nature biotechnology

Article

https://doi.org/10.1038/s41587-023-01910-9

Strand-preferred base editing of organellar and nuclear genomes using CyDENT

Received: 19 April 2023

Accepted: 19 July 2023

Published online: 28 August 2023

Check for updates

Jiacheng Hu[1,4], Yu Sun[1,2,4], Boshu Li[1,2,4], Zhen Liu[3], Zhiwei Wang[3], Qiang Gao[3], Mengyue Guo[3], Guanwen Liu[1], Kevin Tianmeng Zhao[3] ✉ & Caixia Gao[1,2] ✉

Transcription-activator-like effector (TALE)-based tools for base editing of nuclear and organellar DNA rely on double-stranded DNA deaminases, which edit substrate bases on both strands of DNA, reducing editing precision. Here, we present CyDENT base editing, a CRISPR-free, strand-selective, modular base editor. CyDENT comprises a pair of TALEs fused with a FokI nickase, a single-strand-specific cytidine deaminase and an exonuclease to generate a single-stranded DNA substrate for deamination. We demonstrate effective base editing in nuclear, mitochondrial and chloroplast genomes. At certain mitochondrial sites, we show editing efficiencies of 14% and strand specificity of 95%. Furthermore, by exchanging the CyDENT deaminase with one that prefers editing GC motifs, we demonstrate up to 20% mitochondrial base editing at sites that are otherwise inaccessible to editing by other methods. The modular nature of CyDENT enables a suite of bespoke base editors for various applications.

Mitochondria are essential semi-autonomous organelles that possess their own genomes. The human mitochondrial genome is a 16.6 kb double-stranded circular DNA (mitochondrial DNA (mtDNA)) that contains 37 genes, encoding 13 subunits of the respiratory chain complexes, two rRNAs and 22 transfer RNAs (tRNAs)[1]. According to mitomap data (https://www.mitomap.org/MITOMAP), more than 19,500 single-nucleotide variations (SNVs) have been found in mtDNA and more than 250 are believed to be disease-associated[2], including Leber's hereditary optic neuropathy, Leigh's syndrome, progressive encephalomyopathy, MELAS and more[2,3]. CRISPR-based genome editing technologies, including CRISPR–Cas nucleases[4–6], base editors[7–9] and prime editors[10–13], are powerful tools that enable fixing of different nuclear genomic variations; however, these tools fail to edit efficiently in mtDNA owing to poor single guide RNA delivery into the mitochondria[14]. Recently, two protein-based tools, double-stranded DNA deaminase (DddA)-derived cytosine base editor (DdCBE)[14–17] and TALE-linked deaminase (TALED)[18], were developed to enable precise base editing in mtDNA. Both DdCBE and TALED rely on DddA, a double-stranded DNA (dsDNA) deaminase, fused in a split format to a pair of TALE proteins to direct programmable base editing. Although both DdCBE and TALED can perform efficient C•G-to-T•A and A•T-to-G•C base conversions in mtDNA, respectively, these editors lack any precision for DNA target strand specificity. As the DddA protein is essential in directing any base conversion, it targets substrate bases on both strands of DNA so that cytosine and adenine bases on both strands of targeted DNA would be edited, leading to extra undesired mutations[14,18]. Therefore, there is a need to develop strand-specific CRISPR-free base editors to enable precise base conversions in organellar and nuclear DNA based on a protein-only system. Here, we present a modular base editing system, cytidine deaminase–exonuclease–nickase–TALE (CyDENT), which performs CRISPR-free, strand-selective DNA editing. CyDENT is a modular assembly comprised of a pair of programmable TALE proteins, a single-stranded DNA (ssDNA) deaminase, a nickase and an exonuclease. Importantly, we show that CyDENT achieves efficient and strand-selective mitochondrial base editing in human cells.

Results

CyDENT base editing strategy

DdCBE[14–17] and TALED[18] rely on the dsDNA deaminase DddA to perform base editing. DddA has been shown to transiently melt single-stranded

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. [2]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]Qi Biodesign, Beijing, China. [4]These authors contributed equally: Jiacheng Hu, Yu Sun, Boshu Li. ✉e-mail: kzhao@qi-biodesign.com; cxgao@genetics.ac.cn

DNA, which impairs hydrogen bond interactions between both strands of target DNA and allows for transiently exposed ssDNA bases to serve as substrates for deamination[19]. In the Cas9-based CBE[8] and adenine base editor (ABE)[7] systems, ssDNA-specific deaminases operate on the non-targeting DNA single-stranded region contained within the R-loop structure created by Cas9, thus enabling strand-selective base editing. We envisioned that generating a ssDNA substrate would be essential to perform strand-specific base editing and proposed that the elimination of one strand of DNA through DNA-modifying enzymes would facilitate the creation of such a substrate. To partially eliminate one of the paired dsDNA strands at a target site of interest, we first combined a TALE-fused FokI nickase and a nick-compatible exonuclease. We envisioned that after the nickase nicked one strand of target DNA, the exonuclease would recognize the nicked region and digest the nicked DNA strand, thereby exposing a short ssDNA fragment to serve as a substrate for ssDNA-specific modifying enzymes. We then used a locally fused ssDNA-specific deaminase protein to edit this region to perform strand-specific, protein-based base editing. Furthermore, we reasoned that the use of single-strand-specific cytidine deaminase in conjunction with a fused uracil glycosylase inhibitor (UGI) could further enhance overall cytosine base editing efficiencies.

Wild-type FokI is a bipartite restriction endonuclease that recognizes a non-palindromic DNA sequence (GGATG (9/13)) using its recognition domain and cuts the dsDNA outside of the recognition sequence with its endonuclease domain[20–23]. FokI nickases were generated by introducing a D450A mutation to either of the TALE carboxy-terminal-fused FokI cleavage domains (FokICD-L and FokICD-R, abbreviated to FokI-L and FokI-R hereafter) to abolish cleavage activity[24], resulting in two different potential FokI nickases, FokI-Lnickase (with a D450A mutation introduced in the FokI-R) and FokI-Rnickase (with a D450A mutation introduced in the FokI-L), respectively. We envisioned that the use of such a sequence-agnostic nickase would be essential as a first step for performing strand-selective DNA trimming and exposing a substrate for subsequent deamination to achieve base editing (Fig. 1a).

Validating CyDENT base editing in plants

We first evaluated CyDENT base editing in the nuclear genome of rice protoplasts. Four pairs of TALE proteins were custom designed to target the endogenous rice genomic sites *OsDEP1*, *OsCKX2*, *OsBADH2* and *OsSD1*. Exonucleases with 5′ → 3′ (mExoI[25]) or 3′ → 5′ (Trex2 (ref. 26,27)) digestion preferences were used to evaluate the effects of using an exonuclease in conjunction with a nickase to generate a ssDNA intermediate. We chose the highly efficient cytosine deaminase hAPOBEC3A[28,29] (hA3A) to perform cytidine deamination specifically on the ssDNA intermediate and fused a UGI peptide to its C termini to further enhance editing efficiencies by minimizing the effects of DNA base excision repair. Nuclear localization signals were fused to the amino termini of all components to direct editing to the nuclear genome. The combination of these constructs is hereafter designated as nuCyDENT (Fig. 1b). nuCyDENT, targeting rice *OsDEP1*, *OsCKX2*, *OsBADH2* and *OsSD1* sites, was delivered into rice protoplasts, and editing efficiencies were evaluated after 2 days. Following next-generation sequencing, targeted cytosine base editing was evaluated within the 18 bp spacer region between the TALE binding sites at all four nuclear genomic sites. We observed editing efficiencies ranging from 3% to 18% (Fig. 1c and Supplementary Fig. 1a) with low levels of indel formation. Importantly, we noticed that efficient editing frequencies were dependent on the presence of the exonuclease protein (Supplementary Fig. 1a). These results demonstrate that CyDENT base editing can achieve efficient CRISPR-free base editing with low levels of indel byproduct formation.

We envisioned that the base editing window of nuCyDENT would be affected by the choice of nickase. We reasoned that the FokI-Lnickase-containing (nuCyDENT-L) and FokI-Rnickase-containing (nuCyDENT-R) modules could perform base editing on different target strands, as they are biased towards nicking one strand of DNA[30]. When we analyzed the base editing window of the treated groups, we indeed observed that all four nuCyDENT-R-treated targets showed base editing primarily on the top DNA strand (Supplementary Fig. 1b). To confirm these results, both nuCyDENT-L and nuCyDENT-R were delivered into rice protoplasts separately with a pair of TALE arrays targeting the rice genomic sites *OsCKX2* and *OsSD1*. At 48 h post polyethylene glycol (PEG)-mediated protoplast transformation, base editing was evaluated. We observed that the top DNA strand was edited when transformed with nuCyDENT-R, whereas only the cytosine bases on the bottom strand were selectively edited in the presence of nuCyDENT-L, albeit with a lower average editing efficiency compared to that of the top strand (Fig. 1d). These results indicated that the CyDENT base editing strategy could perform efficient strand-selective base editing in nuclear genomes.

