

Supporting Information

for *Adv. Sci.*, DOI 10.1002/adv.202304113

Enhanced C-To-T and A-To-G Base Editing in Mitochondrial DNA with Engineered DdCBE and TALE

*Yinghui Wei**, Ming Jin, Shuhong Huang, Fangyao Yao, Ningxin Ren, Kun Xu, Shangpu Li, Pengfei Gao, Yingsi Zhou, Yulin Chen, Hui Yang, Wen Li*, Chunlong Xu*, Meiling Zhang* and Xiaolong Wang*

Supporting Information for

Enhanced C-to-T and A-to-G base editing in mitochondrial DNA with engineered DdCBE and TALE

Yinghui Wei^{1,2,#,}, Ming Jin^{3,#}, Shuhong Huang^{1,#}, Fangyao Yao^{1,#}, Ningxin Ren⁴, Kun Xu¹, Shangpu Li¹, Pengfei Gao¹, Yingsi Zhou⁴, Yulin Chen^{1,2}, Hui Yang^{4,5}, Wen Li^{6,*}, Chunlong Xu^{5,*}, Meiling Zhang^{6,*}, Xiaolong Wang^{1,2,*}*

This PDF file includes:

Figure S1 to Figure S11

Table S1 to Table S5

Supplementary sequences



Figure S1. Schematic showing the architectures of DddA11 fused ten cytosine deaminases with ssDNA activity used in this study. MTS, mitochondrial targeting sequence; UGI, uracil glycosylase inhibitor; BE3 = rat APOBEC1; YE1 = BE3^{W90Y+R126E}; R132E = BE3^{R132E}; A130F = human APOBEC3A^{A130F}; N57G = human APOBEC3A^{N57G}; FE1 = BE3^{W90F+R126E}; evoYE1 = engineered BE3^{W90Y+R126E}; xAID = *Xenopus laevis* activation-induced cytidine deaminase (AID); hAID = human activation-induced cytidine deaminase (AID); PmCDA1 = *petromyzon marinus* cytidine deaminase 1.

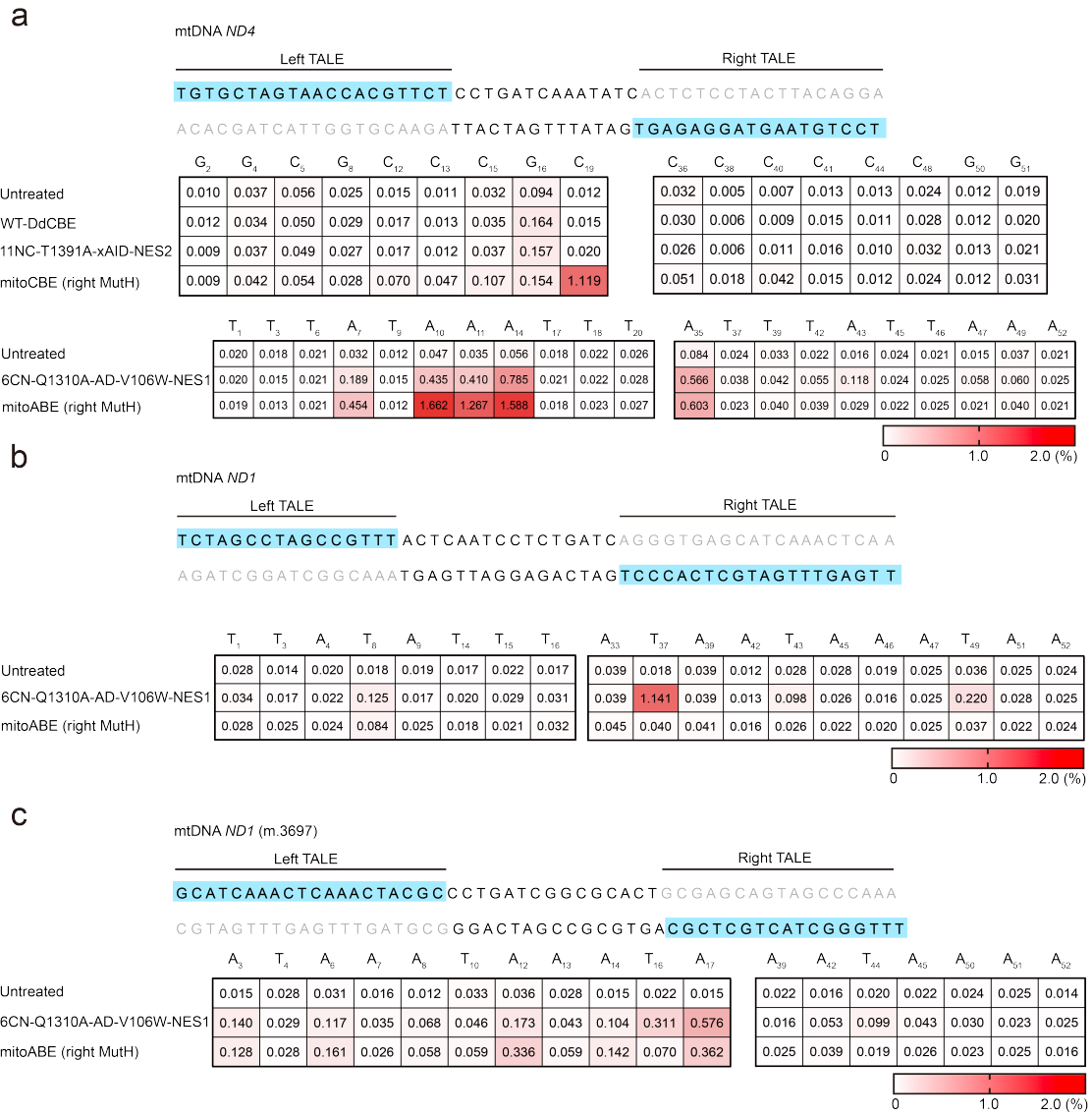


Figure S2. Bystander edits in the TALE binding sites and positions flanking the binding sites. a) Bystander C-to-T and A-to-G edits in the TALE binding sites and positions flanking the binding sites for WT-DdCBE, 11NC-T1391A-xAID-NES2, mitoCBE (right MutH), 6CN-Q1310A-AD-V106W-NES1, and mitoABE (right MutH) tools at mitochondrial *ND4* site. **b-c)** Bystander A-to-G edits in the TALE binding sites and positions flanking the binding sites for CN-Q1310A-AD-V106W-NES1 and mitoABE (right MutH) tools at mitochondrial *ND1* and *ND1* (m.G3697) sites. The nucleotide adjacent to the start of the left-TALE-recognition sequence was numbered “1,” and C, G, T, or A were sequentially numbered. Top 10% of EGFP- and mCherry-double positive cells were harvested from fluorescence-activated cell sorting (FACS) 48 h after transfection. The editing efficiency was tested by targeted deep sequencing. The number is given in units of %. All values are presented as the average of n = 3 biologically independent experiments.

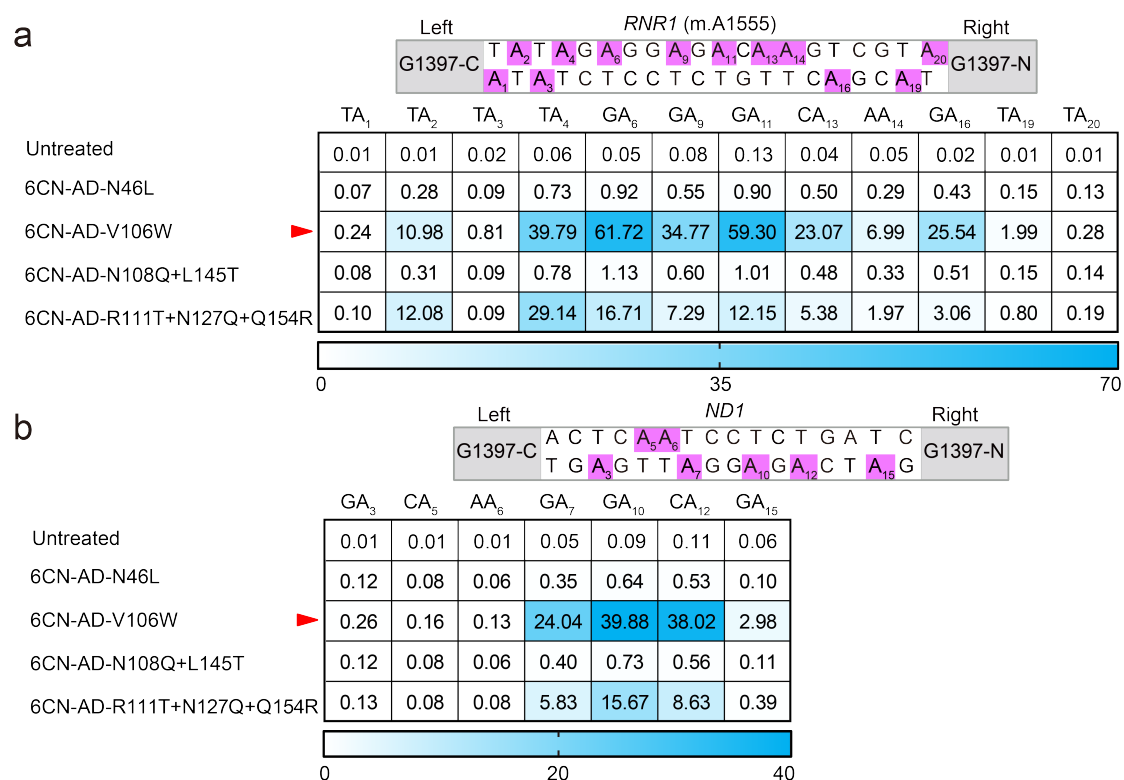


Figure S3. Target A-to-G conversions using engineered TALE variant (6CN-AD) and recently reported several mutations of TadA8e in the mitochondrial *RNR1* (m.A1555) and *ND1* genes. a-b) Heat map showing A·T-to-G·C editing efficiencies induced by 6CN-AD with N46L, V106W, N108Q + L145T, and R111T + N127Q + Q154R mutations in HEK293T cells at two mitochondrial target sites, including *RNR1* (m.A1555) (a) and *ND1* (b). The adenines in the top strands or bottom strands are presented as A-to-G conversion frequencies. The nucleotide adjacent to the end of the left-TALE-recognition sequence was numbered “1,” and A was sequentially numbered. Top 10% of EGFP- and mCherry-double positive cells were harvested from fluorescence-activated cell sorting (FACS) 48 h after transfection. The targeting efficiency was tested by targeted deep sequencing. All values are presented as the average of n = 3 biologically independent experiments. The number is given in units of %.

