene function by introducing specific mutations or insertions at endogenous loci. Recent developments in genome editing technology using programmable site-specific nucleases, especially the CRISPR-Cas system, have made it possible to perform gene knock-in in a broader range of species and cell lines [1]. Cas9 and Cas12a nucleases, which are used for CRISPR-mediated genome editing, are targeted to specific genomic loci with short guide RNA to induce double-strand breaks (DSBs) [2–4]. These DSBs can be repaired by two major pathways. The first is the re-ligation of the broken DNA ends through non-homologous end joining (NHEJ). This pathway is error-prone and often introduces insertions or deletions (indels), which can lead to gene knockout [5]. The second pathway is homology-directed repair (HDR), in which DSBs are repaired precisely by using homologous DNA sequences as a repair template [6]. In HDR, exogenously introduced DNA for gene knock-in can also serve as a repair template when the sequence contains so-called homology arms (HAs)—elements homologous to the region flanking the target site.

The optimal type of DNA donors used as HDR templates for gene knock-in would vary with the length of the sequence to be inserted at the targeted site. For knock-in of short sequences or base substitutions such as point mutations, single-stranded oligodeoxynucleotides (ssODNs) are frequently used due to their high knock-in efficiency and ease of synthesis [7–9]. However, the optimal type of HDR donors for insertion of longer transgenes such as fluorescent reporters remains unclear. Various forms of donors, including plasmids, linear dsDNA produced by PCR, and ssDNA, are applicable for knock-in of long sequences. Although plasmids have been used as a conventional HDR donor, it has been reported that plasmids are less efficient than linear dsDNA or ssDNA for fluorescent tagging in human cell lines. Moreover, their preparation requires time-consuming cloning steps, which also prevents plasmids from being the first choice for knock-in donor [10, 11].

Besides efficiency, the specificity and accuracy of the knock-in are also key factors that determine donor performance [12]. It is known that exogenous DNA can be non-specifically inserted via non-HDR pathways into unintended locations of the genome, such as off-target cleavage sites introduced by the Cas nuclease [13–15]. Homology-independent donor integration can also occur at the target site of knock-in, which results in inaccurate insertion of the transgenes [16–18]. While both linear dsDNA and ssDNA donors can be inserted into the genome in a homology-independent manner, it has been reported that long ssDNA donors are less prone to

off-target integration than long dsDNA [10, 19]. In terms of accuracy, another report suggests that the frequency of precise insertion of long dsDNA and ssDNA donors via HDR varies by cell line [18]. Thus, it remains controversial whether linear dsDNA or ssDNA templates are more suitable as HDR donors for insertion of long transgenes.

In this study, we compare the performance of dsDNA and ssDNA as long HDR donors for the endogenous tagging with fluorescent proteins in the hTERT-immortalized RPE1 cell line and the HCT116 colon cancer cell line, which are widely used as human non-transformed and transformed diploid cell lines in the field of cell biology, respectively. Quantitative analyses of the endogenous tagging in different genes show that ssDNA tends to have lower knock-in efficiency than dsDNA. It also turns out that ssDNA is not superior to dsDNA in terms of the specificity and accuracy of long transgene insertions. Taken together, our findings indicate that dsDNA is more suitable than ssDNA as long HDR donors for endogenous gene tagging with long sequences in human diploid RPE1 and HCT116 cells.

Results

An optimized CRISPR knock-in method using long dsDNA donors for efficient tagging with fluorescent proteins in human diploid cells.

Given the widespread use of CRISPR-mediated generation of knock-in cell lines, it is informative to compare the performance of long dsDNA and ssDNA donors in a simple and practical knock-in method. To this end, we first established a simple, long-dsDNA-based method for endogenous gene tagging with Cas12a or Cas9 by following conventional approaches with slight optimization (Fig. 1a) [15, 20]. Long dsDNA donors were amplified by a one-step PCR using a pair of primers containing 90 bases of HA sequences. To avoid plasmid construction, we synthesized the guide RNA (crRNA for Cas12a and sgRNA for Cas9) via *in vitro* transcription from PCR-assembled DNA templates. The guide RNA was mixed *in vitro* with recombinant Cas12a or Cas9 proteins to form ribonucleoprotein (RNP) complexes, which were then electroporated into cells together with the long dsDNA donors.

To test whether this cloning-free method with the long dsDNA donors can be applied to efficient knock-in in RPE1 cells, we performed Cas12a-mediated endogenous tagging of the nuclear protein HNRNPA1 and the mitochondrial protein TOMM20 with the green fluorescent protein mNeonGreen (mNG) (Fig. S1a). Fluorescence imaging revealed the expected localization of each mNG-fused endogenous protein (Fig. 1b). Similarly, Cas9-based mNG knock-in was carried out to target three different proteins (HNRNPA1, CAMSAP2, and p53) (Fig. S1b). For

each protein, a specific localization pattern of mNG was observed in all mNG-positive cells, indicating successful tagging of the targets (Fig. 1c). The Cas12a or Cas9-mediated insertion of the mNG sequence into the target locus was confirmed by genomic PCR (Fig. 1d). The specificity of the mNG tagging of HNRNPA1 was further confirmed by western blotting with antibodies against HNRNPA1 and mNG (Fig. 1e). The dsDNA-based knock-in method was also applicable to HCT116 cells for mNG tagging of HNRNPA1 with either Cas12a or Cas9 (Fig. S1c).

To evaluate the tagging efficiency of our approach, we conducted a flow cytometric analysis, which allows the detection of mNG-positive cells in a high-throughput and quantitative manner. The analysis, performed on an RPE1 cell population in which HNRNPA1 or TOMM20 was targeted for mNG tagging by Cas12a, revealed that the knock-in efficiency was 3 to 5% for each gene (Fig. 1f, g). A comparable level of mNG-positive cells was detected using a similar strategy with Cas9 (Fig. 1h, i). In HCT116 cells, the knock-in efficiency was about 5% for HNRNPA1 (Fig. S1d, e). Taken together, our optimized cloning-free knock-in method with long dsDNA donors enables efficient endogenous gene tagging in human diploid cells.

Production of long ssDNA donors with high purity using an optimized T7 exonuclease-based method.

For endogenous tagging with long sequences using ssDNA donors, long ssDNA should be produced with high yield and purity. Although synthesized ssDNA of high purity is commercially available, it is costly and thus not a sustainable option for frequent knock-in experiments. As a suitable alternative for routine laboratory work, we optimized an ssDNA production method using dsDNA-specific T7 exonuclease (Fig. 2a) [21, 22]. First, dsDNA was amplified by PCR using HA-containing long primers, whereas one of them has five sequential phosphorothioate (PS) bonds at the 5' end. The amplified dsDNA was then column-purified and mixed with T7

exonuclease. Since the consecutive PS bonds block the digestion by T7 exonuclease, the strand with a non-modified 5' end would be selectively digested, and the other strand would remain as intact ssDNA.

To verify that this strategy is effective enough to produce long ssDNA donors, dsDNA of 943 bp encoding the HDR donor sequence for mNG tagging of HNRNPA1 was amplified using different combinations of PS-modified (PS) and non-modified (noPS) primers and subsequently subjected to the T7 exonuclease reaction. The gel electrophoresis analysis showed that an additional DNA band with higher mobility was detected in the “PS-noPS” and “noPS-PS” conditions, where one of the two primers was PS-modified (Fig. 2b). This DNA band was eliminated when reacted with ssDNA-specific exonuclease I, suggesting a successful production of ssDNA (Fig. S2a). However, we identified two major drawbacks of this ssDNA production method: first, dsDNA remained partially undigested even after the T7 exonuclease reaction. The second problem is that “PS-PS” dsDNA, in which both of the 5' ends are PS-modified, seemed partially degraded by T7 exonuclease, suggesting that the protection by the PS modification was imperfect. Since the quality of chemically synthesized oligonucleotides deteriorates with their length [23, 24], we assumed that the latter problem of insufficient protection was due to the low efficiency of PS-modification of the long primers.

To resolve this issue, we adopted a two-step PCR procedure to produce modified dsDNA, allowing the usage of short PS-modified primers, the purity of which is supposed to be higher than that of long primers (Fig. 2c). Compared to the initial one-step PCR method, the degradation of the PS-PS dsDNA by T7 exonuclease was significantly reduced with the new two-step approach (Fig. 2d). When the two-step PCR method was applied for long ssDNA preparation (i.e., with one modified 5' end primer), the amount of undigested dsDNA was decreased, and the yield of ssDNA was higher (Fig. 2e). Furthermore, an annealing-based analysis revealed that

(See figure on next page.)