We then sought to use the CyDENT base editing strategy to install base edits into the plant chloroplast DNA, an important plant-specific organelle with its own genomic DNA (cpDNA) that is also inaccessible to CRISPR-derived base editors[16,31,32]. We replaced the nuclear localization signals in a chloroplast transit peptide[16] in nuCyDENT (Fig. 1e), resulting in cpCyDENT. We transformed rice protoplasts with cpCyDENT-L (harboring a FokI-Lnickase) and cpCyDENT-R (harboring a FokI-Rnickase) with TALE proteins targeting the endogenous ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (*rbcL*). We observed low but detectable levels of base editing in the *rbcL* target when treated with cpCyDENT-L (Fig. 1f). Importantly, only cytosine bases on the bottom strand of the target DNA were edited, with a frequency of about 1.67% at G_1 (counting the nucleotide closest to the 5′ end of the spacer region as position 1; Fig. 1f). We speculate that the low editing frequencies could be caused by low delivery efficiencies of PEG-mediated transformation into rice protoplasts, and subsequently into chloroplasts once inside cells for cpDNA editing. Importantly, these results highlight that cpCyDENT is able to selectively and precisely base edit on a strand of DNA in chloroplast genomes.

Mitochondria base editing in HEK293T cells using CyDENT

Encouraged by the results in plant chloroplasts, we sought to evaluate CyDENT base editing for mtDNA base editing in human cells. We first generated mtCyDENT constructs by substituting the nuclear localization signals of nuCyDENT with a mitochondrial targeting signal and replacing the promoters and terminators for expression in HEK293T cells. We speculated that recruiting the individually expressed deaminase, exonuclease and UGI to the TALE-FokI nickase could further enhance base editing frequencies. Therefore, we leveraged the use of a small peptide known as γb, fused to the N terminus of each module in mtCyDENT, to drive protein recruitment (Fig. 2a). γb is a viral RNA silencing suppressor[33] derived from the barley stripe mosaic virus that is known to self-interact. Furthermore, as the Trex2 exonuclease is much smaller (236 amino acids versus 837 amino acids in mExoI) and showed editing efficiencies comparable to mExoI (Fig. 1c and Supplementary Fig. 1a), we focused on Trex2 for further tests.

We transfected mtCyDENT and γb-fused mtCyDENT plasmids into HEK293T cells targeting the endogenous *ND6* gene in mtDNA. At 3 days post transfection, cells were collected and genomic DNA was prepared for next-generation sequencing analysis. We observed that although targeted base editing was detected in the *ND6* target, the γb-fused mtCyDENT constructs showed higher editing frequencies. Upon further modular protein assemblies, we found that γb fused to the UGI domain exhibited the greatest increase in editing efficiencies, followed by fusion to Trex2; however, fusing γb to the TALE domain did not elevate editing frequency, and fusing γb to the N terminus of the deaminase hA3A negatively impacted editing efficiency (Supplementary Fig. 2a,b). Therefore, we refer to the mtCyDENT variant with γb fused only to the UGI and Trex2 domains as mtCyDENT1b.

https://doi.org/10.1038/s41587-023-01910-9

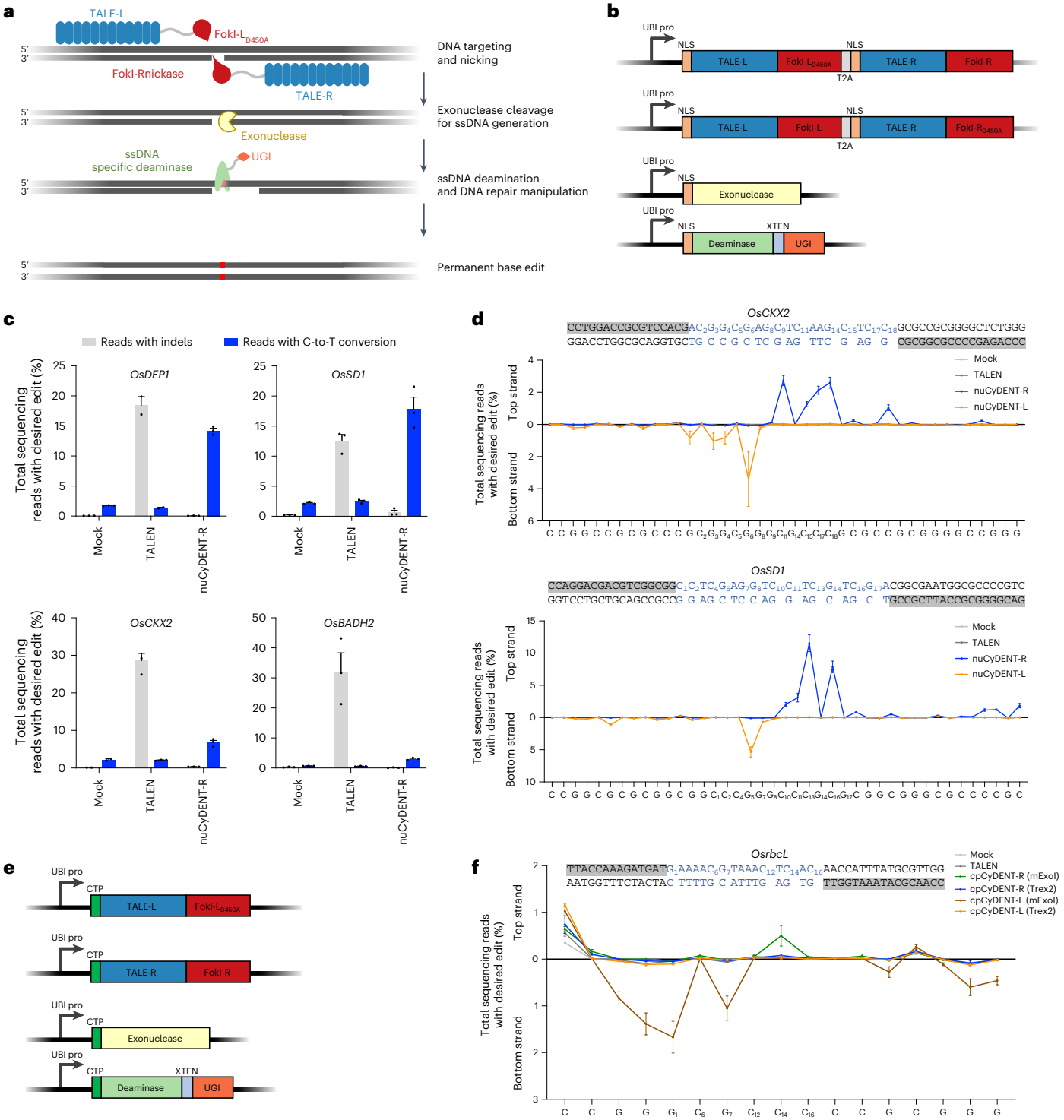


Fig. 1 | Strand-selective base editing using CyDENT in plant cells. a, Schematic overview of CyDENT base editing. A FokI nickase is directed by TALE proteins. An exonuclease uses the nicked site to excise a short patch of DNA, serving as a substrate for deamination and base editing, exemplified by CyDENT-R as shown in the schematic. **b**, Schematic overview of modular CyDENT constructs used in nuclear genome editing. NLS, nuclear localization signal; UBI pro, ubiquitin promoter. **c**, Frequencies of C-to-T conversions and indels by nuCyDENT-R and TALE nuclease (TALEN) at the *OsDEP1*, *OsSD1*, *OsCKX2* and *OsBADH2* sites in rice protoplasts. Mean ± s.e.m. for three (or two for TALEN in *OsDEP1*) independent biological replicates. **d**, CyDENT base editing at each base position at sites *OsCKX2* and *OsSD1* in rice protoplasts. TALE binding sites and spacers are shown. The position of the C base is named the same as its position in the sense strand. Mean ± s.e.m. for three independent biological replicates. **e**, Schematic overview of modular CyDENT constructs used in chloroplast genome editing exemplified by cpCyDENT-R in the schematic. CTP, chloroplast transit peptide. **f**, CyDENT base editing at each base position at the *OsrbcL* site in rice protoplasts. TALE binding sites and spacers are shown. The position of the C base is named the same as its position in the sense strand. Mean ± s.e.m. for two independent biological replicates.