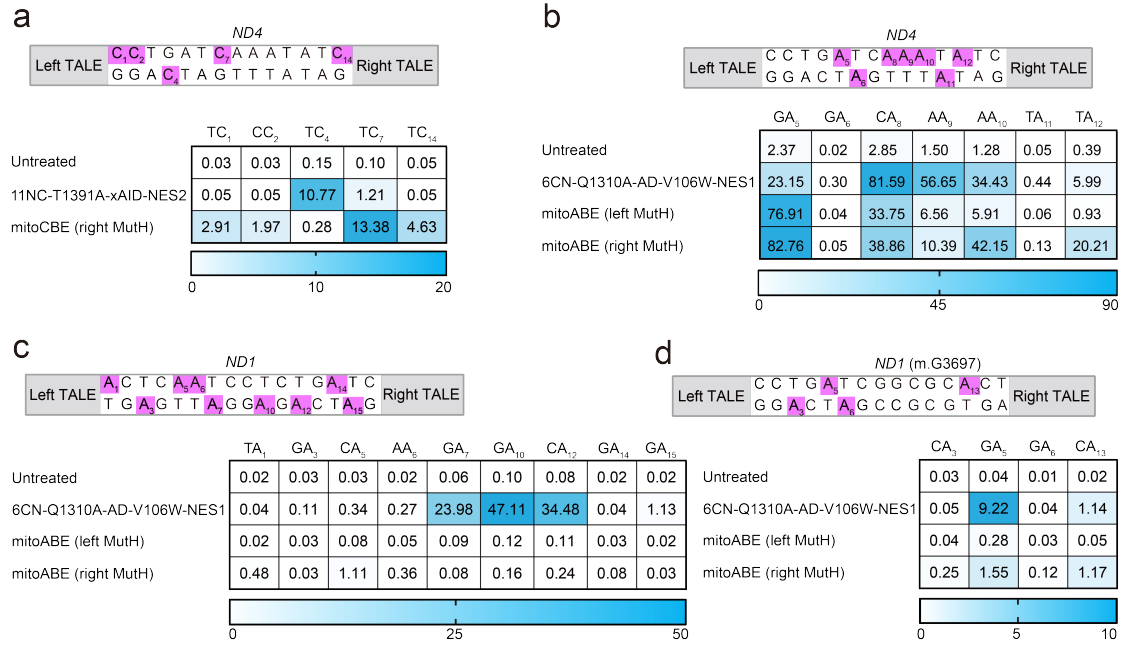


Figure S4. Comparison of the editing efficiency of our engineered variants with the reported mitoBEs (PMID: 37217751). **a)** The C-to-T editing efficiency of 11NC-T1391A-xAID-NES2 and mitoCBE targeting *ND4* in HEK293T cells. **b-d)** The A-to-G editing efficiency of 6CN-Q1310A-AD-V106W-NES1 and mitoABE targeting *ND4* (**b**), *ND1* (**c**), and *ND1* (m.G3697) (**d**) in HEK293T cells. 11NC-T1391A-xAID-NES2, Right-G1397-C + Left-G1397-N orientation; 6CN-Q1310A-AD-V106W-NES1, Right-G1397-N + Left-G1397-C orientation. All values are presented as the average of n = 3 biologically independent experiments. The number is given in units of %.

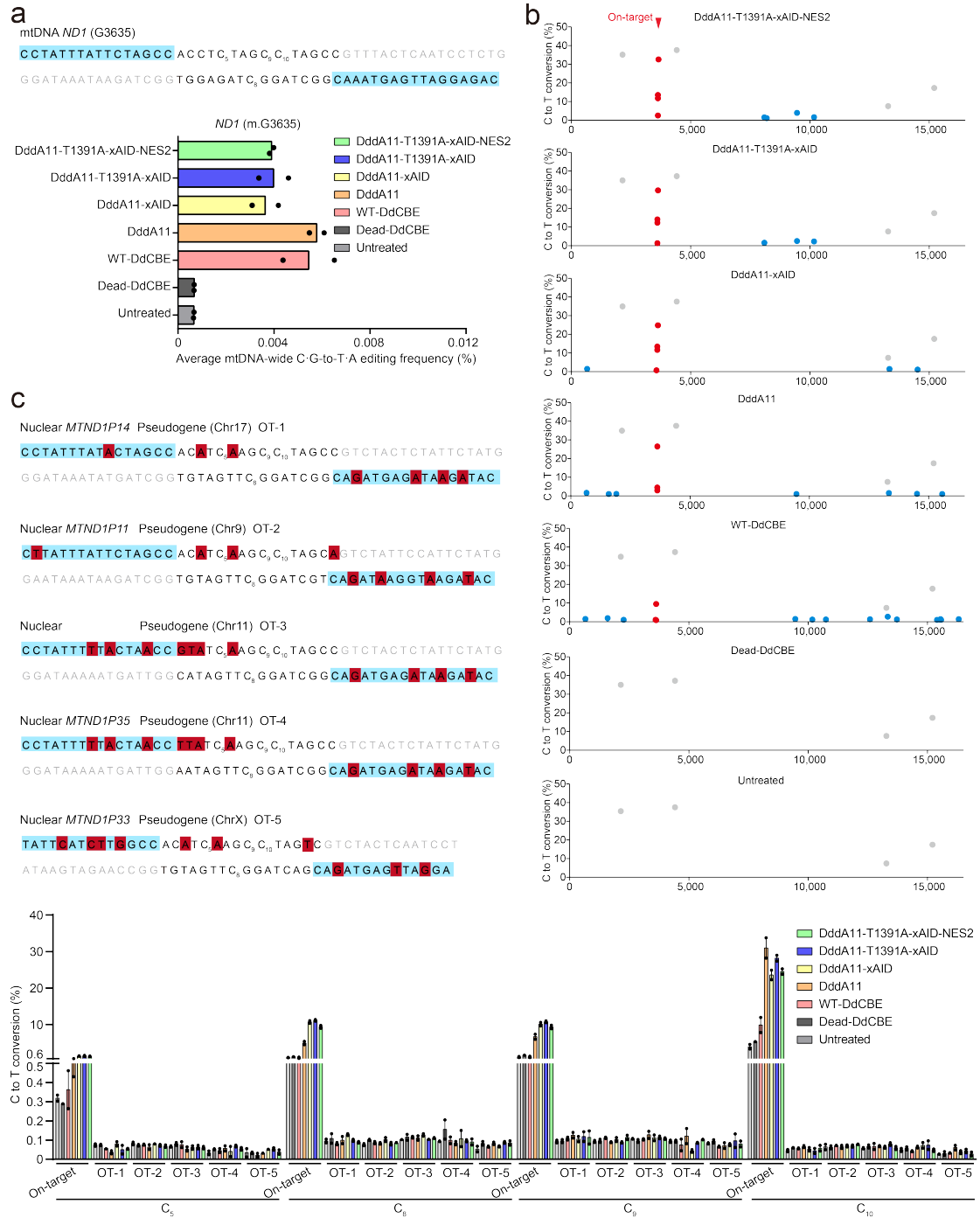


Figure S5. Off-target analysis for engineered DdCBEs specific to the *ND1* (m.G3635) site. a) The average frequencies of mitochondrial genome-wide off-target editing induced by Dead-DdCBE, wild-type DdCBE (WT-DdCBE), DddA11, DddA11-xAID, DddA11-T1391A-xAID, and DddA11-T1391A-xAID-NES2 specific to the *ND1* (m.G3635) site. Error bars are s.e.m. for n=2 biologically independent samples. **b)** Mitochondrial genome-wide plots for C-to-T point mutations with frequencies $\geq 1\%$. Naturally occurring SNVs, on-target edits (including bystander edits in the editing window) and off-target edits are shown in blue and red, respectively. All data points from n=2 biologically independent experiments are shown. **c)** The corresponding nuclear DNA sequences with the high homology are shown for the

ND1 (m.G3635) site. TALE binding sites begin at N0 and are shown in blue. Nucleotide mismatches between the mtDNA and nuclear pseudogene are in red. Editing efficiencies are measured by targeted deep sequencing (see Table S4 for primer sequences) (Supporting Information). Data are presented as means $\pm$ SEM.

