nature biotechnology

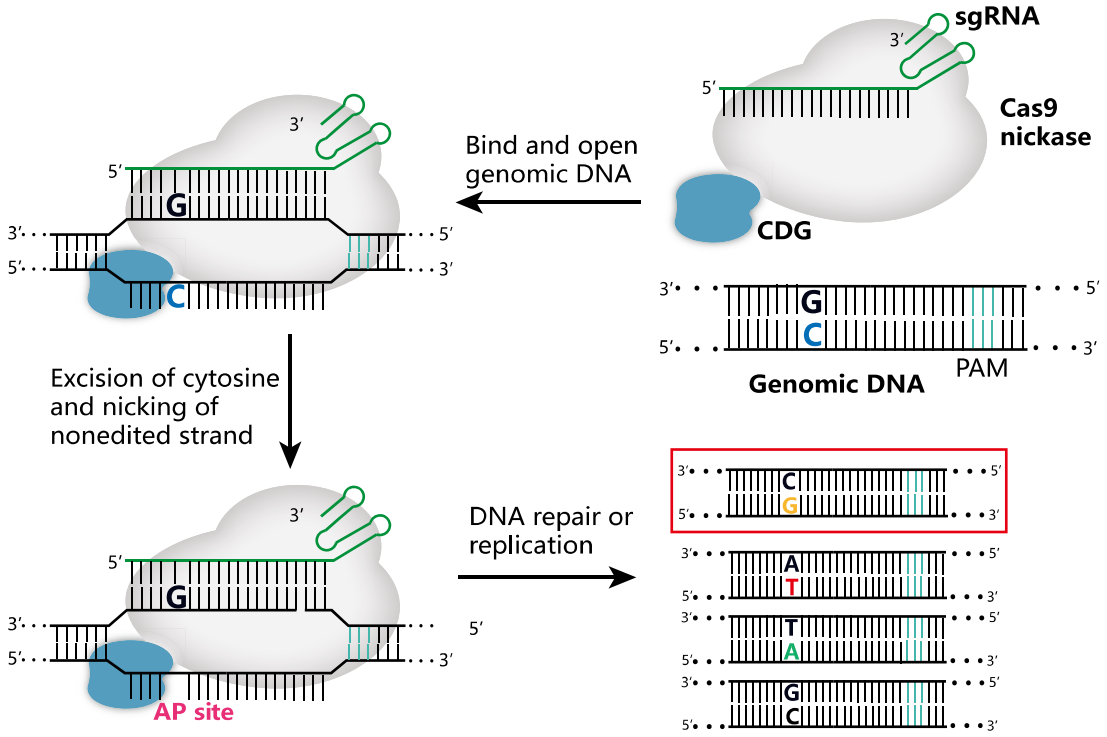
Article

https://doi.org/10.1038/s41587-023-02050-w

# Glycosylase-based base editors for efficient T-to-G and C-to-G editing in mammalian cells

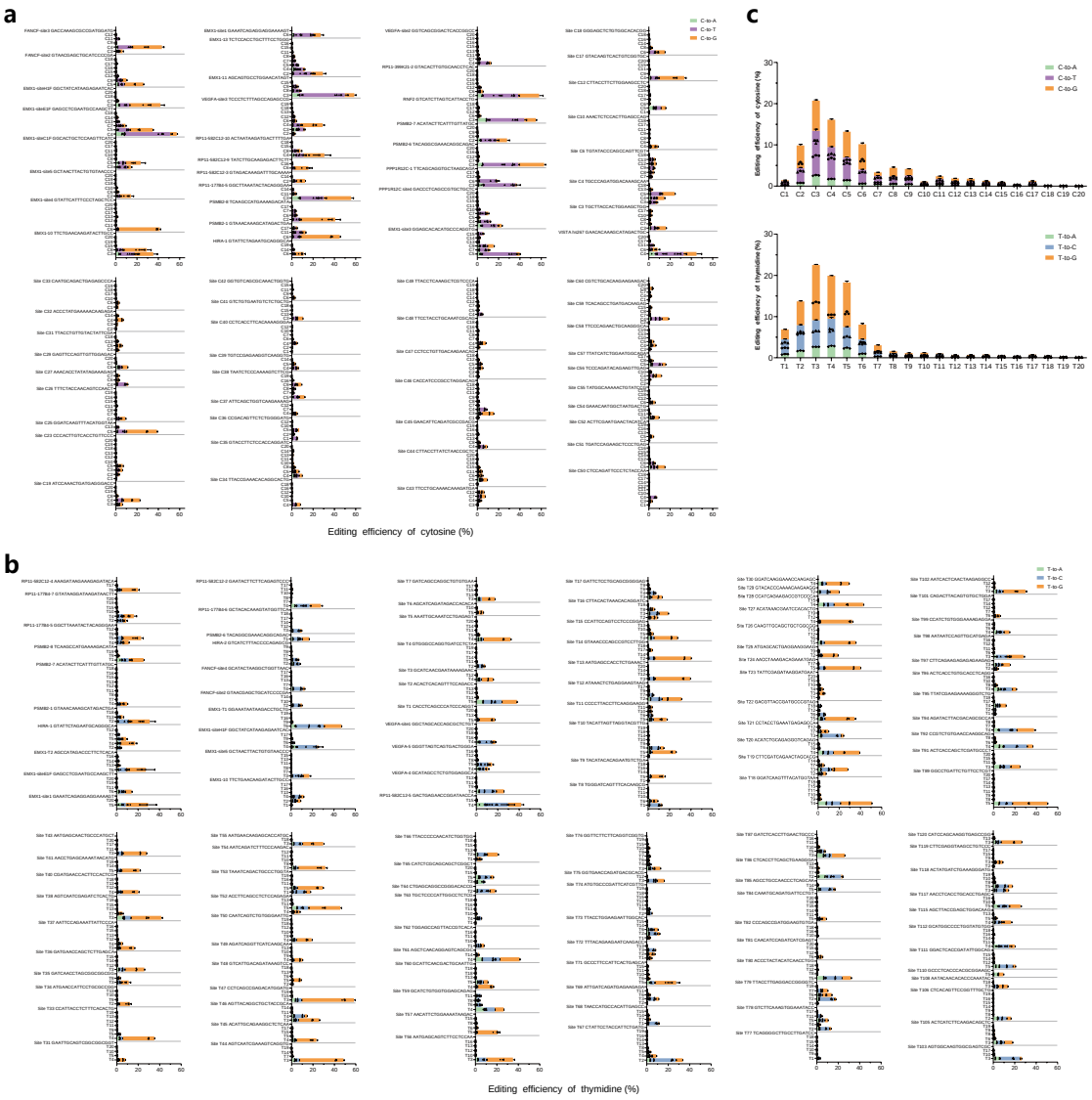
In the format provided by the authors and unedited