https://doi.org/10.1038/s41587-023-01910-9

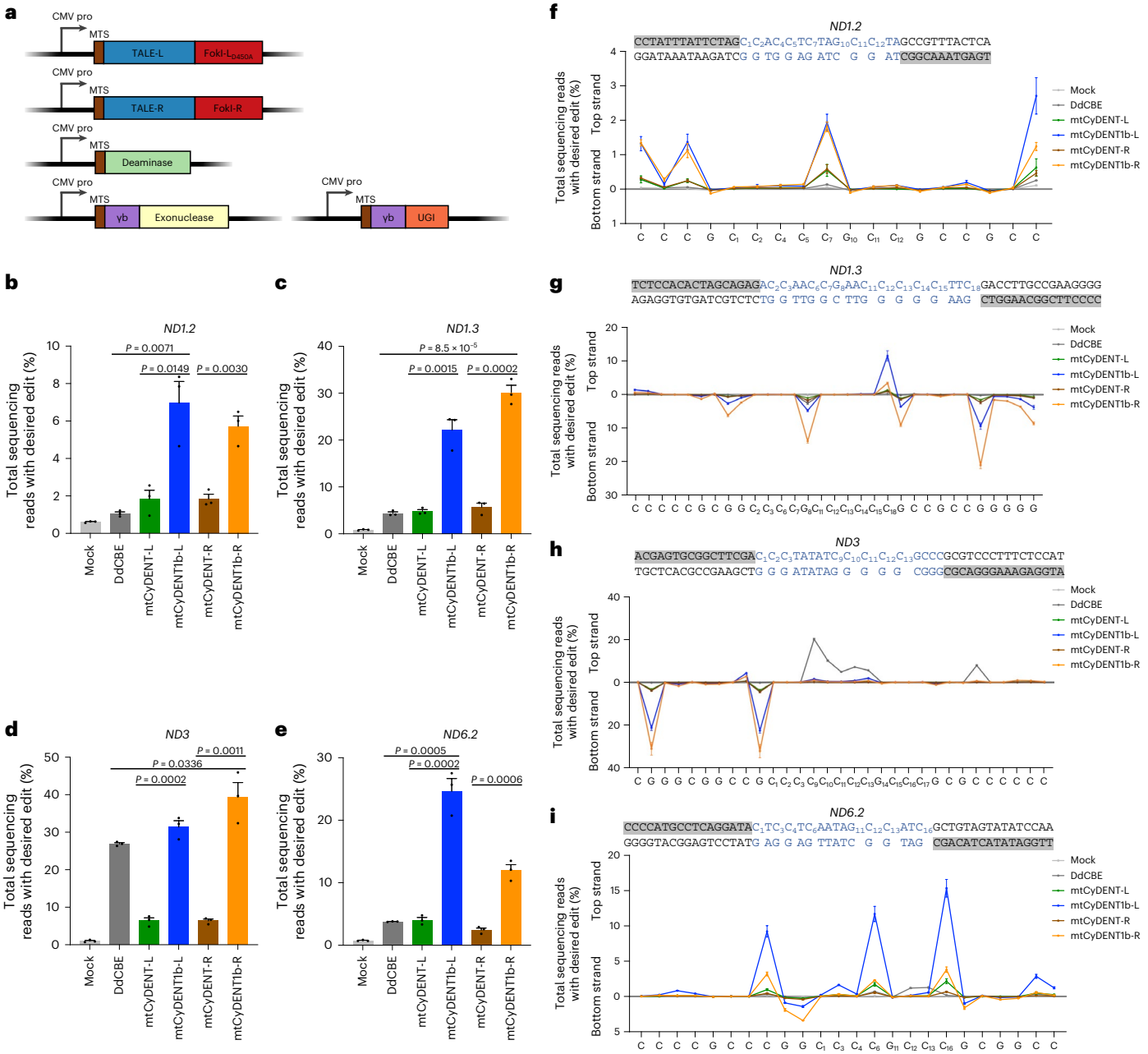


Fig. 2 | Strand-selective mtDNA base editing using CyDENT in HEK293T cells. **a**, Schematic overview of modular mtCyDENT constructs used in mitochondrial genome editing, exemplified by mtCyDENT-R. MTS, mitochondrial targeting signal; CMV pro, cytomegalovirus promoter. **b**–**e**, Frequencies of desired edits by DdCBE, mtCyDENT-R, mtCyDENT1b-R, mtCyDENT-L and mtCyDENT1b-L at the *ND1.2* (**b**), *ND1.3* (**c**), *ND3* (**d**) and *ND6.2* (**e**) sites in the mitochondria of HEK293T cells. Mean ± s.e.m. for three independent biological replicates. Statistical significance between groups was calculated using a two-tailed Student's *t*-test. **f**–**i**, Evaluation of editing strand preferences by DdCBE, mtCyDENT-R, mtCyDENT1b-R, mtCyDENT-L and mtCyDENT1b-L at the *ND1.2* (**f**) *ND1.3* (**g**), *ND3* (**h**) and *ND6.2* (**i**) sites in the mitochondria of HEK293T cells. TALE binding sites and spacers are shown. The position of the C base is named the same as its position in the sense strand. Mean ± s.e.m. for three independent biological replicates.

We next evaluated mtCyDENT and mtCyDENT1b across seven additional endogenous mtDNA genomic sites. We observed average editing frequencies of mtCyDENT ranging from 1.16–11.7%, with mtCyDENT1b further increasing the average editing efficiencies by 2.42–6.18-fold, reaching 4.55–39.3% (Fig. 2b–e and Supplementary Fig. 2c). Importantly, we observed that editing frequencies of mtCyDENT1b were higher than that of DdCBE at the *ND1.2*, *ND1.3*, *ND3* and *ND6.2* target sites with the same TALE arrays. Furthermore, we noted the absence of indel formation when using CyDENT base editing on mtDNA target sites (Supplementary Fig. 2d). These results indicate that mtCyDENT and mtCyDENT1b can achieve efficient base editing in human mtDNA.

We next sought to evaluate the editing strand preferences of CyDENT base editing in mtDNA. We found that the mtCyDENT-R (harboring a FokI-Rnickase) and mtCyDENT1b-R (harboring a FokI-Rnickase) constructs preferred editing the bottom DNA strand at six out of the seven (except for *ND1.2*) genomic sites tested, whereas the mtCyDENT-L (harboring a FokI-Lnickase) and mtCyDENT1b-L (harboring a FokI-Lnickase) constructs showed higher editing on the top strand at the *ND1.2* and *ND6.2* sites (Fig. 2f–i and Supplementary Fig. 2e). We noticed that mtCyDENT1b-L edited both strands at the *ND1.3* site, which could be a result of the D450A mutation in FokI not fully abolishing its cleavage activity[34], which affects the overall strand-nicking preference

(Supplementary Fig. 2f). We next sought to evaluate any phenotypic result mediated by CyDENT editing. We used mtCyDENT1b-R to target the *ND3* gene, which encodes for an NADH dehydrogenase complex in mitochondria, and subsequently measured ATP production rates using a Seahorse assay. As this complex is directly implicated in mitochondrial function and overall ATP production, we found that edited cells displayed decreased ATP levels compared to the untreated group (Supplementary Fig. 2g). These results demonstrate that the modular CyDENT base editing strategy is effective at strand-preferred base editing in mitochondrial genomes and that further efforts can continue to identify factors that can enhance strand precision and editing efficiencies.

Enhancing CyDENT editing efficiency and precision

Given that CyDENT is modularly assembled, we reasoned that the deaminase domain in CyDENT could be exchanged to leverage unique deaminase properties that enhance activity or promote editing strand precision. Recently, a newly identified ssDNA-specific cytidine deaminase, Sdd7, was found to exhibit much higher editing activity compared to other deaminases[35]. We introduced Sdd7 into mtCyDENT1b and evaluated editing efficiencies at the mtDNA target sites *ND5.1*, *ND6* and *ND1.3*. We observed that 87.5% of the alleles base-edited by Sdd7-mtCyDENT1b-L occurred on a single strand of DNA and 93.0% of the alleles base-edited by Sdd7-mtCyDENT1b-R also occurred on a single strand, highlighting the superior base editing strand specificity of CyDENT (Fig. 3a–i). These two editors resulted in the bottom target strand of DNA being edited with average editing efficiencies ranging from 4.88–9.13% (Fig. 3a,d,g and Supplementary Fig. 3a–d). The results particularly showcase the ability to exchange deaminase domains during the modular assembly of mtCyDENT.