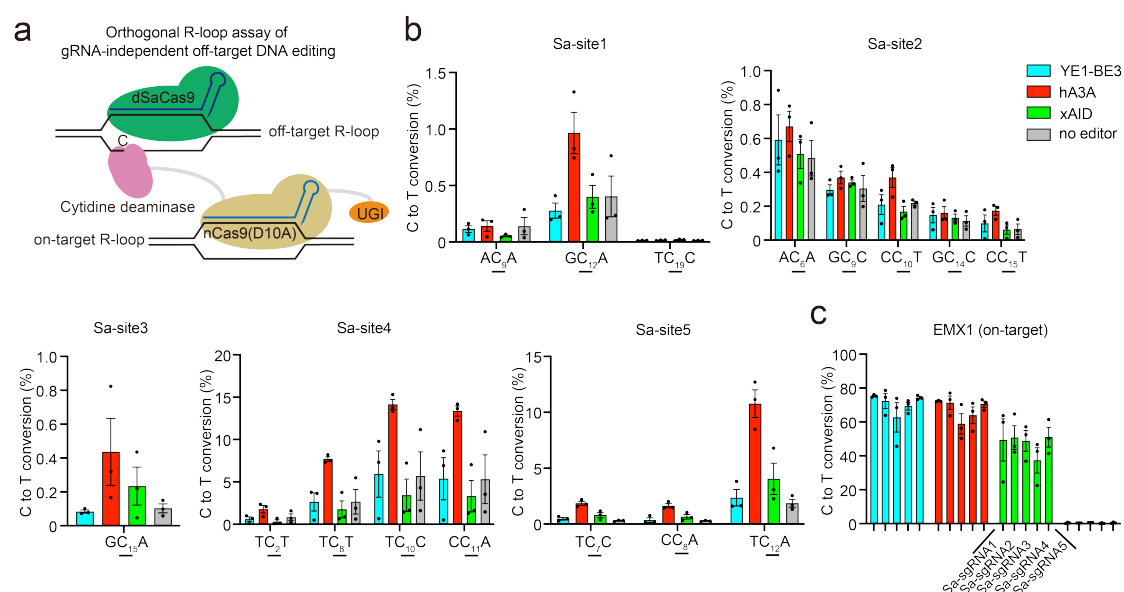


Figure S6. Sequence-independent off-target analysis for different cytosine deaminases in HEK293T cells via R-loop assay. **a)** Schematic of sequence-independent deamination of cytosines within dSaCas9-induced R-loop sites by deaminase fused with SpCas9. **b)** Sequence-independent off-target C•G-to-T•A editing frequencies detected by targeted high-throughput sequencing of five dSaCas9-induced R-loop loci following co-transfection with different CBEs made from fusion of SpCas9 with YE1-BE3, hA3A, and xAID. Editing efficiencies are measured by targeted deep sequencing (see Table S5 for protospacer and primer sequences) (Supporting Information). **c)** On-target DNA base editing efficiencies at the *EMX1* genomic locus corresponding to the SpCas9 sgRNA. Editing efficiencies are measured by targeted deep sequencing (see Table S5 for protospacer and primer sequences) (Supporting Information). All data points from $n = 3$ biologically independent experiments are shown. Data are presented as means $\pm$ SEM.

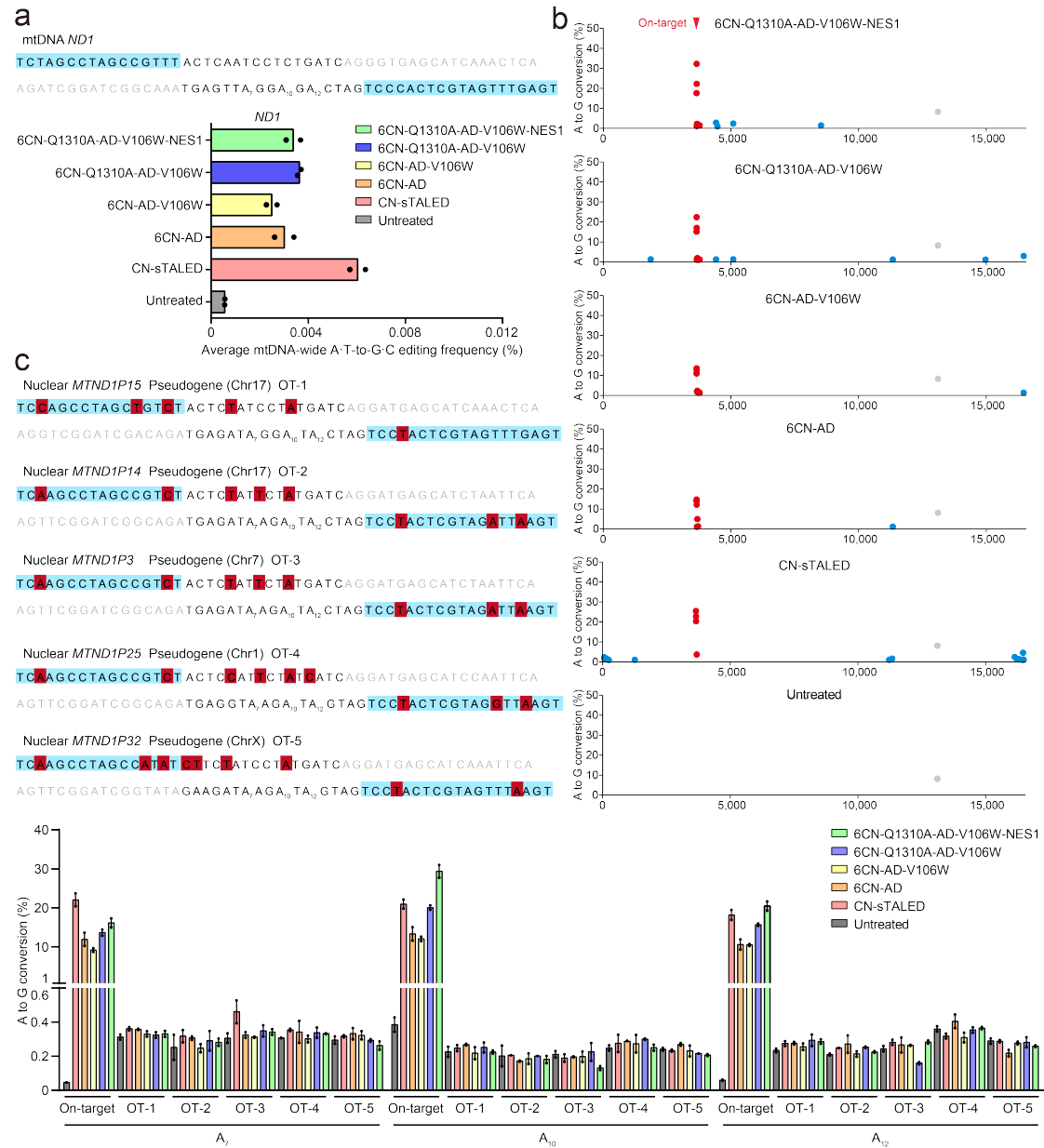
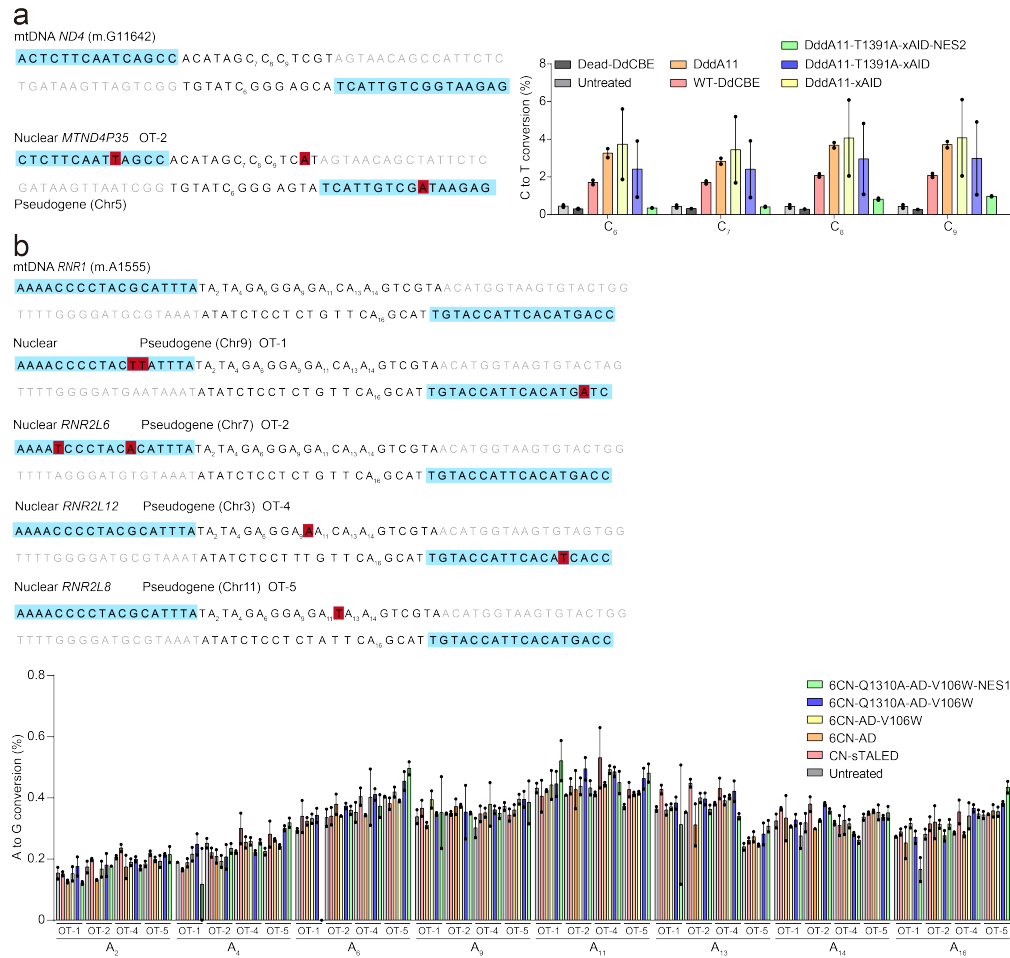


Figure S7. Off-target analysis for engineered TALEs specific to the *ND1* site. a) The average frequencies of mitochondrial genome-wide off-target editing induced by CN-sTALED, 6CN-AD, 6CN-AD-V106W, 6CN-Q1310A-AD-V106W, and 6CN-Q1310A-AD-V106W-NES1 specific to the *ND1* site. Error bars are s.e.m. for $n=2$ biologically independent samples. **b)** Mitochondrial genome-wide plots for A-to-G point mutations with frequencies $\geq 1\%$. Naturally occurring SNVs, on-target edits (including bystander edits in the editing window) and off-target edits are shown in blue and red, respectively. All data points from $n=2$ biologically independent experiments are shown. **c)** The corresponding nuclear DNA sequences with the high homology are shown for the *ND1* site. TALE binding sites begin at N0 and are shown in blue. Nucleotide mismatches between the mtDNA and nuclear pseudogene are in red. Editing efficiencies are measured by targeted deep sequencing (see Table S4 for primer sequences) (Supporting Information). Data are presented as means $\pm$ SEM.



