이학박사학위논문

Nuclear and mitochondrial DNA editing with programmable DNA binding proteins

2022년 8월

서울대학교 대학원

화학부 생화학 전공

정 유 진

Nuclear and mitochondrial DNA editing with programmable DNA binding proteins

지도교수 김 석 희

이 논문을 이학박사학위논문으로 제출함

2022년 8월

서울대학교 대학원

화학과 생화학전공

정 유 진

정유진의 박사학위논문을 인준함

2022년 8월

위 원 장 이 연 (인)

부 위 원 장 김 석 희 (인)

위 원 김 성 연 (인)

위 원 이 현 우 (인)

위 원 김 형 범 (인)

Abstract

Nuclear and mitochondrial DNA editing with programmable DNA binding proteins

Eugene Chung

Department of Chemistry

The Graduate School

Seoul National University

Over the decades, genome engineering techniques with programmable DNA binding proteins had made remarkable progress. Transcription activator-like effector (TALE) proteins originated from the plant pathogenic bacteria genus *Xanthomonas* and naturally alter the transcription of genes in host plant cells. The TALE repeats comprise tandem arrays with 10 to 30 repeats that bind and recognize extended DNA sequences. CRISPR (Clustered Regulatory Interspaced Short Palindromic Repeats)/CRISPR-associated (Cas) system provides acquired immunity against invading foreign DNA via RNA-guided DNA cleavage in bacteria. These programmable DNA-binding proteins are versatile and robust tools that revolutionize biological research.

Programmable nucleases engineer the target genome by inducing double strand breaks (DSBs) relying on non-homologous end joining (NHEJ) and homologous recombination (HR) to induce genetic alterations. Therefore, this technology is limited by the availability of the desired DNA repair mechanism. On the other side, base editors (BEs) introduce targeted point mutations without inducing DSBs. $DddA_{tox}$, which originated from *Burkholderia cenocepacia,* catalyzes the deamination of cytidines within double strand DNA. Conventional $DddA_{tox}$ is split into two inactive halves to avoid its cytotoxicity and fused to TALE to make DddA-derived cytosine base editors (DdCBEs).

In this thesis, I will present the applications of the programmable DNA-binding proteins to engineer nuclear DNA and mitochondria DNA. In Chapter 1, I will demonstrate that the CRISPR-Cas9 technique combined with the somatic cell nuclear transfer can produce a dystrophin mutant dog. I validate the dystrophin mutant dog has proper phenotypes for a disease model.

In Chapter 2, I will describe modifying dimeric DdCBE to non-toxic, full-length monomeric DdCBE (mDdCBE). I apply mDdCBEs for mitochondrial base editing and validate off-target effects. Furthermore, I show the advantages of mDdCBEs by AAV experiment and targeting site for which only one TALE can be designed.

Keywords: CRISPR-Cas9, Gene Knockout, Dog disease model, TALE, DddA toxin, Base-editing, Mitochondrial DNA, Protein Engineering

Student Number: 2013-20282

Table of Contents

Chapter 2. Base editing in human mitochondrial DNA with monomeric DddA-TALE fusion deaminases

List of Figures

Chapter 1

Chapter 2

List of Tables

Chapter 1

Chapter 2

Chapter 1.

Generation of a dystrophin mutant dog using CRISPR-Cas9

Ⅰ. Introduction

Over the last decade, the astonishing development of genome engineering has revolutionized research on the mammal genome. Programmable nucleases are redefining the boundaries of biological research. In the early stage of genome editing, the engineered nucleases were composed of sequence-specific DNA-binding domains fused to a non-specific DNA cleavage module (Carroll, 2011; Urnov *et al*., 2010). Zinc finger nucleases (ZFNs), one of these early engineered nucleases, are assembled by fusing a non-sequence-specific cleavage domain to a site-specific DNA-binding domain that is loaded on the zinc finger (Bibikova *et al*., 2003; Doyon *et al*., 2008; Kim *et al*., 1996; Maeder *et al*., 2008; Urnov *et al*., 2005). The functional specificity of the designed zinc-finger domain comprises an array of Cys2His2 zinc fingers (ZFs), which are derived by highly conserved interactions of their zinc-finger domains with homologous DNA sequences (Beerli and Barbas, 2002). Each zinc-finger unit selectivity binds three base pairs (bp) of DNA (Fairall *et al*., 1993). After binding of the ZFs on either side of the site, the pair of FokI domains dimerize and cleave the DNA at the site (Bitinaite *et al*., 1998; Urnov *et al.*, 2005). Transcriptional activator-like effector nucleases (TALENs) are another type of engineered nuclease that demonstrate better specificity and efficiency than ZFNs in general. Similar to ZFNs, TALENs comprise a nonspecific DNA cleavage domain (FokI) fused to a customizable sequence-specific DNA-binding domain to generate DSBs (Boch *et al*., 2009; Cermak *et al*., 2011; Kim *et al*., 2013; Miller *et al*., 2011; Moscou and

Bogdanove, 2009). The DNA-binding domain of TALE consists of a highly conserved repeat protein sequence from a protein discovered in the phytopathogenic *Xanthomonas* bacteria that naturally alters the transcription of genes in host plant cells (Boch *et al.*, 2009). The central DNA-binding domain of TALE is composed of repeat monomers of 34 amino acids. The amino acids of each monomer are highly conserved, except for amino acid residues at positions 12 and 13 which is called repeat-variable diresidues (RVDs) (Boch *et al.*, 2009; Moscou and Bogdanove, 2009).

The recent discovery of bacterial adaptive immune systems known as clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems as programmable nuclease was a breakthrough in genome engineering. CRISPR was originally discovered in *E. coli*. in 1987, but the function of the short repeat sequences remained unclear for many years before several studies in 2005 characterized their similarities to phage DNA. Subsequent research revealed that these repeat sequences took a role in bacterial and archaea adaptive immune defense against offending foreign DNA by inducing RNA-guided DNA cleavage (Bolotin *et al*., 2005; Jinek *et al*., 2012; Pourcel *et al*., 2005). In early 2013, research groups demonstrated that heterologous expression of a CRISPR-Cas system from *Streptococcus pyogenes*, comprising the Cas9 protein along with guide RNA in mammalian cells results in DSBs at target sites (Cho *et al*., 2013; Cong *et al*., 2013; Hwang *et al*., 2013; Jiang *et al*., 2013; Jinek *et al.*, 2012; Kim *et al*., 2014; Mali *et al*., 2013). The CRISPR/Cas9 system has two components, a guide RNA which can be either two separate RNAs (crRNA and tracrRNA) or a single guide RNA (sgRNA)

and a Cas9 endonuclease. The sgRNA contains a unique 20 bp sequence designed to complement the target DNA site and this must be followed by a short DNA sequence downstream essential for the compatibility with the Cas9 protein. This adjacent essential sequence is termed the protospacer adjacent motif (PAM) in which SpCas9 has an “NGG” or “NAG” (Sternberg *et al*., 2014). DSBs occur via the formation of a ternary complex in which Cas9 binds to the PAM in the DNA, then binds to the nonprotospacer portion of the guide RNA, upon which the protospacer of the guide RNA hybridizes with one strand of the genomic DNA (Jiang *et al*., 2016). SpCas9 protein catalyzes the DSB in the DNA at a position of three base pairs upstream of the PAM (Jinek *et al.*, 2012; Kim and Kim, 2014; Shmakov *et al*., 2017). Unlike ZFNs and TALENs, which require cloning of protein sequences using large and repetitive DNA segments for each new target site, the CRISPR-Cas9 system can be easily adapted to target any genomic sequence by changing the protospacer of the guide RNA.

Programmable nucleases, including ZFN, TALEN, and CRISPR-Cas9, modify a target gene by DSB. When DSB is created precisely at the specific region, repair pathways get activated. There can be two cases; (a) The non-homologous end joining (NHEJ) and (b) homologous recombination (HR). Both DSB repair pathways take important parts in mammalian DSB repair (Chapman *et al*., 2012; Shrivastav *et al*., 2008). NHEJ modifies the broken DNA ends, and ligates them with no regard for homology, generating deletions or insertions (indels) (Lieber, 2008). So NHEJ is preferred for gain or loss of function experiments due to its mutagenic behavior of possible indels resulting in altered reading frames or large deletions. HR uses an

undamaged DNA template to repair the break (Thompson and Schild, 2001). HR usually has a reduced efficiency in editing a genome than NHEJ because of its need for a template and cell cycle dependence. HR is usually preferred for gene knock-out or knock-in applications (Pardo *et al*., 2009).

Dog, *Canis lupus familiaris,* has drawn considerable attention as a model for studying human diseases. Dogs show over 450 naturally occurring diseases, of which approximately 360 are analogous to human diseases (Wayne and Ostrander, 2007). Based on their size, biological features, and ease of behavioral evaluation and handling, dogs can be good animal models. Since humans and dogs share a common environment, food, and carcinogenic load, it is not surprising that the dog has emerged as a viable model for human disease (Glickman *et al*., 2004). The dog genome is less divergent from the human genome than from the mouse genome and more of the human genomic sequence can be aligned to the genome of the dog than that of the mouse (Hytonen and Lohi, 2016; Lindblad-Toh *et al*., 2005).