# Supplementary Information



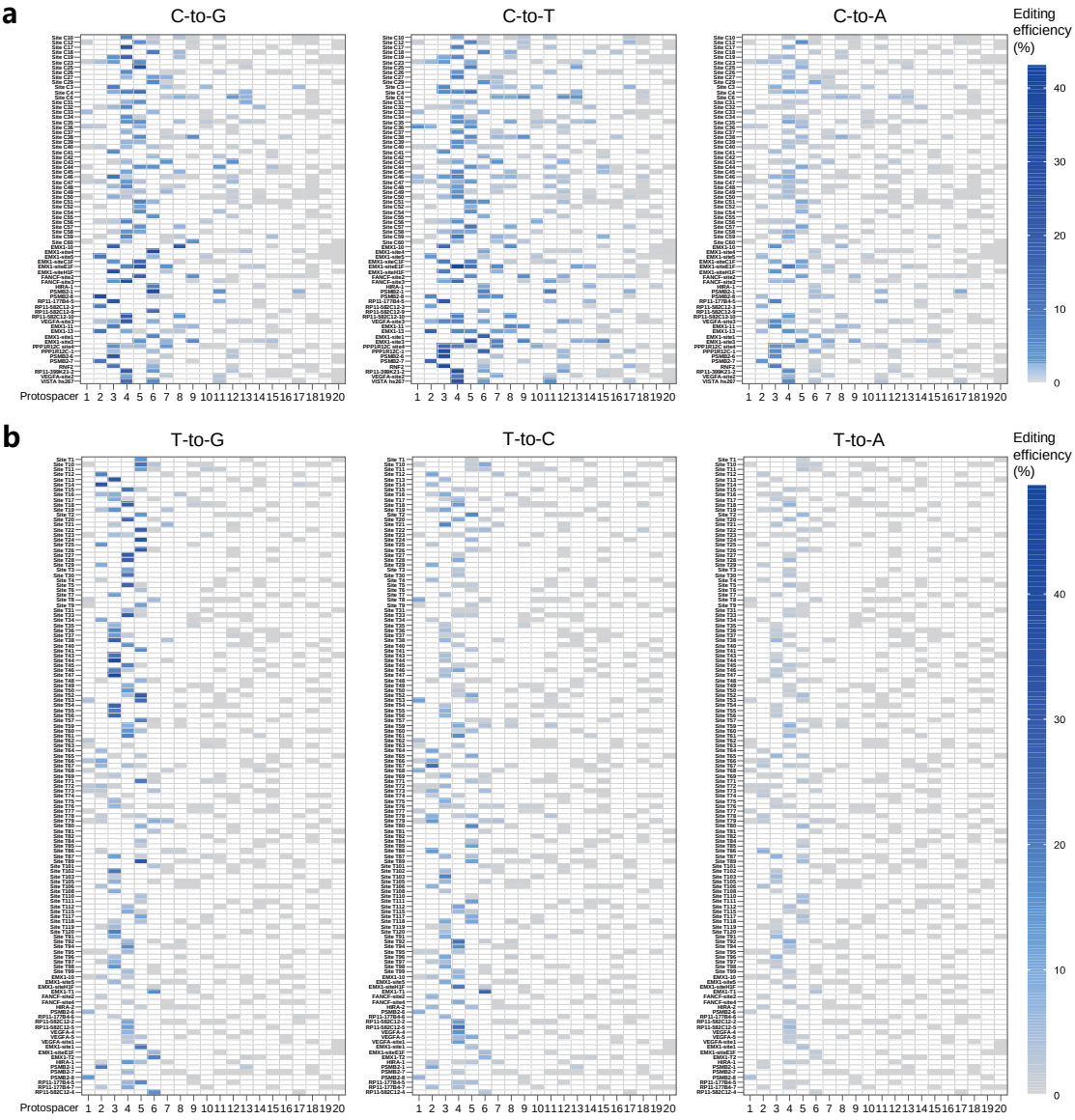

**Supplementary Figure 1. CDG-Cas9 mediated base editing process.**

The CDG-nCas9 complex specifically binds to target genomic DNA loci. CDG directly excises cytosine (C) to create an apurinic/apyrimidinic (AP) site, and nCas9 cleaves the target strand. Finally, after DNA repair or replication, C can be converted to various bases.



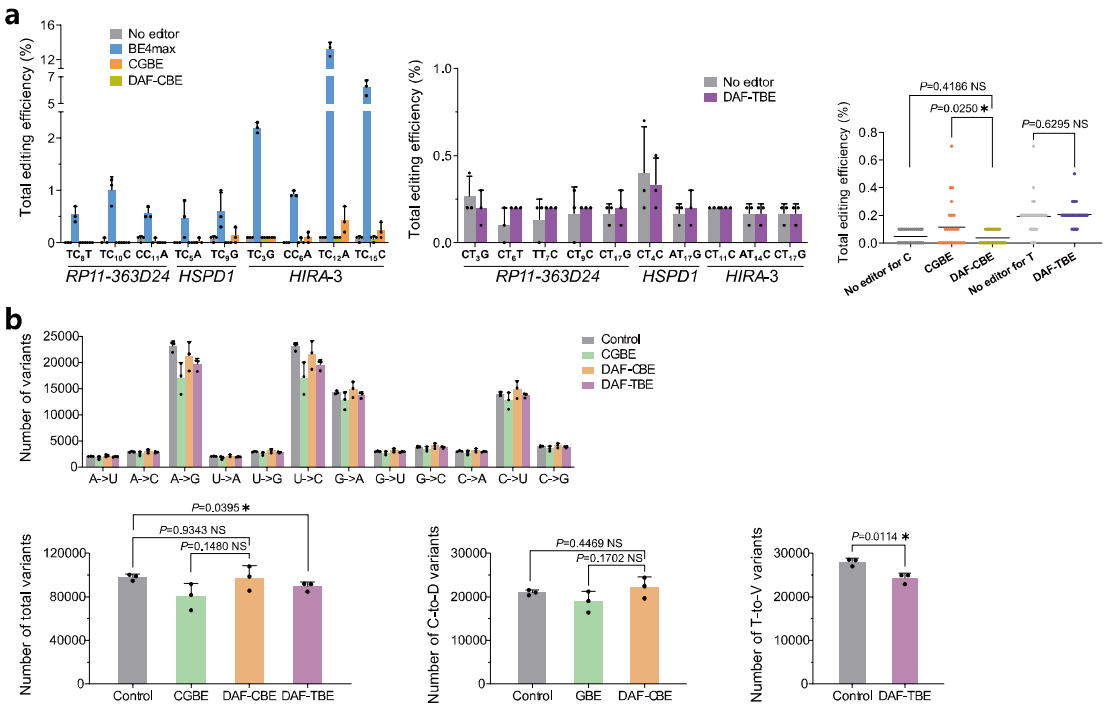

**Supplementary Figure 2. The editing results of DAF-CBE and TBE at each locus in HEK293T cells.**

**a**. Cytosine editing outcomes of DAF-CBE at the protospacers from 70 endogenous genomic loci in HEK293T cells. **b**. Thymidine editing outcomes of DAF-TBE at the protospacers from 127 endogenous genomic loci in HEK293T cells. **c**. The editing window of DAF-CBE and DAF-TBE. The columns represent the average efficiency of cytosine or thymidine at 20 positions of protospacers from 70 and 127 loci in **a** and **b**. Dot, triangle, and diamond symbols represent C-to-G/T-to-G, C-to-T/T-to-C and C-to-A/T-to-A editing of an individual replicate, respectively. Data represent the mean ± s.d. of $n = 3$ independent replicates.