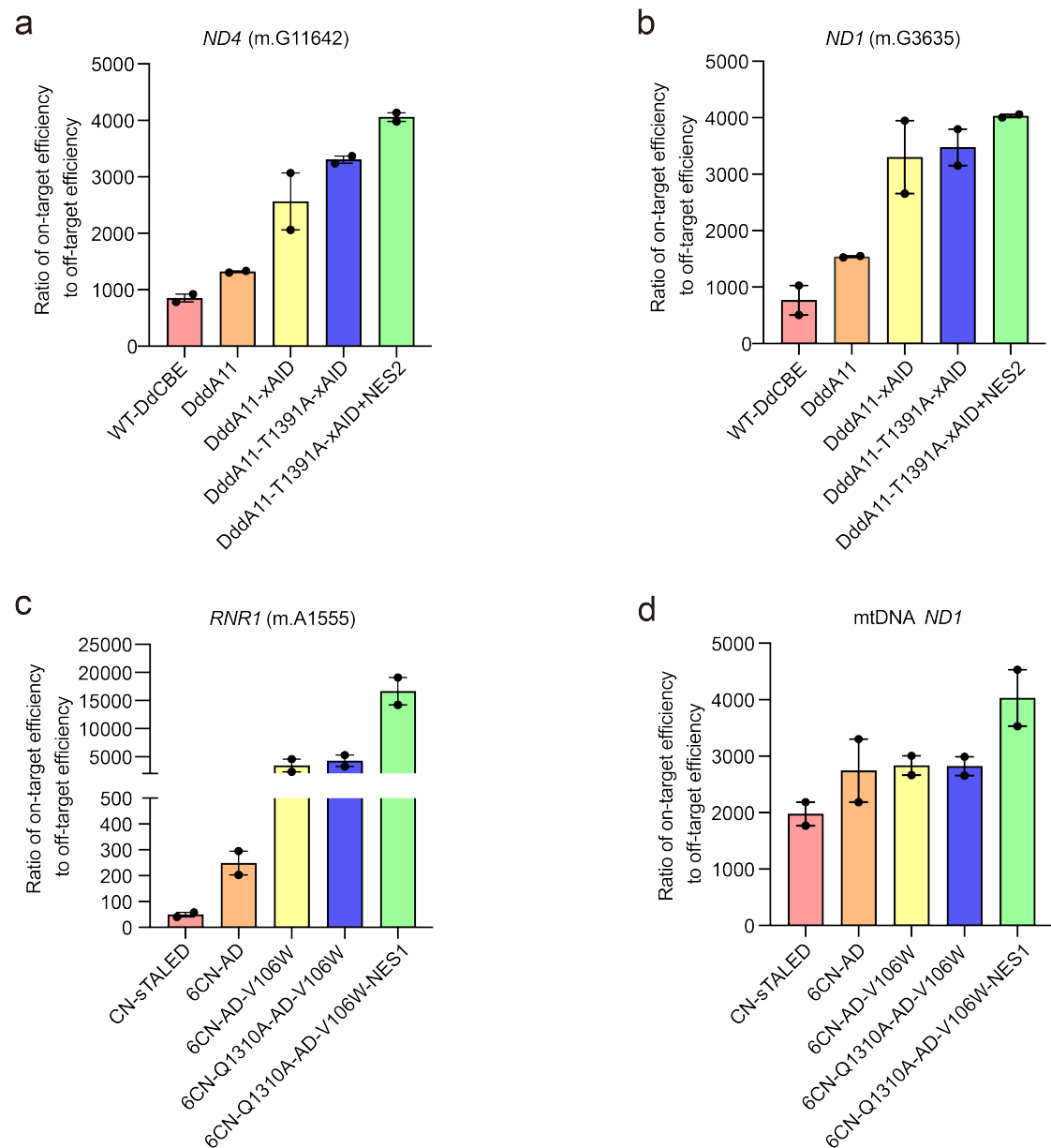


Figure S9. The ratio of on-target efficiency to off-target efficiency for each variant. a-b) The ratio of on-target efficiency to off-target efficiency for our engineered DdCBE variants at *ND4* (m.G11642) and *ND1* (m.G3635) sites. **c-d)** The ratio of on-target efficiency to off-target efficiency for our engineered TALED variants at *RNR1* (m.A1555) and *ND1* sites. All data points from $n=2$ biologically independent experiments are shown. Data are presented as means $\pm$ SEM.

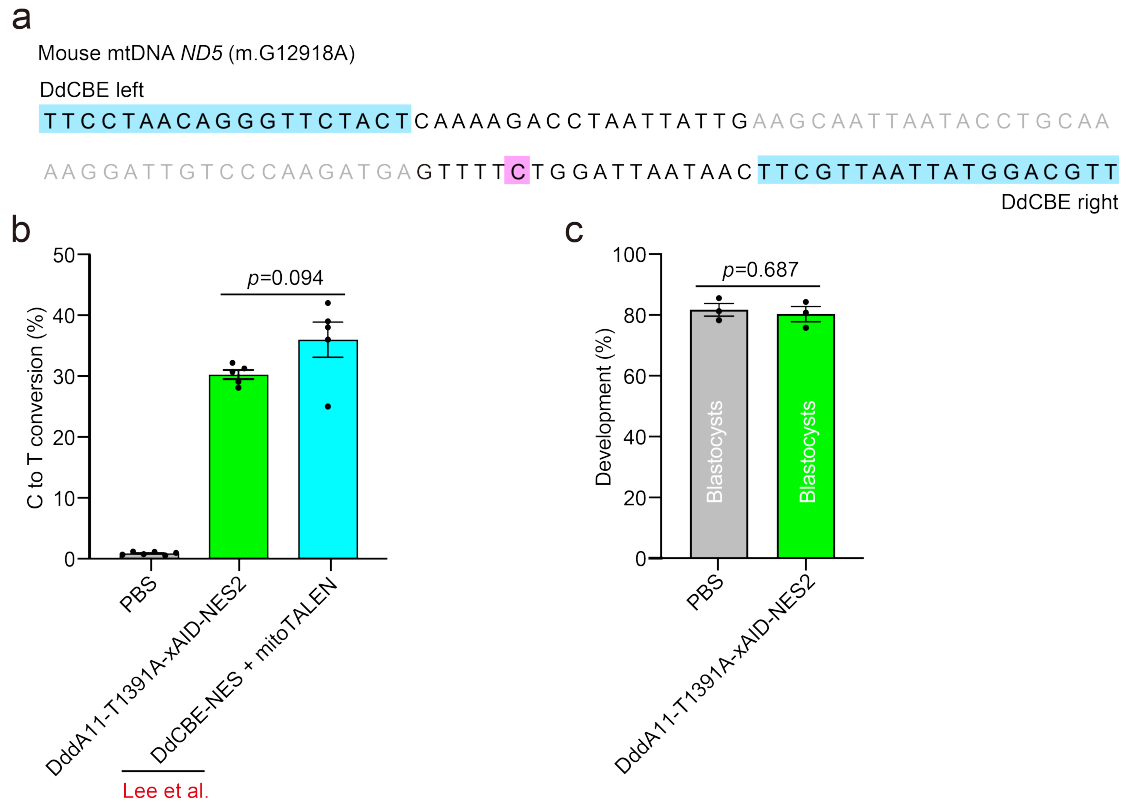


Figure S10. Engineered DdCBE variant mediates high editing efficiency in mouse blastocyst. a) Base editing target for generating the m.G12918A mutation. The TALE binding sequences for the DdCBE are highlighted in blue and for the base editing position in purple. **b)** The base editing efficiency of DddA11-T1391A-xAID-NES2 targeted *ND5* (m.G12918) in mouse blastocysts. The data of the DdCBE-NES + mitoTALEN group were cited from Lee et al.. **c)** Developmental rate for PBS-injected embryos and injected embryos with DddA11-T1391A-xAID-NES2 targeting *ND5* (m.G12918). Data are presented as the mean $\pm$ SEM. p values were evaluated with unpaired Student's t-test.