In the mtCyDENT1b architecture, we used γb to recruit individual modules together. We next sought to further enhance editing efficiency by engineering fusion constructs surmising that local recruitment of modules by direct protein fusion would enhance local effective concentration. We evaluated a variety of protein architectures (Supplementary Fig. 3e) at the endogenous mtDNA *ND6* site, as the performance of mtCyDENT1b was poor at that site relative to its editing ability at other sites. We identified a particular construct in which the deaminase and exonuclease domains are fused to the N terminus of TALE-L and TALE-R using a 48-amino-acid flexible linker, and the UGI was fused to the C terminus and N terminus of FokI-L and FokI-R. This construct architecture is hereby referred to as mtCyDENT2 (Fig. 3j). We observed strand-selective base editing at *ND6* using mtCyDENT2-L (harboring a FokI-Lnickase), with a total of 94.5% of the edited alleles exhibiting base conversions selectively on the top strand (Fig. 3k,l).

In designing TALE nucleases, the length of the spacer region between the TALE binding sites has a substantial influence on gene knock-out efficiencies[36,37]. While evaluating CyDENT base editing, we noticed that a wider spacer region seemed to benefit editing efficiency at the *ND6* site, with average editing efficiencies increasing from 17.8–25.9% when the spacer expended from 15–26 bp (Fig. 3m,n); however, future efforts will need to more systematically evaluate spacer length effects across multiple sites and cell types.

mtCyDENT base editing at GC sequence motifs

The DddA-dependent DdCBE system demonstrated a strict TC sequence-context constraint for cytidine deamination and found that editing in GC sequence contexts occurred at a lower frequency[38]. Recently, phage-assisted non-continuous and continuous evolution was used to evolve the wild-type DddA, resulting in a DddA11 variant that exhibited broadened AC and CC sequence compatibility, but editing GC sequence motifs by DddA11 remained challenging[39]. A DddA homolog from *Simiaoa sunii* (Ddd_Ss) was found to show efficient cytidine deamination in DC contexts, highlighting the ability to modularly use different components with desired properties[40]. We reasoned that CyDENT could enable efficient and strand-selective GC sequence-motif base editing through a modular exchange of the deaminase domain.

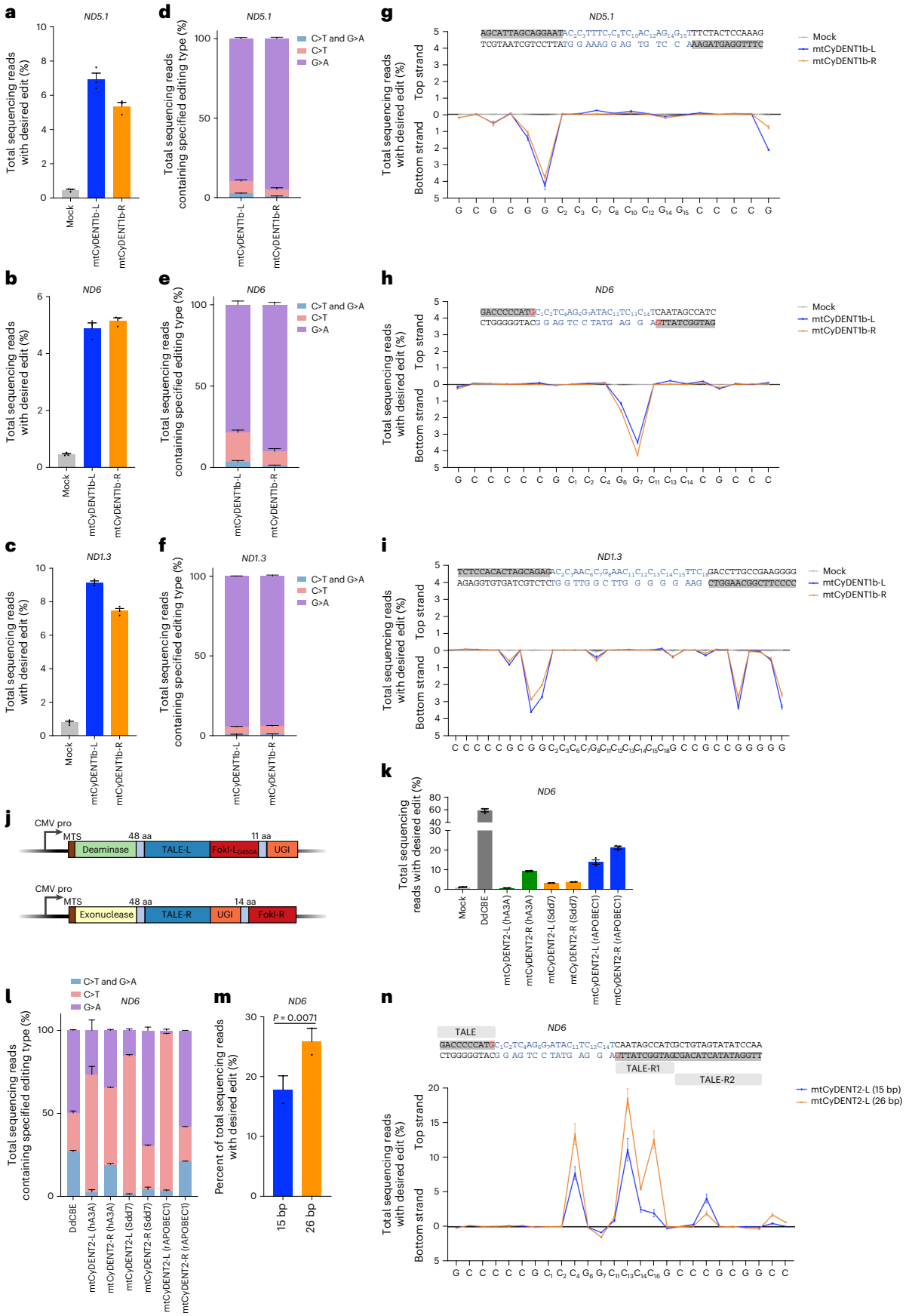
We incorporated a ssDNA-specific cytidine deaminase that edits efficiently at GC sequence motifs to develop a GC-compatible mtCyDENT base editor. Recently, a newly identified ssDNA-specific, GC-compatible and AC-compatible cytidine deaminase, Sdd3, was found to exhibit much higher editing activity than other deaminases at GC sequence motifs[35]. We designed TALE arrays to target the *ND4*, *ND1.2*, *ND3* and *ND6.2* sites in HEK293T cells (Supplementary Fig. 4a–d and Fig. 4a–f) to evaluate editing preferences at sequence motifs that were previously difficult to edit. Strikingly, strand-specific cytosine base editing was observed at GC contexts with efficiencies reaching 21.0% and 20.0% for the *ND1.2* and *ND6.2* sites, respectively, which were previously inaccessible to editing by DdCBE; at *ND1.2*, 96.9% of the edits occurred selectively on the top DNA strand, and at *ND6.2*, 92.0% of the edits occurred selectively on the bottom DNA strand (Fig. 4a–f). Notably, we observed 2.06% editing by Sdd3-mtCyDENT at the *ND6.2* site after adjusting the TALE binding sites (Fig. 4g,h). This particular mutation (m.14453G>A) was reported to be directly associated with the development of Leigh's syndrome[41] but is inaccessible to editing by DdCBE in these two TALE constructs. Therefore, mtCyDENT and potential future enhancements will serve as a superior approach to base editing, allowing for more precise modifications of additional mtDNA pathogenic variants.

Off-target analysis of mtCyDENT base editing

Mitochondrial editing using DdCBE was shown to induce substantial nuclear off-target editing[17,42]. To evaluate the off-target activity of CyDENT in both nuclear and mitochondrial genomes, we transfected HEK293T cells with mtCyDENT1b-R (hA3A) and DdCBE plasmids targeting *ND3*, and with mtCyDENT2-L (Sdd3) plasmids targeting *ND6.2*, which is able to install edits at GC context motifs. We obtained a total of 2.25 Tb of clean bases, with an average of 281.13 Gb for each sample and an average depth of sequencing the mitochondrial genome at approximately 6,362-fold (Fig. 5a). We found that the average mapping