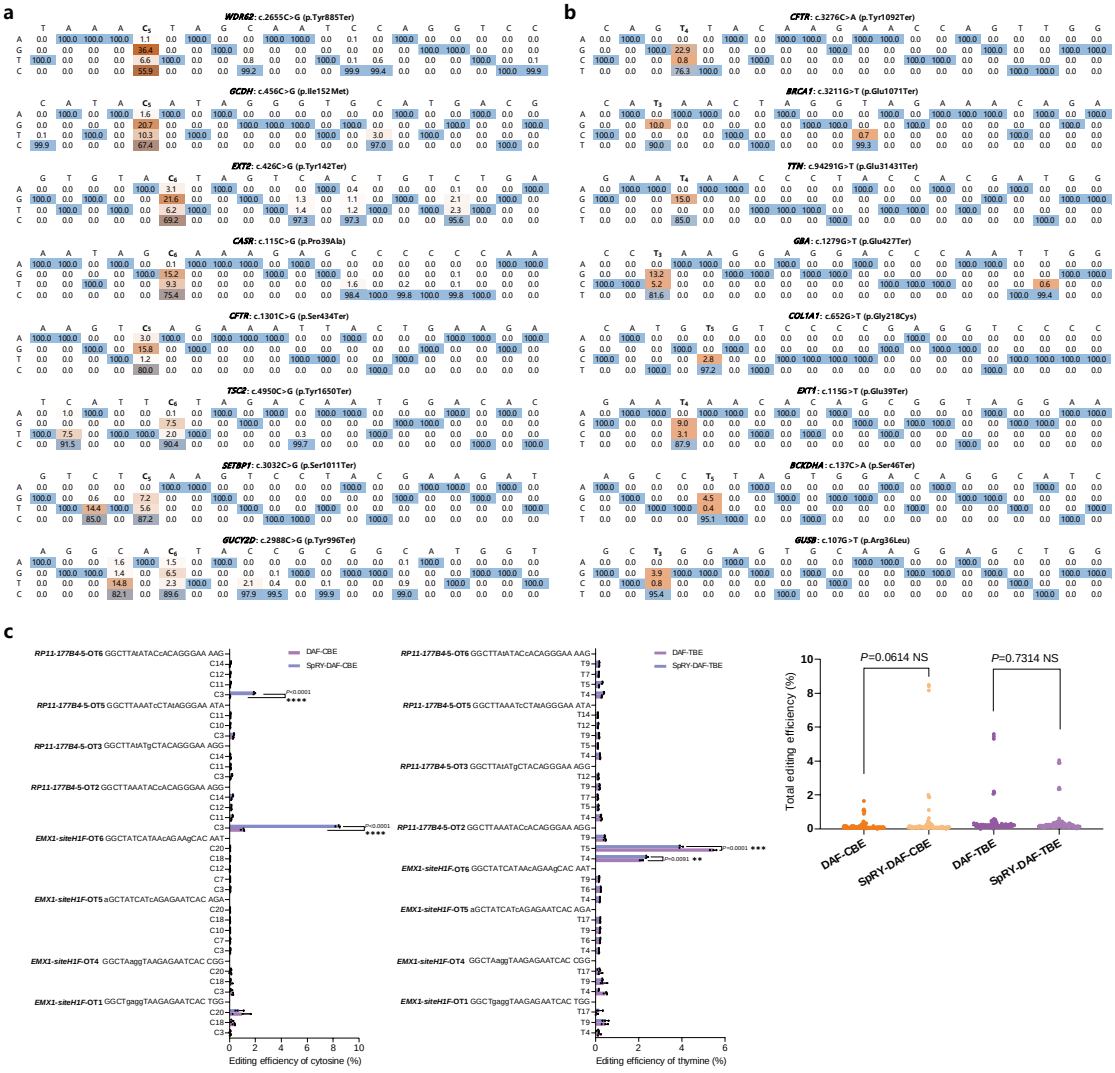

**Supplementary Figure 3. The editing efficiency and position of DAF-CBE and TBE at each locus in HEK293T cells.**

**a**. The color-coded cells in the heatmap represent the C-to-G, C-to-T and C-to-A editing of DAF-CBE across the protospacers (PAM is at positions 21–23) from the 70 endogenous genomic loci in HEK293T cells. **b**. The color-coded cells in the heatmap represent the T-to-G, T-to-C and T-to-A editing of DAF-TBE across the protospacers from the 127 endogenous genomic loci in HEK293T cells. Each cell in the heatmaps represents the mean of three independent biological replicates. The editing loci we randomly selected for DAF-CBE/ DAF-TBE editing were designated SiteCx/ SiteTx or gene names.



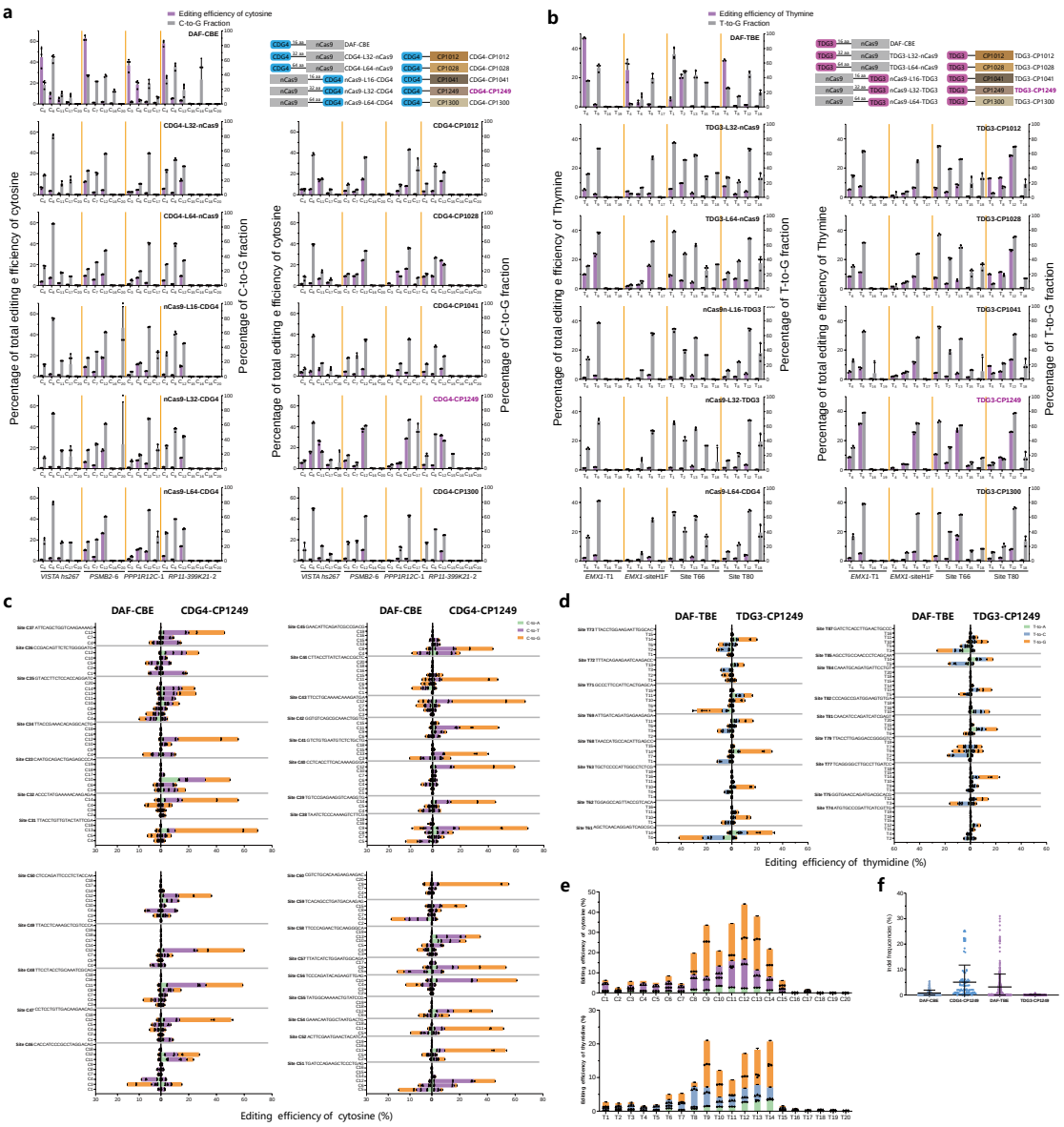

**Supplementary Figure 4. Evaluation of off-Target effects of DAF-CBE and DAF-TBE.**

**a**. Cas9-independent off-target editing (CIOE) detected by the orthogonal R-loop assay at each R-loop site induced by DAF-CBE and DAF-TBE in HEK293T cells. **b**. RNA off-target editing induced by DAF-CBE and DAF-TBE. *P* values were calculated by the Paired two-sided *t*-test, and $P < 0.05$ was considered significant. NS $P > 0.05$; *$P < 0.05$. The bars represent the average values, and the error bars represent the SD of $n = 3$ independent biological replicates.