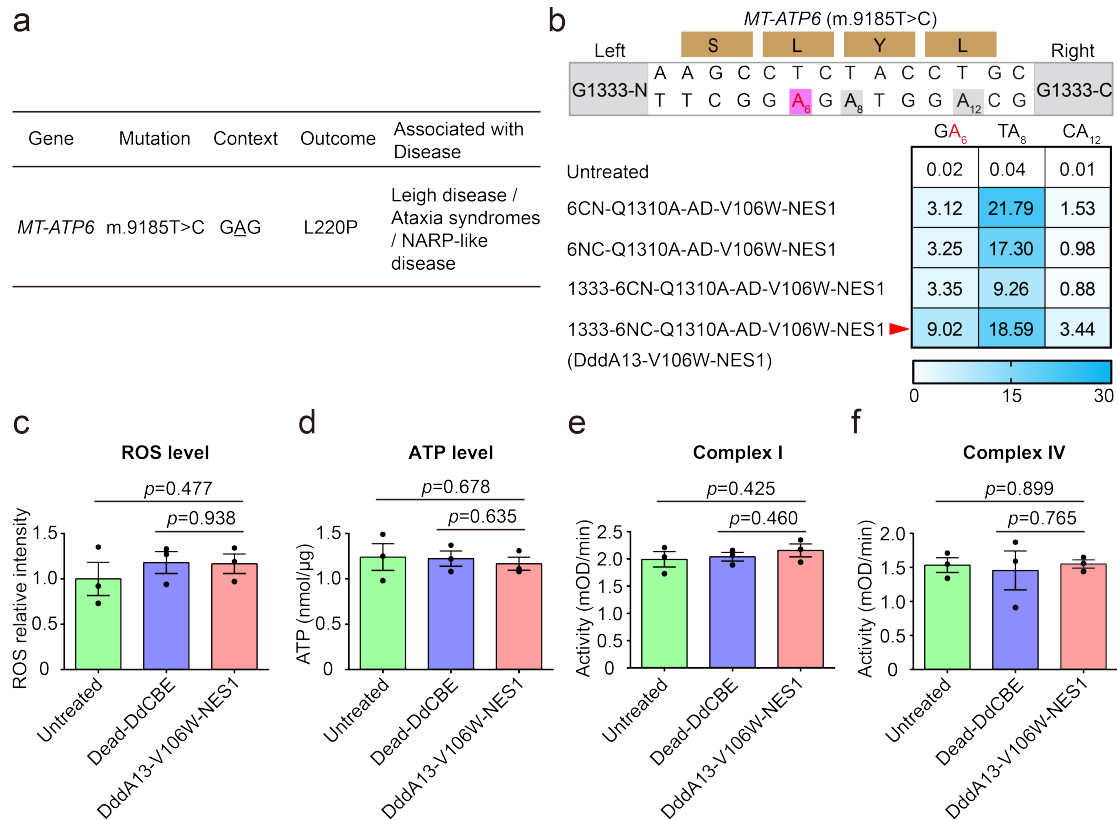


Figure S11. Application of high-fidelity TALED variant to install pathogenic mutations in HEK293T cells. **a)** Use high-fidelity TALED variant to install disease-associated target mutations in human mtDNA (L, leucine; P, proline). **b)** Mitochondrial A-to-G editing efficiencies of HEK293T cells treated with 6CN-Q1310A-AD-V106W-NES1 variant at G1397 and G1333 orientation of split DddA_{tox} for the mitochondrial *ATP6* (m.T9185) site. On-target adenines are colored red or gray, respectively. Top 30% of EGFP- and mCherry-double positive cells expressing the TALED variants were isolated by FACS for targeted deep sequencing. The split orientation, target spacing region and corresponding encoded amino acids are shown. 6CN-Q1310A-AD-V106W-NES1, Right-G1397-N + Left-G1397-C orientation; 6NC-Q1310A-AD-V106W-NES1, Right-G1397-C + Left-G1397-N orientation; 1333-6CN-Q1310A-AD-V106W-NES1, Right-G1333-N + Left-G1333-C orientation; 1333-6NC-Q1310A-AD-V106W-NES1 (DddA13-V106W-NES1), Right-G1333-C + Left-G1333-N orientation. Shown are means $\pm$ SEM; $n = 3$ independent experiments. The transfection time was 2 days. **c-f)** The levels of intracellular reactive oxygen species (ROS) (**c**), ATP (**d**) and the activities of complex I (**e**), complex IV (**f**) in sorted HEK293T cells treated with the DddA13-V106W-NES1 or Dead-DdCBE for the *ATP6* (m.T9185) site. Data are presented as means $\pm$ SEM. p values were evaluated with the unpaired student's t-test (two-tailed). All data points from $n = 3$ biologically independent experiments are shown.

Table S1. TALE binding sites for DdCBEs, TALEs, and mitoBEs.

Site	Left-mitoTALE target sequence	Right-mitoTALE target sequence
TRNL1	5'-T TTGTTAAGATG-3'	5'-T TAAGTTTATGCGAT-3'
ND1 (m.G3635)	5'-T CCTATTTATTCTAGCC-3'	5'-T CAGAGGATTGAGTAAAC-3'
ND1 (m.G3697)	5'-T CAAACTCAAACACGC-3'	5'-T TTGGGCTACTGCTCGC-3'
ND4 (m.G11642)	5'-T ACTCTTCAATCAGCC-3'	5'-T GAGAATGGCTGTTACT-3'
ND4	5'-T GCTAGTAACACGTTCT-3'	5'-T CCTGTAAGTAGGAGAGT-3'
RNR1	5'-T AAAACCCCTACGCATTTA-3'	5'-T CCAGTACACTTACCATGT-3'
ND6	5'-T GACCCCCATG-3'	5'-C GATGGCTATTG-3'
ND1	5'-T CTAGCCTAGCCGTTT-3'	5'-T GAGTTTGATGCTCACCCCT-3'
ATP6 (m.T9185)	5'-T TTTCACACTTCTAGT-3'	5'-T CATTATGTGTTGTCGT-3'
TRNG/TG (m.T10010)	5'-T CTTACTCTTTTAGTAT-3'	5'-T AGTTAATTGGAAGT-3'
mouse ND5 (m.G12918)	5'-T TCCTAACAGGGTTCTACT-3'	5'-T TGCAGGTATTAATTGCTT-3'

Table S2. Primers for HTS of on-target sites targeted by DdCBEs, TALEDs, and mitoBEs.

Site	Forward primer	Reverse primer
TRNL1	GAATTCNNNNNNGGATCCCTCAACTT AGTATTATACCC	GGGTATGTTGTTAAGAAGAG
ND1 (m.G3635)	GAATTCNNNNNNGGATCCCCTGGTC AACCTCAACCTAG	TTTGGGCTACTGCTCGCAGT
ND1 (m.G3697)	ACCTCTAGCCTAGCCGTTTA	GAATTCNNNNNNGGATCCGGCTAG GGTGACTTCATATG
ND4 (m.G11642)	GAATTCNNNNNNGGATCCGCCTACG ACAAACAGACCTA	ATTATGAGAATGACTGCGCC
ND4	GAATTCNNNNNNGGATCCTAGCAAG CCTCGCTAACCTC	GGTAAATATGTAGAGGGAGT
RNR1	GAATTCNNNNNNGGATCCCCTCAAG TATACTTCAAAGG	GCTTTGTGTTAAGCTACACT
ND6	GAATTCNNNNNNGGATCCTCGCTAA CCCCACTAAAACA	TATTTAGGGGGAATGATGGT
ND1	GAATTCNNNNNNGGATCCCCTGGTC AACCTCAACCTAG	TTTGGGCTACTGCTCGCAGT
ATP6 (m.T9185)	GAATTCNNNNNNGGATCCCTATCCTA GAAATCGCTGTC	GGGTTTACTATATGATAGG
TRNG/TG (m.T10010)	GAATTCNNNNNNGGATCCGACTATTT CTGTATGTCTCC	TAAGGCGAAGTTTATTACTC
mouse ND5 (m.G12918)	GAATTCNNNNNNGGATCCGACGAAC AAGACATCCGAAA	GAATTCNNNNNNGGATCCGCTGTT ATAGAAGTGGCGATTA

Table S3. Primers for long-range PCR of whole mitochondrial genome as two amplicons.

1st PCR	Range	1nd PCR forward (5'-3')	2nd PCR reverse (5'-3')
mitoWGS1	2478-10858 (8381bp)	GCAAATCTTACCCCGCCTG	AATTAGGCTGTGGGTGGTTG
mitoWGS2	10653-2688 (8605bp)	GCCATACTAGTCTTTGCCGC	GGCAGGTCAATTTCACTGG

Table S4. Primers for HTS of off-target sites in pseudogenes.