Fig. 3 | Optimization of CyDENT for mtDNA base editing. a,**b**,**c**, Frequencies of desired edits by mtCyDENT1b-L and mtCyDENT1b-R with the Sdd7 deaminase at the *ND5.1* (**a**), *ND6* (**b**) and *ND1.3* (**c**) sites of HEK293T cells. Mean ± s.e.m. for three independent biological replicates. **d**,**e**,**f**, Ratio of specified editing events to total editing events generated by mtCyDENT1b-L and mtCyDENT1b-R with the Sdd7 deaminase at the *ND5.1* (**d**), *ND6* (**e**) and *ND1.3* (**f**) sites of HEK293T cells. Mean ± s.e.m. for three independent biological replicates. **g**,**h**,**i**, Evaluation of editing strand preferences by mtCyDENT1b-L and mtCyDENT1b-R with the Sdd7 deaminase at the *ND5.1* (**g**), *ND6* (**h**) and *ND1.3* (**i**) sites of HEK293T cells. TALE binding sites and spacers are shown. The position of the C base is named the same as its position in the sense strand. The bases in red indicate mismatched bases in the TALE. Mean ± s.e.m. for three independent biological replicates. **j**, Schematic overview of mtCyDENT2 constructs used in mitochondrial genome editing. **k**, Frequencies of desired edits by DdCBE, mtCyDENT2-L and mtCyDENT2-R with three different corresponding deaminases at the *ND6* site in HEK293T cells. Mean ± s.e.m. for three (mock, DdCBE or mtCyDENT2 (rAPOBEC1)) or two (mtCyDENT2 (hA3A or Sdd7)) independent biological replicates. **l**, Ratio of specified editing events to total editing events generated by DdCBE, mtCyDENT2-L and mtCyDENT2-R with three different corresponding deaminases at the *ND6* site in HEK293T cells. Mean ± s.e.m. for three (mock, DdCBE or mtCyDENT2 (rAPOBEC1)) or two (mtCyDENT2 (hA3A or Sdd7)) independent biological replicates. **m**, Frequencies of desired edits by mtCyDENT2-L with 15 or 26 bp spacers at the *ND6* site in HEK293T cells. Mean ± s.e.m. for two independent biological replicates. Statistical differences between groups were tested using a two-tailed Student's *t*-test. **n**, Evaluation of editing strand preferences by mtCyDENT2-L with the rAPOBEC1 deaminase at the *ND6* site in HEK293T cells. TALE binding sites and spacers are shown. Mean ± s.e.m. for two independent biological replicates.

https://doi.org/10.1038/s41587-023-01910-9



https://doi.org/10.1038/s41587-023-01910-9

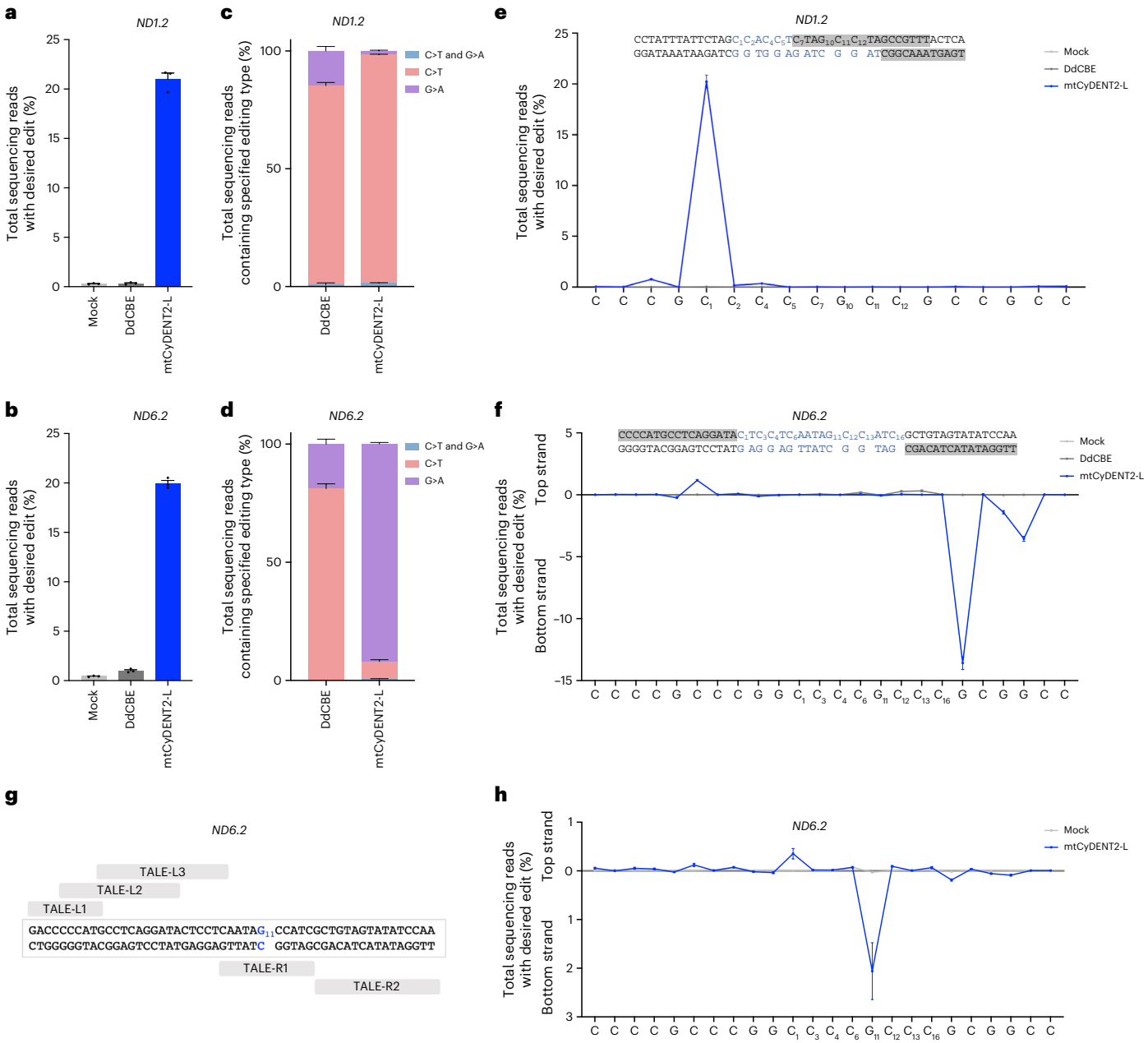


Fig. 4 | mtCyDENT enables efficient strand-selective base editing under GC contexts. a,**b**, Frequencies of desired edits by DdCBE and mtCyDENT2-L with the Sdd3 deaminase at the *ND1.2* (**a**) and *ND6.2* (**b**) sites of HEK293T cells. Mean ± s.e.m. for three independent biological replicates. **c**,**d**, Ratio of specified editing events to total editing events generated by DdCBE and mtCyDENT2-L with the Sdd3 deaminase at the *ND1.2* (**c**) and *ND6.2* (**d**) sites of HEK293T cells. Mean ± s.e.m. for three independent biological replicates. **e**,**f**, Evaluation of editing strand preferences by DdCBE and mtCyDENT2-L with the Sdd3 deaminase at the *ND1.2* (**e**) and *ND6.2* (**f**) sites of HEK293T cells. TALE binding sites and spacers are shown; in the case of *ND1.2*, the left TALE shifted to a new binding site. The position of the C base is named the same as its position in the sense strand. Mean ± s.e.m. for three independent biological replicates. **g**, TALE designs to target the Leigh's syndrome pathogenic mutation at site *ND6.2*. **h**, Evaluation of editing strand preferences by mtCyDENT2-L with the Sdd3 deaminase and TALE-L1 + TALE-R1 at the *ND6.2* site in HEK293T cells. TALE binding sites and spacers are shown. The position of the C base is named the same as its position in the sense strand. Mean ± s.e.m. for three independent biological replicates.

ratio to the human reference genome (hg19) was 99.57% with an average depth of approximately 89.61-fold (Supplementary Table 4).