Duchenne muscular dystrophy (DMD) is a lethal progressive pediatric muscle disorder and is genetically inherited as an X-linked disease that caused by mutations in *DMD*. When *DMD* gene is disrupted for any reason, the muscle isoform of dystrophin cannot be produced (DMD) or truncated dystrophins are produced (causing Becker muscular dystrophy; BMD) (Wein *et al*., 2015; Yiu and Kornberg, 2015). Over time, DMD patients experience progressive functional impairments leading to loss of ambulation (LOA), pulmonary insufficiency, cardiomyopathy, and early mortality. For many years, researchers tried to find an effective therapy method,

but there is no absolute cure currently for patients with DMD (Mirski and Crawford, 2014; Szabo *et al*., 2021; Wein *et al.*, 2015; Yiu and Kornberg, 2015). Dystrophin-deficient mdx mice have been most commonly used for DMD research, but this model has limitations; mdx mice exhibit minimal clinical symptoms and have only a 25% reduction in longevity, unlike DMD patients, who have a 75% reduction in life span (Chamberlain *et al*., 2007) In addition, there is a weak correlation between the effect of therapeutic interventions in the rodent model and the effect observed in humans (McGreevy *et al*., 2015). Recently, canine models have been suggested as a suitable link between rodents and humans because those dystrophin-deficient dogs are genetically and phenotypically similar to human patients (Duan, 2011; 2015; Valentine *et al*., 1988). Canine X-linked muscular dystrophy (CXMD) models have been reported over the last 50 years. Mutations in the canine dystrophin gene have been identified in golden retrievers, German short-hair pointers, and Cavalier King Charles spaniels (Schatzberg *et al*., 1999; Sharp *et al*., 1992; Walmsley *et al*., 2010) However, it is difficult to produce dog pups with the same genetic background as the CXMD model using conventional breeding methods and maintain individuals for use in preclinical studies. For the availability of canine dystrophinopathy models, a corresponding loss-of-function model using targeted mutagenesis of the *DMD* gene was necessary.

Here I generate the dog model of dystrophinopathy using CRISPR-Cas9 system. It is shown that mutations can be introduced into canine fetal fibroblasts in either method of RNP or plasmids via electroporation. I present a report that the mutant dog pup can be produced by nuclear transfer using CRISPR-Cas9 treated

somatic cells. Furthermore, I demonstrate that the clinical characteristics of a dystrophin mutant dog are similar to that of human dystrophinopathy (Oh *et al*., 2022). These results propose that the CXMD model can be generated by CRISPR-Cas9 and is a more appropriate disease model than existing models.

Ⅱ. Materials and Methods

1. Ethics statement

In this study, female mixed dogs from 2 to 4 years of age were used as oocyte donors and embryo transfer recipients. The dogs were housed indoors and fed once a day with water ad libitum. All experiments involving animals, methods and protocols were approved by the Committee for Accreditation of Laboratory Animal Care and the Guideline for the Care and Use of Laboratory Animals of Seoul National University (SNU-1700310-14-1). All methods and protocols were carried out in accordance with the relevant guidelines and regulations.

2. Cell culture and transfection conditions

Canine fetal fibroblasts were derived from the torso of a 27-day fetus post coitum and were cultured in DMEM (Gibco) with 10% fetal bovine serum (FBS). To introduce DSBs using an RNP complex, Cas9 protein was premixed with *in vitro* transcribed sgRNA and incubated for 10 min at room temperature. To introduce DSBs using plasmids, p3s-Cas9-2A-GFP and pRG2-sgRNA were premixed. Canine fetal fibroblasts were electroporated with the Amaxa P3 Primary Cell 4D-Nucleofector Kit using Program EN-150 (Lonza). The transfected cells were transferred onto 12 well plates and further expansion was performed.

3. Purification of Cas9 protein

The pET plasmid that encodes His-tagged Cas9 was transformed into BL21(DE3). Expression of Cas9 was induced using 0.5 mM IPTG for 4 h at 25°C. The His-tagged Cas9 protein was purified using Ni-NTA agarose resin (Qiagen) and dialyzed against 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT and 10% glycerol (Jinek *et al*., 2012). The Cas9 protein was concentrated with centrifugal filter units (Millipore).

4. *In vitro* transcription of sgRNAs

RNAs were *in vitro* transcribed through run-off transcription reactions by T7 RNA polymerase (New England BioLabs). The template DNA contains a T7 RNA promoter sequence followed by protospacer and sgRNA scaffold sequence. The sgRNA transcription templates (200 mM) were mixed with $MgCl_2$ (14 mM), T7 RNA polymerase (500 units), RNase inhibitor (10 units) (New England BioLabs), NTP (4 mM) in a reaction mix volume of 100 μl for 9 h at 37 °C. To remove the template DNA after RNA synthesis, transcribed RNA mixture was pre-incubated with DNase I (4 units) (New England BioLabs) and purified using RNeasy Mini Kit (Qiagen) according to the manufacturer's manual.

5. *In vivo* matured oocytes collection

After vaginal bleeding was first shown, blood was collected daily from the cephalic vein and sera were separated by centrifuging for 10 min. Serum progesterone concentration was monitored by an IMMULITE 1000 (Siemens Healthcare Diagnostics Inc., Flanders, NJ, USA). The day when the progesterone concentration reached 4.0 ng/ml to 10.0 ng/ml was considered the day of ovulation. 72 h after ovulation, *in vivo* matured oocytes were collected by oviductal flushing using HEPES-buffered tissue culture medium-199 (TCM, Invitrogen) supplemented with 10% bovine serum albumin and 2 mM NaHCO3.

6. Somatic cell nuclear transfer and embryo transfer

Cumulus cells were removed from *in vivo* matured oocytes by gentle pipetting in tissue culture medium-199 supplemented with 0.1% hyaluronidase. Metaphase chromosomes and extruded first polar bodies were removed under ultraviolet light by aspiration in HEPES-buffered TCM drops containing cytochalasin B and Hoechst 33442. Single donor cells were inserted into the perivitelline space of oocytes. Each donor cell-cytoplast couplet was fused with two pulses of DC 72 V for 15 μs using an Electro-Cell Fusion apparatus (NEPA GENE Co. Chiba, Japan). Fused embryos were activated in modified synthetic oviductal fluid (mSOF) medium containing 10 μM calcium ionophore (Sigma-Aldrich Corp). After chemical activation for 4 min, cloned embryos were transferred into 40 μl of

mSOF with 1.9 mM 6-dimethylaminopurine for 2 h. Reconstructed cloned embryos were transferred into the oviducts of synchronized recipients. Under laparotomy with general anesthesia, embryos were placed into the ampullary part of the oviduct using a 3.5 Fr Tom-Cat catheter (Sherwood, St. Louis, MO, USA).

7. T7 endonuclease I assay and sequencing

Genomic DNA was isolated using a genome isolation kit (Promega, Madison, USA) according to the manufacturer's instructions. Target genomic DNA was amplified with a KAPA HiFi HotStart PCR kit (KAPA Biosystems #KK2501) and PCR primers for T7E1 assay (Table 1). PCR amplicons were denatured at 95℃, reannealed at 16℃ to form heteroduplex DNA using a thermal cycler and then digested with 5 units of T7 endonuclease 1 (New England Biolabs, Ipswich, MA) for 20 min at 37°C and then analyzed using agarose gel electrophoresis.

8. Targeted deep sequencing analysis

Genomic DNA was isolated from transfected cells and tail tissue of the cloned puppy. On-target and potential off-target sites were amplified with a Phusion polymerase (New England Biolabs) for deep sequencing library generation (Table 1). Target genomic DNA was amplified by PCR with pre-index tailed primers. After pre-indexed amplification, PCR products were amplified with universal index primers for deep sequencing (Table 2). Amplicons were purified using purification

kit (MGmed). PCR amplicons were sequenced using MiniSeq or Miseq with TruSeq HT Dual Index system (Illumina). Potential off-target sites were found by Cas-OFFinder (Bae *et al*., 2014). Targeted deep sequencing data were analysis by Cas-analyzer (Park *et al*., 2017).

9. Creatine kinase analysis

Blood samples for serum CK analysis were obtained from the jugular vein beginning at 10 days after birth. Samples were collected at 1 week intervals until 8 weeks. No attempt was made to limit exercise prior to sampling. Biochemistry analyses were performed on IDEXX Catalyst Dx (IDEXX VetLab Analysers, Westbrook, Maine, USA). The normal reference range is 99- 436 U/L.

10. Biopsy and histopathological analysis

For biopsies of biceps femoris muscles, the dystrophin mutant dog and the control dog were anesthetized with ketamine and xylazine via intravenous injection, and anesthesia was maintained with isoflurane. The dogs were positioned in left lateral recumbency, and he biopsy region was prepared aseptically. After a 3 cm skin incision, a sample of the right biceps femoris muscle (1 cm x 1 cm x 0.5 cm) was collected from each dog. Immediately, the biopsies were flash frozen in isopentane precooled in liquid nitrogen. A standard panel of histochemical stains and reactions was performed on 5 μm muscle cryosections. Additional cryosections were used for

immunohistochemical staining using monoclonal antibodies against the dystrophin carboxy terminal, rod domain and utrophin (Novocastra, Newcastle-upon-Tyne, UK).

11. Western blotting

Proteins of the dystrophin mutant dog and the control dog were extracted from each muscle sample. After measuring protein concentration, equal amounts of proteins were resolved by 10% SDS-PAGE and transferred to nitrocellulose membranes (Hybond; Amersham Biosciences). Membranes were blocked for 1 h in Tris-buffered saline Tween (TBST) containing 5% powdered skim milk and incubated overnight with the following primary antibodies: anti-dystrophin NCL-DYS1 (1:500, Novocastra Laboratories, Newcastle, UK), anti-dystrophin NCL-DYS2 (1:100, Novocastra Laboratories), and anti-utrophin NCL-DRP (1:100, Novocastra Laboratories). Horseradish peroxidase (HRP)-conjugated secondary antibody (1:3000, Santa Cruz Biotechnologies, Piscataway, NJ) was used to detect bound antibodies with the Imaging System from FUSION-Solo (6x, Vilber Lourmat, France).