Site	Forward primer (5'-3')	Reverse primer (5'-3')
ND1 (m.G3635) Pseudogene (OT1)	GAATTCNNNNNNGGATCC CCCTATACCAGATCCTCTAA	GCGCCGATTAGTGCATAATT
ND1 (m.G3635) Pseudogene (OT2)	GAATTCNNNNNNGGATCC CCTATACCCCATCCCCTAAT	GTGTGCCGATCAGTACATAA
ND1 (m.G3635) Pseudogene (OT3)	GAATTCNNNNNNGGATCC CCCTATACCAGATCCTCTAA	TGTAATGCGCCAAATAGTGC
ND1 (m.G3635) Pseudogene (OT4)	GAATTCNNNNNNGGATCC CCCTATACCAGACCCTCTAA	GTAATGTGCCAATTAGTACA
ND1 (m.G3635) Pseudogene (OT5)	GAATTCNNNNNNGGATCC TCCCCATACCTTACCCCTTG	GCAATGTGCCGATCAAGGCA
ND4 (m.G11642) Pseudogene (OT1)	GAATTCNNNNNNGGATCC ACTGCGCCGGTGAAGCTTCA	CATCTGCCTGCGACAAACAG
ND4 (m.G11642) Pseudogene (OT2)	GAATTCNNNNNNGGATCC AGAATGACTGCGCCGGTGAA	TCTGCCTACGACAAACAGAC
RNR1 Pseudogene (OT1)	GAATTCNNNNNNGGATCC CACCCCTCCTCAAATATTACT	GCTTTGTGTTAAGCTACACC
RNR1 Pseudogene	GAATTCNNNNNNGGATCC	CTTTGTGTTAAGCTACACCT

(OT2)	TCACCCTCCTCAAGTATCAC	
RNR1 Pseudogene (OT3)	GAATTCNNNNNNGGATCC AGCTACACCTTGGTTCGTCC	GTCAATGATTTTGCATCAA
RNR1 Pseudogene (OT4)	GAATTCNNNNNNGGATCC AGCTACACTCTGGTTCGTCC	CCCGTCACCCTCCTCAAGTA
RNR1 Pseudogene (OT5)	GAATTCNNNNNNGGATCC GCTTTGTGTTAAGCTACACT	CGCCCGTCACCCTCCTCAAG
ND1 Pseudogene (OT1)	GAATTCNNNNNNGGATCC CTTAACTTAGGCCTTCTATT	GTTTGGGCAACTGCTCATAA
ND1 Pseudogene (OT2)	GAATTCNNNNNNGGATCC TATAGGCCTCCTATTTATAC	CCTGGGCCACAGCTCGTAAT
ND1 Pseudogene (OT3)	GAATTCNNNNNNGGATCC AGCTCGTAATGCATCGATTA	CCAAATCCCCTAATTAATTT
ND1 Pseudogene (OT4)	GAATTCNNNNNNGGATCC GGCTCGTAATATGCCGATCA	CCGGATCCCCTAATTAATTT
ND1 Pseudogene (OT5)	GAATTCNNNNNNGGATCC GCCTAAACATAGGAATGCTA	GTTTGTGCTGCTGCTCATAG

Table S5. Target protospacers and amplicons used in the orthogonal R-loop assays with corresponding primers used for genomic DNA amplification.

Site	Protospacer	PAM	Amplicon	HTS_fwd	HTS_rev
EMX1	GAGTCCGAGCA GAAGAAGAA	GGG	<u>GAATTCNNNNNNGGATCCGTTCCA</u>	GAATTCNNNNNNG	TTGGCCTGCTTC
			<u>GAACCGGAGGACAAA</u> AGTACAAAC	GATCCGTTCCAGA	GTGGCAAT
			GGCAGAAGCTGGAGGAGGAAGGG	ACCGGAGGACAA	
			CCTGAGTCCGAGCAGAAGAAGAA		
			GGGCTCCCATCACATCAACCGGTG		
Sa-site1	GTGGTAGACAGC ATGTGTCCTA	AAG GGT	<u>GAATTCNNNNNNGGATCCTTGGGT</u>	GAATTCNNNNNNG	AGAAGGACCCC
			<u>CTACCTTACTGAGA</u> AAAATGGCCCCA	GATCCTTGGGTCTA	TGTATTTT
			GGTCATTGTTCATGTCCAGTTGTGGT	CCTTACTGAGA	
			AGACAGCATGTGTCCTAAAGGGTAT		
			ATTCACATGCATGTGC <u>AAAAATACA</u> <u>GGGGTCCTTCT</u>		
Sa-site2	ATTACAGCCTG GCCTTTGGGG	TCG GGT	<u>GAATTCNNNNNNGGATCCAGAAGG</u>	GAATTCNNNNNNG	TCACCTCACGGC
			<u>GCAGGTTCCCCGAGG</u> CGCCAGAC	GATCCAGAAGGGC	CCCGCCCTT
			ACCCAATCCTCCCGGTGACATTTAC	AGGTTCCCCGAG	
			AGCCTGGCCTTTGGGGTCGGGTCA		
			ACGCTAGGCTGGCAGGGG <u>AAGGGC</u> <u>GGGGCCGTGAGGTGA</u>		
Sa-site3	GGTGGAGGAGG GTGCATGGGGT	CAG AAT	<u>GAATTCNNNNNNGGATCCTCATAG</u>	GAATTCNNNNNNG	CACTAGAAATTT
			<u>AATCCTGGACAAGG</u> TTGAAGGAC	GATCCTCATAGAAT	TCTCTATCCACC
			AGGTAGGATTTGGGTGGGTGGAGG	CCTGGACAAGG	
			AGGGTGCATGGGGTCAGAATTGTA		
			ACCGAAAACATTC <u>CCAGGTGGAT</u> <u>AGAGAAAATTCTAGTG</u>		
Sa-site4	TCTGCTTCTCCA GCCCTGGC	CTG GGT	<u>GAATTCNNNNNNGGATCCGCTATT</u>	GAATTCNNNNNNG	CTTCCTTTCCTC
			<u>TCTGCTGCAAGTA</u> AGCATGCATTG	GATCCGTCTATTTC	TGCCATCAC
			TAGGCTTGATGCTTTTTTTCTGCTTC	TGCTGCAAGTA	
			TCCAGCCCTGGCCTGGGTCAATCCT		
			TGGGGCCCAGACTGAGCAC <u>GTGAT</u> <u>GGCAGAGGAAAGGAAG</u>		
Sa-site5	GATGTTCCAATC AGTACGCA	GAG AGT	<u>GAATTCNNNNNNGGATCCGATTGA</u>	GAATTCNNNNNNG	GGTGCGCAAG
			<u>CATGCATTTGAC</u> CAATAGCATTGC	GATCCGATTGACAT	GCCCTACTT
			AGAGAGGCGTATCATTTTCGCGGATG	GCATTTGAC	
			TTCCAATCAGTACGCAGAGAGTCG		
			CCGTCTCCAAGGTGAAAGCGG <u>AAG</u> <u>TAGGGCCTTCGCGCAC</u>		

Supplementary sequences.

Sequences used to construct DdCBE, TALE, and mitoBE architectures, and TALE amino acid sequences.

***SOD2* MTS**

LSRAVCGTSRQLAPVLGYLGSRQKHSLPD

***COX8A* MTS**

SVLTPLLLRGLTGSARRLPVPRAKIHSL

3×HA

YPYDVDPDYAGYPYDVDPDYAGYPYDVDPDYA

3×FLAG

DYKDHDGDYKDHDIDYKDDDDK

TALE N-terminal

MDIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPA
ALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPL
QLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLN

TALE C-terminal

SIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLG

G1333 DddA_{tox}-N

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGG

G1333 DddA_{tox}-C

PTYPNYANAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLP
ENAKMTVVPPEGAIPVKRGATGETKVFTGNSNSPKSPTKGGC

G1397 DddA_{tox}-N

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMETLLPENAKMTV
VPPEG

G1397 DddA_{tox}-C

AIPVKRGATGETKVFTGNSNSPKSPTKGGC

***SOD2* 3'UTR**

ACCACGATCGTTATGCTGATCATACCCTAATGATCCCAGCAAGATAATGTCC
TGTCTTCTAAGATGTGCATCAAGCCTGGTACATACTGAAAACCCTATAAGGT
CCTGGATAATTTTTGTTTGATTATTCATTGAAGAAACATTTATTTTCCAATTG
TGTGAAGTTTTTGACTGTTAATAAAAGAATCTGTCAACCATCAAAAAAAAAA
AAAAAA

***ATP5B* 3'UTR**

ACCACGATCGTTATGCTGATCATACCCTAATGATCCCAGCAAGATAATGTCC
TGTCTTCTAAGATGTGCATCAAGCCTGGTACATACTGAAAACCCTATAAGGT
CCTGGATAATTTTTGTTTGATTATTCATTGAAGAAACATTTATTTTCCAATTG

TGTGAAGTTTTTGGACTGTTAATAAAAGAATCTGTCAACCATCAAAAAAAAAA
AAAAAA

UGI

TNLSDIIEKETGKQLVIQESILMLPEEVVEEVIGNKPESDILVHTAYDESTDENVM
LLTSDAPEYKPWALVIQDSNGENKIKML

P2A

ATNFSLLKQAGDVEENPGP

EGFP

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKFICTTGK
LPVPWPPTLVTTLTLYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSK
DPNEKRDHMLLEFVTAAGITLGMDELYK

mCherry

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLK
VTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFE
DGGVVTVTQDSSLQDGEFIYKVKLKRGTFNPSDGPVMQKKTMGWEASSERMY
PEDGALKGEIKQRLKLDGGHYDAEVKTTYKAKKPVQLPGAYNVNIKLDITS
HNEDYTIVEQYERAEGRHSTGGMDELYK

MutH

MSQPRPLLSPPETEEQLLAQAQQLSGYTLGELAALVGLVTPENLKRDKGWIG
VLEIWLGLASAGSKPEQDFAALGVELKTIPVDSLGRPLETTFVCVAPLTGNSG
VTWETSHVRHKLKRWLVIPVEGERSIPLAQRVGSPLLWSPNEEEDRQLRED
WEELMDMIVLGQVERITARHGEYLQIRPKAANAKALTEAIGARGERILTLP
RG FYLKKNFTSALLARHFLIQ

2aa linker

GS

3aa linker

GSG

4aa linker

SGGS

16aa linker

SGSETPGTSESATPES

The sequences of cytosine deaminase with ssDNA activity.