We first confirmed efficient target editing in these samples (Fig. 5b) and then analyzed whole mitochondrial genome and nuclear genome-wide off-target frequencies. The average mitochondrial genome-wide C•G-to-T•A and G•C-to-A•T base conversion frequency was 4.8%, 6.9%, 16.5% and 5.9% for the mock, DdCBE, mtCyDENT1b-R (hA3A) and mtCyDENT2-L (Sdd3)-treated groups, respectively (Fig. 5c). We identified an average of 32, 678 and 16 SNVs in the mitochondrial genome when treated with the DdCBE, mtCyDENT1b-R (hA3A) and mtCyDENT2-L (Sdd3)-treated groups, respectively, compared to the mock treatment (Fig. 5d). After analyzing the 5 bp regions flanking each potential off-target SNV, a consensus TC motif was found from the DdCBE and mtCyDENT1b-R (hA3A) groups, and a consensus GC–AC motif was found in the mtCyDENT2-L (Sdd3) group (Fig. 5e). Through this off-target analysis, we highlight the ability to mitigate CyDENT-mediated off-target effects through the modular exchange of CyDENT components and through optimized editor architectures.

https://doi.org/10.1038/s41587-023-01910-9

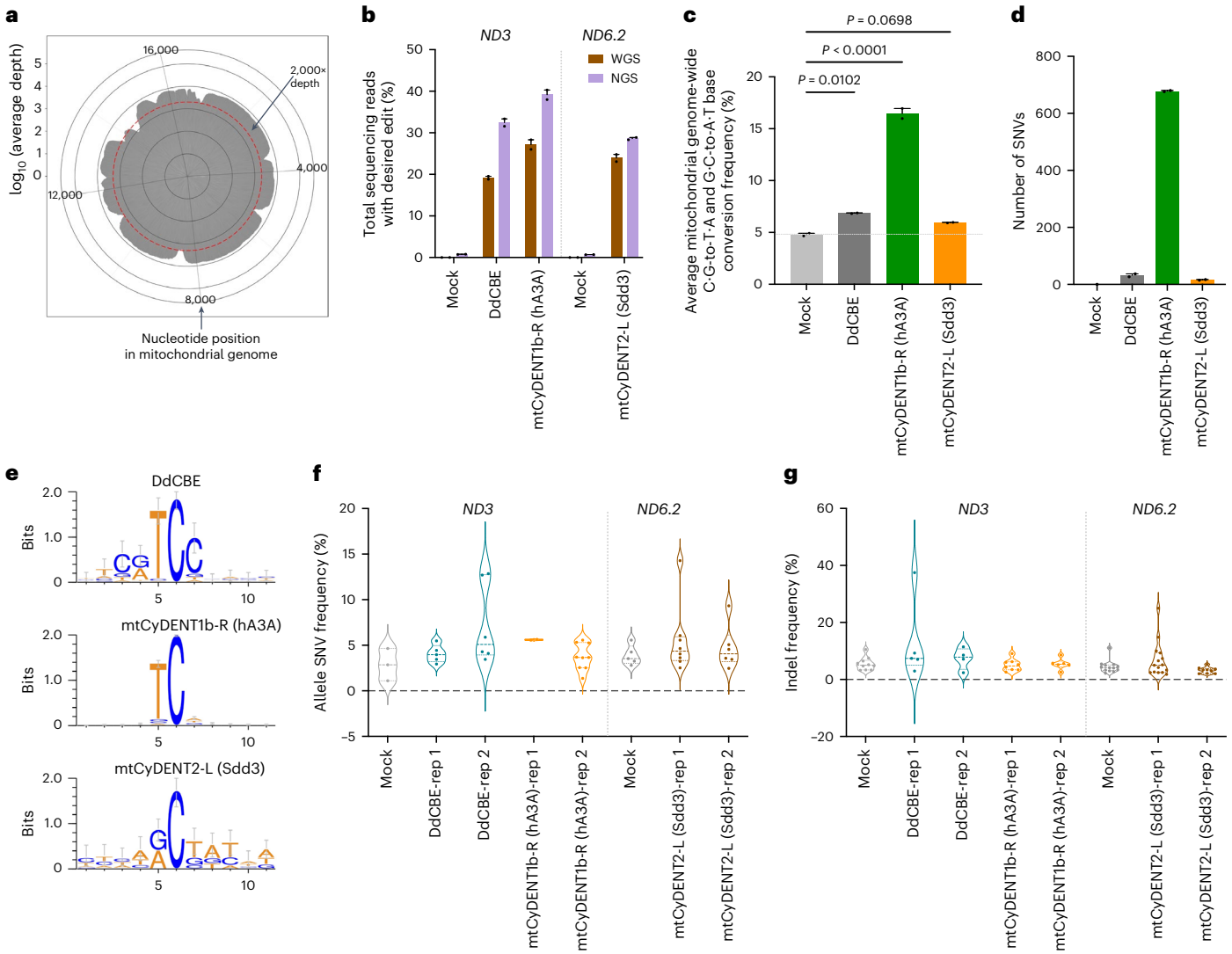


Fig. 5 | Whole nuclear genome and mitochondrial genome-wide off-target analysis of mtCyDENT. a, Average depth of the mitochondrial genome obtained by whole-genome sequencing (WGS). **b**, On-target editing frequencies of *ND3* and *ND6.2* by WGS and next-generation sequencing (NGS). Mean ± s.e.m. for two independent biological replicates. **c**, Average mitochondrial genome-wide C•G-to-T•A and G•C-to-A•T base conversion frequency (per cent). Mean ± s.e.m. for two independent biological replicates. **d**, Number of identified SNVs by different treatments with the mock set as baseline levels. Mean ± s.e.m. for two independent biological replicates. Statistical differences between controls and treatments were defined using Dunnett's multiple comparisons test after ANOVA. **e**, Sequence logos generated from off-target C•G-to-T•A and G•C-to-A•T base conversion by the indicated editors. Bits reflect sequence conservation at a given position. **f**, Allele SNV frequency identified in the potential-TALE-dependent off-target sites. Mean ± s.e.m. for two independent biological replicates. Statistical differences between controls and treatments were defined using Dunnett's multiple comparisons test after ANOVA. **g**, Allele indel frequency identified in the potential-TALE-dependent off-target sites. Mean ± s.e.m. for two independent biological replicates. Statistical differences between controls and treatments were defined using Dunnett's multiple comparisons test after ANOVA.

In the nuclear genome, we analyzed TALE-dependent off-target effects. A total of 74,963 potential off-target regions (containing 0–3 mismatches to the *ND3* and *ND6.2* TALE binding sites) were identified (Supplementary Table 5). We observed no significant difference in the allele SNV frequencies between the mock, DdCBE, mtCyDENT1b-R (hA3A) and mtCyDENT2-L (Sdd3)-treated groups for the *ND3* or *ND6.2* sites (Fig. 5f). Furthermore, we observed no difference of indel formations at these TALE-dependent off-target sites (Fig. 5g). These results indicate that modular assembly and optimization of CyDENT can minimize off-target effects in both the mitochondrial and nuclear genomes.

To assess any negative cellular impact of mtDNA editing using mtCyDENT, we evaluated cell viability at 2, 4 and 6 days post transfection and found that the performance of mtCyDENT was similar to DdCBE. Notably, 6 days after transfection, cell viabilities of both CyDENT and DdCBE-treated cells tended to recover to levels similar to that of untreated controls. The viabilities of cells transfected with solely Trex2 and γb-Trex2 were also similar to that of the untreated group (Supplementary Fig. 5a). These results indicate that mtCyDENT could be a valuable tool for mitochondrial genome editing.

Discussion

CRISPR-dependent base editors fundamentally rely on a transient ssDNA R-loop structure created by a Cas protein and a single guide RNA[43], whereas the DdCBE and TALED systems require a special dsDNA deaminase to target bases on both strands of a particular dsDNA region for effective base editing[14,18]. In this study, we present a base editing mechanism that separates DNA binding, ssDNA exposure and DNA editing into individual modular protein components. CyDENT base editing enables the easy exchange of protein domains with unique properties to perform strand-specific, CRISPR-free base editing.

This suite of CyDENT base editors and potential future additions will expand the suite of precision genome editing technologies for both nuclear and organelle genome editing.

Owing to its modular assembly, CyDENT enables bespoke editor designs, as demonstrated by the exchange of exonucleases with different trimming directions and deaminases with different sequence context biases and editing activities. In addition to cytidine deaminases, we speculated that a similar mechanism could be used for ABE by substituting the cytidine deaminase with an adenosine deaminase, resulting in adenine deaminase–exonuclease–nickase–TALE (AdDENT). When we evaluated AdDENT base editing in the nuclear genome using the evolved ssDNA-specific adenosine deaminase TadA-8e[44], we observed targeted ABE, both in plant and human cells at frequencies of around 0.90% and 2.10%, respectively (Supplementary Fig. 5b,c). As TadA-8e and other *E. coli* TadA variants were directly evolved from the tRNA-targeting wild-type adenosine deaminase TadA fused with Cas9 (refs. 7,44), we surmised that additional protein engineering and efforts would be required to perform ABE with TALEs in the AdDENT editor.

Given that CyDENT base editing relies on FokI nickase-mediated ssDNA nicking, we speculate that the overall construct architecture, deaminase choice and specific genome modifications could influence dimer formation and nicking activity, which may substantially affect the levels of CyDENT base editing. Future work to elucidate these factors and the use of other nickases could improve the editing activity and specificity of the CyDENT and AdDENT base editors. As the field of precision genome editing advances from the lab to the clinic, we recognize the need for bespoke editors that have greater precision and are designed to target a particular base. As DdCBE and TALED fundamentally rely on a dsDNA deaminase, the lack of precision for editing a particular base on one strand of DNA is a great limitation to this suite of protein-based base editors. Recently, mtDNA base editors were used for strand-biased mtDNA editing by combining sequence-specific nickases MutH or Nt.BspD6I(C) and adenine deaminase or cytosine deaminase for mtDNA base editing[45]. In addition to a nickase and a deaminase, the presence of an exonuclease in the CyDENT strategy presented here boosted editing frequency (Supplementary Fig. 1a). Owing to their modularity, CyDENT and AdDENT also serve as initial forays into the realm of bespoke, modular, CRISPR-free, precision genome editing technologies, and future work will further explore the potential of these unique protein assemblies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-023-01910-9.