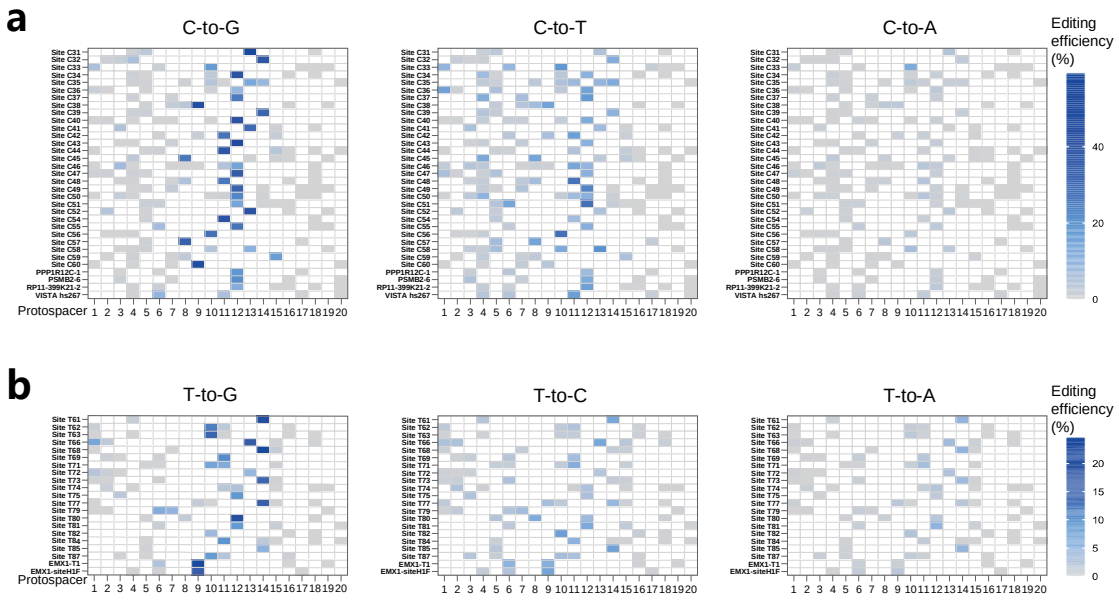

**Supplementary Figure 5. Correction of disease-relevant SNVs using DAF-CBE and TBE.**

**a**. The editing results of 8 disease-relevant SNVs using DAF-CBE. **b**. The editing results of 8 disease-relevant SNVs using DAF-TBE. **c**. Cas9-dependent off-target DNA editing in HEK293T cells induced by DAF-CBE, TBE compared with SPRY- DAF-CBE, TBE. In **a** and **b**, the target base is bolded and has a subscript indicating its position number. In **c**, mismatched nucleotides in off-target sequences are indicated in lowercase letters. Data represent the mean ± s.d. of $n = 3$ independent replicates. $P$ values were calculated by the Paired two-sided $t$-test, and $P < 0.05$ was considered significant. NS $P > 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.




**Supplementary Figure 6. Editing activities of DAF-CBE and TBE variants in HEK293T cells.** **a**. The editing efficiency and purity of reconstructed DAF-CBEs. **b**. The editing efficiency and purity of reconstructed DAF-TBEs. The composition diagram of DAF-CBEs /DAF-TBEs is shown in the upper right corner. The bars represent the editing efficiency and the T-to-C/T-to-G fractions of all cytosines in protospacers. **c**. Comparison of cytosine editing outcomes of DAF-CBE and CDG4-CP1249 at the protospacers from 29 endogenous genomic loci in HEK293T cells, excluding the four genomic loci shown in Fig. 6a. **d**. Comparison of thymidine editing outcomes of DAF-TBE and TDG3-CP1249 at the protospacers from 17 endogenous genomic loci in HEK293T cells, excluding the four genomic loci shown in Fig. 6b. **e**. The editing window of CDG4-CP1249 and TDG3-CP1249. The columns represent the average efficiency of cytosine or thymidine at 20

positions of protospacers from 33 and 21 loci. Dot, triangle, and diamond symbols represent C-to-G/T-to-G, C-to-T/T-to-C, and C-to-A/T-to-A editing outcomes of an individual replicate in c, d and e, respectively. The data presented are the mean ± s.d. of $n = 3$ independent replicates. **f.** The indel frequencies of DAF-CBE, CDG4-CP1249, DAF-TBE and TDG3-CP1249 at 70, 33, 127 and 21 genomic loci.




**Supplementary Figure 7. Editing activities of DAF-CBE and TBE variants in HEK293T cells.** **a**. The color-coded cells in the heatmap represent the C-to-G, C-to-T and C-to-A editing of CDG4-CP1249 (DAF-CBE2) across the protospacers (PAM is at positions 21–23) from the 33 endogenous genomic loci in HEK293T cells. **b**. The color-coded cells in the heatmap represent the T-to-G, T-to-C and T-to-A editing of TDG3-CP1249 (DAF-TBE2) across the protospacers from the 21 endogenous genomic loci in HEK293T cells. Each cell in the heatmaps represents the mean of three independent biological replicates.

**Supplementary Table 1. The protospacer sequences of gRNA expression plasmids in *E. coli***

| Plasmids | Protospacer sequences |
|---|---|
| pTrpB96stop | GCAACGCCTGCCCCAGCA CC TGG |
| pTrpB230stop | ACACAGGCGATAACGGCA TC CGG |
| pTrpB96reverse | GTCAGGCCTGCCCCAGCA CC TGG |
| pTrpB230reverse | ACTCAGGCGATAACGGCA TC CGG |

pTrpB96stop and pTrpB230stop were used for construction of *E. coli* (TrpB Leu96*/Cys230*); pTrpB96reverse and pTrpB230reverse were used for correction of stop mutations in *E. coli* (TrpB Leu96*/Cys230*).

**Supplementary Table 2. Amino acid mutations of CDG variants from 25 rounds evolution**

| Rounds | Variants | CDG1 Amino acid | | | | | | | | | | | | | | | | | | | | | | | | | Linker | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | E | G | G | V | T | T | K | G | P | K | T | D | R | Q | S | Y | G | A | R | T | V | Y | K | D | K | E | S | E | T | T |
| | | 85 | 87 | 95 | 107 | 109 | 127 | 129 | 135 | 149 | 150 | 175 | 179 | 183 | 220 | 235 | 237 | 248 | 253 | 255 | 258 | 266 | 274 | 275 | 297 | 300 | 302 | 303 | 1 | 4 | 5 | 13 |
| 1 | Variant-1 | | | | | | | | | | | | | | | | | | | | | | | | | N | | | | | | |
| 1 | Variant-2 | | | | | | | | | | | | | G | | | | | | | | | | | | | | | | | | |
| 3 | Variant-3 | | | | | | | | | | | | | | | | | | C | | | | A | | | | | | | | | |
| 3 | Variant-4 | | | | | | | | | | | | | | | | | | | | | | A | | | | | | | | | |
| 5 | Variant-5 | | | | | I | | | | | | | | | | | | | | | | A | A | | | | | | | | | |
| 5 | Variant-6 | | | | | | | | | | | | | | Q | | | C | | | | A | A | | | | | | | | | |
| 5 | Variant-7 | | | | | | | | | | | | | G | | K | | | | T | | | A | | | | | | | | | |
| 7 | Variant-8 | | G | | R | | | | | | | | | | | | | | | | | | A | | | | | | | | | |
| 7 | Variant-9 | | | | | | | | | | | | | | | | | | | | | | A | | | | | V | | | | |
| 7 | Variant-10 | | | | | | | | | | | | | | | | | C | | | | | A | | | | | | | | | |
| 9 | Variant-11 | | | | | | A | | | | | | | | | | | | | | | | A | | | | | | | | | |
| 9 | Variant-12 | | | | | | | | | | | | | G | | | L | C | | | | | A | | | | | | | | | |
| 9 | Variant-13 | | | | E | | | | | | | | | | | | | | | | | | A | | | | | | G | | | |
| 11 | Variant-14 | | | | | | | | | | | | | | | | | | | | | A | A | | | | | | G | | | |
| 11 | Variant-15 | | | | | | | | | | | | | | | | L | | | | | | A | | | | | | G | | A | |
| 13 | Variant-16 | | | | | | | | | | | | | | | | | C | | | | A | A | | | | | | | | | |
| 13 | Variant-17 | | | | E | | | | | | | | | | | | | | | | | | A | | | | | | | | | |
| 15 | Variant-18 | | | | E | | | | | | L | E | | | | | | | | | | | A | | | | | | | | | |
| 15 | Variant-19 (**CDG2**) | | | | E | | | | | | | | | | | | | C | | | | A | A | | | | | | | | | |
| 17 | Variant-20 | | | | | | | | | | | | | G | | | L | C | | | | A | | C | | | | | | | | |
| 17 | Variant-21 (**CDG3**) | L | | | E | | | | | | | | | | Q | | L | C | | | | A | A | | | | | | G | | | |
| 19 | Variant-22 | | | | R | | | | | | | | | G | | | | | | | | | A | | | | | V | | | | |
| 19 | Variant-23 | L | | | | | | | | | | | | G | | | | C | | | | A | A | | | | | | | | | |
| 21 | Variant-24 | L | | V | E | | | | | | | E | | | | | L | C | | | | A | A | | | | | | | | | |
| 21 | Variant-25 | | | | E | | | | | | | | | | | | L | H | | | | A | A | | | | | | | | | |
| 23 | Variant-26 (**CDG4**) | | | | E | | | | | | | E | | G | Q | | | H | | | | A | A | H | | T | | G | | | | |
| 23 | Variant-27 (**CDG5**) | | | | E | | | S | N | E | | E | A | | | | L | C | | | G | A | A | H | E | | V | | | | | |
| 23 | Variant-28 | L | | | E | | | | | | | E | | | | | | C | | | | A | A | | | N | | | K | | | |
| 25 | Variant-29 (**CDG6**) | | | | E | | | | | | | E | | | Q | | | C | | | | A | A | | | | | | | | | |
| 25 | Variant-30 | | | | E | | A | | | | | E | | G | | | L | C | | | | A | A | | | | | | C | | | A |

Conserved amino acids are highlighted. The five variants shown in Fig. 2a were bolded. The mutations of linker were list behind CDG.

**Supplementary Table 3. Amino acid mutations of TDG variants from 33 rounds evolution**

| Rounds | Variants | TDG1 Amino acid | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | Linker | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | E | K | S | G | K | T | F | T | D | L | Q | R | V | K | Q | Q | N | S | N | A | K | T | A | V | Y | R | K | D | E | S | T | A | S | M |
| | | 85 | 87 | 90 | 94 | 107 | 114 | 117 | 126 | 129 | 136 | 142 | 161 | 162 | 164 | 175 | 224 | 235 | 236 | 237 | 238 | 255 | 259 | 266 | 267 | 274 | 275 | 276 | 297 | 300 | 303 | 1 | 5 | 12 | 16 | 17 |
| 1 | Variant-1 | L | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Variant-2 | | | | | | | | Y | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 1 | Variant-3 | | | | | | | | | | | | | | | | | | | | | | | | | | | | E | | | | | | | |
| 2 | Variant-4 | | | | | | | | | | | | R | | | | | | | | | | | | S | | | | | | | | | | | |
| 2 | Variant-5 | | | | | | | | | | | | | | | | | L | | | | | | | | | | | | | | | | | | |
| 2 | Variant-6 | | | | | | | | | | | | | | | | | | S | | | | | | | | | | | | | | | | | |
| 4 | Variant-7 | | | E | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Variant-8 | | | | | | | | | | | | R | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | Variant-9 | | | | | | | | | | | | | G | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Variant-10 | | Q | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 6 | Variant-11 | | | | | | | A | | | | | | | | | | | | | | | | | | A | | | | | | | | | | |
| 6 | Variant-12 | | | | | | | | | | | | | | | | | | | L | | | | | | | | | | | | | | | | |
| 8 | Variant-13 | | | | | | | | | | | V | | | | | | | | | | | | | | | | | | | | | | | | |
| 8 | Variant-14 | | | | | | | | | | | V | | | | | | | | | | | | | | | | | | | | G | | | | |
| 8 | Variant-15 | | | | | R | | | | | | V | | | | | | | | | | | | | | A | | | | | | | | | | |
| 9 | Variant-16 | | | | | R | | | | | | V | | | | | | | | | | | | | | | | | | | | G | | | | |
| 9 | Variant-17 | | | | | R | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 9 | Variant-18 | | | | | R | | | | | | V | | | | | | | | L | | | | | | | | | | | | G | | | | |
| 10 | Variant-19 | | | | | R | | | | | | V | | | | | | | | | | | | | | A | | | | | | G | | | | V |
| 10 | Variant-20 | | | | | R | | | | | | V | | | | | | | | L | | V | | A | | | | | | | | G | | | | |
| 10 | Variant-21 | | | | | R | | | | | | V | | | | | | | | L | | | | | | A | | | | | G | G | | | | |
| 12 | Variant-22 | | | | | E | R | | | | | V | | | | | | | | | | | | | | | | | | | | | | V | | |
| 12 | Variant-23 | | | | | R | | | | | | V | | | | | | | | L | | | | | | A | | | | | | | | | | |
| 14 | Variant-24 | | | | | R | | | | A | | V | | | | | | | | | | | | A | | A | | | | | | | | | | V |
| 17 | Variant-25 | | | | | E | | | | | G | V | | | | | | | | | | | | | | A | | | | | | G | | | | V |
| 18 | Variant-26 | L | | | G | R | E | | | | | V | | | | | | | | L | | V | E | A | | A | C | | | | | G | | | | |
| 24 | Variant-27 (**TDG2**) | L | | | | R | | | | | | V | | | | | | | | L | | V | E | A | | A | C | G | | | | G | | | | T |
| 25 | Variant-28 | L | | | | R | | | | | | V | | | | | | | | L | | V | E | A | | A | | G | | N | | G | | | | |
| 26 | Variant-29 | L | | | | E | R | | | | | V | | | A | E | | | | L | | | | | | | H | | | | | G | A | | N | T |
| 28 | Variant-30 | L | | | | E | R | | | | | V | | | | E | | | | L | | V | E | A | | | | S | | | | G | | | | V |
| 32 | Variant-31 (**TDG3**) | L | | | | E | | | | | | V | | | | E | | | | L | | V | E | A | | A | S | | | | | G | | | | T |
| 33 | Variant-32 (**TDG4**) | L | | | | E | | | | | | V | | | | E | | | | L | | V | E | A | | A | | G | | N | | G | | | G | |
| 33 | Variant-33 (**TDG5**) | L | | | | E | | | | | | V | | | | E | R | | | L | | V | E | A | | A | S | | | | | G | | | | T |
| 33 | Variant-34 (**TDG6**) | L | | | | E | | | | | | V | | | | E | | | | L | S | V | E | A | | A | | G | | N | | G | | | | T |

Conserved amino acids are highlighted. The five variants shown in Fig. 2b were bolded. The mutations of linker were list behind TDG.

**Supplementary Table 4. The editing efficiency of PE5max.**

| | position | PE5max | | | | | | DAF-TBE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Total editing efficiency (%) | | | C-to-G purity (%) | | | Total editing efficiency (%) | | | C-to-G purity (%) | | |
| | | rep. 1 | rep. 2 | rep. 3 | rep. 1 | rep. 2 | rep. 3 | rep. 1 | rep. 2 | rep. 3 | rep. 1 | rep. 2 | rep. 3 |
| *EMX1*-T2 | $T_6$ | 22.5 | 22.6 | 21.7 | 100 | 100 | 99.5 | 34.9 | 33.8 | 12.9 | 68.2 | 68.4 | 68.1 |
| *PSMB2*-1 | $T_2$ | 0.1 | 0.1 | 0.1 | 0 | 0 | 0 | 34.1 | 36.3 | 21.9 | 64.4 | 66.1 | 64.3 |
| *RP11-582C12*-4 | $T_6$ | 0 | 0.1 | 0.1 | 0 | 100 | 100 | 19.7 | 18.3 | 21.6 | 80.7 | 79.9 | 79.2 |
| *HIRA*-1 | $T_4$ | 30.2 | 29.3 | 30.5 | 99.7 | 99.7 | 99.7 | 18.0 | 18.9 | 14.8 | 78.1 | 80.0 | 82.2 |
| *EMX1*-site1 | $T_5$ | 1.1 | 1.3 | 1.2 | 100 | 92.3 | 91.7 | 32.3 | 39.2 | 16.4 | 66.8 | 72.2 | 77.9 |
| *RP11-177B4*-5 | $T_5$ | 25.6 | 24.2 | 25.6 | 100 | 100 | 100 | 20.5 | 25 | 23 | 74.3 | 77.1 | 74.7 |

**Supplementary Table 5. The Cas9-independent off-target sites.**

| Loci | R-loop sequences[a] | Sequence ID |
|---|---|---|
| *RP11-363D24* | TCTGCTTC$_8$TC$_{10}$C$_{11}$AGCCCTGGC CTGGGT | AL359552.16 |
| *HSPD1* | ATCTC$_5$ACTC$_9$GGGCTTATGCC AAAGAT | NG_008915.1 |
| *HIRA*-3 | ATC$_3$GCC$_6$AAGCTC$_{12}$ATC$_{15}$CTGGG AGAAGT | NG_009231.2 |
| *RP11-363D24* | TCT$_3$GCT$_6$T$_7$CT$_9$CCAGCCCT$_{17}$GGC CTGGGT | AL359552.16 |
| *HSPD1* | ATCT$_4$CACTCGGGCTTAT$_{17}$GCC AAAGAT | NG_008915.1 |
| *HIRA*-3 | ATCGCCAAGCT$_{11}$CAT$_{14}$CCT$_{17}$GGG AGAAGT | NG_009231.2 |

[a]Subscript numbers indicate edited cytosines or thymines. All these sequence were designed using Cas-OFFinder[1]

**Supplementary Table 6. Disease-associated SNVs used for creating disease models**

| Diseases | Disease-associated SNVs | References |
|---|---|---|
| Hereditary pulmonary alveolar proteinosis | NM_001089.3(*ABCA3*):c.2880G>C (p.Leu960Phe) | Garmany et al.[2] |
| Hypertrophic cardiomyopathy | NM_000256.3(*MYBPC3*):c.2526C>A (p.Tyr842Ter) | Marston et al.[3] |
| Familial hypercholesterolemia | NM_000527.5(*LDLR*):c.1467C>G (p.Tyr489Ter) | Marduel et al.[4] |
| Primary ciliary dyskinesia | NM_080860.4(*RSPH1*):c.1A>C (p.Met1Leu) | Kott et al.[5] |
| Spastic paraplegia | NM_014231.5(*VAMP1*):c.340+2T>G | Bourassa et al.[6] |
| Thrombophilia due to protein C deficiency | NM_000312.4(*PROC*):c.541T>G (p.Phe181Val) | Inoue et al.[7] |

**Supplementary Table 7. Disease-associated SNVs targeted for correction using SpRY-DAF-CBE/TBE**

| Diseases | Disease-associated SNVs | References |
|---|---|---|
| WDR62 primary microcephaly (WDR62-MCPH) | NM_001083961.2(*WDR62*):c.2655C>G (p.Tyr885Ter) | Verloes et al.[8] |
| glutaric acidemia type 1 (GA-1) | NM_000159.4(*GCDH*):c.456C>G (p.Ile152Met) | Hedlund et al.[9] |
| Hereditary multiple osteochondromas (HMO) | NM_207122.2(*EXT2*):c.426C>G (p.Tyr142Ter) | Wuyts et al.[10] |
| Familial hypocalciuric hypercalcemia (HHC) | NM_000388.4(*CASR*):c.115C>G (p.Pro39Ala) | Christensen et al.[11] |
| Cystic fibrosis (CF) | NM_000492.4(*CFTR*):c.1301C>G (p.Ser434Ter) | Elborn[12] |
| Tuberous sclerosis complex (TSC) | NM_000548.5(*TSC2*):c.4950C>G (p.Tyr1650Ter) | Northrup et al.[13] |
| SETBP1 haploinsufficiency disorder (SETBP1-HD) | NM_015559.3(*SETBP1*):c.3032C>G (p.Ser1011Ter) | Morgan et al.[14] |
| Leber congenital amaurosis (LCA) | NM_000180.4(*GUCY2D*):c.2988C>G (p.Tyr996Ter) | Tsang et al.[15] |
| Cystic fibrosis (CF) | NM_000492.4(*CFTR*):c.3276C>A (p.Tyr1092Ter) | Elborn[12] |
| BRCA1- and BRCA2-associated hereditary breast and ovarian cancer (HBOC) | NM_007294.4(*BRCA1*):c.3211G>T (p.Glu1071Ter) | Petrucelli et al.[16] |
| Dilated cardiomyopathy (DCM) | NM_001267550.2(*TTN*):c.94291G>T (p.Glu31431Ter) | Vaikhanskaya et al.[17] |
| Gaucher disease (GD) | NM_000157.4(*GBA*):c.1279G>T (p.Glu427Ter) | Stirnemann et al.[18] |
| COL1A1/2 osteogenesis imperfecta (COL1A1/2-OI) | NM_000088.4(*COL1A1*):c.652G>T (p.Gly218Cys) | Steiner et al.[19] |
| Hereditary multiple osteochondromas (HMO) | NM_000127.3(*EXT1*):c.115G>T (p.Glu39Ter) | Wuyts et al.[10] |
| Maple syrup urine disease (MSUD) | NM_000709.4(*BCKDHA*):c.137C>A (p.Ser46Ter) | Blackburn et al.[20] |
| Mucopolysaccharidosis type VII (MPS VII) | NM_000181.4(*GUSB*):c.107G>T (p.Arg36Leu) | Holtz et al.[21] |

**Supplementary Table 8. Induction conditions of CDG-nCas9/TDG-nCas9 in *E. coli* (with *trpB* Leu96*/Cys230*, gRNA)**

| Rounds | Induction conditions of CDG-nCas9 in *E. coli* (TrpB Leu96*, gRNA) | Rounds | Induction conditions of TDG-nCas9 in *E. coli* (TrpB Cys230*, gRNA) |
|---|---|---|---|
| **1~3** | L-arabinose (2 g/L), 24h | **1~4** | L-arabinose (2 g/L), 24h |
| **4~5** | L-arabinose (2 g/L), 12h | **5~6** | L-arabinose (2 g/L), 12h |
| **6~7** | L-arabinose (2 g/L), 6h | **7~8** | L-arabinose (2 g/L), 6h |
| **8~9** | L-arabinose (2 g/L), 3h | **9~10** | L-arabinose (2 g/L), 3h |
| **10~12** | L-arabinose (1 g/L), 3h | **11~12** | L-arabinose (1 g/L), 3h |
| **13** | L-arabinose (0.4 g/L), 3h | **13~14** | L-arabinose (0.4 g/L), 3h |
| **14** | L-arabinose (0.2 g/L), 3h | **15~18** | L-arabinose (0.2 g/L), 3h |
| **15~18** | L-arabinose (0.1 g/L), 3h | **19~26** | L-arabinose (0.1 g/L), 3h |
| **19~25** | L-arabinose (0.05 g/L), 3h | **27~30** | L-arabinose (0.05 g/L), 3h |
| | | **31~33** | No L-arabinose, 3h |

The concentration of L-arabinose and induction duration decreased as evolution progressed.

**Supplementary Table 9. The protospacer, 3' extension and linker sequences of pegRNA**

| Target sites | pegRNA spacer | 3' extension sequences | Linker | Nicking gRNA spacer |
|---|---|---|---|---|
| ***EMX1*-T2** | AGCCTTGTGAGA AGGGTCTA | GCTCTGAGCCAGAGACCCTT CTCACAAG | TAATACGA | CAGACCGTCA CATTAGCTCC |
| ***PSMB2*-1** | ATGAGGAAAGG GACTAGAGT | GAGGAAACAAAGCATAGACT GAGGGGTACAATCCTACTCT AGTCCCTTTCCTC | TAATTTAA | ACCATCTTTTG TACACTCAG |
| ***RP11-582C12*-4** | GCTGGAGAAAA GACAGAAAG | TCTTCTCTTTCCTCTTTCTGT CTTTTCTCCA | TACCCGTT | TGTCTTTTCTC CAGCTACTG |
| ***HIRA*-1** | CTGCCCTGCATT CTAGAATA | ATACAAACCGTAGTCTAGAAT GCAGGGC | CCTCAAAT | TCTCTGAACAG AAAGCTGTG |
| ***EMX1*-site1** | GGGTGGGGAAA AGGAGGAGG | TGCTTTCTTCCCCCTCCTCCT TTTCC | AATGTAAA | TTTCCCCCTCT CCTGTCTCC |
| ***RP11-177B4*-5** | ACCATTACGCCC ATGACCCA | TATTTCAGCCCTGGGTCATGG GCGTAA | CTCCTCAT | CCCTGTAGTAT TTCAGCCCT |

All of these sequences were designed using pegFinder[22], except for linker sequence which was designed using pegLIT[23].

**Supplementary Sequences 1. DNA sequences of inactive *trpB* genes in *E. coli* genome used in this study.** Sop mutations are highlighted in yellow and the complementary bases which should be edited are shown in red.

Gene sequence of TrpB Leu96*:

ATGACAACATTACTTAACCCCTATTTTGGTGAGTTTGGCGGCATGTACGTGC
CACAAATCCTGATGCCTGCTCTGCGCCAGCTGGAAGAAGCTTTTGTCAGTG
CGCAAAAAGATCCTGAATTTCAGGCTCAGTTCAACGACCTGCTGAAAAAC
TATGCCGGGCGTCCAACCGCGCTGACCAAATGCCAGAACATTACAGCCGG
GACGAACACCACGCTGTATCTCAAGCGTGAAGATTTGCTGCACGGCGGCG
CGCATAAAACTAACCAGGTGCTGGGGCAGGCCTGACTGGCGAAGCGGATG
GGTAAAACCGAAATCATCGCCGAAACCGGTGCCGGTCAGCATGGCGTGGC
GTCGGCCCTTGCCAGCGCCCTGCTCGGCCTGAAATGCCGTATTTATATGGGT
GCCAAAGACGTTGAACGCCAGTCGCCTAACGTTTTTCGTATGCGCTTAATG
GGTGCGGAAGTGATCCCGGTGCATAGCGGTTCCGCGACGCTGAAAGATGC
CTGTAACGAGGCGCTGCGCGACTGGTCCGGTAGTTACGAAACCGCGCACT
ATATGCTGGGCACCGCAGCTGGCCCGCATCCTTATCCGACCATTGTGCGTGA
GTTTCAGCGGATGATTGGCGAAGAAACCAAAGCGCAGATTCTGGAAAGAG
AAGGTCGCCTGCCGGATGCCGTTATCGCCTGTGTTGGCGGCGGTTCGAATG
CCATCGGCATGTTTGCTGATTTCATCAATGAAACCAACGTCGGCCTGATTGG
TGTGGAGCCAGGTGGTCACGGTATCGAAACTGGCGAGCACGGCGCACCGC
TAAAACATGGTCGCGTGGGTATCTATTTCGGTATGAAAGCGCCGATGATGCA
AACCGAAGACGGGCAGATTGAAGAATCTTACTCCATCTCCGCCGGACTGG
ATTTCCCGTCTGTCGGCCCACAACACGCGTATCTTAACAGCACTGGACGCG
CTGATTACGTGTCTATTACCGATGATGAAGCCCTTGAAGCCTTCAAAACGCT
GTGCCTGCACGAAGGGATCATCCCGGCGCTGGAATCCTCCCACGCCCTGGC
CCATGCGTTGAAAATGATGCGCGAAAACCCGGATAAAGAGCAGCTACTGG
TGGTTAACCTTTCCGGTCGCGGCGATAAAGACATCTTCACCGTTCACGATAT
TTTGAAAGCACGAGGGGAAATCTGA

Gene sequence of TrpB Cys230*:

ATGACAACATTACTTAACCCCTATTTTGGTGAGTTTGGCGGCATGTACGTGC
CACAAATCCTGATGCCTGCTCTGCGCCAGCTGGAAGAAGCTTTTGTCAGTG
CGCAAAAAGATCCTGAATTTCAGGCTCAGTTCAACGACCTGCTGAAAAAC
TATGCCGGGCGTCCAACCGCGCTGACCAAATGCCAGAACATTACAGCCGG
GACGAACACCACGCTGTATCTCAAGCGTGAAGATTTGCTGCACGGCGGCG
CGCATAAAACTAACCAGGTGCTGGGGCAGGCGTTGCTGGCGAAGCGGATG
GGTAAAACCGAAATCATCGCCGAAACCGGTGCCGGTCAGCATGGCGTGGC
GTCGGCCCTTGCCAGCGCCCTGCTCGGCCTGAAATGCCGTATTTATATGGGT
GCCAAAGACGTTGAACGCCAGTCGCCTAACGTTTTTCGTATGCGCTTAATG
GGTGCGGAAGTGATCCCGGTGCATAGCGGTTCCGCGACGCTGAAAGATGC
CTGTAACGAGGCGCTGCGCGACTGGTCCGGTAGTTACGAAACCGCGCACT
ATATGCTGGGCACCGCAGCTGGCCCGCATCCTTATCCGACCATTGTGCGTGA
GTTTCAGCGGATGATTGGCGAAGAAACCAAAGCGCAGATTCTGGAAAGAG
AAGGTCGCCTGCCGGATGCCGTTATCGCCTGAGTTGGCGGCGGTTCGAATG
CCATCGGCATGTTTGCTGATTTCATCAATGAAACCAACGTCGGCCTGATTGG
TGTGGAGCCAGGTGGTCACGGTATCGAAACTGGCGAGCACGGCGCACCGC
TAAAACATGGTCGCGTGGGTATCTATTTCGGTATGAAAGCGCCGATGATGCA
AACCGAAGACGGGCAGATTGAAGAATCTTACTCCATCTCCGCCGGACTGG
ATTTCCCGTCTGTCGGCCCACAACACGCGTATCTTAACAGCACTGGACGCG
CTGATTACGTGTCTATTACCGATGATGAAGCCCTTGAAGCCTTCAAAACGCT
GTGCCTGCACGAAGGGATCATCCCGGCGCTGGAATCCTCCCACGCCCTGGC
CCATGCGTTGAAAATGATGCGCGAAAACCCGGATAAAGAGCAGCTACTGG
TGGTTAACCTTTCCGGTCGCGGCGATAAAGACATCTTCACCGTTCACGATAT
TTTGAAAGCACGAGGGGAAATCTGA

**Supplementary Sequences 2. Amino acid sequences of DAF-CBE and TBE**

SV40 NLS are shown in blue;

CDG4 or TDG3 are shown in green;

Linker are shown in orange;

nCas9 are shown in purple;

Nucleoplasmin NLS are shown in black;

The mutations generated in evolution are shown in red

**DAF-CBE** (SV40 NLS-CDG4-linker (17 aa)-nCas9-nucleoplasmin NLS-SV40 NLS)

MPKKKRKVFGESWKKHLSGEFGKPYFIKLMEFVAEERKHYTVYPPPHQVFT
WTQMCDIKDVKVVILGQDPYHGPNQAHGLCFSVQRPVPPPPSLENIYEELSTD
IEGFVHPGHGDLSGWAKQGVLLLDAVLTVRAHQANSHKEQGWEQFTDAVVS
WLNQNSNGLVFLLWGSHAQKKGSAIDRKRHHVLQAAHPSPLSAHRGFFGCR
HFSKTNELLQKSGKKPIDWTELGGSETPGTSESATPESMDKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR
RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP
DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP
GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKA
LVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF
RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD

NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKKPKKKRKV*

**DAF-TBE** (SV40 NLS-TDG3-linker (17 aa)-nCas9-nucleoplasmin NLS-SV40 NLS)

MPKKKRKVLGESWKKHLSGEFGKPYFIKLMEFVAEERKHYTVYPPPHQVFT
WTQMCDIKDVKVVIVGQDPAHGPNQAHGLCFSVQRPVPPPPSLENIYEELSTD
IEDFVHPGHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVS
WLNQNLNGLVFLLWGSYAQKKGSVIDRERHHVLQAAHPSPLSASRGFFGCRH
FSKTNELLQKSGKKPIDWKELGGSETPGTSESATPESTDKKYSIGLAIGTNSVG
WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR
RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD
NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG
EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD
QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKAL
VRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTF
RIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDL
LFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK
DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQK
AQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEM
ARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYL
QNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSD
NVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF
YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA
KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPK
KYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF

LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK
VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL
DATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKKPKKKRKV*

**Supplementary Sequences 3. Linker amino acid sequences used in Fig. 6a, b and Supplementary Fig. 6a, b.**

**16 aa Linker:**

SGSETPGTSESATPES

**32 aa Linker:**

SGGSSGGSSGSETPGTSESATPESSGGSSGGS

**64 aa Linker:**

GGGGSGGGGSAEYVRALFDFNGNDEEDLPFKKGDILRIRDKPEEQWWNAED
SEGKRSGGSSGGS

# References


1. Bae, S., Park, J. & Kim, J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473-1475 (2014).
2. Garmany, T.H. et al. Surfactant composition and function in patients with ABCA3 mutations. *Pediatr Res* **59**, 801-805 (2006).
3. Marston, S. et al. Evidence from human myectomy samples that MYBPC3 mutations cause hypertrophic cardiomyopathy through haploinsufficiency. *Circ Res* **105**, 219-222 (2009).
4. Marduel, M. et al. Molecular spectrum of autosomal dominant hypercholesterolemia in France. *Hum Mutat* **31**, E1811-1824 (2010).
5. Kott, E. et al. Loss-of-function mutations in RSPH1 cause primary ciliary dyskinesia with central-complex and radial-spoke defects. *Am J Hum Genet* **93**, 561-570 (2013).
6. Bourassa, C.V. et al. VAMP1 mutation causes dominant hereditary spastic ataxia in Newfoundland families. *Am J Hum Genet* **91**, 548-552 (2012).
7. Inoue, H. et al. The clinical presentation and genotype of protein C deficiency with double mutations of the protein C gene. *Pediatr Blood Cancer* **64** (2017).
8. Verloes, A., Ruaud, L., Drunat, S. & Passemard, S. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle
Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA); 1993).
9. Hedlund, G.L., Longo, N. & Pasquali, M. Glutaric acidemia type 1. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* **142c**, 86-94 (2006).
10. Wuyts, W., Schmale, G.A., Chansky, H.A. & Raskind, W.H. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle
Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA); 1993).
11. Christensen, S.E., Nissen, P.H., Vestergaard, P. & Mosekilde, L. Familial hypocalciuric hypercalcaemia: A review. *Current Opinion in Endocrinology, Diabetes, and Obesity* **18**, 359-370 (2011).
12. Elborn, J.S. Cystic fibrosis. *Lancet* **388**, 2519-2531 (2016).
13. Northrup, H., Koenig, M.K., Pearson, D.A. & Au, K.S. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle
Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA); 1993).
14. Morgan, A., Srivastava, S., Duis, J. & van Bon, B. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle
Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved., Seattle (WA); 1993).
15. Tsang, S.H. & Sharma, T. Leber Congenital Amaurosis. *Advances in Experimental Medicine and Biology* **1085**, 131-137 (2018).
16. Petrucelli, N., Daly, M.B. & Pal, T. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle
Copyright © 1993-2022, University of Washington, Seattle. GeneReviews is a registered trademark

of the University of Washington, Seattle. All rights reserved., Seattle (WA); 1993).
17. Vaikhanskaya, T. et al. LMNA-related dilated cardiomyopathy. *Oxford Medical Case Reports* **2014**, 102-104 (2014).
18. Stirnemann, J. et al. A review of Gaucher disease pathophysiology, clinical presentation and treatments. *International Journal of Molecular Sciences* **18** (2017).
19. Steiner, R.D. & Basel, D. in GeneReviews(®). (eds. M.P. Adam et al.) (University of Washington, Seattle

20. Blackburn, P.R. et al. Maple syrup urine disease: Mechanisms and management. *The Application of Clinical Genetics* **10**, 57-66 (2017).
21. Holtz, M., Montaño, A.M. & Sly, W.S. Association between mucopolysaccharidosis Type VII and hydrops fetalis. *Ultrasound in Obstetrics & Gynecology* **55**, 416-417 (2020).
22. Chow, R.D., Chen, J.S., Shen, J. & Chen, S. A web tool for the design of prime-editing guide RNAs. *Nature Biomedical Engineering* **5**, 190-194 (2021).
23. Nelson, J.W. et al. Engineered pegRNAs improve prime editing efficiency. *Nature Biotechnology* **40**, 402-410 (2022).