Fig. 1 An optimized knock-in method using long dsDNA donors for efficient endogenous tagging with fluorescent proteins. **a** Schematic overview of long-dsDNA-based endogenous gene tagging in human RPE1 cells. **b** Representative images of RPE1 cells with Cas12a-mediated endogenous mNG tagging of the indicated genes. Cells at 7–12 days after electroporation were fixed and analyzed. Scale bar: 10 μ m. **c** Representative images of RPE1 cells with Cas9-mediated endogenous mNG tagging of the indicated genes. Cells at 12–17 days after electroporation were fixed and analyzed. Scale bar: 10 μ m. **d** Genomic PCR detecting the mNG insertion into the indicated genes with Cas12a or Cas9-mediated knock-in in RPE1 cells. The primers were designed to amplify the 5' junction of the mNG insertion for each gene. *TUBB5*: loading control. LHA: left HA, RHA: right HA. **e** Western blotting confirming the fusion of mNG to HNRNPA1 in RPE1 cells via the Cas12a-mediated knock-in method. The FACS-enriched, mNG-positive knock-in cells were used. HSP90: loading control. WB: Western blotting. **f** Flow cytometric analysis of Cas12a-mediated HNRNPA1-mNG and TOMM20-mNG knock-in RPE1 cells. Cells at 8 days after electroporation were analyzed. Percentages of cells with mNG signal are shown in the plots. **g** Quantification of percentages of mNG-positive cells from (f). Data from three biological replicates are shown. > 5,000 cells were analyzed for each sample of HNRNPA1 and TOMM20. Data are represented as mean \pm S.D. **h** Flow cytometric analysis of Cas9-mediated HNRNPA1-mNG knock-in RPE1 cells. The donor concentration was 33 nM. Cells at 5 days after electroporation were analyzed. Percentages of cells with mNG signal are shown in the plots. **i** Quantification of percentages of mNG-positive cells from (h). Two different concentrations of the dsDNA donor were analyzed. Data from three biological replicates are shown. 10,000 cells were analyzed for each sample. Data are represented as mean \pm S.D. Full-length blots and gels are presented in Fig. S6

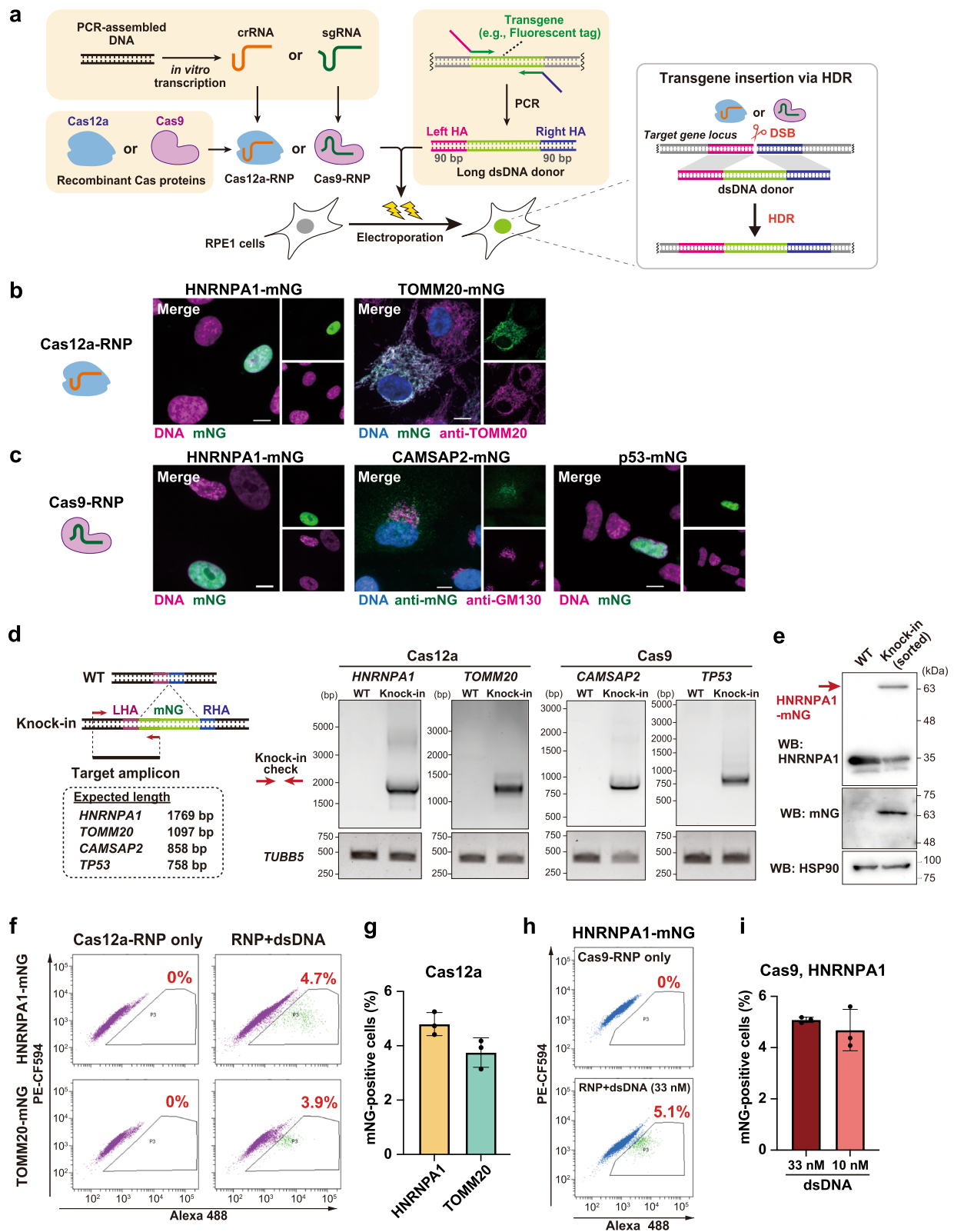


Fig. 1 (See legend on previous page.)

the two-step PCR method improved the strand selectivity of ssDNA production compared to the one-step PCR method (Fig. S2b).

Even though our improved T7 exonuclease-based method (hereafter referred to as T7 method) enables robust production of long ssDNA donors, a faint band of undigested dsDNA was still observed in agarose gel electrophoresis (Fig. 2e). For further improvement of the ssDNA purity, we added a restriction enzymes (REs) reaction after the digestion by T7 exonuclease (referred to as T7RE method) (Fig. 2f). We selected four REs (HpyCH4III, Hpy188I, NlaIII, and RsaI) whose recognition sequences are short so that they can be applied to various DNA sequences. Indeed, by combining these four REs, DNA sequences encoding typical fluorescent proteins can be mostly cleaved to fragments of 150 bp or less (Fig. S2c). Since these REs are fully active in the same buffer as T7 exonuclease, a sequential one-pot reaction can be applied to T7 exonuclease and the REs. Importantly, all the four REs produce blunt or 3'-protruding ends, which serve as substrates for T7 exonuclease. Therefore, dsDNA fragments produced by these REs are supposed to be degraded to even smaller ssDNA fragments in the presence of T7 exonuclease activity. Indeed, the T7RE method resulted in the successful removal of dsDNA remnants below the detection limit (Fig. 2g). In summary, our optimized T7RE method enables the preparation of long ssDNA donors with high yield and purity.

Comparison of knock-in efficiency between dsDNA and ssDNA long donors

Using long ssDNA donors produced by the T7 and the T7RE methods (referred to as T7 donors and T7RE donors, respectively), we performed endogenous gene tagging in RPE1 cells. Electroporation of Cas12a-RNP and ssDNA donors for mNG tagging of HNRNPA1 or TOMM20 resulted in successful knock-in in each gene, as confirmed by the correct subcellular localization of the mNG signal (Fig. 3a). ssDNA donors prepared with

our optimized methods were also applicable to Cas9-mediated mNG tagging of HNRNPA1 (Fig. 3a). We further performed flow cytometric analysis to evaluate the knock-in efficiency of ssDNA donors compared to that of dsDNA donors. For mNG tagging of HNRNPA1 using Cas12a, knock-in efficiency tended to be lower with T7RE donors than with dsDNA donors, especially in the case of the sense strands (Fig. 3b, c). Similarly, for mNG tagging of TOMM20, the ratio of mNG-positive cells introduced with T7RE donors was less than one-third of that with dsDNA donors (Fig. 3d). Moreover, for mNG tagging of HNRNPA1 using Cas9, the percentage of mNG-positive cells was also lower with T7RE donors than with dsDNA donors (Fig. 3e). In HCT116 cells, the lower knock-in efficiency of T7RE donors was also observed for Cas12a-mediated mNG tagging of HNRNPA1 (Fig. S3a-c). These data indicate low knock-in efficiency of T7RE donors compared to dsDNA donors across different Cas nucleases, target genes and cell types.

Interestingly, the knock-in rate with T7RE donors tended to be also lower than that with T7 donors in all the tested conditions (Figs. 3c-e, S3c). The difference in efficiency between the two donors might be attributed to the amount of residual dsDNA. To estimate whether a small amount of dsDNA remnant in T7 donors would impact the knock-in performance, we conducted titration of donor concentration for dsDNA for Cas12a-mediated mNG tagging of HNRNPA1. Analysis by flow cytometry showed that dsDNA donors retained more than half of their maximum efficiency at the concentration of 1 nM and about 0.5% efficiency even at 0.1 nM (Fig. 3f). In contrast, the efficiency of 1 nM of T7RE donors was reduced to less than one-tenth of their maximum. The result suggests that a small amount of dsDNA remnant among T7 donors might work as a template for HDR together with the ssDNA. Importantly, the concentration of dsDNA donors required to reach their maximum efficiency was lower than that of ssDNA. Collectively, these

(See figure on next page.)