References

1. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
2. Tuppen, H. A., Blakely, E. L., Turnbull, D. M. & Taylor, R. W. Mitochondrial DNA mutations and human disease. *Biochim. Biophys. Acta* **1797**, 113–128 (2010).
3. Greaves, L. C., Reeve, A. K., Taylor, R. W. & Turnbull, D. M. Mitochondrial DNA and disease. *J. Pathol.* **226**, 274–286 (2012).
4. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
6. Shan, Q. et al. Targeted genome modification of crop plants using a CRISPR–Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
7. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
8. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
9. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
10. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
11. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
12. Yarnall, M. T. N. et al. Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat. Biotechnol.* **41**, 500–512 (2023).
13. Sun, C. et al. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01769-w (2023).
14. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
15. Lee, S., Lee, H., Baek, G. & Kim, J. S. Precision mitochondrial DNA editing with high-fidelity DddA-derived base editors. *Nat. Biotechnol.* **41**, 378–386 (2022).
16. Kang, B. C. et al. Chloroplast and mitochondrial DNA editing in plants. *Nat. Plants* **7**, 899–905 (2021).
17. Lei, Z. et al. Mitochondrial base editor induces substantial nuclear off-target mutations. *Nature* **606**, 804–811 (2022).
18. Cho, S. I. et al. Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell* **185**, 1764–1776 (2022).
19. Yin, L., Shi, K. & Aihara, H. Structural basis of sequence-specific deamination by double-stranded DNA deaminase toxin DddA. *Nat. Struct. Mol. Biol.* https://doi.org/10.1038/s41594-023-01034-3 (2023).
20. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97–100 (1997).
21. Bitinaite, J., Wah, D. A., Aggarwal, A. K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA* **95**, 10570–10575 (1998).
22. Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25**, 778–785 (2007).
23. Szczepek, M. et al. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat. Biotechnol.* **25**, 786–793 (2007).
24. Waugh, D. S. & Sauer, R. T. Single amino acid substitutions uncouple the DNA binding and strand scission activities of Fok I endonuclease. *Proc. Natl Acad. Sci. USA* **90**, 9596–9600 (1993).
25. Lee, B. I., Shannon, M., Stubbs, L. & Wilson, D. M. 3rd Expression specificity of the mouse exonuclease 1 (mExo1) gene. *Nucleic Acids Res.* **27**, 4114–4120 (1999).
26. Certo, M. T. et al. Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat. Methods* **9**, 973–975 (2012).
27. Chen, M. J., Ma, S. M., Dumitrache, L. C. & Hasty, P. Biochemical and cellular characteristics of the 3′ → 5′ exonuclease TREX2. *Nucleic Acids Res.* **35**, 2682–2694 (2007).
28. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).
29. Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).

30. Sanders, K. L., Catto, L. E., Bellamy, S. R. & Halford, S. E. Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res.* **37**, 2105–2115 (2009).
31. Li, R. et al. High-efficiency plastome base-editing in rice with TAL cytosine deaminase. *Mol. Plant* **14**, 1412–1414 (2021).
32. Nakazato, I. et al. Targeted base editing in the plastid genome of *Arabidopsis thaliana*. *Nat. Plants* **7**, 906–913 (2021).
33. Jiang, Z., Yang, M., Zhang, Y., Jackson, A. O. & Li, D. In *Encyclopedia of Virology* 4th edn, Vol. 3 (eds Bamford, D. H. & Zickerman, M.) 420–429 (Academic Press, 2021).
34. Bonillo, M., Pfromm, J. & Fischer, M. D. Challenges to gene editing approaches in the retina. *Klin. Monbl. Augenheilkd.* **239**, 275–283 (2022).
35. Huang, J. et al. Discovery of deaminase functions by structure-based protein clustering. *Cell* **186**, 3182–3195 (2023).
36. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757–761 (2010).
37. Kim, Y. et al. A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* **31**, 251–258 (2013).
38. Nakazato, I. et al. Targeted base editing in the mitochondrial genome of *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **119**, e2121177119 (2022).
39. Mok, B. Y. et al. CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat. Biotechnol.* **40**, 1378–1387 (2022).
40. Mi, L. et al. DddA homolog search and engineering expand sequence compatibility of mitochondrial base editing. *Nat. Commun.* **14**, 874 (2023).
41. Shimura, M. et al. Development of Leigh syndrome with a high probability of cardiac manifestations in infantile-onset patients with m.14453G > A. *Mitochondrion* **63**, 1–8 (2022).
42. Wei, Y. et al. Mitochondrial base editor DdCBE causes substantial DNA off-target editing in nuclear genome of embryos. *Cell Discov.* **8**, 27 (2022).
43. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
44. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
45. Yi, Z. et al. Strand-selective base editing of human mitochondrial DNA using mitoBEs. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01791-y (2023).

Methods

Plasmid construction

Genes encoding the modules for cpCyDENT and part of the mtCyDENT were codon-optimized and synthesized by GenScript. The genes for expression in plant and HEK293T cell lines were PCR-amplified using 2× Phanta Max Master Mix (Vazyme Biotech) and cloned into the pGW3 (ref. 46) (rice codon-optimized) backbone and pCMV[44] backbone, respectively, using ClonExpress II One Step CloningKit (Vazyme Biotech).

Plasmids for HEK293T cell line transfection were extracted and purified using EndoFree Plasmid Kits (Qiagen) or the FastPure EndoFree Plasmid Mini Kit DC203 (Vazyme Biotech). Plasmids for protoplast transfection were purified using the Wizard Plus Maxipreps DNA Purification System according to the manufacturer's instructions. The amino acid sequences of all CyDENT vectors used are provided in the Supplementary Information (Supplementary Sequences). Primer sets for plasmid construction in this work were synthesized by the Beijing Genomics Institute and are listed in Supplementary Table 3.

PEG-mediated protoplast transformation

The *Japonica* rice cultivar Kitaake was used to isolate protoplasts and was transformed as previously described[46]. In brief, plasmids (5 μg per construct) were mixed and introduced into the protoplast by PEG-mediated transfection. After 48 h incubation at 26 °C in the dark, total genomic DNA was extracted with the CTAB method and used for amplicon deep sequencing.

Cell culture and transfection

HEK293T cells were cultured in DMEM (Gibco) supplemented with 10% (vol/vol) fetal bovine serum (FBS, Gibco) and 1% (vol/vol) Penicillin–Streptomycin (Gibco) in a humidified incubator at 37 °C with 5% CO_2. All cells were routinely tested for *Mycoplasma* contamination with a Mycoplasma Detection Kit (TransGen Biotech). A total of 40,000 cells per well were seeded into 48-well poly-D-lysine-coated plates (Corning) in the absence of antibiotics. After 16–24 h, cells were transfected with 1 μl jetPRIME transfection reagent (Polyplus), 400 ng (for mtCyDENT2, 200 ng each construct) or 450 ng (for mtCyDENT1b, 50 ng for UGI expression vector and 100 ng for each of the other four vectors) mtCyDENT plasmids per well at a 60–80% cell confluency. For DdCBE transfection, cells were incubated with 1 μl jetPRIME transfection reagent, 200 ng left DdCBE and 200 ng right DdCBE. Cells were washed with PBS and followed by DNA extraction 72 h after transfection.

Genomic DNA isolation from mammalian cell culture

Genomic DNA extraction was performed by the addition of 100 μl freshly prepared lysis buffer (10 mM Tris-HCl (pH 8.0), 0.05% SDS and 25 μg ml^{-1} proteinase K (Thermo Fisher Scientific)) directly into the 48-well culture plate after cells were washed once with 1× Dulbecco's PBS (Thermo Fisher Scientific). The mixture was incubated at 37 °C for 60 min and then at 80 °C for 20 min.

Amplicon deep sequencing and data analysis

Nested PCR was used for the amplification of amplicons spanning the target sites. In the first round of PCR, the target region was amplified from genomic DNA with site-specific primers. In the second round, both forward and reverse barcodes were added to the ends of the PCR products for library construction. Equal amounts of PCR product were pooled and purified with a GeneJET Gel Extraction Kit (Thermo Scientific) and quantified with a Qubit 4 Fluorometer (Thermo Scientific). The purified products were sequenced using the MiSeq platform, and the sequences around the target regions were examined for editing events. Amplicon sequencing was repeated three times for each target site using genomic DNA extracted from three independent samples unless specifically mentioned. Sequences of the amplicons are listed in Supplementary Table 1, and DNA sequences recognized by TALE proteins are listed in Supplementary Table 2. Analysis of base editing outcomes by MiSeq was performed as described previously[47].