Table 1. PCR primers for T7E1 analysis and Deep sequencing

Primers used for T7E1 assay and Deep Seqeuncing of on-target site

Fwd_1st	GACACCTACCAATCAGAGTAGATTCC
Rev_1st	CAGTGGATAGTCAGATCAGTATGG
Fwd_2nd	GAATGGACTCCGTCCTGGTAG
Rev_2nd	GAAATGAGCTGGAACCACACTGG
Fwd_DS	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGATTGCAACAAACCAACA
Rev_DS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGGAACCACACTGGTGAAAA

Primers used for Deep Sequencing of off-target site

DogDMD_Off_1_O_F	AGGACACCCAGGTCCTTACC
DogDMD_Off_1_O_R	CCACTGTGTTCTTCCCAGGT
DogDMD_Off_2_O_F	TGCACCTGGTCTTAGCTCCT
DogDMD_Off_2_O_R	TCCTCCTTCTGCCTGTGTCT
DogDMD_Off_3_O_F	GACTCAATCCCAGGTCTCCA
DogDMD_Off_3_O_R	CCATTTCGGCAGAAAATCAT
DogDMD_Off_4_O_F	GCTGGATCCTGGGTAGACAA
DogDMD_Off_4_O_R	CCAATGTGGGACTTGATCCT
DogDMD_Off_5_O_F	TCACCTAGGCATGCATTCAA
DogDMD_Off_5_O_R	CCCCCTGGTTGACTAATGTG
DogDMD_Off_6_O_F	CAGCTCTGAAGGACCACACA
DogDMD_Off_6_O_R	CAGCAGGTAAACACCAAGCA
DogDMD_Off_7_O_F	CATGGAGGACACGATGCTAA
DogDMD_Off_7_O_R	GTGCAGGCAGCAACTTGTAA
DogDMD_Off_8_O_F	AAACTCCTTTTTGGGGCCTA
DogDMD_Off_8_O_R	TCCTCCTCTCATCTGGGCTA
DogDMD_Off_9_O_F	GGGATTTTGTTCCCAAGGTT
DogDMD_Off_9_O_R	GAGAAGAATTTGGGGGAAGC
DogDMD_Off_10_O_F	AACCGCAAACCAATACAAGG
DogDMD_Off_10_O_R	GGGAAAGACAGACCTTCACG
DogDMD_Off_11_O_F	TAGGGTTGCCATCACAAGGT
DogDMD_Off_11_O_R	TTCAGGAATGTCCTGGCTCT
DogDMD_Off_12_O_F	AGAGTACATGGCCTGGATGG
DogDMD_Off_12_O_R	GATGGAGACTGGGCTCTCTG
DogDMD_Off_13_O_F	TCCCATGAAAGAGGATCCAG
DogDMD_Off_13_O_R	TGTGTGGTCTCCCCTGTGTA
DogDMD_Off_14_O_F	TTGCAGTGGTGGAGAGAGTG
DogDMD_Off_14_O_R	TGTGGGTCACAGAACTGGAA
DogDMD_Off_15_O_F	TTGGTTTGGTTGGTGCAGTA
DogDMD_Off_15_O_R	CAGTTGACGGCAGTTGAGAA
DogDMD_Off_16_O_F	CAGTTGATGGCAGTTGAGAA
DogDMD_Off_16_O_R	TTGGTTTGGTTGGTGCAGTA
DogDMD_Off_17_O_F	CCATCCATCATCCTTTGCTT
DogDMD_Off_17_O_R	TGAAAGCGTTTTGCAGATTG
DogDMD_Off_18_O_F	GCTTCTTGGAGGAGCAACAC

DogDMD_Off_18_O_R	CAGGGAACAGGCTGTGGTAT
DogDMD_Off_1_i_F	CCAGGACAGCAGGAAGAAAG
DogDMD_Off_1_i_R	GCCTCCTGGTGGATTATGAA
DogDMD_Off_2_i_F	CCTCTTGATGAGTCCCATTGA
DogDMD_Off_2_i_R	GGGAAACTTTTGTTTGCCTTT
DogDMD_Off_3_i_F	CTCACTTCAGGCTTGCCTCT
DogDMD_Off_3_i_R	TCCCCTCACCTTGACTTCAG
DogDMD_Off_4_i_F	TGCCTTAACCATTTCTATCCTCA
DogDMD_Off_4_i_R	GCGTCCCCCTGTTTAATCTT
DogDMD_Off_5_i_F	TTCTACCCAAAGCAGCCACT
DogDMD_Off_5_i_R	TCCAAACCTTTCCTTGTGCT
DogDMD_Off_6_i_F	CCTGTGGTCTGGTGACATTG
DogDMD_Off_6_i_R	CACACCTGACACTTGGCTGT
DogDMD_Off_7_i_F	CAAGCCAGACACAAAAGCAA
DogDMD_Off_7_i_R	CACCCAATGCCTAGCTCTTC
DogDMD_Off_8_i_F	AGGAAAAGTTGGGACGGTCT
DogDMD_Off_8_i_R	CAGGTTCCTACCTCGCTCTG
DogDMD_Off_9_i_F	TTCACAACCCCGTGGATAAT
DogDMD_Off_9_i_R	TGTGTGGATCAGAGCAGAGG
DogDMD_Off_10_i_F	CCCCAGTCACTGATCCCTTA
DogDMD_Off_10_i_R	GAGCTCACGGCTGGACTTAC
DogDMD_Off_11_i_F	CTTTTGAGGCTTCCCTCCTT
DogDMD_Off_11_i_R	GACACTGCACTGTGGCATTT
DogDMD_Off_12_i_F	GGGTCCTCGCATCTATGTGT
DogDMD_Off_12_i_R	CCCTTCTTTCTTTCCCCATC
DogDMD_Off_13_i_F	CATCCCTGAACACTCCTGGT
DogDMD_Off_13_i_R	CAGGTTTGGAGGATCTTGGA
DogDMD_Off_14_i_F	CTGGGAGACACTGTGCGTAG
DogDMD_Off_14_i_R	GAGGGAAACGGACAGCACTA
DogDMD_Off_15_i_F	CCTGTCAAAGGTGGAAGCAT
DogDMD_Off_15_i_R	TCACGCCAGTCATTTCTGAG
DogDMD_Off_16_i_F	TCACACCAGTCATTTCTGAG
DogDMD_Off_16_i_R	CCTGTCAAAGGTGGAAGCAT
DogDMD_Off_17_i_F	CTGTCCCTGCACCCCTATTA
DogDMD_Off_17_i_R	CCCAGCTCTTGTCTGGATGT
DogDMD_Off_18_i_F	CCCTCGTCCTCCAACAAATA
DogDMD_Off_18_i_R	GGGGTCACACACCAGGTACT
DogDMD_Off_1_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACCAGAATCCAAGCCTCCT
DogDMD_Off_1_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCTGGACTGGTCAACTCAGCA
DogDMD_Off_2_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAAGACCTGGGATTCCAAA
DogDMD_Off_2_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCATTTAACTCAAACACAACG
DogDMD_Off_3_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTGTCCCTGTCTCAAGGTT
DogDMD_Off_3_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATGGGTTACAAAGGGCTGT
DogDMD_Off_4_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAAAAGAGAGAAGGAGGGAAA

(*Continued*)

DogDMD_Off_4_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGATTTGGCTCTTTGGCTGAA
DogDMD_Off_5_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCATTTTTCCTAAATGCCAGA
DogDMD_Off_5_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCATTGGGGTAAAGACTGG
DogDMD_Off_6_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGGCCAAGGTTCTGTTTGT
DogDMD_Off_6_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTGTGTGTCTGGGCTGAGTTC
DogDMD_Off_7_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGATGGAGTGCCAGTTTTA
DogDMD_Off_7_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCAGGGAGTCCAACACAGGAC
DogDMD_Off_8_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCACAGCAAACTACCCCAAA
DogDMD_Off_8_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGCTTGCCAGTGAAACATC
DogDMD_Off_9_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCTTGCTCCAAAGCTGTC
DogDMD_Off_9_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCTGGTAGGCTGCTCATCTC
DogDMD_Off_10_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCAAAGCCAGAAGATGGAAGC
DogDMD_Off_10_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGACCACCCTGGTTTCTA
DogDMD_Off_11_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTCAGACCTGCAGCAGACAA
DogDMD_Off_11_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGGGGCTTTAGTCAGGGTAG
DogDMD_Off_12_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGGTCAGTGGGTACCTGTGC
DogDMD_Off_12_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCTATCCACCCTCCCATCC
DogDMD_Off_13_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTACCCTGGCTTGTACCCATTT
DogDMD_Off_13_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTGCAAGCTGGAGTGTCCT
DogDMD_Off_14_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGTTTACCTGCGTCCTCTCA
DogDMD_Off_14_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCCCCAGAAACACAGCTAAG
DogDMD_Off_15_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAAAGCAGGCTCCCTGTAA
DogDMD_Off_15_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCATGTGAAAGTGGAGGTCCATA
DogDMD_Off_16_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATGTGAAAGTGGAGGTCCATA
DogDMD_Off_16_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGAAAAGCAGGCTCCCTGTAA
DogDMD_Off_17_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCTGGGCCCTACTCACTAT
DogDMD_Off_17_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCCATCTCATGGCTCGTCT
DogDMD_Off_18_DS_F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAGCCACTTGGTCTCCTCTG
DogDMD_Off_18_DS_R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTCCATGCAGTCCTCACAC

Table 2. Universal index primer set for deep sequencing.