Fig. 2 Optimization of long ssDNA production using T7 exonuclease and restriction enzymes. **a** Schematic of long ssDNA production using T7 exonuclease (one-step PCR method). **b** T7 exonuclease reaction on dsDNA amplified using three different combinations of primers (PS-modified (PS) or non-modified (noPS) for the forward and reverse primers). The DNA sequence of the donor for Cas12a-mediated mNG tagging of HNRNPA1 was used in this experiment. The bottom image is of the same gel as the top one, with higher brightness and contrast. ssDNA shows higher mobility than dsDNA of the same length. Asterisks show undigested dsDNA remnants. **c** Schematic of ssDNA production by two-step PCR and T7 exonuclease (T7 method). **d** "PS-PS" dsDNA was prepared with one-step or two-step PCR and subsequently subjected to the T7 exonuclease reaction. Plot profiles for each lane are shown below the gel electrophoresis images. The two images are cropped from the same gel image. **e** Production of long ssDNA donors using the one-step and the two-step PCR methods. The bottom image is of the same gel as the top one, with higher brightness and contrast. **f** Schematic of ssDNA production using T7 exonuclease and restriction enzymes (T7RE method). After two-step PCR and T7 exonuclease reaction, the indicated four restriction enzymes digest dsDNA remnants to produce short dsDNA fragments which can be further degraded by T7 exonuclease. **g** ssDNA production by the T7 and the T7RE methods. The DNA sequence (sense strand) of the donor for Cas9-mediated mNG tagging of HNRNPA1 was used in this experiment. The last two lanes contain column-purified DNA products of both reactions. Plot profiles for the last two lanes are shown below the gel electrophoresis image. The two images are cropped from the same gel image. Full-length gels are presented in Fig. S6

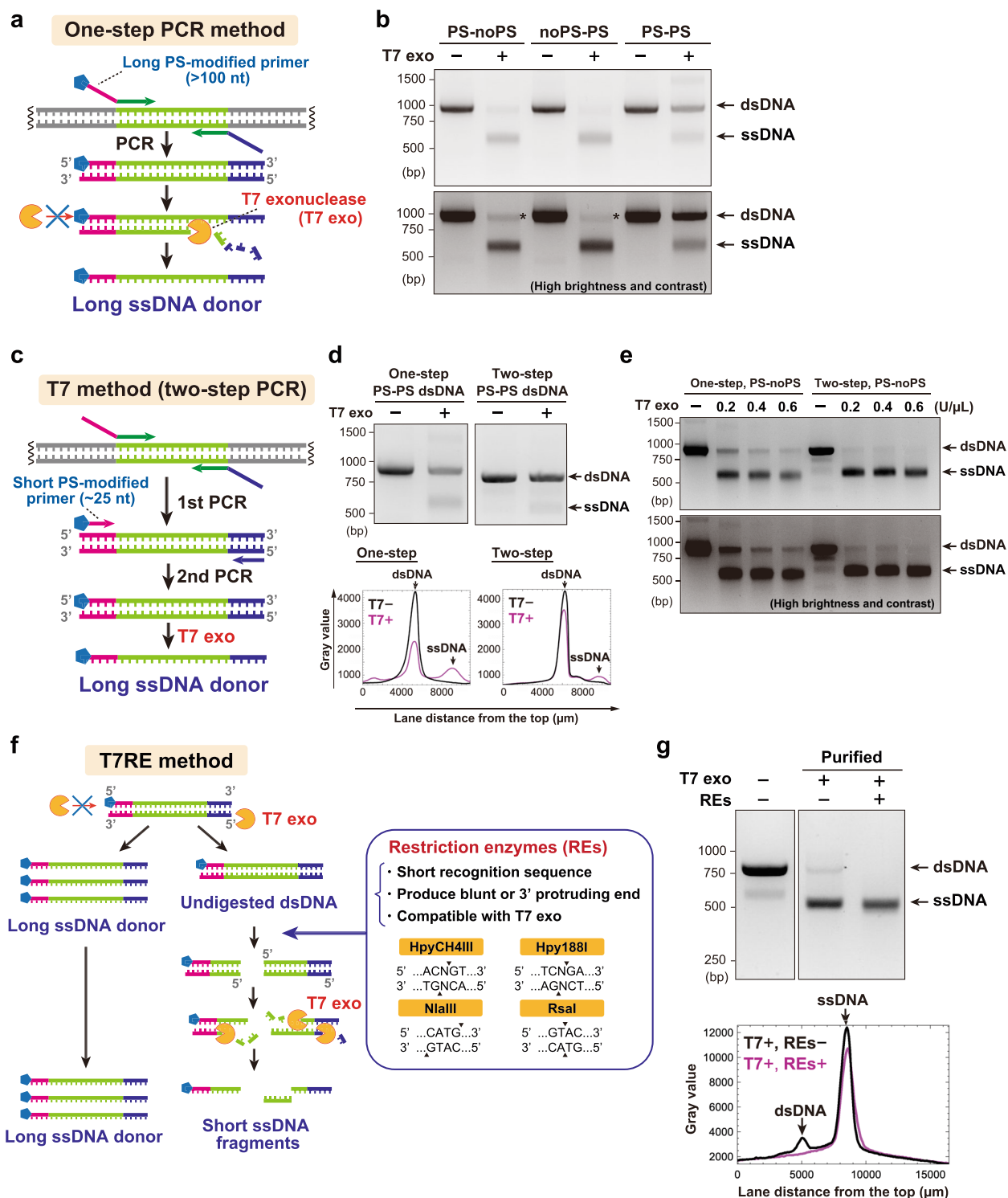


Fig. 2 (See legend on previous page.)

data indicate the superiority of dsDNA donors to ssDNA donors in terms of knock-in efficiency.

Evaluation of knock-in accuracy using long-read amplicon sequencing and knock-knock pipeline

Next, we compared the frequency of precise insertion of transgenes between dsDNA and ssDNA donors in RPE1 cells. To this end, we performed the long-read amplicon sequencing by PacBio and analysis by *knock-knock*, a computational framework that allows a high-throughput genotyping of knock-in alleles [18]. We applied this approach to mNG tagging of HNRNPA1 using Cas12a-RNP (Fig. 4a). After electroporation and subsequent cell expansion for two to three weeks, we collected mNG-positive cells by fluorescence-activated cell sorting (FACS) to enrich knock-in cells, mimicking the selection process in the establishment of endogenously tagged cell lines with fluorescent proteins. Genomic DNA was isolated after several days of culture in a 96-well plate. The specific DNA sequence at the target locus was then amplified for library preparation of sequencing. After sequencing by PacBio and Circular Consensus Sequence (CCS) generation, *knock-knock* classified each read into a specific category of the knock-in outcome, such as WT, indels, HDR, or subtypes of mis-integration.

The result of *knock-knock* analysis revealed that various mis-integration events occurred in addition to precise insertion via HDR for both dsDNA and ssDNA donors, as previously described [18] (Fig. 4b, c, S4a). *Knock-knock* classified these mis-integration events into the following categories: blunt (at least one of the donor ends is directly ligated to the DSB site), incomplete (only one side of the donor is integrated via HDR), concatemer (multiple donors are inserted), donor fragment (both ends of the donor are integrated in a non-HDR manner), and complex (not classified into the other four mis-integration categories).

We then calculated the percentage of each category to total integration events. The results show that the rate of perfect HDR tends to be lower in ssDNA donors than in dsDNA donors (Fig. 4c, d). The blunt integration, which

is assumed to be an outcome of NHEJ-based-ligation, was less likely to occur in T7RE (pure ssDNA) donors than in dsDNA donors (Fig. 4c). On the contrary, integration of donor fragments and complex integration were prominent in T7RE donors compared to dsDNA donors. Across the conditions, only a small percentage of reads were classified into the concatemer category. We further compared sense and antisense ssDNA donors for the directionality of HDR in the “asymmetric HDR” events, in which only one of the ssDNA ends is inserted via HDR (Fig. S4b). We found that HDR exhibited a directional bias towards the 3' end for the sense strand, while displaying a contrary trend towards the 5' end for the antisense strand (Fig. S4c). To put it differently, HDR was likely to occur at the 3' side of ssDNA donors. This trend is consistent with what was shown in the previous studies [17, 18], indicating that our *knock-knock* analysis reflects the distribution of repair outcomes at least to some extent.

Taken together, these data obtained from *knock-knock* analysis suggest that dsDNA outperforms ssDNA in the frequency of perfect HDR for long transgene insertion.

Comparison of a propensity for homology-independent integration between dsDNA and ssDNA donors

Finally, we compared the propensity for off-target insertion between dsDNA and ssDNA donors, which is considered to occur via non-homologous pathways. To do this, we evaluated the frequency of non-homologous integration by using donors without HAs for Cas12a-mediated mNG tagging of TOMM20 (Fig. 5a). When analyzed by flow cytometry, mNG-positive cells were detected for the condition where they were transfected with dsDNA donors together with the Cas12a/crRNA complex, but not when the crRNA was omitted (Fig. S5a). This indicates that homology-independent donor insertion into Cas12a-RNP-induced DSBs can be quantitatively evaluated with this approach. In the comparison of dsDNA and ssDNA donors, mNG-positive cells were observed with T7RE donors at a

(See figure on next page.)

Fig. 3 Comparison of knock-in efficiency between dsDNA and ssDNA long donors. **a** Representative images of RPE1 cells with mNG tagging to the indicated genes using T7 or T7RE donors. The indicated Cas nuclease was used for each condition. For ssDNA donors, sense strands were used. Cells at 7–13 days after electroporation were fixed and analyzed. Scale bar: 10 μ m. **b** Flow cytometric analysis of Cas12a-mediated HNRNPA1-mNG knock-in RPE1 cells, using dsDNA, T7, and T7RE donors. Cells at 12 days after electroporation were analyzed. Percentages of cells with mNG signal are shown in the plots. **c** Quantification of percentages of mNG-positive cells from (b). Data from three biological replicates are shown. Approximately 10,000 cells were analyzed for each sample. **d** Flow cytometric quantification of mNG-positive cells in Cas12a-mediated TOMM20-mNG knock-in RPE1 cells, using the indicated donors. Cells at 9 days after electroporation were analyzed. Data from three biological replicates are shown. > 5,000 cells were analyzed for each sample. **e** Flow cytometric quantification of mNG-positive cells in Cas9-mediated HNRNPA1-mNG knock-in RPE1 cells, using the indicated donors at 33 nM. Cells at 6 days after electroporation were analyzed. Data from three biological replicates are shown. Approximately 10,000 cells were analyzed for each sample. **f** Titration of the indicated donors for mNG tagging of HNRNPA1 using Cas12a in RPE1 cells. Cells at 11 days (dsDNA) or 10 days (T7 and T7RE) after electroporation were analyzed. For ssDNA donors, sense strands were used. Data from three biological replicates are shown. > 8,000 cells were analyzed for each sample. n.a.: Not analyzed. Data are presented as mean \pm S.D. P-value was calculated by the Tukey–Kramer test. *** $P < 0.001$, n.s.: Not significant

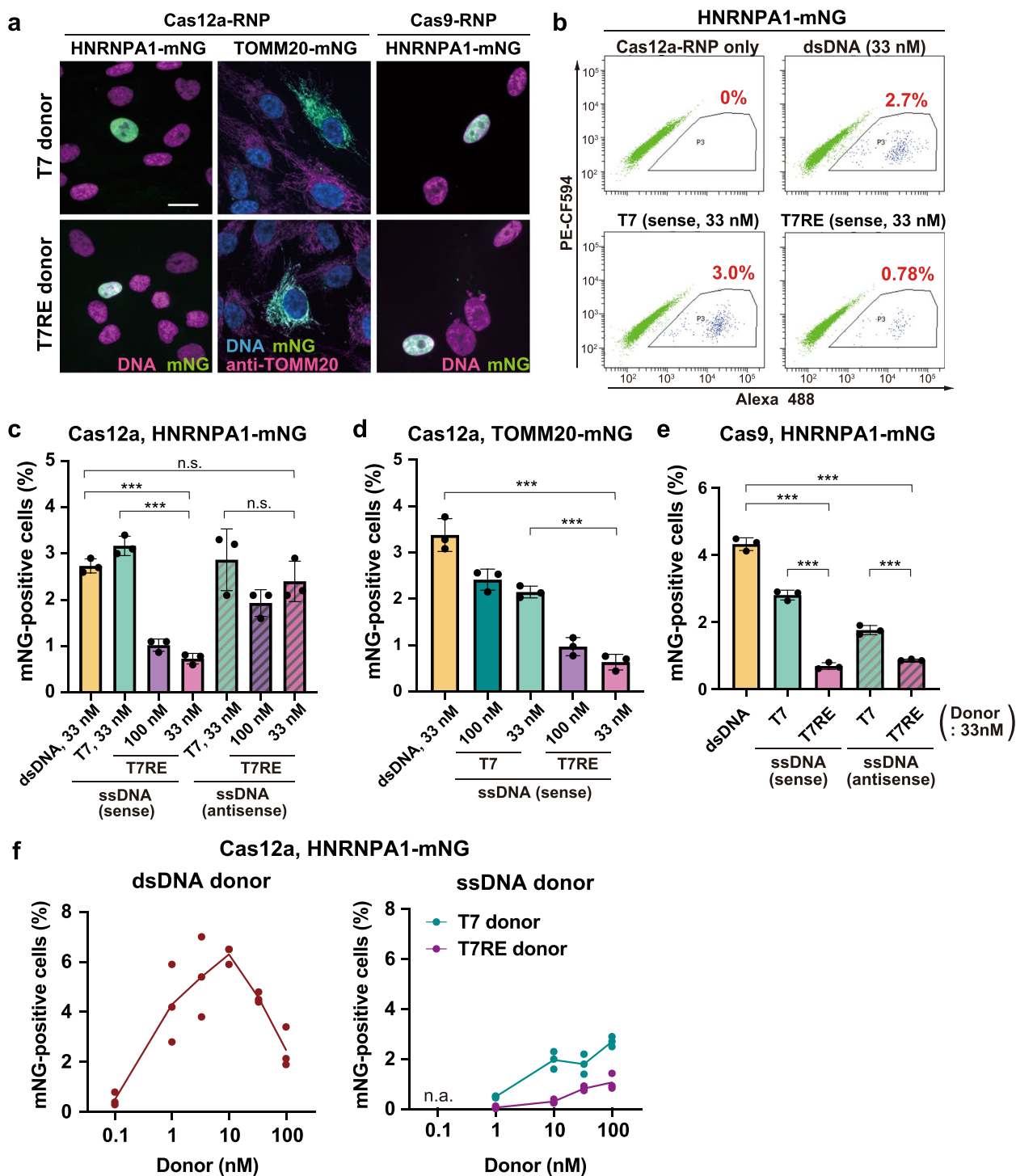


Fig. 3 (See legend on previous page.)

similar level as dsDNA donors in RPE1 cells (Fig. 5b). The tendency of a similar frequency of fluorescent cells with dsDNA and T7RE donors was also observed in HCT116 cells (Fig. S5b). Furthermore, mitochondrial localization of mNG fluorescence was confirmed

by microscopic observation for all the conditions, suggesting integration of mNG donors into the targeted *TOMM20* locus via non-homologous pathways (Fig. 5c, S5c). It should be noted that the fraction of mNG-positive cells is expected to be much lower than

the frequency of total donor integration events at the cut site because the mNG signal can be observed only when mNG is inserted in the correct orientation and a correct reading frame. Indeed, insertion of the donors in reverse orientation was detected for all three donor types in RPE1 cells (Fig. S5d).

To further confirm the occurrence of homology-independent integration, we prepared a DNA cassette of 1774 bp encoding TRE3G-GALNT2-mNG-polyA, which allows doxycycline-dependent expression of mNG-fused Golgi protein GALNT2 (Fig. 5d). This approach was expected to be more sensitive than the former strategy (Fig. 5a) for detecting integration events because GALNT2-mNG expression occurs regardless of how the cassette is inserted into the genome due to its promoter. The electroporated RPE1 cells were cultured for more than 10 days so that the non-integrated cassettes could be cleared from the cells. When analyzed by flow cytometry, the control population that was electroporated with Cas12a and dsDNA cassettes but without crRNA exhibited about 0.2% of fluorescent cells, presumably due to random integration of dsDNA cassettes into the genome, as reported previously (Fig. 5e) [15]. Nevertheless, a higher population of mNG-positive cells (4 to 10%) was detected from the cells electroporated with Cas12a-crRNA RNP complexes and dsDNA cassettes, suggesting the integration into the Cas12a-induced DSBs. In the case of ssDNA donors, the fractions of mNG-positive cells were comparable to those of dsDNA, consistent with the result of the former analysis (Fig. 5b). The mNG signals detected by flow cytometry were confirmed to be derived from the GALNT2-mNG cassette since the fluorescence was doxycycline-dependent and a Golgi-like localization of the mNG signal was observed by microscopy (Fig. 5f). Therefore, it is likely that there are no significant differences between dsDNA and ssDNA donors in their propensity for homology-independent insertion into the Cas12a-induced DSBs. In conclusion, our comprehensive analyses indicate that ssDNA donors are not superior to dsDNA for endogenous gene tagging with long transgenes in RPE1 and HCT116 cells.

(See figure on next page.)