For strand-specificity analysis, the top strands of the targets were used as reference sequences. Top strand-specific editing was calculated as the ratio of sequencing reads with only C-to-T conversion to the total edited reads, and bottom strand-specific editing was calculated as the ratio of sequencing reads with only G-to-A conversion to the total edited reads. The ratio of sequencing reads with both C-to-T and G-to-A conversions to the total edited reads were sorted as editing events on both strands. A custom perl script was written for this analysis and is included in the Supplementary Information.

Cell viability assays

Cell viability was measured 2, 4 and 6 days after transfection using the CellTiter 96 AQ_{ueous} One Solution Kit (Promega) according to the manufacturer's protocol. Cells were transfected at 80–90% confluency. Absorbance at 490 nm was measured using a 96-well plate reader (Tecan).

Oxygen consumption analysis by Seahorse XF Analyzer

Cells were seeded on the Seahorse plate and analyzed in the Seahorse XFe96 Analyzer (Agilent) when they reached 70–80% cell confluency. Analysis was performed in the Seahorse XF DMEM Medium pH 7.4 (Agilent) supplemented with 10 mM glucose (Agilent), 2 mM L-Glutamine (Gibco) and 1 mM sodium pyruvate (Gibco). The mito stress protocol was applied with the use of 1.5 mM oligomycin, 1 mM FCCP and 0.5 mM rotenone plus antimycin.

Analysis of nuclear genome-wide off-target editing

Genomic DNA was extracted from transfected cells using the FastPure Cell/Tissue DNA Isolation Mini Kit (Vazyme) according to the manufacturer's instructions. Whole-genome sequencing was performed by using the MGI DNBSEQ-T7 platform. The sequencing data was cleaned using fastp[48] (v.0.23.2) to obtain high-quality clean data. The clean data were then aligned to hg19 (ftp://gsapubftp-anonymous@ftp.broadinstitute.org/bundle) through BWA[49] (v.0.7.17-r1188) using the default parameters. Read group information was also added (-R) during the process. The data were subsequently processed through MarkDuplicates with Sambamba[50] and quality-calibrated with GATK[51] (v.4.4.0.0) BaseRecalibrator and ApplyBQSR modules. Finally, the Mutect2 module was used to obtain the SNVs.

The allele frequency was calculated using the DP tag from the results obtained from the VCF file. The depth calculation was performed using the depth module of SAMtools[52] (v.1.8).

Whole-genome-wide off-target analysis was performed using Cas-OFFinder[53], which predicted potential off-target sites with three or fewer mismatches to the TALE binding sites within the genome.

Analysis of mitochondrial genome-wide off-target editing

Analysis of C•G-to-T•A and G•C-to-A•T conversion frequency on the mitochondrial genome was conducted by using SAMtools[52] (v.1.8) to extract all mitochondrial data from the whole-genome sequencing data. Firstly, we extracted the data of the mock and obtained all the homozygous SNV mutations of the mock on hg19 using SAMtools. Then, we corrected all mutations to create the pseudo mtDNA genome. Subsequently, the mitochondrial sequencing data from all samples were extracted and aligned back to the pseudo mtDNA genome using the BW-aln algorithm. A perl script (in-house) was employed to parse the MD tag from BAM files in order to obtain all reads with C•G-to-T•A or G•C-to-A•T variants. SAMtools was used to extract variant reads in the on-target region. Reads outside the on-target region were excluded from all mitochondrial data to obtain the final off-target results. The result was calculated as follows:

$$Freq_{\text{base conversion frequency}} = \frac{N_{Reads\ contain\ \text{CG to TA mutation}}}{N_{total\ reads\ on\ chrM}}$$

Statistical analysis

GraphPad Prism 9 software was used to analyze the data. All numerical values are presented as means ± s.e.m. Statistical differences between controls and treatments were tested using the Student's *t*-test and one-way ANOVA; $P < 0.05$ was considered statistically significant and $P < 0.01$ was considered highly significant.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The deep amplicon sequencing data are deposited at the NCBI as Bioproject PRJNA957099 and PRJNA957096 ref. 54. All other data are available in the paper or Supplementary Information.

References

46. Shan, Q. et al. Rapid and efficient gene modification in rice and *Brachypodium* using TALENs. *Mol. Plant* **6**, 1365–1368 (2013).
47. Jin, S., Lin, Q., Gao, Q. & Gao, C. Optimized prime editing in monocot plants using PlantPegDesigner and engineered plant prime editors (ePPEs). *Nat. Protoc.* **18**, 831–853 (2023).
48. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2**, e107 (2023).
49. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
50. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
51. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
52. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
53. 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).
54. Hu, J. et al. Strand-selective base editing of organellar and nuclear genomes with TALE fusions. *National Center for Biotechnology Information* (*NCBI*) https://www.ncbi.nlm.nih.gov/bioproject/PRJNA957096 (2023).

Acknowledgements

We thank Y. Ai for kindly providing the HEK293T cell line. We also thank C. Yi for providing DdCBE vectors targeting *ND1*, *ND4*, *ND5.1* and *ND6*. We are grateful to Y. Wang for the helpful discussions. We acknowledge Y. Li for assistance with figure drawing. We thank J. Huang at BGI Genomics for the whole-genome sequencing service. This work was supported by the Ministry of Agriculture and Rural Affairs of China, the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, XDA24020102), the National Key Research and Development Program (2022YFF1002802) and the National Natural Science Foundation of China (32388201).

Author contributions

J.H., K.T.Z. and C.G. conceived the project and designed the experiments; J.H. designed vectors and tested mtCyDENT in HEK293T cells; Y.S. and B.L. designed vectors and tested nuCyDENT and cpCyDENT in plants; J.H. and Y.S. collected and analyzed MiSeq data; J.H., Y.S. and G.L. prepared MiSeq samples; Z.L. and M.G. prepared plasmids and HEK293T cells; Z.W. and Q.G. wrote scripts and processed the MiSeq RAW data; Q.G. processed the whole-genome sequencing RAW data; J.H., Y.S. and B.L. prepared the figures; J.H., Y.S., B.L., K.T.Z. and C.G. wrote the manuscript with input from all authors; and C.G. and K.T.Z. supervised the study.

Competing interests

The authors have submitted a patent application based on the results reported in this paper. K.T.Z. is a founder and employee at Qi Biodesign. Z.L., Z.W., Q.G. and M.G. are employees of Qi Biodesign.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-023-01910-9.

Correspondence and requests for materials should be addressed to Kevin Tianmeng Zhao or Caixia Gao.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature research

Corresponding author(s): Caixia Gao, Kevin Tianmeng Zhao

Last updated by author(s): Jul 18, 2023

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina Miseq platform was used to collect the amplicon deep sequencing data. The WGS data was collected using the MGI DNBSEQ-T7 platform.
Data analysis	Amplicon sequencing data and WGS data were analyzed using the published code as previously described and noted in the methods. A custom perl script was used to perform strand specific analysis and this code is included in the supplementary information. SAMtools (v1.8), fastp (v0.23.2), BWA (v0.7.17-r1188), Sambamba, GATK (v4.4.0.0), Cas-OFFinder, and BW-aln algorithm were used to process and analyze the WGS data. Graphpad prism 9 was used to analyze the data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The authors declare that all data supporting the findings of this study are available in the article and its Supplementary Information files or are available from the corresponding author on request. Datasets of high-throughput sequencing experiments will be deposited with the National Center for Biotechnology Information (NCBI) under PRJNA957099 and PRJNA957096.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The experiments of protoplasts were performed with three biological repeats. About 100,000 protoplasts were used for each transfection. The number of protoplasts in each transfection was measured by thrombocytometry. HEK293T cell transfections were performed as described in the methods.
Data exclusions	No data was excluded from the study.
Replication	All attempts for replication were successful. For the experiments in rice protoplasts and HEK293T cells, a minimum of two independent experiments were included.
Randomization	Rice protoplasts were isolated and randomly separated for each transformation biological replicate. HEK293T cells were separately passaged and transfected for each replicate.
Blinding	Not applicable. As samples were processed identically through standard and in some cases automated procedures (DNA sequencing, transfection, DNA isolation) that should not bias outcomes.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	ATCC
Authentication	Validated by the company
Mycoplasma contamination	Validated by the company and tested in our studies to ensure no mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	N/A