BE3

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKRETCCLYEINWGGRHSIWRHT
SQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYP

HVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYS
NEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ
RLPPHILWATGLK

YE1

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHT
SQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSYSPCGECSRAITEFLSRYPH
VTLFIYIARLYHHADPENRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYS
EAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQR
LPPHILWATGLK

R132E

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHT
SQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYP
HVTLFIYIARLYHHADPRNRQGLEDLISSGVTIQIMTEQESGYCWRNFVNYS
NEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQ
RLPPHILWATGLK

A3A-130F (A130F)

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHR
GFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWG
CAGEVRAFLQENTHVRLRIFAARIFDYDPLYKEALQMLRDAGAQVSIMTYDE
FKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

A3A-57G (N57G)

MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHR
GFLHGQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFISWSPCFSWG
CAGEVRAFLQENTHVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDE
FKHCWDTFVDHQGCPFQPWDGLDEHSQALSGRLRAILQNQGN

FE1

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHT
SQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSFSPCGECSRAITEFLSRYPH
VTLFIYIARLYHHADPENRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYS
EAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQR
LPPHILWATGLK

evoYE1

MSSGTGPVAVDPTLRRRIEPHEFEVFFDPRELKTCCLYEINWGGRHSIWRHT
SQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSYSPCGECSRAITEFLSRYPN
VTLFIYIARLYHLANPENRQGLRDLISSGVTIQIMTEQESGYCWHNFVNYS
ESHWPYPHILWVRLYVLELYCIILGLPPCLNILRRKQSQLTSFTIALQSCHYQRL
PPHILWATGLK

xAID

MTMDSMLLKRNKFIYHYKNLRWARGRHETLYCYIVKRRYSSVSCALDFGYL
RNRNGCHAEMFLRYLSIWVGHDPHRNYRVTFSSWSPCYDCAKRTLEFLK

GHPNFSLRIFSARLYFCEERNAEPEGLRKLQKAGVRLSVMSYKDYFYCWNTF
VETRESGFEAWDGLHENSURLARKLRRILQPPYDMEDLREVFVLLGL

hAID

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRN
KNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP
NLSLRIFTARLYFCEDRKAPEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENH
ERTFKAWEGLHENSURLSRQLRRILLPLYEVDDLRFDAFRTLGL

PmCDA1

MAGSTDAEYVRIHEKLDIYTFKKQFFNNKKSVMHRCYVLFELKRRGERRACF
WGYAVNKPQSGTERGHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAE
KILEWYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVS
EHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPA
V

Nuclear export signal (NES) sequences.

NES1

LPPLERLTL

NES2

LQKKLEELELD

NES3

VDEMTKKFGTLTIHDTEK

NES4

VDEMTKKFGTLTIHDTEKGSLQLPPLERLTL

NES5

VDEMTKKFGTLTIHDTEKGSLQKKLEELELD

The sequences of TadA8e and their variants.

TadA8e (AD)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNSKRGAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQ
VFNAQKKAQSSIN

TadA8e-V106W (AD-V106W)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGWRNSKRGAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQ
VFNAQKKAQSSIN

TadA8e-N46L (AD-N46L)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWLRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNSKRGAAGSLMNVLNYPGMNHRVEITEGILADECAALLCDFYRMPRQ
VFNAQKKAQSSIN

TadA8e-N108Q+L145T (AD-N108Q+L145T)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRQSKRGAAGSLMNVLNYPGMNHRVEITEGILADECAALTCDFYRMPRQ
VFNAQKKAQSSIN

TadA8e-R111T+N127K+Q154R (AD-R111T+N127K+Q154R)

MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGWNRAIGL
HDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCAGAMIHSRIGRVV
FGVRNSKTGAAGSLMNVLNYPGMKHRVEITEGILADECAALLCDFYRMPRR
VFNAQKKAQSSIN

The sequences of full-length DddA_{tox} and their variants.

DddA_{tox}

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTV
VPPEGAIPVKGATGETKVFTGNSNSPKSPTKGGC

DddA_{tox}-Q1310A

GSYALGPYQISAPQLPAYNGATVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTV
VPPEGAIPVKGATGETKVFTGNSNSPKSPTKGGC

DddA_{tox}-K1389A

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAAMTV
VPPEGAIPVKGATGETKVFTGNSNSPKSPTKGGC

DddA_{tox}-T1391A

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMAV
VPPEGAIPVKGATGETKVFTGNSNSPKSPTKGGC

DddA_{tox}-V1411A

GSYALGPYQISAPQLPAYNGQTVGTFYYVNDAGGLESKVFSSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNM TETLLPENAKMTV
VPPEGAIPVKGATGETKAFTGNSNSPKSPTKGGC

DddA6

GSYALGPYQISAPQLPAYNGRTVGTFFYYVNDAGGLESKVFISSGGPTYPNYA
NAGHVEGQSALFMRDNGISEGLVFHNNPEGTCGFCVNMIIETLLPENAKMTV
VPPEGAIPVKRGATGETKVFIGNSNSPKSPTKGGC

DddA11

GSYALGPYQISAPQLPAYNGQTVGTFFYYVNDAGGLESKVFISSGGPTYPNYV
SAGHVEGQSALFMRDNGISEGLVFHNNPKGTCTGFCVNMIIETLLPENAKMTV
VPPEGAIPVKRGATGETKVFIGNSNSPKSPTKGGC

The general architecture of left-side halves of DdCBEs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - DddA_{tox} half - 4aa linker - 1×UGI - (4aa linker - NES) - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

The general architecture of right-side halves of DdCBEs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - DddA_{tox} half - 16aa linker - cytosine deaminase with ssDNA activity - 4aa linker - 1×UGI - (4aa linker - NES) - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

or

The general architecture of left-side halves of DdCBEs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - DddA_{tox} half - 16aa linker - cytosine deaminase with ssDNA activity - 4aa linker - 1×UGI - (4aa linker - NES) - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

The general architecture of right-side halves of DdCBEs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - DddA_{tox} half - 4aa linker - 1×UGI - (4aa linker - NES) - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

The general architecture of left-side halves of TALEDs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - DddA_{tox} half - 16aa linker - TadA8e - (4aa linker - NES) - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

The general architecture of right-side halves of TALEDs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - DddA_{tox} half - (4aa linker - NES) - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

or

The general architecture of left-side halves of TALEDs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - DddA_{tox} half - 3aa linker - (4aa linker - NES) - P2A - EGFP - *ATP5B* 3'UTR**

The general architecture of right-side halves of TALEDs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - DddA_{tox} half - 16aa linker - TadA8e - (4aa linker - NES) - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

mitoCBE (right MutH)

The general architecture of left-side halves of mitoCBEs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - rAPOBEC1 - 4aa linker - 2×UGI - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

The general architecture of right-side halves of mitoCBEs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - MutH - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

mitoABE (left MutH)

The general architecture of left-side halves of mitoABEs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - MutH - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

The general architecture of right-side halves of mitoABEs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - TadA8e-V106W - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

mitoABE (right MutH)

The general architecture of left-side halves of mitoABEs (from N- to C-terminus):
***SOD2* MTS - 3×HA - mitoTALE - 2aa linker - TadA8e-V106W - 3aa linker - P2A - EGFP - *ATP5B* 3'UTR**

The general architecture of right-side halves of mitoABEs (from N- to C-terminus):
***COX8A* MTS - 3×FLAG - mitoTALE - 2aa linker - MutH - 3aa linker - P2A - mCheery - *SOD2* 3'UTR**

TRNL1-DdCBE Left mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASNGGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNGGGGKQALETVQRLLPVLCQAHGLTPDQVVAIA
SNGGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGGKQALETVQRLLP
VLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGGKQALETVQRLLPVLC
QAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG**

GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGRPALESIVAQLSRPDP
ALAALTNDHLVALACLGGRPALDAVKKGLG

TRNL1-DdCBE Right mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIA
SNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLP
VLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN
GGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASN
NNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QAHGLTPDQVVAIASNNGGRPALESIVAQLSRPDPALAALTNDHLVALACL
GGRPALDAVKKGLG

ND1 (m.G3635)-DdCBE/TALED Left mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
CQAHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN
GGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAH
GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQA
LETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLT
PDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGRPALE
SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG

ND1 (m.G3635)-DdCBE/TALED Right mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIA
SNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASN
GGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQ

AHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAH
GLTPDQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALET
VQRLLPVLCQAHGLTPDQVVAIASHDGGRPALESIVAQLSRPDPALAALTND
HLVALACLGGRPALDAVKKGLG

ND1 (m.G3697)-DdCBE/TALED/mitoABE Left mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVL
CQAHGLTPDQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHD
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGK
QALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQAL
ETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTP
AQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGRPALESI
VAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG

ND1 (m.G3697)-DdCBE/TALED/mitoABE Right mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASNNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVL
CQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG
GGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGG
KQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAH
GLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGL
TPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGRPAL
ESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG

ND4 (m.G11642)-DdCBE Left mitoTALE repeat:

DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA

IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG~~G~~GKQALETVQRLLPVLC
QAHGLTPDQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHGLTPAQVVAIASHD
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG~~G~~G
KQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG~~G~~GKQ
ALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGL
TPAQVVAIASHDGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPAL
DAVKKGLG

ND4 (m.G11642)-DdCBE Right mitoTALE repeat:

DIADLRTLGYSQQQQEKIKPKVRS**TVAQH**HEALVGHGF**THAHIVALSQH**
PAALGTVAVKYQDMIAAL**PEATHEAIVGVGKQWSGARALEALLTVAGE**
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALT**GAPLN**LT**PDQVVA**
IASNN~~G~~GKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNN~~G~~GKQALETVQRLLPVLCQAHGLTPDQVVAIAS
SNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QAHGLTPDQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHGLTPDQVVAIASNN
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN~~G~~GKQALETVQRLLPVLCQ
AHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG~~G~~G
KQALETVQRLLPVLCQAHGLTPDQVVAIASNN~~G~~GKQALETVQRLLPVLCQAH
GLTPDQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHGLTPDQVVAIASNG~~G~~GKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLT
PDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNG~~G~~GGRPALES
IVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG

ND4-DdCBE/TALED/mitoBE Left mitoTALE repeat:

DIADLRTLGYSQQQQEKIKPKVRS**TVAQH**HEALVGHGF**THAHIVALSQH**
PAALGTVAVKYQDMIAAL**PEATHEAIVGVGKQWSGARALEALLTVAGE**
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALT**GAPLN**LT**TPEQVVA**
IASNN~~G~~GKQALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLL
PVLCQAHGLTPQQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHGLT**TPEQVVAIA**
SNIGGKQALETVQALLPVLCQAHGLT**TPEQVVAIAS**NN~~G~~GKQALETVQALLPV
LCQAHGLT**TPEQVVAIAS**NG~~G~~GKQALETVQALLPVLCQAHGLT**TPEQVVAIAS**N
IGGKQALETVQRLLPVLCQAHGLT**TPEQVVAIAS**NIGGKQALETVQALLPVLCQ
AHGLT**TPEQVVAIAS**HDGGKQALETVQRLLPVLCQAHGLT**TPEQVVAIAS**HDGG
KQALETVQALLPVLCQAHGLT**TPEQVVAIAS**NIGGKQALETVQALLPVLCQAH
GLT**TPEQVVAIAS**HDGGKQALETVQALLPVLCQAHGLT**TPEQVVAIAS**NN~~G~~GKQ
ALETVQRLLPVLCQAHGLTPQQVVAIASNG~~G~~GKQALETVQRLLPVLCQAHG
LT**TPEQVVAIAS**NG~~G~~GKQALETVQALLPVLCQAHGLT**TPEQVVAIAS**HDGGKQA
LETVQALLPVLCQAHGLTPQQVVAIASNG~~G~~GGRPALESIVAQLSRPDPALAAL
TNDHLVALACLGGRPALDAVKKGLG

ND4-DdCBE/TALED/mitoBE) Right mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTPEQVVA
IASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLL
PVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIA
SNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLP
VLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPV
CQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNI
GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGGKQALETVQALLPVLC
QAHGLTPEQVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIG
GKQALETVQALLPVLCQAHGLTPEQVVAIASNNGGKQALETVQRLLPVLCQ
AHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
AALTNDHLVALACLGGRPALDAVKKGLG**

mouse ND5 (m.G12918)-DdCBE Left mitoTALE repeat:

**GIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALS
QHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVA
GELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTTPAQV
VAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQ
RLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQV
VAIASNIGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNNGGKQALETVQR
LLPVLCQDHGLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVV
AIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRL
LPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAI
ASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP
VLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
NGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASH
DGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPV
CQDHGLTPEQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALAC
LGGRPALDAVKKGLGGS**

mouse ND5 (m.G12918)-DdCBE Right mitoTALE repeat:

**GIRIQDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALS
QHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVA
GELRGPPLQLDTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTTPAQV
VAIASNGGGKQALETVQRLLPVLCQDHGLTPAQVVAIASHDGGKQALETVQ
RLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVV
AIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRL
LPVLCQDHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVAI
ASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLP**

VLCQAHGLTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NGGKQALETVQRLLPVLCQDHGLTPAQVVAIASNGGKQALETVQRLLPV
LCQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASN
GGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPV
CQDHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPEQVVAIASN
GGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLC
QDHGLTPEQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLVALACL
GGRPALDAVKKGLGGS

RNR1-DdCBE/TALED Left mitoTALE repeat:

DIADLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTTPAQVVAI
ASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPV
CQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPDQVVAIASH
GGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGK
QALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQA
LETVQRLLPVLCQAHGLTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTP
DQVVAIASNGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGKQALET
VQRLLPVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPAQ
VVAIASNGGRPALSIVAQLSRPDPALAALTNDHLVALACL**GGRPALDAV**
KKGLG

RNR1-DdCBE/TALED Right mitoTALE repeat:

DIADLRTLGYSSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
NGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPV
CQAHGLTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDG
GKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGKQALETVQRLLPVLCQA
HGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGK
QALETVQRLLPVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQAL
ETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGLTP
DQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGKQALET
VQRLLPVLCQAHGLTPDQVVAIASNGGKQALETVQRLLPVLCQAHGLTPDQ
VVAIASNGGRPALSIVAQLSRPDPALAALTNDHLVALACL**GGRPALDAV**
KKGLG

ND6-TALED Left mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTTPQQVVA
IASNNGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASNIGGKQALETVQALL
PVLCQAHGLTPEQVVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVVAIA
SHDGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASHDGGKQALETVQRLLP
VLCQAHGLTPEQVVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVVAIAS
HDGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASNIGGKQALETVQALLPVL
CQAHGLTPQQVVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVVAIASN
NNGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLG**

ND6-TALED Right mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTTPQQVVA
IASNNGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASNIGGKQALETVQALL
PVLCQAHGLTPQQVVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVVAIA
SNNGGKQALETVQRLLPVLCQAHGLTPQQVVVAIASNNGGKQALETVQRLLP
VLCQAHGLTPEQVVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVVAIAS
NNGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASNIGGKQALETVQALLPVL
CQAHGLTPQQVVVAIASNNGGKQALETVQRLLPVLCQAHGLTPQQVVVAIASN
GGGGKQALETVQRLLPVLCQAHGLTPQQVVVAIASNNGRPALESIVAQLSRPD
PALAALTNDHLVALACLGGRPALDAVKKGLG**

ND1-TALED/mitoABE Left mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVVAI
ASHDGGKQALETVQALLPVLCQAHGLTPQQVVVAIASNNGGKQALETVQRLL
PVLCQAHGLTPQQVVVAIASNIGGKQALETVQRLLPVLCQAHGLTPEQVVVAIAS
NNGGKQALETVQALLPVLCQAHGLTPEQVVVAIASHDGGKQALETVQALLPVL
CQAHGLTPEQVVVAIASHDGGKQALETVQALLPVLCQAHGLTPEQVVVAIASNG
GGKQALETVQRLLPVLCQAHGLTPEQVVVAIASNIGGKQALETVQALLPVLCQ
AHGLTPEQVVVAIASNNGGKQALETVQRLLPVLCQAHGLTPEQVVVAIASHDGG
KQALETVQALLPVLCQAHGLTPEQVVVAIASHDGGKQALETVQALLPVLCQAH
GLTPEQVVVAIASNNGGKQALETVQALLPVLCQAHGLTPEQVVVAIASNNGGGKQ
ALETVQRLLPVLCQAHGLTPQQVVVAIASNNGGKQALETVQRLLPVLCQAHGL
TPQQVVVAIASNNGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPA
LDAVKKGLG**

ND1-TALED/mitoABE Right mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE**

**LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPEQVVAI
ASNGGGKQALETVQALLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIAS
NGGGKQALETVQALLPVLCQAHGLTPEQVVAIASNGGGKQALETVQALLPVL
CQAHGLTPEQVVAIASNGGGKQALETVQALLPVLCQAHGLTPEQVVAIASNN
GGKQALETVQRLLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQ
AHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNNGG
KQALETVQALLPVLCQAHGLTPEQVVAIASHDGGKQALETVQALLPVLCQAH
GLTPEQVVAIASNGGGKQALETVQALLPVLCQAHGLTPEQVVAIASHDGGKQ
ALETVQRLLPVLCQAHGLTPQQVVAIASNIGGKQALETVQRLLPVLCQAHGLT
PEQVVAIASHDGGKQALETVQALLPVLCQAHGLTPQQVVAIASHDGGKQALE
TVQRLLPVLCQAHGLTPQQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQ
QVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALDA
VKKGLG**

ATP6 (m.T9185)-TALED Left mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVL
CQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIG
GKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQA
HGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGK
QALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQAL
ETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTP
DQVVAIASNGGGRPALESIVAQLSRPDPALAALTNDHLVALACLGGRPALD
AVKKGLG**

ATP6 (m.T9185)-TALED Right mitoTALE repeat:

**DIADLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVL
CQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNN
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGG
KQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPAQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQ
ALETVQRLLPVLCQAHGLTPAQVVAIASHDGGKQALETVQRLLPVLCQAHGL**

TPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGRPAL
ESIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLG

TRNG/TG (m.T10010)-TALED Left mitoTALE repeat:

**DIADLRTLGYSSQQQQEKIKPKVIRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPDQVVA
IASHDGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLL
PVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIAS
NIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHDGGKQALETVQRLLPVLC
QAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASHD
GGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPAQVVAIASNGGG
KQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAH
GLTPAQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQA
LETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLT
PDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGRPAL
ESIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLG**

TRNG/TG (m.T10010)-TALED Right mitoTALE repeat:

**DIADLRTLGYSSQQQQEKIKPKVIRSTVAQHHEALVGHGFTHAHIVALSQH
PAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGE
LRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTDPDQVVA
IASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQALETVQRLL
PVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPDQVVAI
SNGGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNIGGKQALETVQRLLPV
LCQAHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNG
GGKQALETVQRLLPVLCQAHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQ
AHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGG
KQALETVQRLLPVLCQAHGLTPAQVVAIASNIGGKQALETVQRLLPVLCQAH
GLTPDQVVAIASNIGGKQALETVQRLLPVLCQAHGLTPDQVVAIASNNGGKQ
ALETVQRLLPVLCQAHGLTPDQVVAIASNGGGRPAL
ESIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLG**