Correction of substitution, deletion, and insertion mutations by 5'-tailed duplexes

Hidehiko Kawai, Kento Sato, Taiki Kato, and Hiroyuki Kamiya*

Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553, Japan

* Corresponding author.

E-mail address: hirokam@hiroshima-u.ac.jp (H. Kamiya).

Short title: Gene correction by short 5'-tailed duplexes

(**Abstract)**

Germline and somatic mutations cause various diseases, including cancer. Clinical applications of genome editing are keenly anticipated, since it can cure genetic diseases. Recently, we reported that a 5'-tailed duplex (TD), consisting of an approximately 80-base editor strand oligodeoxyribonucleotide and a 35-base assistant strand oligodeoxyribonucleotide, could edit a target gene on plasmid DNA and correct a single-base substitution mutation without an artificial nuclease in human cells. In this study, we assessed the ability of the TD to correct base substitution mutations located consecutively or separately, and deletion and insertion mutations. A TD with an 80-base editor strand was co-introduced into human U2OS cells with plasmid DNA bearing either a wild-type or mutated copepod green fluorescent protein (*copGFP*) gene. Among the mutations, three-base consecutive substitutions were efficiently repaired. The correction efficiencies of deletion mutations were similar to those of substitution mutations, and two to three times higher than those of insertion mutations. Up to three-base substitution, deletion, and insertion mutations were excellent targets for correction by TDs. These results suggested that the TDs are useful for editing disease-causing genes with small mutations.

Key words: Gene correction; Gene editing; 5'-Tailed duplex; *copGFP* gene; Base substitution; Deletion; Insertion

(INTRODUCTION)

Mutations are formed by various factors, such as chemical modifications of DNA (1-4). Germline and somatic mutations cause numerous diseases, including cancer, and are possibly involved in aging and neurodegeneration. In addition to the conventional low molecular weight drugs, new drug modalities such as oligonucleotide therapeutics, proteolysis-targeting chimeras (PROTACs), and gene therapy products have been proposed and tested to treat these types of diseases (5-7).

Genome editing, the rewriting of genetic information, is expected to be applied as another type of treatment for diseases caused by mutations. Unlike gene therapy using non-integrating viruses such as recombinant adeno-associated virus, genome editing therapy is considered to be permanent. Moreover, the mutated gene is not expressed in the cells where both alleles are correctly edited. Currently, the major genome editing methods utilize artificial nucleases to cleave the DNA near the target site (8-12). The donor DNA containing the normal (or desired) sequence that is co-introduced into the target cells can correct the gene by homology-dependent repair. However, these artificial nucleases do not have extremely high sequence specificity and the intracellular availability of the donor nucleic acid is insufficient. Consequently, unintended on- and off-target mutations are observed (13,14). Thus, more precise and safer editing methods are required for the clinical applications of this technology.

Previously, we prepared several hundred-base single-stranded (ss) DNA and 5'-tailed duplex (TD) DNA as gene-editing tools (15-17). The TD consists of a long ss DNA (editor strand) plus an approximately 35-base oligodeoxyribonucleotide (ODN) hybridized to the 3'-

region of the ss DNA (assistant strand). When plasmid DNA carrying a mutant gene and the ss DNA/TD containing the right sequence are co-introduced into mammalian cells, the base substitution mutation in the gene is corrected. The correction efficiency by the TD is higher than that by the ss DNA alone. However, frameshift mutations are relatively resistant to the correction by these DNAs (18,19).

Recently, we found that TDs with a short, approximately 80-base editor strand corrected a base substitution mutation more efficiently than those with a longer editor strand (20,21). Moreover, higher correction efficiency was observed when using TDs with an antisense editor strand, rather than a sense editor strand. These short editor strands are available by chemical synthesis, and thus TDs may become more suitable for clinical applications.

In this study, we examined the editing abilities of the new type of TDs for various mutations: corrections of multi-base substitution, deletion, and insertion mutations. The novel TDs could repair consecutive (up to three base) substitution/deletion/insertion mutations with efficiencies comparable to that for a single-base substitution mutation.

MATERIALS AND METHODS

Plasmids and plasmid constructions

The TD-mediated gene editing efficiencies in cells were assessed by using copepod green fluorescent protein (copGFP), as described in our previous papers (20,21). The gene

editing target plasmids, which contain both an *mPlum* gene, encoding a far-red fluorescent protein, and a series of green fluorescence-deficient mutant *copGFP* genes linked with the T2A peptide-coding sequence, were used for the assay (Fig. 1A). All target plasmids, except for pHR720PA-1_mPlum_copGFP[Y/H] (21), were reconstructed from the pHR720PA-1_copGFP[Y/H]_mPlum plasmid (20), by assembling the plasmid digested with a restriction enzyme and fragment(s) amplified by PCR, using KOD One PCR Master Mix (Toyobo, Osaka, Japan). Assembly reactions were performed using an NEBuilder HiFi DNA Assembly Cloning kit, according to the manufacturer's protocol (New England Biolabs, Ipswich, MA, USA). The ODNs used for the construction were obtained from Fasmac (Atsugi, Japan) in purified forms.

Plasmids were amplified in *Escherichia coli* HST08 cells (Takara Bio, Kusatsu, Japan), and purified with either a GenElute HP Plasmid DNA Miniprep kit (Sigma–Aldrich, St. Louis, MO, USA) or a NucleoSpin Plasmid Transfection-grade plasmid purification kit (Takara Bio). The sequences of all plasmids were confirmed by Sanger sequencing.

Preparation of TDs

The TDs were prepared with an 80-base editor ODN and a 10-fold molar excess of a 35-base assistant ODN (Table 1 and Fig. 1B), as described previously (20,21). The mixture of both ODNs was heated at 98°C for 5 min, immediately placed on ice, and then heated again at 85°C for 5 min and cooled down at a rate of 1°C/4 s to room temperature.

Cell culture and transfection

U2OS cells were obtained from American Type Culture Collection (Manassas, VA, USA) and cultured in α-modified MEM (Sigma–Aldrich) supplemented with 10% fetal bovine serum and 5 mM glutamine, at 37°C under a 5% CO_2 atmosphere. Cells were prepared at 1×10^5 cells/ml, and seeded at 40 μl per well in a CellCarrier-384 Ultra microplate (PerkinElmer, Waltham, MA, USA). After 18 h, 10 ng (6.25 fmol) of each target plasmid was mixed with a 10-fold molar excess of the prepared TD, and transfected into cells using Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's instructions.

Analysis of gene editing efficiency

The efficiency of gene editing mediated by TDs was analyzed as described in our previous studies (20,21). Briefly, at 48 h after transfection, the cells were stained with 1 μg/ml Hoechst 33342 (Thermo Fisher Scientific) for 30 min and then fixed with 4% paraformaldehyde (Nacalai Tesque, Kyoto, Japan). The three confocal z-stack fluorescence images were acquired using the Opera Phenix High Content Screening System (PerkinElmer). The fluorescence intensities of individual nuclei were analyzed by the Harmony Analysis Software (PerkinElmer) and the Spotfire Software (TIBCO, Palo Alto, CA, USA).

Next-generation sequencing

The co-introduction of plasmid DNA plus TD was performed similarly except that the experimental format was changed to a 6-well plate. The plasmid DNA was extracted from the cells at 48 h after transfection as described previously (15). The DNA was subsequently

116 introduced into *E. coli* DH10B by electroporation and the bacterial cells were incubated on agar
117 plates containing kanamycin at 37°C for 24 h. The colonies (approximately 10^4 colonies) on the
118 plates were suspended in LB medium (10 ml). A 300-µl aliquot was centrifuged, and the plasmid
119 DNA was extracted from the pelleted bacterial cells and purified by a NucleoSpin Plasmid
120 Transfection-grade plasmid purification kit. The DNA was used as the template and a 283-bp
121 DNA fragment including the *copGFP* gene was amplified by PCR using 2× Platinum SuperFi
122 II PCR Master Mix (Thermo Fisher Scientific). The consensus sequences of PCR primers with
123 different N_6-index sequences used for the multiplexed next-generation sequencing (NGS) were
124 5'-TGAACGGCGTGGAGTTCG-3' and 5'-TGAAGCTCACGTGCAGCAC-3' (Table 2). The
125 PCR products were purified by a NucleoSpin Gel and PCR Clean-up purification kit (Takara
126 Bio), and combined in equal proportion. Library preparation and paired-end sequencing with 2×
127 150 bp read length using the MGI DNBSEQ-G400RS platform were performed by Genome-
128 Lead (Takamatsu, Japan). The FASTQ files were generated from each N_6-index sample using
129 Je demultiplex (22). Each generated FASTQ file was used for the following analysis. (i)
130 Sequences containing the letter N or with more than 10% of bases having a PHRED score less
131 than 20 were excluded. (ii) The paired-end reads from the FASTQ data were merged with a
132 minimum overlap of 10 bp and a maximum overlap of 65 bp using FLASH (23). (iii) The
133 merged sequences were aligned to a reference sequence. (iv) The editing efficiency is calculated
134 by dividing the number of desired mutations among the detected mutations by the total number
135 of reads.
136

RESULTS

Correction of consecutive base substitution mutations

First, we examined the correction of the target gene containing consecutive base substitution mutations. The target plasmid carried an *mPlum-T2A-copGFP(mutant)* gene driven by the human EF1α promoter (Figs. 1A and 2A) (21). The mutant genes have one to nine substitution(s) in series and the mutations cause alteration(s) of at least one of the amino acids of the protein chromophore (Fig. 2A). Codons 64-66 (corresponding to codons 56-58 in the original gene) encode the amino acid residues of the chromophore, glycine-tyrosine-glycine. We introduced the TDs carrying an 80-base sense or antisense editor strand with the wild-type sequence, depicted in Fig. 1B, together with plasmid DNA carrying one of the mutant *copGFP* genes, into U2OS cells by lipofection. The target region, corresponding to the mutated bases in the plasmid DNA, was located in the ss portion of both the sense and antisense TDs. The green and red fluorescences were observed at 48 h post-transfection.

As shown in Fig. 2B, some multiple base substitutions were corrected with efficiencies similar to the single-base substitution. For the sense TD, up to three-base substitution mutations were repaired comparably and the five- to nine-base substitutions were corrected less efficiently. Meanwhile, similarly effective repair was observed for up to five-base substitutions in the case of the antisense TD. The correction of the seven-base mutations was moderate and that of the

nine-base mutations was inefficient. Overall, the correction efficiencies by the antisense TD were higher than those by the sense TD.

Correction of deletion and insertion mutations

Next, we examined corrections of deletion and insertion mutations. The deletion of the C base of codon 65 (TAC) results in termination codon formation (TAG, Fig. 3A). The deletions of two to seven bases destroy the chromophore (and except for the three-base deletion, truncate the protein). The addition of an A base as the third letter of codon 65 (the formation of TAA) also stops the translation. Plasmid DNAs containing the gene with one of these mutations were used as the target molecules. Once again, the mutated region in the plasmid DNAs corresponds to the ss portion of the TDs.

The sense and antisense TDs corrected a single-base deletion with approximately 15% efficiencies (Fig. 3B). These values were comparable to those of a single-base substitution (Fig. 2B), indicating that the TDs with an approximately 80-base editor strand have the ability to correct both single-base substitution and deletion mutations with similar efficiencies. This is completely different from the TDs with a long editor strand, which correct a single-base deletion much less efficiently than a single-base substitution (19). Surprisingly, the 80-base TDs seemed to correct the three-base deletion mutation more efficiently than the single-base deletion. The repair of the seven-base deletion was less efficient.

The TDs were less active for editing the insertion mutations than the deletion mutations (Fig. 3B). The repair efficiencies of a single-base insertion were two to three times lower than

178 those of a single-base substitution/deletion. The corrections of the three- and five-base insertion
179 mutations seemed to be comparable to, but slightly less efficient than, those of a single-base
180 insertion mutation.
181
182 **Correction of separated base substitution mutations**
183 Next, we examined the repair of two base substitution mutations separated by 10 or 21
184 bases. In addition to codon 65, codons 57 and 68 are TAC and their substitutions to TAG cause
185 protein truncation. These two codons were independently altered to TAG. Two of the mutant
186 genes have two base substitution mutations: TAG at codon 57 plus CAC (His) at codon 65,
187 located 21 bases apart, and CAC (His) at codon 65 plus TAG at codon 68, located 10 bases apart
188 (Fig. 4A). We expressed mutant genes such as XHY, where X, H, and Y mean TAG, CAC, and
189 TAC at codons 57, 65, and 68, respectively. The recovery of green fluorescence is achieved by
190 the dual corrections of two substitutions, in the cases of the XHY and YHX mutants. The region
191 corresponding to codon 57 in the sense TD and that corresponding to codon 68 in the antisense
192 TD are in the ss portions (Fig. 4B). Since the region corresponding to codon 68 in the sense TD
193 and that corresponding to codon 57 in the antisense TD are within the double-stranded portions,
194 the mutants of these codons were excluded in the analysis.
195 As shown in Fig. 4B, the dual correction efficiencies were lower than the single-base
196 YHY correction efficiencies. However, the repair of the XHY mutant by the sense TD was
197 similar to that of the XYY mutant. For the antisense TD, the YYX and YHX mutants were
198 repaired with equivalent efficiencies. Thus, the correction of separated base substitution

mutations was comparable to that of the single-base substitution mutations with lower efficiency (at more distant positions from the double-stranded region of TDs). Again, the antisense TD more efficiently edited these mutations than the sense TD.

Confirmation of corrections by NGS

We finally examined the corrections (sequence changes) at the DNA level. We chose the plasmid DNAs containing three-base substitution, deletion, and insertion mutations as the correction targets. We co-transfected the target plasmids and the antisense TD and recovered the plasmid DNAs from the cells. The DNAs were then introduced into recombination-deficient (*recA*$^-$) *E. coli* cells to exclude the possible influence of TD that might be recovered from the human cells. The plasmid DNAs isolated from the bacterial cells were used as the templates and the regions including the *copGFP* gene were amplified by high fidelity DNA polymerase. We analyzed the duplicated PCR products by NGS and repeated the series of operations two times.

As shown as the closed bars in Fig. 5, the sequence corresponding to the wild-type gene was detected for the TD experimental groups, indicating the corrections at the DNA level. However, the ratios of the wild-type sequence (the correction efficiencies) were <0.15%, much lower than the correction efficiencies observed by the fluorescence method (Figs. 2B and 3B). Moreover, the ratios were similar for the three-base substitution and insertion groups in contrast to the results obtained by the fluorescence assay (see the Discussion section).

DISCUSSION

In this study, we examined the TD-mediated correction of deletion and insertion mutations. Importantly, the efficiencies of deletion correction were similar to those of substitution correction (Figs. 2B and 3B). The insertion mutation was two to three times less efficiently repaired than the substitution and deletion mutations. Thus, the TD is a potent candidate for gene correction therapeutics. As described above, the correction of frameshift mutations by the TD with a long editor strand is much less efficient as compared to that of substitution mutations (19). This discrepancy probably reflects the difference in the gene correction pathways between TDs with short and long editor strands, although the precise mechanisms remain unknown.

As shown in Figs. 2B and 3B, the TD could correct multi-base substitution and indel (insertion and deletion) mutations. In particular, two- and three-base mutations were repaired with efficiencies comparable to single-base mutations. One typical cystic fibrosis mutation in the Caucasian population is the three-base deletion of the cystic fibrosis transmembrane conductance regulator (CFTR) gene (24). Thus, many disease-causing mutations like this deletion could be targets of the TD.

We confirmed the corrections at the DNA level by NGS (Fig. 5). However, the calculated correction efficiencies were highly different from those determined by the green fluorescence assay. We introduced DNAs (plasmid plus TD) into U2OS cells by lipofection and this method delivers multiple copies of plasmid into a single cell. In the case of three-base deletion, the correction efficiencies obtained by the GFP assay and NGS were approximately

241 25% and 0.13%, respectively (Figs. 3B and 5). This difference might be explained by the positive
242 green fluorescence from one copy of corrected gene in a cell containing 200 copies of
243 uncorrected genes. Thus, the fluorescence assay overestimated the correction efficiencies. Indeed,
244 approximately 20% of green-fluorescent cells were observed when the mixture of 0.14% of
245 wild-type copGFP plasmid and 99.86% of inactive plasmid was transfected (data not shown).

246 In addition, the orders of the correction efficiency were deletion > mismatch > insertion
247 and deletion > mismatch = insertion in the fluorescence assay and NGS, respectively. If plasmid
248 DNA(s) with an unexpected sequence (other than the wild-type sequence) contributed to the
249 green fluorescent cell formation, such sequence should be frequently detected in NGS. However,
250 we did not find such sequences in the TD-treated groups. One possible reason is that biased
251 amplification of the shorter (corrected) DNA molecule over the longer (uncorrected) molecule
252 during PCR in the insertion group. We could not exclude multiple analysis of the PCR products
253 amplified from an identical plasmid. However, only three-base difference does not seem to affect
254 the amplification efficiency.

255 Please note that the open bar in the 3b-Ins (–TD) group represents the ratio of corrected
256 mutations (Fig. 5). We cannot exclude the sequence information derived from the errors during
257 PCR and in sequencing reaction and signal detection at this time. Thus, detailed condition
258 settings would be necessary when the gene correction efficiency is determined by the NGS.

259 In previous studies, we showed that the TDs with an antisense editor strand are more
260 potent than those with a sense editor strand in editing single-base substitutions (20,21). Likewise,
261 in the present study, the correction efficiencies with the antisense TD were generally higher than

those with the sense TD for all types of mutations examined (Figs. 2 and 3). One possible explanation is that the transcription of a target gene enhances the gene editing by TDs. As illustrated in our previous study, the strand invasion/displacement of the editor strand, the hypothesized first step of the editing, might preferentially occur during transcription (21). The antisense editor strand might pair with the sense (non-template) strand of the gene, and this could be a reason for the higher correction efficiencies with antisense TDs.

The TD consists of approximately 80-base editor and 35-base assistant ODNs (Fig. 1B). At first glance, it may resemble the ss ODNs used as gene editing tools. However, these ss ODNs edit DNA in a replication-dependent manner (25-27). Since the plasmid DNAs used in this study lack an origin sequence required for replication in human cells, the detected copGFP correction occurred in a replication-independent manner. Thus, the ss DNAs used by other researchers and the TDs are completely different editing tools.

In this study, we examined TD-mediated gene corrections with indel mutations and multi-base substitutions as the targets. Our results indicate that the TD could be applied to the treatment of diseases caused by relatively small (one- to three-base) mutations. Our final goal is to use TD to correct mutated genes on chromosomes. Elucidations of the gene correction mechanism and the factors affecting the correction will be required to improve the editing efficiency, and experiments toward these goals are in progress in our laboratory. The reasons for the different correction efficiencies of mismatch, deletion, and insertion mutations will be explained after understanding of the correction mechanism.

283

284 **ACKNOWLEDGMENTS**

285 This work was supported in part by the Japan Society for the Promotion of Science
286 (JSPS) KAKENHI grant number JP 17K19491. Part of this study was conducted through the
287 Joint Usage/Research Center Program of the Research Institute for Radiation Biology and
288 Medicine (RIRBM), Hiroshima University.

289

290 **REFERENCES**

291

292 1. **Basu, A. K. and Essigmann, J. M.:** Establishing linkages among DNA damage,
293 mutagenesis, and genetic diseases, Chem. Res. Toxicol., **35**, 1655–1675 (2022).
294 2. **Kawai, H., Iwata, R., Ebi, S., Sugihara, R., Masuda, S., Fujiwara, C., Kimura, S., and**
295 **Kamiya, H.:** Development of a versatile high-throughput mutagenesis assay with
296 multiplexed short read NGS using DNA-barcoded *supF* shuttle vector library amplified in
297 *E. coli*, eLife, **11**, e83780 (2022).
298 3. **Fukushima, R., Suzuki, T., Komatsu, Y., and Kamiya, H.:** Biased distribution of action-
299 at-a-distance mutations by 8-oxo-7,8-dihydroguanine, Mutation Res. (Fundam. Mol. Mech.
300 Mutagen.), **825**, 111794 (2022).
301 4. **Suzuki, T., Y. Zaima, Y. Fujikawa, R. Fukushima, , and Kamiya, H.:** Paradoxical role
302 of the major DNA repair protein, OGG1, in action-at-a-distance mutation induction by 8-
303 oxo-7,8-dihydroguanine, DNA Repair, **111**, 103276 (2022).
304 5. **Yamada, Y.:** Nucleic acid drugs–Current status, issues, and expectations for exosomes.
305 Cancers (Basel), **13**, 5002 (2021).
306 6. **Hu, Z. and Crews, C. M.:** Recent developments in PROTAC-mediated protein
307 degradation: From bench to clinic. ChemBioChem. **23**, e202100270 (2022).
308 7. **Fortunato, F., Farnè, M., and Ferlini, A.:** The DMD gene and therapeutic approaches to
309 restore dystrophin. Neuromuscul. Disord., **31**, 1013–1020 (2021).

8. **Kim, Y.-G., Cha, J., and Chandrasegaran, S.:** Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain, Proc. Natl. Acad. Sci. USA, **93**, 1156–1160 (1996).
9. **Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., and other 10 authors:** A TALE nuclease architecture for efficient genome editing, Nat. Biotechnol., **29**, 143–148 (2011).
10. **Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F.:** Multiplex genome engineering using CRISPR/Cas systems, Science, **339**, 819–823 (2013).
11. **Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M.:** RNA-guided human genome engineering via Cas9, Science, **339**, 823–826 (2013).
12. **Hartenian, E. and Doench, J. G.:** Genetic screens and functional genomics using CRISPR/Cas9 technology, FEBS J., **282**, 1383–1393 (2015).
13. **Kosicki, M., Tomberg, K., and Bradley, A.:** Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements, Nat. Biotechnol., **36**, 765–771 (2018).
14. **Fu, Y., Foden, J. A., Khayter, C., Maeder, M. L., Reyon, D., Joung, J. K., and Sander, J. D.:** High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells, Nat. Biotechnol., **31**, 822–826 (2013).
15. **Tsuchiya, H., Harashima, H., and Kamiya, H.:** Increased SFHR gene correction efficiency with sense single-stranded DNA, J. Gene Med., **7**, 486−493 (2005).

16. **Tsuchiya, H., Uchiyama, M., Hara, K., Nakatsu, Y., Tsuzuki, T., Inoue, H., Harashima, H., and Kamiya, H.:** Improved gene correction efficiency with a tailed duplex DNA fragment, Biochemistry, **47**, 8754−8759 (2008).

17. **Kawai, H., Sato, K., Shirahama, W., Suzuki, T., and Kamiya, H.:** Single-stranded DNA versus tailed duplex in sequence conversion of *lacZα* DNA, Nucleosides Nucleotides Nucleic Acids, **39**, 1245–1250 (2020).

18. **Tsuchiya, H., Sawamura, T., Harashima, H., and Kamiya, H.:** Correction of frameshift mutations with single-stranded and double-stranded DNA fragments prepared from phagemid/plasmid DNAs, Biol. Pharm. Bull., **28**, 1958–1962 (2005).

19. **Morita, Y., Tsuchiya, H., Harashima, H., and Kamiya, H.:** Correction of frameshift mutations with tailed duplex DNAs, Biol. Pharm. Bull., **34**, 1465–1468 (2011).

20. **Kawai, H., Yazama, K., Yanai, Y., Kamitsubo, R., and Kamiya, H.:** Gene correction by 5'-tailed duplexes with short editor oligodeoxyribonucleotides, J. Biosci. Bioengng., **132**, 552–559 (2021).

21. **Kawai, H., Kamitsubo, R., and Kamiya, H.:** Correction of monomeric enhanced green fluorescent protein (*mEGFP*) gene by short 5'-tailed duplexes, *J. Biosci. Bioengng.*, **134**, 175–181 (2022).

22. **Girardot, C., Scholtalbers, J., Sauer, S., Su, S.-Y., and Furlong, E. E. M.:** Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers, *BMC Bioinformatics*, **17**, 419 (2016).

23. **Magoč, T. and Salzberg, S. L.:** FLASH: fast length adjustment of short reads to improve genome assemblies, *Bioinformatics*, **27**, 2957–2963 (2011).

24. **Kerem, B.-S., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., and Tsui, L.-C.:** Identification of the cystic fibrosis gene: Genetic analysis, Science, **245**, 1073–1080 (1989).

25. **Brachman, E. E. and Kmiec, E. B.:** Gene repair in mammalian cells is stimulated by the elongation of S phase and transient stalling of replication forks, DNA Repair (Amst.), **4**, 445–457 (2005).

26. **Olsen, P. A., Randol, M., and Krauss, S.:** Implications of cell cycle progression on functional sequence correction by short single-stranded DNA oligonucleotides, Gene Ther., **12**, 546–551 (2005).

27. **Huen M. S. Y., Li, X., Lu, L.-Y., Watt, R. M., Liu, D.-P., and Huang, J.-D.:** The involvement of replication in single stranded oligonucleotide-mediated gene repair, Nucleic Acids Res., **34**, 6183–6194 (2006).

366 **Figure legends**

367

368 FIG. 1. (A) Schematic map of the target plasmid DNAs used in this study. The plasmid
369 encodes the mPlum protein linked to the copGFP protein (wt or mutant) with the T2A peptide.
370 *kan*^{r} represents the kanamycin-resistance gene. As examples, the base and amino acid sequences
371 around the chromophore of wt and a single-base substitution mutant (Y65H) are also shown. (B)
372 The structures of the sense (S) and antisense (AS) TDs. The 80-base editor strand ss DNAs are
373 hybridized with the 35-base assistant strands.

374

375 FIG. 2. Correction of base substitution mutations by sense (S) and antisense (AS) TDs. (A)
376 The base and amino acid sequences around the chromophore of the target plasmids with one or
377 more base–base mismatches (MM). The Y65H mutant corresponds to 1b-MM. (B) The
378 correction efficiencies, as determined by microscopic observation. The efficiencies were
379 evaluated as described in the MATERIALS AND METHODS section. The bars indicate the
380 mean values of three independent experiments, with two wells per experiment. All error bars
381 represent standard errors of the mean. One-way ANOVA with Tukey's multiple comparison
382 tests was used to compare means between indicated pairs (*$P < 0.05$, **$P < 0.01$).

383

384 FIG. 3. Correction of deletion and insertion mutations by sense (S) and antisense (AS) TDs.
385 (A) The base and amino acid sequences around the chromophore of the target plasmids with
386 deletion(s) (Del) or insertion(s) (Ins). (B) The correction efficiencies determined by microscopic

observation. The bars indicate the mean values of three independent experiments, with two wells per experiment. All error bars represent standard errors of the mean. One-way ANOVA with Tukey's multiple comparison tests was used to compare means between indicated pairs (*$P <$ 0.05).

FIG. 4. Correction of two separated base substitution mutations by sense (S) and antisense (AS) TDs. (A) The base and amino acid sequences around the chromophore of the target plasmids with one or two substitution(s). The Y65H mutant corresponds to YHY. (B) The correction efficiencies determined by microscopic observation. The bars indicate the mean values of three independent experiments, with two wells per experiment. All error bars represent standard errors of the mean.

FIG. 5. NGS analysis of gene correction. The antisense (AS) TD was co-introduced into U2OS cells together with target plasmid with a three-base mismatch (MM), deletion (Del), or insertion (Ins) mutation. The plasmid DNA was recovered from the transfected cells, amplified in *E. coli* cells, and used as the template for PCR. The efficiencies were evaluated as described in the MATERIALS AND METHODS section. The bars indicate the mean values of four experiments. All error bars represent standard errors of the mean.

Table 1 Oligodeoxyribonucleotides used in this study.

Oligodeoxyribonucleotide	Sequence (5'→3')
oligodeoxyribonucleotides for editor strands	
80-base (S) E-strand	ACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCTACGAGAA
80-base (AS) E-strand	TTCTCGTAGCCGCTGGGGTAGGTGCCGAAGTGGTAGAAGCCGTAGCCCATCACGTGGCTCAGCAGGTAGGGGCTGAAGGT
oligodeoxyribonucleotides for assistant strands	
for 80-base (S) A-strand	TTCTCGTAGCCGCTGGGGTAGGTGCCGAAGTGGTA
for 80-base (AS) A-strand	ACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGG

The regions corresponding to codon 65 are underlined.

Table 2 Primers used in the next-generation sequencing assay.

Primer set (with N_6-index)	Forward primer's sequence (5'→3')	Reverse primer's sequence (5→3')
mut_pOnly_1-1	AAAGCGTGAACGGCGTGGAGTTCG	ATTACGTGAAGCTCACGTGCAGCAC
mut_pOnly_2-1	AAACTGTGAACGGCGTGGAGTTCG	ATTCGGTGAAGCTCACGTGCAGCAC
mut_pOnly_1-2	AAATGGTGAACGGCGTGGAGTTCG	ATTGAGTGAAGCTCACGTGCAGCAC
mut_pOnly_2-2	NAAGCAGTGAACGGCGTGGAGTTCG	NATACTGTGAAGCTCACGTGCAGCAC
del_pOnly_1-1	NNAGCAAGTGAACGGCGTGGAGTTCG	NNAACTTGTGAAGCTCACGTGCAGCAC
del_pOnly_2-1	NNACTAAGTGAACGGCGTGGAGTTCG	NNACGTTGTGAAGCTCACGTGCAGCAC
del_pOnly_1-2	NNATGAAGTGAACGGCGTGGAGTTCG	NNAGATTGTGAAGCTCACGTGCAGCAC
del_pOnly_2-2	CAAGCTTGAACGGCGTGGAGTTCG	CTTACTTGAAGCTCACGTGCAGCAC
ins_pOnly_1-1	NCAGCATTGAACGGCGTGGAGTTCG	NCTACTTTGAAGCTCACGTGCAGCAC
ins_pOnly_2-1	NCACTATTGAACGGCGTGGAGTTCG	NCTCGTTTGAAGCTCACGTGCAGCAC
ins_pOnly_1-2	NCATGATTGAACGGCGTGGAGTTCG	NCTGATTTGAAGCTCACGTGCAGCAC
ins_pOnly_2-2	NNCGCAATTGAACGGCGTGGAGTTCG	NNCACTTTTGAAGCTCACGTGCAGCAC
mut_TD_1-1	CGGATTTGAACGGCGTGGAGTTCG	CCCGTTTGAAGCTCACGTGCAGCAC
mut_TD_2-1	CGGTCTTGAACGGCGTGGAGTTCG	CCCAGTTGAAGCTCACGTGCAGCAC
mut_TD_1-2	CGGCATTGAACGGCGTGGAGTTCG	CCCTATTGAAGCTCACGTGCAGCAC
mut_TD_2-2	NCGATGTTGAACGGCGTGGAGTTCG	NCCGTCTTGAAGCTCACGTGCAGCAC
del_TD_1-1	NNCATGGTTGAACGGCGTGGAGTTCG	NNCGTCCTTGAAGCTCACGTGCAGCAC
del_TD_2-1	NNCTCGGTTGAACGGCGTGGAGTTCG	NNCAGCCTTGAAGCTCACGTGCAGCAC

del_TD_1-2	ACCGTGTGAACGGCGTGGAGTTCG	GAAGCCTGAAGCTCACGTGCAGCAC
del_TD_2-2	ACCAGGTGAACGGCGTGGAGTTCG	GAACTCTGAAGCTCACGTGCAGCAC
ins_TD_1-1	NACGTCGTGAACGGCGTGGAGTTCG	NGAGCACTGAAGCTCACGTGCAGCAC
ins_TD_2-1	NACAGCGTGAACGGCGTGGAGTTCG	NGACTACTGAAGCTCACGTGCAGCAC
ins_TD_1-2	NACTACGTGAACGGCGTGGAGTTCG	NGATGACTGAAGCTCACGTGCAGCAC
ins_TD_2-2	NNAGTCCGTGAACGGCGTGGAGTTCG	NNGGCAACTGAAGCTCACGTGCAGCAC

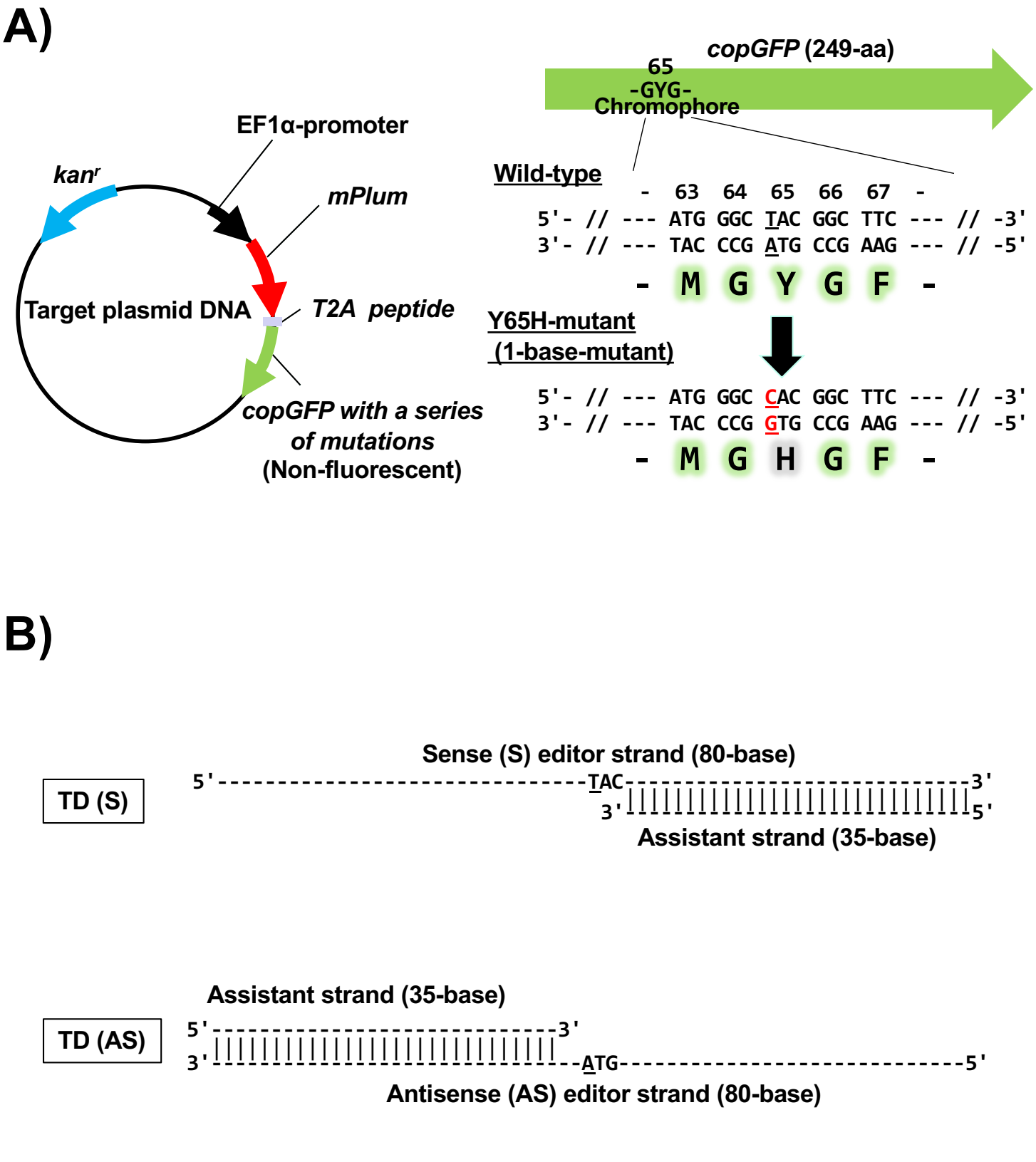
A)
copGFP (249-aa)
65
-GYG-
Chromophore
EF1α-promoter
kanr
mPlum
Target plasmid DNA
T2A peptide
copGFP with a series of mutations
(Non-fluorescent)
Wild-type
- 63 64 65 66 67 -
5'- // --- ATG GGC TAC GGC TTC --- // -3'
3'- // --- TAC CCG ATG CCG AAG --- // -5'
- M G Y G F -
Y65H-mutant
(1-base-mutant)
5'- // --- ATG GGC CAC GGC TTC --- // -3'
3'- // --- TAC CCG GTG CCG AAG --- // -5'
- M G H G F -
B)
Sense (S) editor strand (80-base)
TD (S)
5'----------TAC----------3'
3'||||||||||||||||||||||||||||5'
Assistant strand (35-base)
Assistant strand (35-base)
TD (AS)
5'----------3'
3'||||||||||||||||||||||||||||ATG----------5'
Antisense (AS) editor strand (80-base)

Fig. 1

A)

		63	64	65	66	67	
Wild-type	5'- // ---	ATG	GGC	TAC	GGC	TTC	--- // -3'
	-	M	G	Y	G	F	-
1-base-mutant (1b-MM)	5'- // ---	ATG	GGC	CAC	GGC	TTC	--- // -3'
	-	M	G	H	G	F	-
2-base-mutant (2b-MM)	5'- // ---	ATG	GGC	CTC	GGC	TTC	--- // -3'
	-	M	G	L	G	F	-
3-base-mutant (3b-MM)	5'- // ---	ATG	GGG	CTC	GGC	TTC	--- // -3'
	-	M	G	L	G	F	-
5-base-mutant (5b-MM)	5'- // ---	ATG	GCG	CTG	GGC	TTC	--- // -3'
	-	M	A	L	G	F	-
7-base-mutant (7b-MM)	5'- // ---	ATG	TCG	CTG	AGC	TTC	--- // -3'
	-	M	S	L	S	F	-
9-base-mutant (9b-MM)	5'- // ---	ATC	TCG	CTG	ATC	TTC	--- // -3'
	-	I	S	L	I	F	-

B)

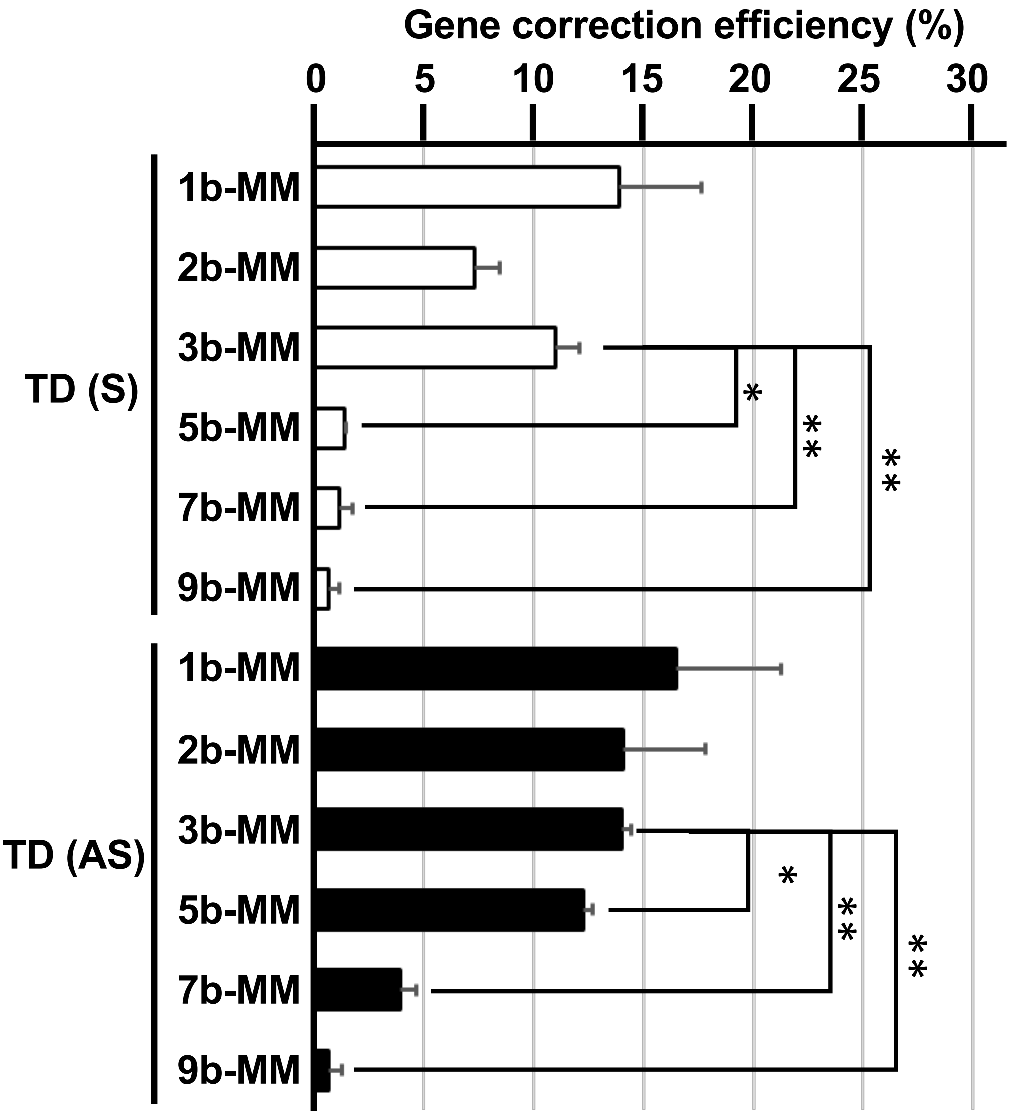

Fig. 2

A)

	- 63 64 65 66 67 -
Wild-type	5'- // --- ATG GGC TAC GGC TTC --- // -3' - M G Y G F -
1-base-deletion (1b-Del)	5'- // --- ATG GGC TA-GGCTTC --- // -3' - M G X - -
2-base-deletion (2b-Del)	5'- // --- ATG GGC T--GG CTT C-- - // -3' - M G W L L -
3-base-deletion (3b-Del)	5'- // --- ATG GGC --- GGC TTC --- // -3' - M G - G F -
5-base-deletion (5b-Del)	5'- // --- ATG G-----GG CTT C --- // -3' - M G V L -
7-base-deletion (7b-Del)	5'- // --- AT-------G GCT TC- --- // -3' - M A S -
1-base-insertion (1b-Ins)	5'- // --- ATG GGC TAA CGGCTTC-- // -3' - M G X - -
3-base-insertion (3b-Ins)	5'- // --- ATG GGC TAA TACGCTTC-- // -3' - M G X - -
5-base-insertion (5b-Ins)	5'- // --- ATG GGC TAA TATACGCTTC // -3' - M G X - -

B)

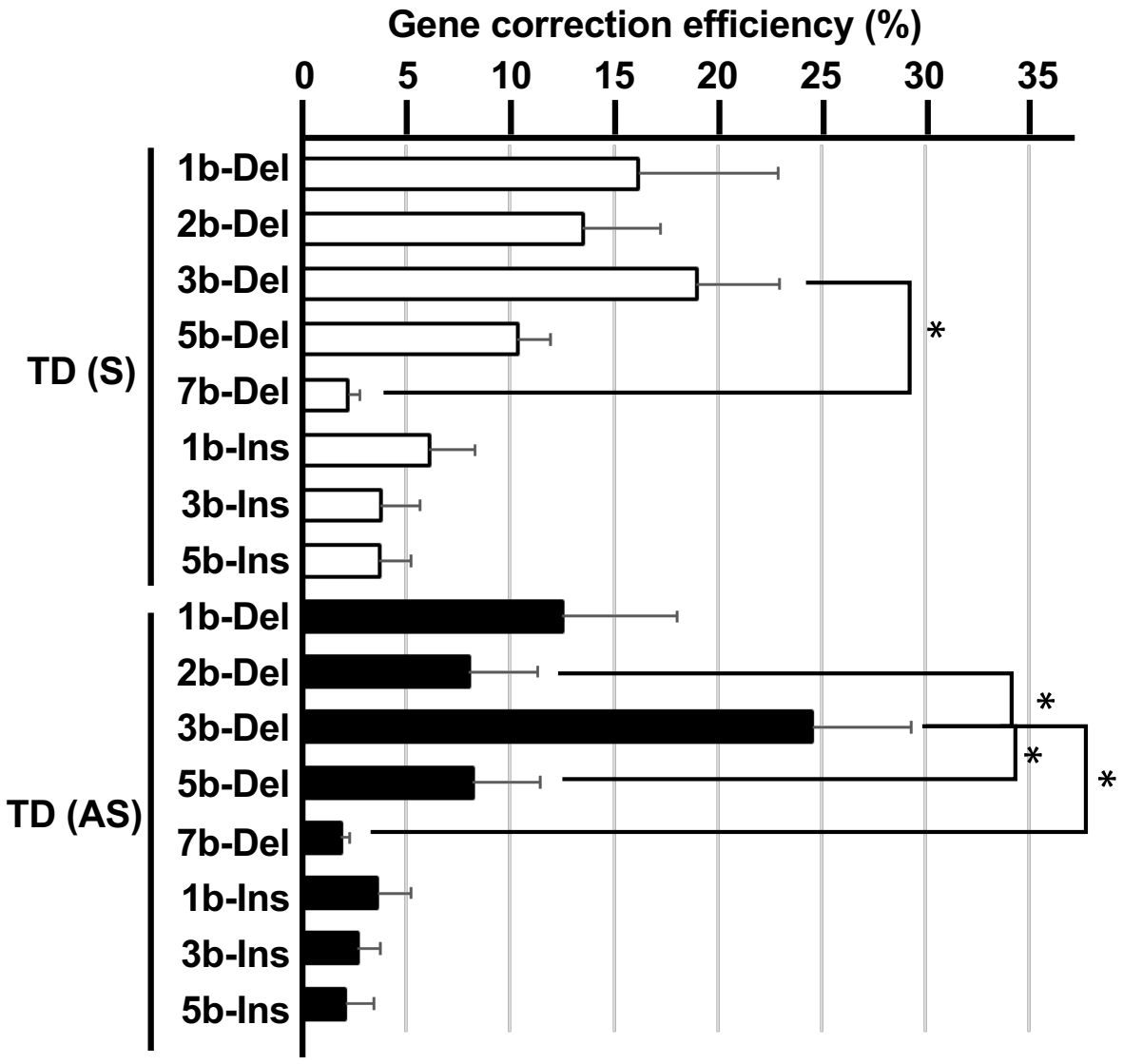

Fig. 3

A)

		57	58	59	60	61	62	63	64	65	66	67	68	
Wild-type (YYY)	5'- // ---TAC	CTG	CTG	AGC	CAC	GTG	ATG	GGC	TAC	GGC	TTC	TAC	--- // -3'	
	-	Y	L	L	S	H	V	M	G	Y	G	F	Y	-
1-base-mutant (YHY)	5'- // ---TAC	CTG	CTG	AGC	CAC	GTG	ATG	GGC	CAC	GGC	TTC	TAC	--- // -3'	
	-	Y	L	L	S	H	V	M	G	H	G	F	Y	-
1-base-mutant (XYY)	5'- // ---TAG	CTG	CTG	AGC	CAC	GTG	ATG	GGC	TAC	GGC	TTC	TAC	--- // -3'	
	-	X	L	L	S	H	V	M	G	Y	G	F	Y	-
2-base-mutant (XHY)	5'- // ---TAG	CTG	CTG	AGC	CAC	GTG	ATG	GGC	CAC	GGC	TTC	TAC	--- // -3'	
	-	X	L	L	S	H	V	M	G	H	G	F	Y	-
1-base-mutant (YYX)	5'- // ---TAC	CTG	CTG	AGC	CAC	GTG	ATG	GGC	TAC	GGC	TTC	TAG	--- // -3'	
	-	Y	L	L	S	H	V	M	G	Y	G	F	X	-
2-base-mutant (YHX)	5'- // ---TAC	CTG	CTG	AGC	CAC	GTG	ATG	GGC	CAC	GGC	TTC	TAG	--- // -3'	
	-	Y	L	L	S	H	V	M	G	H	G	F	X	-

B)

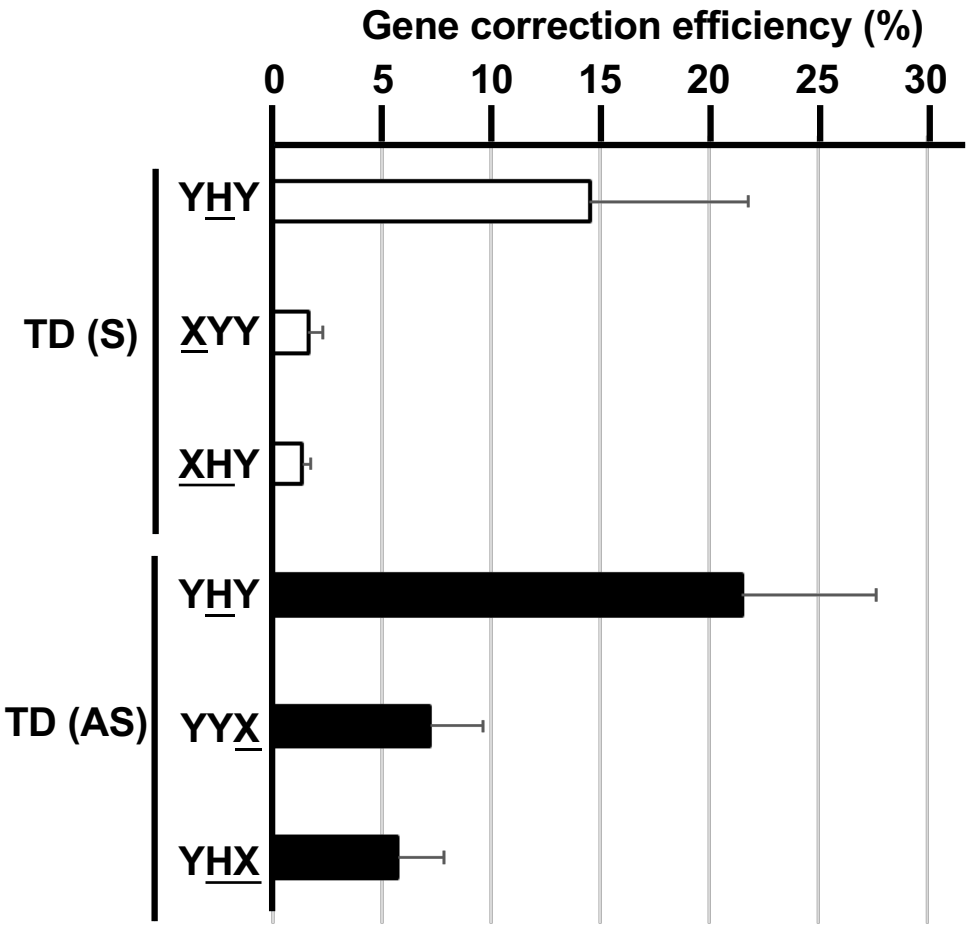

Fig. 4

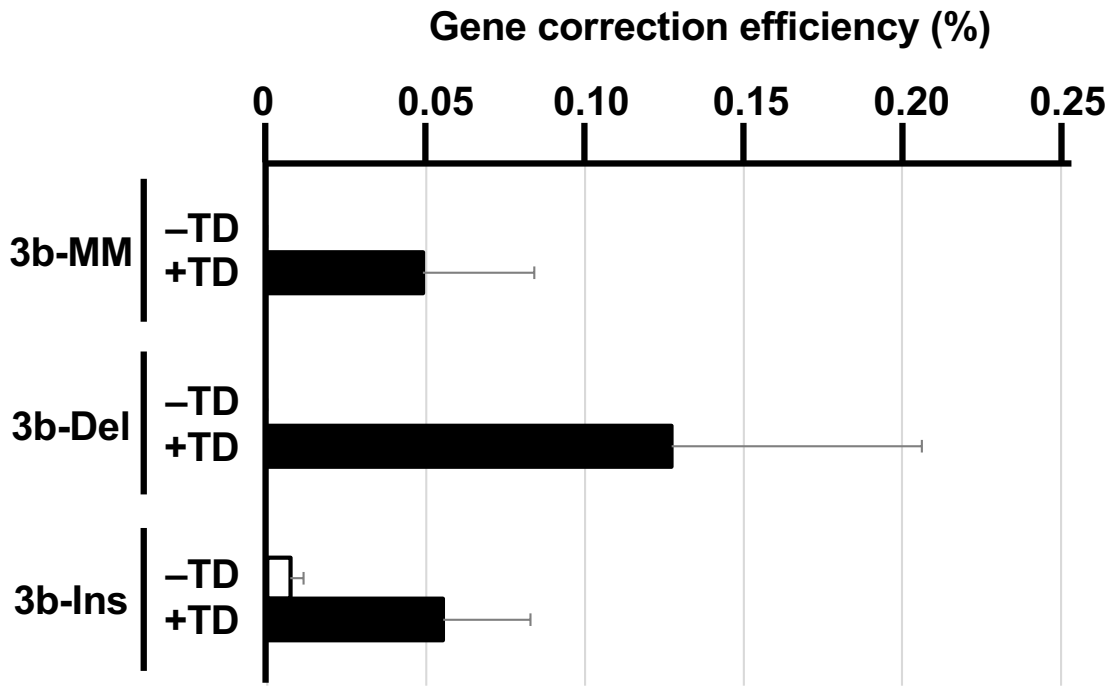
Gene correction efficiency (%)
0
0.05
0.10
0.15
0.20
0.25
3b-MM
–TD
+TD
3b-Del
–TD
+TD
3b-Ins
–TD
+TD

Fig. 5