Fig. 4 Long-read amplicon sequencing to evaluate the knock-in accuracy of dsDNA and ssDNA donors. **a** Schematic overview of the analysis of knock-in outcomes. After electroporation of Cas12a-RNP and HDR donors for mNG tagging of HNRNPA1, RPE1 cells were expanded for two to three weeks. mNG-positive cells were then collected by FACS, and genomic DNA was isolated. Libraries for sequencing were prepared from the amplified target locus and subjected to long-read amplicon sequencing by PacBio. After analysis of sequencing outputs, including CCS generation, *knock-knock* categorized each read into a specific category of a knock-in outcome. **b** Representative plots generated by *knock-knock* showing the distribution frequency of amplicon lengths. The range of read lengths corresponding to WT and indels, perfect HDR, truncated integrations, and duplication of homology arm(s) are indicated. **c** Distribution of integration events across the donor types. For each category, the percentage within total integration events was calculated. Data from three biological replicates are shown. For each sample, 11,559–44,431 reads were categorized as the integration events from 43,697–91,850 total reads. **d** Data from (c), the frequencies of perfect HDR are highlighted. A two-tailed, unpaired Student's t-test was used to obtain the P-value. *P < 0.05. s: sense strand, as: antisense strand. Data are presented as mean \pm S.D

Discussion

In this study, we systematically compared the performance of dsDNA and ssDNA donors for CRISPR-Cas knock-in of long transgenes in two different human diploid cell lines. Our analysis revealed that knock-in efficiency tended to be higher for dsDNA compared to the pure ssDNA (T7RE) donors in these cell lines. Recent studies have shown that long ssDNA donors can be used for efficient knock-in in various species [10, 25–29]. Especially in zebrafish, ssDNA has been shown to be more efficient than dsDNA as long HDR donors [29]. On the other hand, in human cells such as primary T cells, HEK293T cells, and hiPS cells, the knock-in efficiency of long ssDNA donors has been described to be lower than that of dsDNA donors [10, 18, 19]. Thus, our data together with these previous reports indicate that dsDNA outcompetes ssDNA for knock-in efficiency in human cells.

To establish cell lines with accurate gene knock-in, the efficiency of perfect HDR is crucial rather than that of seemingly correct knock-in just assessed by flow cytometry or microscopy. By performing long-read amplicon sequencing and *knock-knock* analysis, we estimated the frequency of precise insertion via perfect HDR among a pool of heterogeneous repair outcomes in a high throughput manner, as described previously [18]. Our data show that long ssDNA donors result in lower percentages of perfect HDR in RPE1 cells than dsDNA donors. This observation is consistent with the previous study by Canaj and colleagues, who developed *knock-knock*, in which the perfect HDR rate for long ssDNA donors was similar to or lower than that for dsDNA donors in three different cell lines. Therefore, dsDNA donors are presumably superior to ssDNA donors in terms of precise knock-in of long transgenes.

The previous *knock-knock* data show that dsDNA donors are more prone to NHEJ-mediated mis-integration into the target locus [18]. In agreement with this, our data from *knock-knock* analysis showed that the percentage of the blunt integration of dsDNA donors was higher than that of T7RE donors (pure ssDNA). On the other hand, the previous study reported that ssDNA donors show more pronounced incomplete mis-integration, in

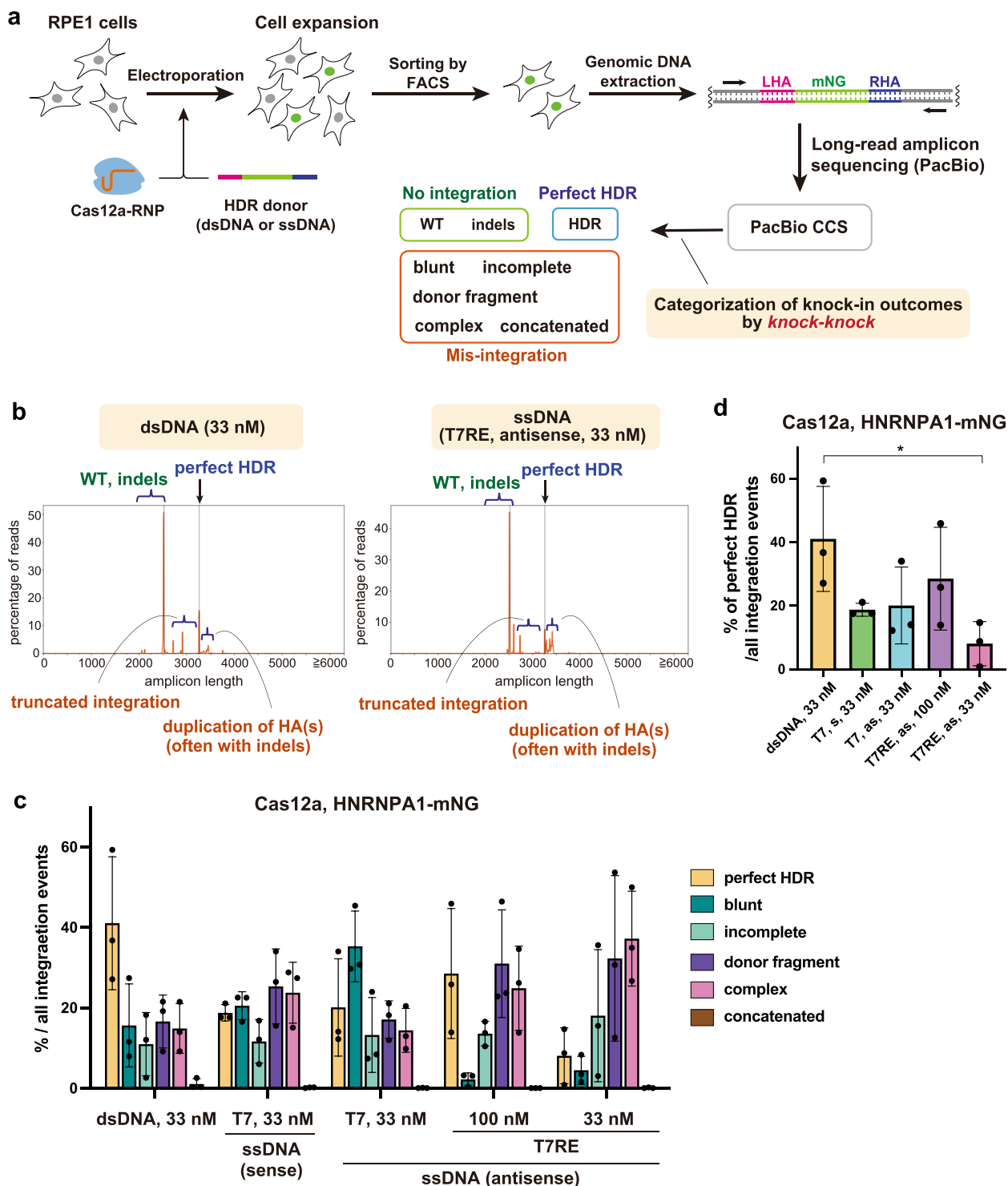


Fig. 4 (See legend on previous page.)

which one end exhibits HDR and the other is repaired imperfectly (often in a truncated manner), which however could not be confirmed by our analysis. This difference might arise due to our enrichment procedure of

fluorescent cells by FACS prior to the sequencing, which eliminates cells with truncated integrations that did not express functional fluorescent proteins. Nevertheless, our *knock-knock* analysis still detected a significant population

of fluorescent cells with truncated integrations. There are two possible explanations for this. The first possibility is that the partially truncated fluorescent proteins could still emit fluorescence. Indeed, GFP1-10, the truncated mutant used in the split GFP system, has been shown to be still weakly fluorescent [30]. The second possibility is that the diploid cells may have two differently edited alleles, of which one has a truncated integration, while in the other mNG is precisely inserted, producing a functional fluorescent protein.

Linear dsDNA is prone to be randomly integrated into the genome via non-HDR pathways at sites of naturally occurring DSBs [14, 31]. In the context of endogenous tagging using long HDR donors, previous reports suggest that non-homologous integration or off-target integration is less likely to occur with ssDNA than with linear dsDNA [10, 19]. However, our two different analyses of homology-independent integration revealed that the integration rate of donors without HAs is almost the same for ssDNA and dsDNA, suggesting a comparable propensity for off-target integration of these donors. Consistently, our *knock-knock* data showed that the frequency of imprecise integration of ssDNA donors was not lower than that of dsDNA. One major difference in our study compared to previous reports is the use of Cas12a, unlike the more commonly applied Cas9. Another source of discrepancy could be the specific cell types used in these experiments. Indeed, it has been shown that the propensity for non-homologous insertion of dsDNA and ssODN varied between HEK293T and hiPS cells [18].

Thus, considering the knock-in efficiency, the insertion accuracy, and the off-target integration frequency, this study indicates that ssDNA is not superior to dsDNA as long HDR donors for knock-in in human diploid RPE1 and HCT116 cell lines. Given that dsDNA donors can be prepared easier than long ssDNA donors, we suggest using dsDNA rather than ssDNA as HDR donors for endogenous tagging with long transgenes in human cells.

Conclusions

Our in-depth evaluation of long ssDNA with relatively short homology arms as a knock-in template indicated that long ssDNA is not superior to dsDNA in CRISPR knock-in in human diploid RPE1 and HCT116 cell lines.

Materials and methods

Cell culture

RPE1 and HCT116 cells obtained from the American Type Culture Collection (ATCC) were maintained in Dulbecco's Modified Eagle Medium/Nutrient Mixture F-12 (DMEM/F-12, Nacalai Tesque) and McCoy's 5A medium (Thermo Fisher Scientific) supplemented with 10% FBS, 100 U/mL penicillin, and 100 µg/mL streptomycin, respectively. Cells were cultured at 37 °C in a humidified 5% CO₂ incubator.

dsDNA donor preparation

dsDNA donors were amplified by PCR from a plasmid encoding the 5xGA linker-mNG sequence using two primers containing 90-base left and right HA sequences, respectively. We used Q5 High-Fidelity 2X Master Mix (New England Biolabs) for PCR. DpnI (0.04 U/µL) and exonuclease I (0.4 U/µL) purchased from New England Biolabs were directly added to the PCR reaction mix and incubated at 37 °C for 30 min, followed by heat inactivation at 80 °C for 20 min. DpnI and exonuclease I were used for digestion of residual template plasmids and primers, respectively. The dsDNA donors were then column-purified using the NucleoSpin Gel and PCR Cleanup kit (Macherey–Nagel) and stored at -20 °C or directly used for electroporation. All primer sequences used in this study are listed in Supplementary Table 1.

ssDNA production using one-step PCR and T7 exonuclease

dsDNA was amplified by PCR as described above with a minor alteration. One of the HA-containing primers was modified with five consecutive phosphorothioate (PS) bonds at the 5' end. The dsDNA was treated with DpnI

(See figure on next page.)

Fig. 5 Comparison of homology-independent integration between dsDNA and ssDNA donors. **a** Schematic for evaluating homology-independent integration of mNG donors into Cas nuclease-induced DSBs. Since the Cas12a cleavage site is located inside the coding region of the *TOMM20* gene, homology-independent integration of mNG into the cleavage site in the correct orientation and a correct reading frame leads to the expression of TOMM20-mNG proteins. **b** Flow cytometric analysis of the homology-independent integration experiment using RPE1 cells. Sense strands were used for T7 and T7RE donors. Cells at 9 days after electroporation were analyzed. Data from three biological replicates are shown. > 5,000 cells were analyzed for each sample. **c** Representative images from the homology-independent integration experiment. Cells at 11 days after electroporation were fixed and analyzed. Scale bar: 10 µm. **d** Schematic overview of the workflow for evaluating homology-independent integration of GALNT2-mNG cassettes into Cas12a nuclease-induced DSBs. Non-integrated cassettes are cleared from cells during long-term culture for more than 10 days. The cassettes inserted into the genome produce doxycycline (dox)-induced expression of the Golgi protein GALNT2-mNG. **e** Flow cytometric analysis of the cassette integration experiment using RPE1 cells. Cells at 20 days after electroporation were treated with 1 µg/mL of doxycycline (dox) for 24 h and analyzed. Data from three biological replicates are shown. Approximately 20,000 cells were analyzed for each sample. **f** Representative images from the cassette integration experiment. Cells at 13 days after electroporation were treated with 1 µg/mL of doxycycline (dox) for 24 h and fixed for analysis. Arrowheads indicate cells with Golgi-like mNG signals. Scale bar: 100 µm (the left panel), 10 µm (the right panel). Data are presented as mean ± S.D. A two-tailed, unpaired Student's t-test was used to obtain the P-value. n.s.: Not significant

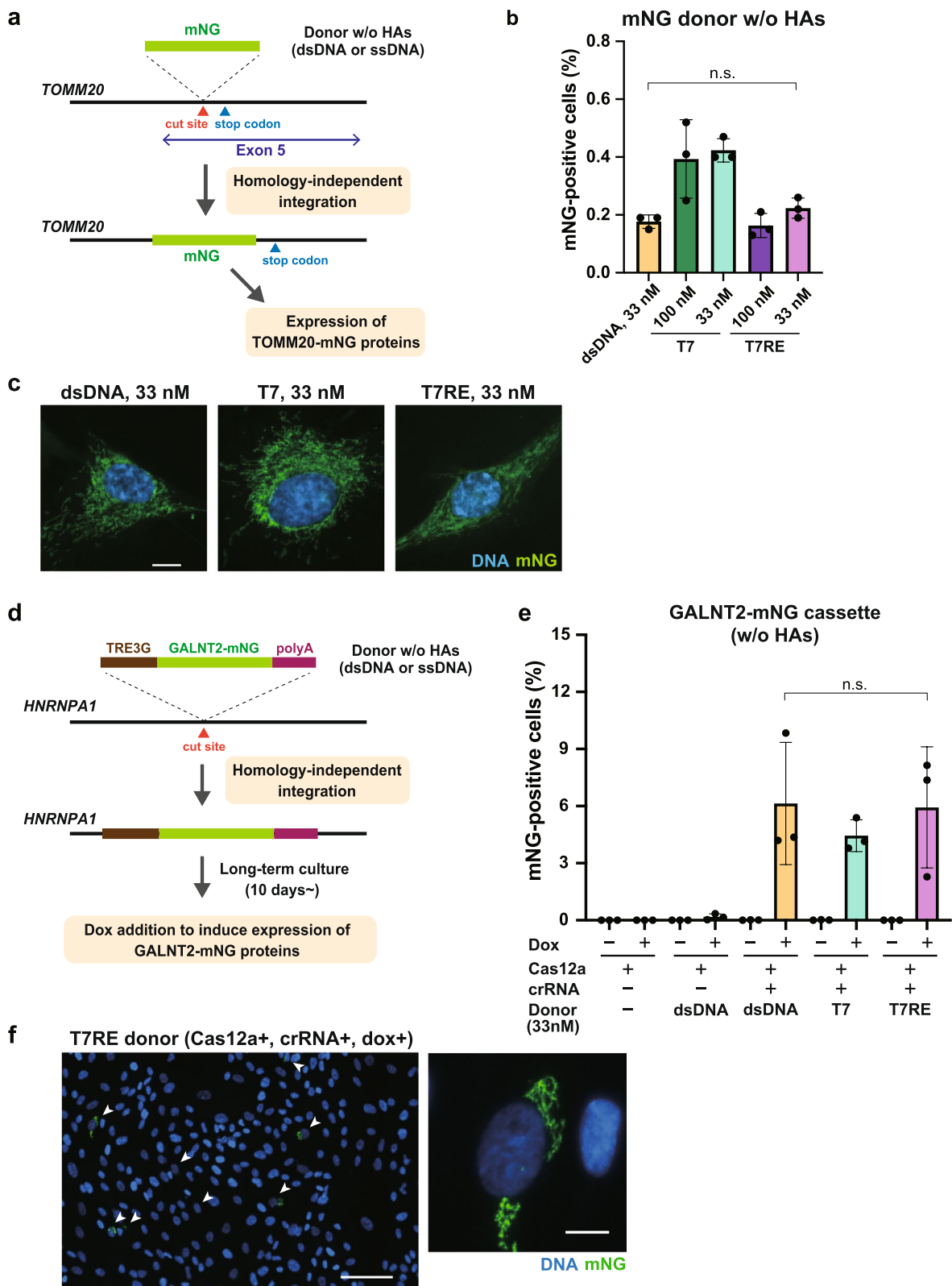


Fig. 5 (See legend on previous page.)

and exonuclease I and column-purified. T7 exonuclease (0.3 U/ μ L, New England Biolabs) was mixed with the purified dsDNA (60 ng/ μ L) in rCutSmart buffer and incubated at 25 °C for 30 min. The reaction mix was directly used for electrophoresis using 2% agarose gel supplemented with Midori Green Advance (Nippon Genetics) at 4 °C for 40 to 60 min to check ssDNA production. Gels were imaged using ChemiDoc XRS+ (Bio-Rad Laboratories). Plot profiles of each lane were generated using FIJI distribution in the ImageJ (NIH) software.

Validation of ssDNA production using exonuclease I

After T7 exonuclease reaction as described above, the reaction mix was column-purified. Exonuclease I (0.6 U/ μ L, New England Biolabs) was mixed with the purified DNA (20 ng/ μ L) in rCutSmart buffer and incubated at 37 °C for 5 min. The reaction mix was directly used for electrophoresis using 2% agarose gel.

Evaluation of the strand selectivity of ssDNA production

The T7 exonuclease reaction products were denatured at 95 °C for 1 min. The products were then annealed by decreasing temperature by 5 °C every 30 s until it reached 25 °C. The annealed products were then subjected to gel electrophoresis to visualize the change in band intensity of dsDNA and ssDNA.

ssDNA donor preparation with the T7 or T7RE method

dsDNA was prepared with two-step PCR. The first-round PCR was performed using non-modified primers followed by treatment with DpnI and exonuclease I, as described above. The specific PCR product was gel purified and then used as a template for the second-round PCR with two short primers (about 25 nt), one of which contains five sequential PS bonds at the 5' end. After column purification, the dsDNA was reacted with T7 exonuclease as described above. For the preparation of T7RE donors, HpyCH4III (0.025 U/ μ L), Hpy188I (0.1 U/ μ L), NlaIII (0.025 U/ μ L), and RsaI (0.05 U/ μ L) purchased from New England Biolabs were added directly to the T7 exonuclease reaction mix and incubated at 37 °C for 15 min. After the enzymatic reactions, ssDNA was column-purified using Buffer NTC (Macherey–Nagel) as a binding buffer. Typically, 4 to 5 μ g of ssDNA was obtained from 15 μ g of dsDNA, when elution with 15 μ L of nuclease-free water was conducted twice.

Synthesis of guide RNA

Guide RNA (sgRNA for Cas9 and crRNA for Cas12a) was transcribed in vitro from PCR-generated DNA templates according to a previous method [32]. Briefly, for sgRNA, template DNA containing T7 promoter and sgRNA

sequence was amplified by PCR from five different oligos. Template DNA for crRNA was likewise assembled by PCR from two different oligos. The purified DNA template was subjected to in vitro transcription by T7 RNA polymerase using the HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs). After being treated with DNase I (Takara Bio), the synthesized guide RNA was purified using the RNA Clean & Concentrator Kit (Zymo Research). All target site sequences of guide RNA used in this study are listed in Supplementary Table 2.

Gene knock-in using CRISPR-Cas12a and CRISPR-Cas9 system

Endogenous gene tagging using the CRISPR-Cas12a system was performed with the electroporation of Cas12a-RNP and HDR donors (dsDNA, T7, or T7RE donors) using the Neon Transfection System (Thermo Fisher Scientific) according to the manufacturer's protocol. A.s. Cas12a Ultra (1 μ M) from Integrated DNA Technologies (IDT) and crRNA (1 μ M) were pre-incubated in resuspension buffer R (Thermo Fisher Scientific) at room temperature and mixed with cells (0.125×10^5 / μ L), Cpf1 electroporation enhancer (1.8 μ M, IDT), and the HDR donors (33 nM). Electroporation was conducted using a 10 μ L Neon tip at a voltage of 1300 V with two 20 ms pulses for RPE1 cells, and 1200 V with one 40 ms pulse for HCT116 cells. The transfected cells were seeded into a 24-well plate.

CRISPR-Cas9-mediated knock-in was performed similarly to the Cas12a-RNP condition described above, with a modification in the electroporation solution. Briefly, HiFi Cas9 protein (1.55 μ M, IDT) and sgRNA (1.84 μ M) were pre-incubated in buffer R and mixed with cells, Cas9 electroporation enhancer (1.8 μ M, IDT), and the HDR donors. Electroporation was conducted at a voltage of 1300 V with two 20 ms pulses for RPE1 cells, and 1400 V with two 20 ms pulses for HCT116 cells.

Quantification of knock-in efficiency by flow cytometry

Flow cytometric analysis was conducted 3 to 12 days after electroporation. Cells were harvested with trypsin/EDTA solution and suspended in DMEM/F-12 medium with HEPES and without phenol red. The cell suspensions were analyzed using BD FACS Aria III (BD Biosciences), equipped with 355/405/488/561/633 nm lasers to detect cells with mNG signal. Data were collected from more than 5,000 gated events.

Amplicon sequencing and analysis by knock-knock Genomic DNA preparation

After electroporation of Cas12a-RNP targeting the *HNRNPA1* locus and HDR donors, cells were expanded for 17 days. mNG-positive cells were sorted using BD

FACS Aria III and seeded into a 96-well plate. Cells were expanded until confluent and genomic DNA was extracted using DNAzol Direct (Molecular Research Center).

Amplicon sequencing

Amplicon libraries were prepared with two-step PCR and subsequent adapter ligation, according to the protocol provided by Pacific Biosciences (Part Number 101-791-800 Version 02 (April 2020)) with slight modifications. The first-round PCR was conducted to amplify a region flanking the target site of mNG insertion from extracted genome DNA. For the amplification, KOD One Master Mix (TOYOBO) was used with primers tailed with universal sequences which serve as an annealing site for a barcoded primer. The amplified DNA was purified using AMPure XP (Beckman Coulter). The purified DNA was re-amplified by PCR using primers from Barcoded Universal F/R Primers Plate-96v2 (Pacific Biosciences) and subsequently purified with AMPure PB beads (Pacific Biosciences). The barcoded amplicons were then analyzed by TapeStation (Agilent Technologies) and Qubit Fluorometer (Thermo Fisher Scientific). All the amplicons were pooled as one sample in equimolar amounts. A pooled sequencing library was prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). One Sequel II SMRT cell was run on the PacBio Sequel II Platform with Binding Kit 2.0/Sequencing Kit 2.0 and 24 h movies, yielding a total of 7,031,124 polymerase reads (328,085,638,677 bp). The consensus reads (1,572,695 HiFi reads with $QV \geq 40$) were generated from the raw full-pass subreads using the PacBio CCS program (SMRT Link v10.2.1.143962) and then 1,319,631 barcoded reads were selected after demultiplexing.

Analysis of knock-in outcomes by knock-knock

Before analysis of knock-in outcomes, the universal primer sequences at both ends were trimmed from the reads. We then analyzed these trimmed reads with *knock-knock*, a computational pipeline developed by Canaj et al. (2019). The source code is available at <https://github.com/jeffhussmann/knock-knock>.

Analysis of homology-independent integration using GALNT2-mNG cassettes

The TRE3G-GALNT2 (1–114 aa)-5xGA-mNG-BGH polyA sequence was amplified by PCR from pRetroX-TRE3G-GALNT2-mNG-polyA plasmid for the preparation of donor cassettes, using primers not having HA sequences. The PCR products were subjected to the preparation of dsDNA and ssDNA donors as described

above. The purified DNA cassettes (33 nM) were electroporated into RPE1-Tet3G cells with Cas12a-RNP targeting the *HNRNPA1* locus. Electroporated cells were cultured for more than 10 days to remove the non-integrated cassettes. The cells at 13 days and 20 days after electroporation were treated with doxycycline (1 $\mu\text{g}/\text{mL}$, Merck) for 24 h and subjected to microscopic and flow cytometric analyses, respectively.

Genomic PCR

Genomic DNA was purified using the NucleoSpin DNA RapidLyse kit (Macherey–Nagel). The knock-in region was amplified by PCR using primers and KOD One PCR Master Mix and the reaction mix was then subjected to agarose gel electrophoresis.

Immunofluorescence

For indirect immunofluorescence, cells cultured on coverslips (Matsunami Glass) were fixed with 4% PFA in PBS at room temperature for 15 min. Fixed cells were blocked with blocking buffer (1% bovine serum albumin in PBS containing 0.05% Triton X-100) for 30 min at room temperature. The cells were then incubated with primary antibodies in the blocking buffer at room temperature for 1 h in a humid chamber. After washing with PBS, the cells were incubated with secondary antibodies in the blocking buffer at room temperature for 30 min. The coverslips were washed with PBS and mounted onto glass slides (Matsunami Glass) using ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific), with the cell side down. The images were acquired using Axio Imager.M2 microscope (Carl Zeiss) equipped with a 63 \times lens objective.

Western blotting

Cells were lysed on ice in lysis buffer containing 50 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1% NP-40, 0.5% Sodium deoxycholate, 0.1% SDS, 5 mM EDTA, 15 mM MgCl_2 , 1:1,000 protease inhibitor cocktail (Nacalai Tesque), and 1:1,000 phosphatase inhibitor cocktail (Nacalai Tesque). After centrifugation, the supernatant mixed with Laemmli sample buffer was boiled and subjected to SDS-PAGE. Separated proteins were transferred onto Immobilon-P PVDF membrane (Merck) using Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (Bio-Rad Laboratories). The membrane was blocked with 2% skim milk in PBS containing 0.02% Tween-20 and probed with the primary antibodies, followed by incubation with their respective HRP-conjugated secondary antibodies. The membrane was soaked with Chemi-Lumi One L or Chemi-Lumi One Super (Nacalai Tesque) for signal detection using ChemiDoc XRS+.

Antibodies

The following primary antibodies were used in this study: anti-TOMM20 (Santa Cruz Biotechnology; sc-17764, IF 1:1,000), anti-mNG (Chromotek, 32f6; IF 1:500), anti-GM130 (Cell Signaling Technology, #12480; IF 1:1,000), anti-HNRNPA1 (Santa Cruz Biotechnology, sc-32301; WB 1:500), anti-mNG (Cell Signaling Technology, #53061, WB 1:100), and anti-HSP90 (BD Biosciences, 610419; WB 1:5,000). The following secondary antibodies were used: donkey anti-mouse IgG Alexa Fluor 555 (Invitrogen, A32773; IF 1:500), donkey anti-mouse IgG Alexa Fluor 647 (Invitrogen, A32787; IF 1:500), donkey anti-rabbit IgG Alexa Fluor 647 (Invitrogen, A32795; IF 1:500), anti-mouse IgG HRP (Promega, WB 1:10,000), and anti-rabbit IgG HRP (Promega, WB 1:10,000).

Statistical analysis

Statistical comparison between the data from different groups was performed in PRISM v.9 software (Graph-Pad) using either the Tukey–Kramer test or a two-tailed, unpaired Student's t-test as indicated in the figure legends. P -value < 0.05 was considered statistically significant. All data shown are mean \pm S.D. Sample sizes are indicated in the figure legends.

Abbreviations

DSB	Double-strand break
dsDNA	Double-stranded DNA
HA	Homology arm
HDR	Homology-directed repair
mNG	mNeonGreen
NHEJ	Non-homologous end joining
PS	Phosphorothioate
RE	Restriction enzymes
ssDNA	Single-stranded DNA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09377-3>.

Additional file 1: Figure S1. dsDNA-based endogenous tagging with mNG in HCT116 cells. **Figure S2.** Validation of the ssDNA production methods. **Figure S3.** Comparison of Cas12a-mediated knock-in efficiency between dsDNA and ssDNA long donors in HCT116 cells. **Figure S4.** Analysis of HDR directionality of ssDNA donors using long-read sequencing and *knock-knock*. **Figure S5.** Homology-independent integration using mNG donors without HAs. **Figure S6.** Uncropped images of gels and blots for Figs. 1, 2, S2, and S5.

Additional file 2: Table 1. Primer sequences for PCR. **Table 2.** Target site sequences of guide RNA.

Acknowledgements

We thank Miho Kiyooka and Wei Chen at National Institute of Genetics for supporting NGS sequencing, Dr. Yusuke Kishi at Institute for Quantitative Biosciences at the University of Tokyo for supporting quality control of NGS library preparation, and the Kitagawa lab members and Dr. Elmar Schiebel from ZMBH in Heidelberg University for technical supports and helpful discussions.

Authors' contributions

S.H. conceived and designed the study. A.M. designed and performed most of the experiments. M.G. optimized the genome editing conditions. C.T. performed tagging of CAMSAP2 and validated knock-in specificity. K.K.I. and A.M. analyzed the PacBio data with *knock-knock*. M.H. performed endogenous tagging of p53. T.K. contributed to quantification by flow cytometry. M.F. and T.C. provided suggestions. A.T. performed PacBio sequencing. A.M., S.H. and D.K. analyzed the data. A.M., S.H. and D.K. wrote the manuscript. All authors contributed to discussions and manuscript preparation. The authors read and approved the final manuscript.

Funding

This work was supported by JSPS KAKENHI grants (Grant numbers: 18K06246, 19H05651, 20K15987, 20K22701, 21H02623, 22H02629, 22K19305, 22K19370) from the Ministry of Education, Science, Sports and Culture of Japan, the PRESTO program (JPMJPR21EC) of the Japan Science and Technology Agency, Takeda Science Foundation, The Uehara Memorial Foundation, The Research Foundation for Pharmaceutical Sciences, Koyanagi Foundation, The Kanai Foundation for the Promotion of Medical Science, Kato Memorial Bioscience Foundation, Tokyo Foundation for Pharmaceutical Sciences, The Naito Foundation, Mochida Memorial Foundation for Medical and Pharmaceutical Research, and The Sumitomo Foundation.

Availability of data and materials

All data supporting the findings of this study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, The University of Tokyo, Bunkyo, Tokyo, Japan. ²Precursory Research for Embryonic Science and Technology (PRESTO) Program, Japan Science and Technology Agency, Honcho Kawaguchi, Saitama, Japan. ³Zentrum Für Molekulare Biologie, Universität Heidelberg, DKFZ-ZMBH Allianz, Heidelberg, Germany. ⁴Comparative Genomics Laboratory and Advanced Genomics Center, National Institute of Genetics, Mishima, Shizuoka, Japan.

Received: 15 November 2022 Accepted: 14 May 2023

Published online: 29 May 2023

References

- Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157:1262–78.
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
- Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A*. 2012;109:2579–86.
- Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. 2015;163:759–71.
- Chang HHY, Pannunzio NR, Adachi N, Lieber MR. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol*. 2017;18:495–506.
- Yeh CD, Richardson CD, Corn JE. Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol*. 2019;21:1468–78.

7. Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods*. 2011;8:753–7.
8. Zhang X, Li T, Ou J, Huang J, Liang P. Homology-based repair induced by CRISPR-Cas nucleases in mammalian embryo genome editing. *Protein Cell*. 2022;13:316–35.
9. Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, et al. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell*. 2013;13:659–62.
10. Li H, Beckman KA, Pessino V, Huang B, Weissman JS, Leonetti MD. Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv*. 2017. <https://doi.org/10.1101/178905>.
11. Paix A, Folkmann A, Goldman DH, Kulaga H, Grzelak MJ, Rasoloson D, et al. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc Natl Acad Sci U S A*. 2017;114:E10745–54.
12. Maggio I, Gonçalves MAFV. Genome editing at the crossroads of delivery, specificity, and fidelity. *Trends Biotechnol*. 2015;33:280–91.
13. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*. 2015;33:187–98.
14. Zelensky AN, Schimmel J, Kool H, Kanaar R, Tijsterman M. Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA. *Nat Commun*. 2017;8:1–7.
15. Fueller J, Herbst K, Meurer M, Gubicza K, Kurtulmus B, Knopf JD, et al. CRISPR-Cas12a-assisted PCR tagging of mammalian genes. *J Cell Biol*. 2020;219:e201910210.
16. Roberts B, Haupt A, Tucker A, Grancharova T, Arakaki J, Fuqua MA, et al. Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol Biol Cell*. 2017;28:2854–74.
17. Renaud JB, Boix C, Charpentier M, De Cian A, Cochenne J, Duvernois-Berthet E, et al. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep*. 2016;14:2263–72.
18. Canaj H, Hussmann JA, Li H, Beckman KA, Goodrich L, Cho NH, et al. Deep profiling reveals substantial heterogeneity of integration outcomes in CRISPR knock-in experiments. *bioRxiv*. 2019. <https://doi.org/10.1101/841098>.
19. Roth TL, Puig-Saus C, Yu R, Shifrut E, Carnevale J, Li PJ, et al. Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature*. 2018;559:405–9.
20. Ghetti S, Burigotto M, Mattivi A, Magnani G, Casini A, Bianchi A, et al. CRISPR/Cas9 ribonucleoprotein-mediated knockin generation in hTERT-RPE1 cells. *STAR Protoc*. 2021;2:100407.
21. Nikiforov TT, Rendle RB, Kotewicz ML, Rogers YH. The use of phosphorothioate primers and exonuclease hydrolysis for the preparation of single-stranded PCR products and their detection by solid-phase hybridization. *PCR Methods Appl*. 1994;3:285–91.
22. Noteborn WEM, Abendstein L, Sharp TH. One-Pot Synthesis of Defined-Length ssDNA for Multiscaffold DNA Origami. *Bioconjug Chem*. 2021;32:94–8.
23. Hecker KH, Rill RL. Error analysis of chemically synthesized polynucleotides. *Biotechniques*. 1998;24:256–60.
24. Ellington A, Pollard JD. Synthesis and Purification of Oligonucleotides. *Curr Protoc Mol Biol*. 1998;2:11.1–25.
25. Miura H, Gurumurthy CB, Sato T, Sato M, Ohtsuka M. CRISPR/Cas9-based generation of knockdown mice by intronic insertion of artificial microRNA using longer single-stranded DNA. *Sci Rep*. 2015;1:2799.
26. Quadros RM, Miura H, Harms DW, Akatsuka H, Sato T, Aida T, et al. Easi-CRISPR: A robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol*. 2017;18:1–15.
27. Nakayama T, Grainger RM, Cha SW. Simple embryo injection of long single-stranded donor templates with the CRISPR/Cas9 system leads to homology-directed repair in *Xenopus tropicalis* and *Xenopus laevis*. *Genesis*. 2020;58:6.
28. Bai H, Liu L, An K, Lu X, Harrison M, Zhao Y, et al. CRISPR/Cas9-mediated precise genome modification by a long ssDNA template in zebrafish. *BMC Genomics*. 2020;21:1–12.
29. Ranawakage DC, Okada K, Sugio K, Kawaguchi Y, Kuninobu-Bonkohara Y, Takada T, et al. Efficient CRISPR-Cas9-Mediated Knock-In of Composite Tags in Zebrafish Using Long ssDNA as a Donor. *Front Cell Dev Biol*. 2021;8:598634.
30. Feng S, Sekine S, Pessino V, Li H, Leonetti MD, Huang B. Improved split fluorescent proteins for endogenous protein labeling. *Nat Commun*. 2017;8:370.
31. Saito S, Maeda R, Adachi N. Dual loss of human POLQ and LIG4 abolishes random integration. *Nat Commun*. 2017;8:16112.
32. Komori T, Hata S, Mabuchi A, Genova M, Harada T, Fukuyama M, et al. A CRISPR-del-based pipeline for complete gene knockout in human diploid cells. *J Cell Sci*. 2023;136:6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