		adapter sequence	
Forwad primer	D501	tatagcct	AATGATACGGCGACCACCGAGATCTACACtatagcctACACTCTTTCCCTACACGAC
	D502	atagaggc	AATGATACGGCGACCACCGAGATCTACACatagaggcACACTCTTTCCCTACACGAC
	D503	cctatcct	AATGATACGGCGACCACCGAGATCTACACcctatcctACACTCTTTCCCTACACGAC
	D504	ggctctga	AATGATACGGCGACCACCGAGATCTACACggctctgaACACTCTTTCCCTACACGAC
	D505	aggcgaag	AATGATACGGCGACCACCGAGATCTACACaggcgaagACACTCTTTCCCTACACGAC
	D506	taatctta	AATGATACGGCGACCACCGAGATCTACACtaatcttaACACTCTTTCCCTACACGAC
	D507	caggacgt	AATGATACGGCGACCACCGAGATCTACACcaggacgtACACTCTTTCCCTACACGAC
	D508	gtactgac	AATGATACGGCGACCACCGAGATCTACACgtactgacACACTCTTTCCCTACACGAC

Reverse primer	D701	cgagtaat	CAAGCAGAAGACGGCATACGAGATcgagtaatGTGACTGGAGTTCAGACGTGT
	D702	tctccgga	CAAGCAGAAGACGGCATACGAGATtctccggaGTGACTGGAGTTCAGACGTGT
	D703	aatgagcg	CAAGCAGAAGACGGCATACGAGATaatgagcgGTGACTGGAGTTCAGACGTGT
	D704	ggaatctc	CAAGCAGAAGACGGCATACGAGATggaatctcGTGACTGGAGTTCAGACGTGT
	D705	ttctgaat	CAAGCAGAAGACGGCATACGAGATttctgaatGTGACTGGAGTTCAGACGTGT
	D706	acgaattc	CAAGCAGAAGACGGCATACGAGATacgaattcGTGACTGGAGTTCAGACGTGT
	D707	agcttcag	CAAGCAGAAGACGGCATACGAGATagcttcagGTGACTGGAGTTCAGACGTGT
	D708	gcgcatta	CAAGCAGAAGACGGCATACGAGATgcgcattaGTGACTGGAGTTCAGACGTGT
	D709	catagccg	CAAGCAGAAGACGGCATACGAGATcatagccgGTGACTGGAGTTCAGACGTGT
	D710	ttcgcgga	CAAGCAGAAGACGGCATACGAGATttcgcggaGTGACTGGAGTTCAGACGTGT
	D711	gcgcgaga	CAAGCAGAAGACGGCATACGAGATgcgcgagaGTGACTGGAGTTCAGACGTGT
	D712	ctatcgct	CAAGCAGAAGACGGCATACGAGATctatcgctGTGACTGGAGTTCAGACGTGT

Ⅲ. Results

1. Production of dystrophin mutant dog cells

First, I designed an sgRNA targeting the exon 6 of the dog dystrophin gene (Figure 1a). The efficiency of sgRNA was validated by electroporation (the Amaxa P3 Primary Cell 4D-Nucleofector Kit, Lonza) with a Cas9/sgRNA RNP into canine male fetal fibroblasts (Figure 1b). As expected, the indel efficiency increased as the amount of Cas9 protein increased when the amount of sgRNA was equal. After confirming the activity of sgRNA through RNP experiments, the subsequent experiments were conducted based on plasmid for cell viability.

Figure 1. Genome editing in Canine fetal fibroblast using CRISPR-Cas9. (a) The target sequence of the sgRNA is underlined within the genomic sequence of the DMD gene. Red characters represent a protospacer-adjacent motif (PAM) sequence. (b) Genome editing in Canine fetal fibroblast via RNP delivery. Mutations were detected by T7E1 assay. A mixture of Cas9 protein and *in vitro* transcribed sgRNA (20 μg) was transfected into $2x10^5$ cells.

2. Generation of a dystrophin mutant cloned dog by somatic cell nuclear transfer

Having confirmed that Cas9/sgRNA was highly active in cultured cells, I performed somatic cell nuclear transfer (SCNT) to generate a dystrophin mutant dog. Oocytes matured *in vivo* were enucleated, injected with a donor cell, and fused by electrical stimulation. The fused couplets were activated with 10 μM calcium ionophore and 6-Dimethylaminopurine and then a total of 26 reconstructed oocytes were transferred into three naturally synchronized surrogate recipients.

In total, 49 matured *in vivo* from 4 oocyte donor dogs were recovered and these oocytes were enucleated, injected with a donor cell, and fused by electrical stimulation. The fused couplets (26/49, 53.1%) were activated with 10 μM calcium ionophore and 6-Dimethylaminopurine and then a total of 26 reconstructed oocytes were transferred into three naturally synchronized surrogate recipients. One of the recipients was pregnant to term (33.33%) and gave birth to one puppy (3.85%) (Table 3).

Table 3. Summary of embryo transfer and generation of the dystrophin mutant cloned pup.

Recipient	No. *in vivo* matured oocytes	No. reconstructed oocytes	No. transferred embryos	Pregnancy	No. births
A	16	6	6	+	1
B	13	7	7	-	0
C	20	13	13	-	0
Total	49	26	26	1 (33.3%) [∫]	1 (3.84%) [∬]

[∫] The percentage is based on the total number of recipient dogs.

[∬] The percentage is based on the total number of transferred embryos.

3. Analysis of the dystrophin mutant cloned dog

a. On-target indel frequency of the mutant cloned dog

To confirm the newborn was the dystrophin mutant clone, I performed targeted deep sequencing of the mutagenic site on the dystrophin mutant puppy. The tail tissue of the cloned puppy was collected two days after birth to detect genomic mutations in the target dystrophin locus. The puppy had a 57 bp deletion in the dystrophin gene (Figure 2a). Although it was an in-frame shift mutation, it was determined that it would be sufficient to disrupt the function of the DMD gene.

b. Creatine kinase analysis

The Creatine kinase (CK) level was recorded in both the dystrophin mutant dog and an age-matched control at two and eight weeks after birth. Until eight weeks of age, the control group had CK levels within the normal range (99–436 U/L), but the CK level (1,019-19,880 U/L) of the dystrophin mutant dog was much higher than the normal range from two weeks to eight weeks (Figure 2b). At 10 months of age, the CK levels of the dystrophin mutant dog and the control dog were 261 and 31,540 U/L, respectively. However, after 30 min of exercise, the dystrophin mutant dog had a CK level 300-fold higher than that of the control group (Figure 2c).

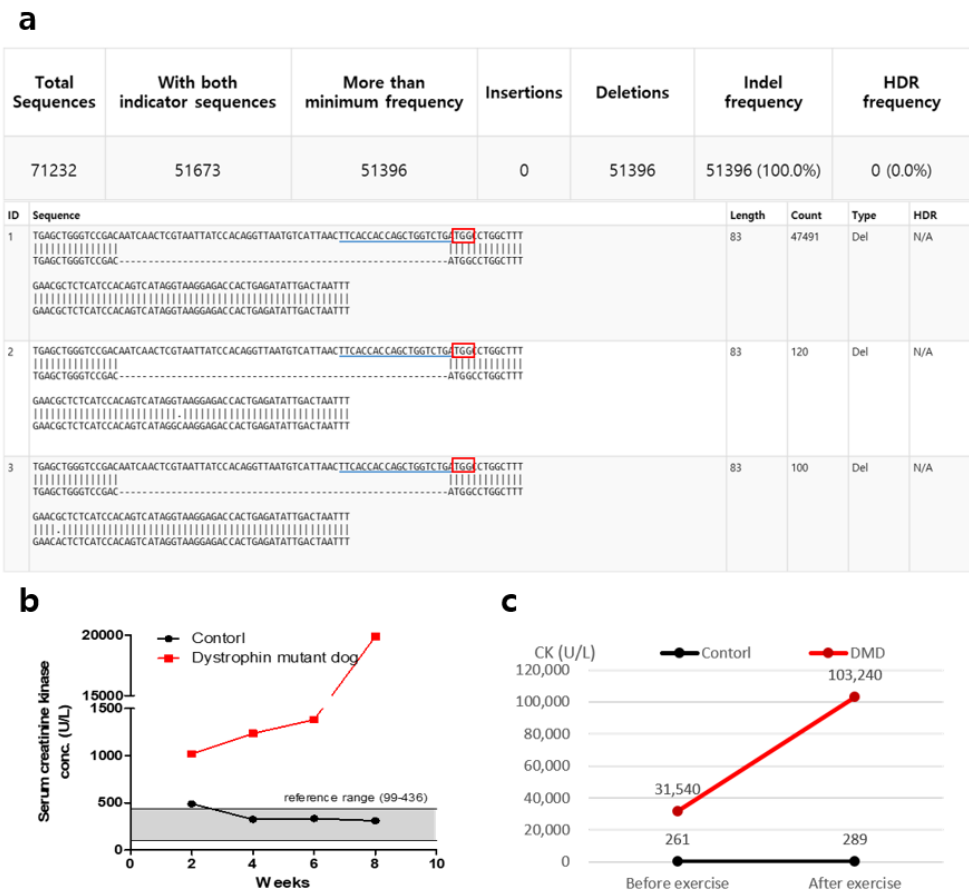


Figure 2. Production of a dystrophin mutant cloned dog by CRISPR-Cas9. (a) The result of deep sequencing of mutant dog cells. Target Sequence of the sgRNA is underlined within the genomic sequence of DMD gene. Red boxes represent a protospacer-adjacent motif (PAM) sequence. (b) Summary of serum creatine kinase (CK) values from two to eight weeks of age in the control and dystrophin mutant dog. (c) Change in creatine kinase in normal and dystrophin mutant dog after exercise.

c. Analysis of off-target effect

To analyze the off-target effect, I first found potential off-target sites with Cas-OFFinder (Bae *et al*., 2014) in the whole dog genome, allowing up to 3-nucleotide mismatches (Table 4). No off-target indel mutations at candidate sites were detected in the dystrophin mutant dog genome (Figure 3). Off-target candidate #6 seems to have a substitution frequency of half, but it was pre-existed single nucleotide polymorphism (SNP) found in the Cas9-untreated control.

Table 4. Potential off-target sites with up to 3 mismatches.

Mismatch 0 ; 1 / Mismatch 1 ; 0 / Mismatch 2 ; 2 / Mismatch 3 ; 16

	Target	Chromosome	Position	Direction	Mismatches	
1	crRNA: TTCACCACCAGCTGGTCTGANGG	chr24	22878368	+	3	
	DNA: cTCtCCACCAGCTGGTCTGgGGG					
2	crRNA: TTCACCACCAGCTGGTCTGANGG	chr8	29061641	+	3	
	DNA: TTCACCtCCAGCTGtTCTaATGG					
3	crRNA: TTCACCACCAGCTGGTCTGANGG	chr12	4938755	+	3	
	DNA: TTCACaAgCAGCTGGaCTGAGGG					
4	crRNA: TTCACCACCAGCTGGTCTGANGG	chr3	26238745	+	3	
	DNA: TTCACCACattCTGGTCTGAAGG					
5	crRNA: TTCACCACCAGCTGGTCTGANGG	chr3	50964839	-	2	
	DNA: TTCACCACCAcCTGGcCTGATGG					
6	crRNA: TTCACCACCAGCTGGTCTGANGG	chr3	57426757	+	3	
	DNA: cgCcCCACCAGCTGGTCTGACGG					
7	crRNA: TTCACCACCAGCTGGTCTGANGG	chr3	62735685	+	3	
	DNA: TTCACCcCCAGCTGaTCTGgGGG					
8	crRNA: TTCACCACCAGCTGGTCTGANGG	chr3	73060142	+	3	
	DNA: TTCACCAgCAGCTGGTaTGgTGG					
9	crRNA: TTCACCACCAGCTGGTCTGANGG	chr25	45357425	-	3	
	DNA: aTCAgCACCAGCTGGgCTGAGGG					
10	crRNA: TTCACCACCAGCTGGTCTGANGG	chr1	16905540	-	3	
	DNA: TTCACCACCAcacGGTCTGAAGG					
11	crRNA: TTCACCACCAGCTGGTCTGANGG	chr1	35850847	+	2	
	DNA: TTCACCACCAGCTGGaCaGATGG					
12	crRNA: TTCACCACCAGCTGGTCTGANGG	chr1	1.13E+08	-	3	
	DNA: TTCACCACCAGCTGGaggGATGG					
13	crRNA: TTCACCACCAGCTGGTCTGANGG	chr33	17707648	+	3	
	DNA: TTCAgCAtCAGCTGGcCTGACGG					
14	crRNA: TTCACCACCAGCTGGTCTGANGG	chr20	56928844	+	3	
	DNA: TTCACCAtCAGCgGGTCTGtAGG					
	crRNA: TTCACCACCAGCTGGTCTGANGG	chrX	27933928	-	0	On target
	DNA: TTCACCACCAGCTGGTCTGATGG					
15	crRNA: TTCACCACCAGCTGGTCTGANGG	chrX	1.06E+08	+	3	
	DNA: aTCACCACCAGCTGagCTGAAGG					
16	crRNA: TTCACCACCAGCTGGTCTGANGG	chrX	1.06E+08	-	3	
	DNA: aTCACCACCAGCTGagCTGAAGG					
17	crRNA: TTCACCACCAGCTGGTCTGANGG	chr18	50645543	+	3	
	DNA: TTCACCACCAGCTGGagaGATGG					
18	crRNA: TTCACCACCAGCTGGTCTGANGG	chr28	32426415	-	3	
	DNA: TgCACCACCAGCaGGTCTGcAGG					

	Total Sequences	With both indicator sequences	More than minimum frequency	Insertions	Deletions	Indel frequency
1	13213	11325	10553	0	68	68 (0.6%)
2	14376	13779	13442	0	7	7 (0.1%)
3	12183	11326	10835	0	9	9 (0.1%)
4	15741	15231	14892	0	18	18 (0.1%)
5	15043	14459	14088	0	2	2 (0.0%)
6	14224	12056	11843	0	4	4 (0.0%)
7	11372	10141	9789	0	3	3 (0.0%)
8	12832	12439	12244	0	2	2 (0.0%)
9	330652	11629	11439	0	2	2 (0.0%)
10	339871	8548	8240	0	2	2 (0.0%)
11	336445	9669	9364	0	0	0 (0.0%)
12	295067	8274	8135	0	0	0 (0.0%)
13	245902	9766	9577	0	0	0 (0.0%)
14	280667	9419	8966	0	0	0 (0.0%)
15	256539	10355	10007	0	0	0 (0.0%)
16	235573	9310	8534	0	0	0 (0.0%)
17	337899	8469	8107	0	0	0 (0.0%)
18	373015	7996	7691	0	0	0 (0.0%)

#6

<table>
<tr><td rowspan="3">1</td><td>GTTTGTTTAGCCGCCCCACCAGCTGGTCTGACGGCCAGGTCCTATGCTCGCTCTTGCTGGCCATCCATTCCCCCAGCA</td><td rowspan="3">78</td><td rowspan="3">6126</td><td rowspan="3">WT or Sub</td></tr>
<tr><td>||||||||||||.|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>GTTTGTTTAGCCTCCCCACCAGCTGGTCTGACGGCCAGGTCCTATGCTCGCTCTTGCTGGCCATCCATTCCCCCAGCA</td></tr>
<tr><td rowspan="3">2</td><td>GTTTGTTTAGCCGCCCCACCAGCTGGTCTGACGGCCAGGTCCTATGCTCGCTCTTGCTGGCCATCCATTCCCCCAGCA</td><td rowspan="3">78</td><td rowspan="3">4872</td><td rowspan="3">WT or Sub</td></tr>
<tr><td>||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||</td></tr>
<tr><td>GTTTGTTTAGCCGCCCCACCAGCTGGTCTGACGGCCAGGTCCTATGCTCGCTCTTGCTGGCCATCCATTCCCCCAGCA</td></tr>
</table>

Figure 3. No off-target mutations at candidate sites in Dmd mutant dog pup. Indel frequencies at potential off-target sites were analyzed using targeted deep sequencing. Off-target candidate #6 had 1 bp substituted allele which is pre-existed SNP not caused by CRISPR/Cas9.

d. Histopathological analysis and western blotting

The present study examined the biceps femoris of the dystrophin mutant dog and a control dog at six months of age. Muscle histopathology examination revealed mild fiber size variations, muscle fiber necrosis, and regeneration in focal muscle groups (Figure 4). Immunohistochemical staining of frozen muscle using monoclonal antibodies against the dystrophin carboxy terminal, rod domain, and utrophin, revealed decreased expression of dystrophin 1 and 2 (Figure 4F and 4G) along with upregulation of utrophin (Figure 4H) compared to the control dog muscles (Figure 4A-D).

Dystrophin 1 and dystrophin 2 were detected in the muscle tissue of the control in western blots, but were barely expressed in the dystrophin mutant dog (Figure 5). Utrophin is markedly upregulated in the dystrophin mutant dog when compared with controls in western blot analysis.

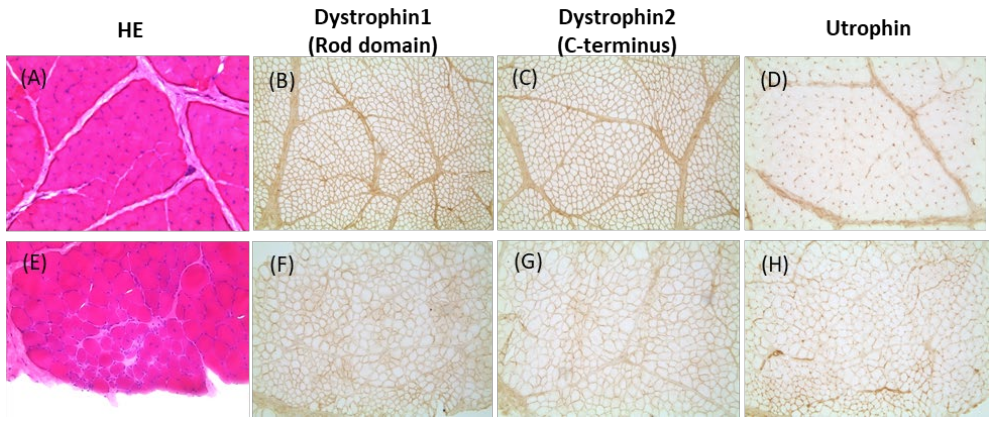


Figure 4. Histopathological analyses of dystrophic muscle. (A-D) A wild-type control dog. (E-H) the dystrophin mutant dog. Muscle pathology showing focal necrosis and regeneration of muscle fibers (HE), immunohistochemical staining using monoclonal antibody against dystrophin rod domain and utrophin, showing decrease in dystrophin 1 and dystrophin 2 expression and increased utrophin expression and increased utrophin expression compared to control muscles.

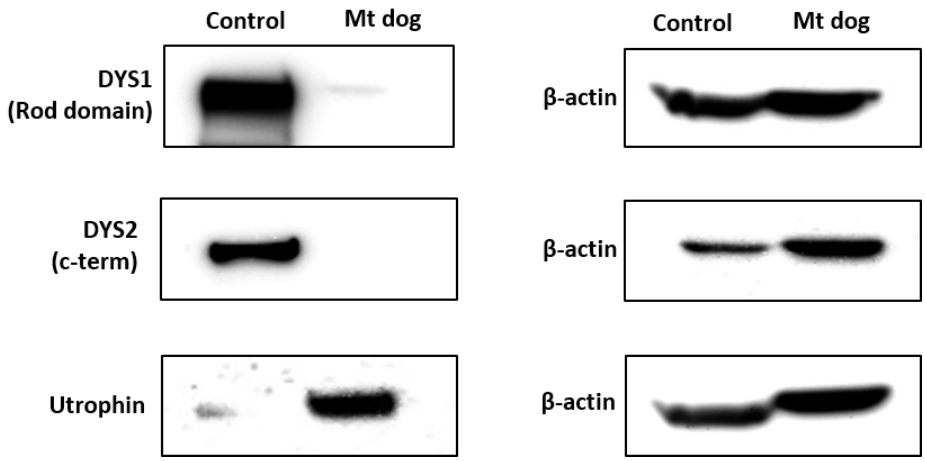


Figure 5. Western blotting of the dystrophin mutant dog. The analysis data confirmed the absence of DYS (dystrophin)1, 2 and upregulation of utrophin in dystrophin mutant dog (Mt dog) and control dog.

Ⅳ. Discussion

The present study demonstrated that 1) donor cells were edited with the dystrophin gene by CRISPR/Cas9 system, 2) a dystrophin mutant dog was successfully generated by SCNT using the donor cells. For this study, first, I designed a sgRNA targeting the exon 6 of the dog dystrophin gene (Aartsma-Rus *et al*., 2006). The remaining approximately 7% dystrophin gene mutations are caused by single or multi-exon duplications, with exons 2 to 20 being the most commonly affected. It is important to note that the GRMD and CXMD models have a frame shift because of a mutation in the splice acceptor of exon 6 that disrupts exon 7, and inducing a mutation in exon 6 could have a therapeutic effect on these canine dystrophinopathy models (Kornegay *et al*., 2003).

I next evaluated whether the dystrophin mutant dog shared a remarkably similar clinical course to that of dystrophinopathy patients. In both GRMD dogs and human DMD patients, serum creatine kinase (CK) activity is markedly elevated (Rowland, 1980). CK is most important for pre-neuter evaluations in young dogs because increased CK activity may be an early indicator of underlying muscle disease.

In addition, I and my colleagues observed walking or exercise-related changes in the dystrophin mutant dog after five months of age. The dystrophin mutant dog started to show "bunny hopping" from six months of age. Over time, its

limbs became stiff, with a decreased range of joint motion while moving, more pronounced bunny hopping, and difficulty in climbing stairs, and avoidance of movement were revealed. In untreated DMD patients, ambulation loss usually occurs during the early teenage years. Unlike in DMD patients, complete loss of ambulation is not a clinical feature in young DMD dogs (Duan, 2015; Valentine *et al.*, 1988). In this study, the 10-month-old dystrophin mutant dog was reluctant to exercise and showed limb muscle atrophy, but was still able to walk. Because of the clinical symptoms seen in DMD, it is necessary to observe whether the dystrophin mutant dog will completely loose ambulation in the future. The present study examined the biceps femoris of the dystrophin mutant dog and a control dog at six months of age. Besides its resemblance to the human clinical cases, the dystrophin mutant dog also exhibited histological lesions similar to dystrophinopathy patients. The present study examined the biceps femoris of the dystrophin mutant dog and a control dog at six months of age.

DMD is caused by out-of-frame mutations and absence of dystrophin protein in skeletal muscles because the dystrophin protein that is produced is truncated as a result of the premature stop codon and, therefore, is unstable *(*Yugeta *et al*., 2006*)*. BMD is caused by an in-frame mutation resulting in insufficient dystrophin protein, and clinical progression can be predicted by whether the deletion or duplication maintains or disrupts the translational reading frame (Arahata *et al*., 1988; Monaco *et al*., 1988). However, there are reports about exceptions to the reading-frame rule occurring in 10% of all DMD-causing mutations (Aartsma-Rus *et al.*, 2006). These findings are similar to the pathologic features in human

dystrophinopathy. The results demonstrated that the clinical characteristics of a dystrophin mutant dog are consistent with the diagnosis of dystrophinopathy in humans. A recent study reported that treatment with Cas9 and sgRNA-51 in spontaneous dystrophin KO dogs showed improved muscle histology (Amoasii *et al*., 2018). The dystrophin mutant dog will also be useful in research for developing therapeutics using the CRISPR/Cas9 system.

In summary, this study demonstrated for the first time that donor cells with CRISPR/Cas9 for a specific gene combined with the SCNT technique can efficiently produce a *dystrophin* knockout dog. Furthermore, this *dystrophin* KO dog showed many features such as CK elevation, *dystrophin* deficiency, skeletal defects and avoidance of ambulation, that are consistent with human dystrophinopathy. On September 19, 2016, the U.S. Food and Drug Administration approved a controversial drug to treat DMD, but the company involved failed to conduct a large scale clinical trial. Now, however, the dystrophinopathy dog will be a more useful model for contributing to preclinical studies aimed at the development of new therapies.

Ⅴ. Reference

Aartsma-Rus, A., Van Deutekom, J.C.T., Fokkema, I.F., Van Ommen, G.J.B., and Den Dunnen, J.T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. Muscle Nerve *34*, 135-144.

Amoasii, L., Hildyard, J.C.W., Li, H., Sanchez-Ortiz, E., Mireault, A., Caballero, D., Harron, R., Stathopoulou, T.R., Massey, C., Shelton, J.M., *et al*. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. Science *362*, 86-90.

Arahata, K., Ishiura, S., Ishiguro, T., Tsukahara, T., Suhara, Y., Eguchi, C., Ishihara, T., Nonaka, I., Ozawa, E., and Sugita, H. (1988). Immunostaining of skeletal and cardiac muscle surface membrane with antibody against Duchenne muscular dystrophy peptide. Nature *333*, 861-863.

Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473-1475.

Beerli, R.R., and Barbas, C.F., 3rd (2002). Engineering polydactyl zinc-finger transcription factors. Nat Biotechnol *20*, 135-141. 10.1038/nbt0202-135.

Bibikova, M., Beumer, K., Trautman, J.K., and Carroll, D. (2003). Enhancing gene targeting with designed zinc finger nucleases. Science *300*, 764.

Bitinaite, J., Wah, D.A., Aggarwal, A.K., and Schildkraut, I. (1998). FokI dimerization is required for DNA cleavage. Proc Natl Acad Sci U S A *95*,

10570-10575.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science *326*, 1509-1512.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology (Reading) *151*, 2551-2561.

Carroll, D. (2011). Genome engineering with zinc-finger nucleases. Genetics *188*, 773-782.

Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J., and Voytas, D.F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res *39*, e82.

Chamberlain, J.S., Metzger, J., Reyes, M., Townsend, D., and Faulkner, J.A. (2007). Dystrophin-deficient mdx mice display a reduced life span and are susceptible to spontaneous rhabdomyosarcoma. FASEB J *21*, 2195-2204.

Chapman, J.R., Taylor, M.R., and Boulton, S.J. (2012). Playing the end game: DNA double-strand break repair pathway choice. Mol Cell *47*, 497-510.

Cho, S.W., Kim, S., Kim, J.M., and Kim, J.S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol *31*, 230-232.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome

engineering using CRISPR/Cas systems. Science *339*, 819-823.

Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., *et al*. (2008). Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. Nat Biotechnol *26*, 702-708.

Duan, D. (2011). Duchenne muscular dystrophy gene therapy: Lost in translation? Res Rep Biol *2011*, 31-42.

Duan, D. (2015). Duchenne muscular dystrophy gene therapy in the canine model. Hum Gene Ther Clin Dev *26*, 57-69.

Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. Nature *366*, 483-487.

Glickman, L.T., Raghavan, M., Knapp, D.W., Bonney, P.L., and Dawson, M.H. (2004). Herbicide exposure and the risk of transitional cell carcinoma of the urinary bladder in Scottish Terriers. J Am Vet Med Assoc *224*, 1290-1297.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.R., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol *31*, 227-229.

Hytonen, M.K., and Lohi, H. (2016). Canine models of human rare disorders. Rare Dis *4*, e1241362.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol *31*, 233-239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet *15*, 321-334.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res *24*, 1012-1019.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., *et al*. (2013). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol *31*, 251-258.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci U S A *93*, 1156-1160.

Kornegay, J.N., Cundiff, D.D., Bogan, D.J., Bogan, J.R., and Okamura, C.S. (2003). The cranial sartorius muscle undergoes true hypertrophy in dogs with golden retriever muscular dystrophy. Neuromuscular Disord *13*, 493-500.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J Biol Chem *283*, 1-5.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal,

M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., *et al*. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803-819.

Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., *et al*. (2008). Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. Mol Cell *31*, 294-301.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

McGreevy, J.W., Hakim, C.H., McIntosh, M.A., and Duan, D. (2015). Animal models of Duchenne muscular dystrophy: from basic mechanisms to gene therapy. Dis Model Mech *8*, 195-213.

Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J., *et al*. (2011). A TALE nuclease architecture for efficient genome editing. Nat Biotechnol *29*, 143-148.

Mirski, K.T., and Crawford, T.O. (2014). Motor and cognitive delay in Duchenne muscular dystrophy: implication for early diagnosis. J Pediatr *165*, 1008-1010.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. (1988). An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. Genomics *2*, 90-95.

Moscou, M.J., and Bogdanove, A.J. (2009). A simple cipher governs DNA recognition by TAL effectors. Science *326*, 1501.

Oh, H.J., Chung, E., Kim, J., Kim, M.J., Kim, G.A., Lee, S.H., Ra, K., Eom, K., Park, S., Chae, J.H., *et al*. (2022). Generation of a Dystrophin Mutant in Dog by Nuclear Transfer Using CRISPR/Cas9-Mediated Somatic Cells: A Preliminary Study. Int J Mol Sci *23*.

Pardo, B., Gomez-Gonzalez, B., and Aguilera, A. (2009). DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci *66*, 1039-1056.

Park, J., Lim, K., Kim, J.S., and Bae, S. (2017). Cas-analyzer: an online tool for assessing genome editing results using NGS data. Bioinformatics *33*, 286-288.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology (Reading) *151*, 653-663.

Rowland, L.P. (1980). Biochemistry of muscle membranes in Duchenne muscular dystrophy. Muscle Nerve *3*, 3-20.

Schatzberg, S.J., Olby, N.J., Breen, M., Anderson, L.V., Langford, C.F., Dickens, H.F., Wilton, S.D., Zeiss, C.J., Binns, M.M., Kornegay, J.N., *et al*. (1999). Molecular analysis of a spontaneous dystrophin 'knockout' dog. Neuromuscul Disord *9*, 289-295.

Sharp, N.J., Kornegay, J.N., Van Camp, S.D., Herbstreith, M.H., Secore, S.L., Kettle, S., Hung, W.Y., Constantinou, C.D., Dykstra, M.J., Roses, A.D., and *et al*. (1992). An error in dystrophin mRNA processing in golden retriever muscular dystrophy, an animal homologue of Duchenne muscular dystrophy. Genomics

13, 115-121.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., *et al*. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol *15*, 169-182.

Shrivastav, M., De Haro, L.P., and Nickoloff, J.A. (2008). Regulation of DNA double-strand break repair pathway choice. Cell Res *18*, 134-147.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62-67.

Szabo, S.M., Salhany, R.M., Deighton, A., Harwood, M., Mah, J., and Gooch, K.L. (2021). The clinical course of Duchenne muscular dystrophy in the corticosteroid treatment era: a systematic literature review. Orphanet J Rare Dis *16*, 237.

Thompson, L.H., and Schild, D. (2001). Homologous recombinational repair of DNA ensures mammalian chromosome stability. Mutat Res *477*, 131-153.

Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646-651.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636-646.

Valentine, B.A., Cooper, B.J., de Lahunta, A., O'Quinn, R., and Blue, J.T. (1988).

Canine X-linked muscular dystrophy. An animal model of Duchenne muscular dystrophy: clinical studies. J Neurol Sci *88*, 69-81.

Walmsley, G.L., Arechavala-Gomeza, V., Fernandez-Fuente, M., Burke, M.M., Nagel, N., Holder, A., Stanley, R., Chandler, K., Marks, S.L., Muntoni, F., *et al*. (2010). A duchenne muscular dystrophy gene hot spot mutation in dystrophin-deficient cavalier king charles spaniels is amenable to exon 51 skipping. PLoS One *5*, e8647.

Wayne, R.K., and Ostrander, E.A. (2007). Lessons learned from the dog genome. Trends Genet *23*, 557-567.

Wein, N., Alfano, L., and Flanigan, K.M. (2015). Genetics and emerging treatments for Duchenne and Becker muscular dystrophy. Pediatr Clin North Am *62*, 723-742.

Yiu, E.M., and Kornberg, A.J. (2015). Duchenne muscular dystrophy. J Paediatr Child Health *51*, 759-764.

Yugeta, N., Urasawa, N., Fujii, Y., Yoshimura, M., Yuasa, K., Wada, M.R., Nakura, M., Shimatsu, Y., Tomohiro, M., Takahashi, A., *et al*. (2006). Cardiac involvement in Beagle-based canine X-linked muscular dystrophy in Japan (CXMDJ): electrocardiographic, echocardiographic, and morphologic studies. BMC Cardiovasc Disord *6*, 47.

Chapter 2.

Base editing in human mitochondrial DNA with monomeric DddA-TALE fusion deaminases

Ⅰ. Introduction

Mitochondria are bacterium-sized essential organelles that are present in almost all eukaryotic cells. Mitochondria have the crucial role in intermediary metabolism in various cellular metabolic pathways, including oxidative phosphorylation, fatty acid oxidation, Krebs cycle, urea cycle, gluconeogenesis, and ketogenesis (Duchen, 2004; Landrum *et al*., 2016; Ratnaike *et al*., 2021). In the endosymbiotic theory, a proteobacterium was engulfed by endocytosis, providing the host with the ability to produce cellular energy in the form of ATP. As a result, mitochondria are double-membrane organelles lacking a nucleus and contain their own DNA (mitochondrial DNA (mtDNA)) (Archibald, 2015). Unlike nuclear DNA (nDNA), mtDNAs are maternally inherited and are present in multiple copies per cell. Therefore, mitochondrial diseases caused by mutations in mtDNA follow the laws of population genetics (Schon *et al*., 2012). The level of heteroplasmy is crucial in determining the extent of cellular dysfunction. mtDNA has a circular structure and lacks an intron–exon structure. In addition, replication, transcription and translation are all controlled by a single non-coding region, termed the displacement loop (D loop) (Gorman *et al*., 2016).

Recently, the identification of pathogenic mtDNA mutations has evolved in parallel with advances in sequencing and cell biology technologies but there were still many shortcomings in the tools for mitochondrial gene correction. A diverse array of DNA repair pathways exists in mammalian mitochondria with the notable

absence of efficient DNA double-strand break (DSB) repair, and either inefficient or absent homologous recombination (HR) (Hagstrom *et al*., 2014; Kazak *et al*., 2012; Moretton *et al*., 2017; Shokolenko *et al*., 2013). As a result, programmable nuclease-based therapy cannot induce or revert a specific mutation in mtDNA, possibly because DNA DSBs are not efficiently repaired in mitochondria. The difficulty in transporting RNAs into mitochondria is another obstacle to adapting CRISPR-Cas9 for mtDNA engineering. The series of studies suggested that endogenous RNA import into mammalian mitochondria is not required for normal cellular functions (Wanrooij *et al*., 2010). There is a report that had claimed CRISPR-Cas9 successfully manipulates mtDNA (Jo *et al*., 2015). Nonetheless, the editing efficiency of the CRISPR-Cas9 system for mtDNA is expected to be either absent or very low. So previous approaches to editing mtDNA relied on RNA-free programmable nucleases resulting in heteroplasmic shifts to favour uncut mtDNA genomes (Bacman *et al*., 2018; Bacman *et al*., 2013; Gammage *et al*., 2014)

Base editors (BE) precisely induce targeted point mutations without requiring DSBs or donor DNA templates, and without reliance on HDR (Gaudelli *et al*., 2017; Komor *et al*., 2017; Komor *et al*., 2016; Nishida *et al*., 2016). CRISPR-Cas9 based BEs are composed of a catalytically dead Cas9 (dCas9) or a nickase Cas9 (nCas9) fused to a deaminase and guided by a single guide RNA (sgRNA) to the locus of interest. The cytosine BEs (CBEs) allow the conversion of a C:G to a T:A base pair (bp), while adenine BEs (ABEs) convert an A:T into a G:C bp (Gaudelli *et al*., 2017; Komor *et al*., 2016). In both CBEs and ABEs with CRISPR-Cas9, the catalytically impaired Cas nuclease domain localizes a ssDNA deaminase enzyme to

a genomic target sequence of interest. In this mechanism, R-loop caused by the hybridization of the guide RNA spacer to the target DNA strand must be formed to exhibit robust base editing activity (Jiang *et al*., 2016; Jiang *et al*., 2015; Jinek *et al*., 2014).

DddA is a double-stranded DNA cytidine deaminase originated from *Burkholderia cenocepacia.* DddA was predicted to function as an antibacterial toxin that is delivered by the type VI secretion system (T6SS) (Coulthurst, 2019; Hood *et al*., 2010; Mok *et al*., 2020). Unlike other deaminases that are known to catalyse the deamination of ssDNA, RNA, and free nucleotide, DddA reduced the viability of *Escherichia coli* when it was produced. The identified domain that confers toxicity in which amino acids 1264-1427 of DddA is termed as $DddA_{tox}$. $DddA_{tox}$ efficiently converts cytosine to uracil within dsDNA, in contrast of the activity of previous deaminase domains for genome engineering that convert within ssDNA or RNA. Because of this characteristic, $DddA_{tox}$ can be used with zinc finger (ZF) or transcriptional activator-like effector nucleases (TALE) as well as CRISPR-Cas9 (Joung and Sander, 2013; Mok *et al.*, 2020; Urnov *et al*., 2010). In addition, $DddA_{tox}$ has s strong preference for 5'-TC contexts as substrates (Mok *et al.*, 2020). Because the expression of full-length $DddA_{tox}$ fused to programmable DNA-binding proteins was toxic to human HEK293T cells, $DddA_{tox}$ is engineered to the split form which maintains high deaminase activity and target specificity. Among the splits tested, splits at G1333 and G1397 yielded the highest editing efficiencies within the target spacing region. Researchers showed that split $DddA_{tox}$ nontoxic halves and an uracil glycosylase inhibitor (UGI) fused to TALE proteins can achieve DNA base editing

in the various organisms (DdCBEs; DddA-derived cytosine base editors) (Guo *et al*., 2022; Kang *et al*., 2021; Lee *et al*., 2021; Mok *et al.*, 2020; Silva-Pinheiro *et al*., 2022). Split halves of $DddA_{tox}$ can be fused to ZF, which is termed zinc finger deaminase (ZFD) (Lim *et al*., 2022).

Although DdCBEs and ZFDs are highly versatile, enabling targeted base editing in both nuclear and organelle DNA, the requirement for two TALE and ZF constructs, respectively, rather than one, to induce base editing is disadvantageous and challenging. Here, I modify $DddA_{tox}$ by random mutagenesis and generate non-toxic, full-length $DddA_{tox}$. In addition, I show that full-length $DddA_{tox}$ variants fused to Cas9 or TALE, termed mDdCBEs (monomeric DdCBEs), have the high deaminase activity in human and mouse mtDNA. I demonstrate that E1347A $DddA_{tox}$ variant which is known to be catalytic inactive in split form is shown to be active at full-length. It is shown that mDdCBEs can be designed to target TC motif which conventional dimeric DdCBE cannot target. I also show that mDdCBEs expressed via AAV in cultured human cells can achieve nearly homoplasmic C-to-T editing in mitochondrial DNA.

Ⅱ. Materials and Methods

1. Plasmid construction

Sequences encoding the $DddA_{tox}$ variants were PCR amplified using the synthesized full-length $DddA_{tox}$-encoding sequence (gBlock, IDT) as a template, the primers listed in Table 1, and Q5 DNA polymerase (NEB). The resulting PCR products were cloned using Gibson assembly (NEB) into p3s-BE3 that had been digested with BamHI and Sma I (NEB) in the Apobec1 sequence. The TALE-$DddA_{tox}$ (Addgene #158093, #158095, #157842, #157841) plasmids were digested with BamH I and Sma I, and sequences encoding the $DddA_{tox}$ variants were PCR amplified using the primers listed in Table 1 and cloned into the digested plasmid using Gibson assembly. TALE arrays were designed to target human *ND1*, the human *MT-TC* gene encoding tRNA-Cys, and mouse *ND5* following the approach used in previous reports (Kang *et al.*, 2021; Kim *et al*., 2013; Lee *et al.*, 2021). Assembled plasmids were chemically transformed into E. coli DH5α, and plasmids from the surviving colonies were analyzed by the Sanger sequencing method. Final plasmids were midi-prepped (Macherey-Nagel) for cell transfection.

To generate an AAV vector encoding mDdCBEs, PCR amplicons containing ND1 L-GSVG, ND1 L-E1347A, and the CMV promoter were assembly with Not I- (NEB) and Apa I- (NEB) digested pAAV vector by Gibson assembly (NEB). pAAV-CMV-ND1 L-GSVG was digested with EocR I (NEB) and Apa I (NEB) and ND4 R-GSVG, ND4 R-E1347A, ND6 L-GSVG, and ND6 L-E1347A

were amplified by PCR for Gibson assembly (NEB). Assembled plasmids were chemically transformed into *E. coli* DH5α, and plasmids from the surviving colonies were analyzed by the Sanger sequencing method. Final plasmids were midi-prepped (Macherey-Nagel) for viral particle production.

2. Cell culture and transfection conditions

HEK293T (ATCC, CRL-11268) cells and HeLa (ATCC, CCL-2) cells were cultured and maintained at 37 °C in a 5 % CO2 atmosphere. Cells were grown in Dulbecco's Modified Eagle Medium supplemented with 10% (v/v) fetal bovine serum (Welgene) and 1% penicillin/streptomycin (Welgene). Twenty-four hours before transfection, cells were seeded in a 48-well plate (Corning) at a density of 3X105 cells/well (HEK293T) and 4x104 cells/well (HeLa), after which they were transfected with a plasmid encoding the Cas9-$DddA_{tox}$ fusion (750 ng) and a single guide RNA- (sgRNA-) encoding plasmid (250 ng) with Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. TALE-$DddA_{tox}$ constructs (200 ng) were transfected into HEK293T cells using Lipofectamine 2000.

3. *In vitro* activity assay

To measure the base editing activity of the $DddA_{tox}$ variants fused to Cas9 *in vitro*, templates were PCR amplified from genomic DNA from HEK293T cells using PrimeSTAR® GXL DNA Polymerase (TAKARA) and the TYRO3-specific

primers listed in Table X. An sgRNA targeted to the TYRO3 site was transcribed *in vitro* using an mMESSAGE mMACHINE™ T7 ULTRA Transcription Kit (Invitrogen) according to the manufacturer's protocol. The PCR amplified template (200 ng) and sgRNA (200 ng) were incubated with purified proteins (300 nM) in NEB buffer 3.1 in a 20 μL reaction at 37 °C for various amounts of time. The reaction was terminated by incubation at 65 °C for 15 min, after which PCR amplification was performed for targeted deep sequencing.

4. Random mutagenesis

Error-prone PCR was performed on the synthesized full-length $DddA_{tox}$-encoding sequence (gBlock, IDT) using a GeneMorph ii Random mutagenesis kit (Agilent) according to the manufacturer's protocol. In brief, 1 ng, 100 ng, and 700 ng of $DddA_{tox}$-encoding DNA fragments were used as template for the introduction of random mutations at a density of 0~16 mutations/kb. The full-length $DddA_{tox}$ gBlock sequence was PCR amplified beforehand with primers listed in Table 1. PCR products were pooled and Gibson assembled (NEB) into p3s-UGI-Cas9(H840A)-$DddA_{tox}$(E1347A) that had been digested with Sma1 and Xho1. The assembled plasmids were transformed into DH5α heat-shock competent cells, after which plasmids from a fraction of the surviving colonies were subjected to Sanger sequencing. p3s-UGI-Cas9(H840A)-$DddA_{tox}$ plasmids with correct in-frame fusions were transfected into HEK293T cells along with a plasmid encoding an appropriate sgRNA to target the selected site. Editing activity was detected through targeted deep

sequencing.

5. Genomic and mtDNA preparation

Cells transfected with plasmids encoding $DddA_{tox}$ variants fused to Cas9 were harvested 2 days post-transfection and cells transfected with plasmids encoding TALE-$DddA_{tox}$ were harvested 3 days post-transfection. Genomic and mtDNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen). For large-scale analysis, DNA was extracted using 100 μL of cell lysis buffer (50 mM Tris-HCl, pH 8.0 (Sigma-Aldrich), 1 mM EDTA (Sigma-Aldrich), 0.005% sodium dodecyl sulfate (Sigma-Aldrich)) that included 5 μL of Proteinase K (Qiagen). The lysate was incubated at 55 °C for 1 h, and then at 95 °C for 10 min.

6. Targeted deep sequencing analysis

To analyze the frequency of edits, on-target sites were amplified via nested primary PCR, a secondary PCR, and a third PCR using TruSeq HT Dual index-containing primers and PrimeSTAR® GXL DNA Polymerase (TAKARA) to generate deep sequencing libraries. The libraries were sequenced using Illumina MiniSeq with paired-end sequencing systems. The computer program used to analyze the frequency of edits is available at https://github.com/ibs-cge/maund.

7. *In vitro* transcription and transfection

The templates for *in vitro* transcription were amplified by PCR using PrimeSTAR® GXL DNA Polymerase (TAKARA) with the primers listed in Table 1. mRNAs were synthesized from PCR amplicons using an mMESSAGE mMACHINE™ T7 ULTRA Transcription Kit (Invitrogen) according to the manufacturer's protocol. Various concentrations of the mRNAs were then transfected into cells with Lipofectamine 2000 (Invitrogen).

8. Analysis of mitochondrial genome-wide off-target effects

The mtDNA was amplified with PrimeSTAR® GXL DNA Polymerase (TAKARA) using the primers listed in Supplmentary Table 6. Amplicons were purified with a Qiaquick PCR Purification Kit according to the manufacturer's protocol. Sequencing libraries were constructed using a Nextera DNA Flex Library Prep Kit (Illumina). Sequencing was performed with an Illumina MiniSeq with the 300 cycles paired end program.

To analyze next-generation sequencing data from whole mitochondrial genome sequencing, I followed methods in previously published reports (Lim *et al.*, 2022; Mok *et al.*, 2020). Fastq files were aligned to the GRCh38.p13 (release v102) reference genome using BWA (v.0.7.17), and BAM files were generated with SAMtools (v.1.9) by fixing read pairing information and flags. Positions with conversion rates ≥ 0.1% were identified among all cytosines and guanines in the

mitochondrial genome using the REDItoolDenovo.py script from REDItools (v.1.2.1) (Picardi *et al*., 2015). In all samples, positions with 50% conversion and target sites considered as single nucleotide mutations in cell lines were excluded. To calculate the average C•G-to-T•A editing frequency for all C•Gs in the mitochondrial genome, the conversion rates were averaged at each base position in the off-target sites.

9. Transduction of AAV into HEK293T cells

AAV serotype 2 was produced by the IBS Virus Facility (https://www.ibs.re.kr/virusfacility/). HEK293T cells were seeded at a density of $8X10^4$ cells/well in a 48-well plate (Corning) 24 h before transduction. They were then infected with different viral doses determined by quantitative PCR, and cultured in DMEM containing with 10% (v/v) fetal bovine serum (Welgene) and 1% penicillin/streptomycin (Welgene). At different time points after transduction, cells were collected for targeted deep sequencing.

10. *In silico* analysis of base-editable TC motifs in organelle DNA

The human mitochondrial genome (NC_012920) was used as reference sequences for this analysis. All TC motifs were extracted from the organelle DNA sequences, after which the split TALE-$DddA_{tox}$ recognition sites for each TC motif were determined according to the following criteria: (i) the TALE array-binding

sequence must contain a thymidine at the 5' and 3' ends, (ii) the length of the TALE array-binding sequence must be 14-20 bp, and (iii) the sequence between a pair of TALE-binding sites must be 14-18 bp long, including at least one TC motif. After this analysis, the remaining TC motifs were analyzed to find locations at which only one TALE array could bind. For this procedure, criteria (i) and (ii) above were used; the third criterion was changed to (iii) TC motifs must be within 18 bp of the thymidine at the 3’ end of a TALE array-binding sequence.

Table 1. Primer list for plasmid construction.

Name	Sequence 5` to 3`
DddA F	cccaagcttgccaccatgggatccggcagctacgccctgggtccgtatcag
DddA R	cccgggagtctcgctgccgctgcaaccgcctttggtcgggctcttcggg
E1347 F	ctatgccaatgccggtcatgtgGCCggtcagagcgccctgttcatg
E1347 R	catgaacagggcgctctgaccGGCcacatgaccggcattggcatag
error_prone PCR F	tgagccagctgggcggcgacagcggcagcgagactcccgggacctcagagtccgccacacccgaaagtggcagctacgccctgggtcc
error_prone PCR R	aatagggccctctagatgcatgctcgagttagcaaccgcctttggtcggg
TALE Right DddA F	tgaagaaaggtctgggcggatccggcagctacgccctgggtccgtatcag
TALE_Left_DddA_F	tgaagaaaggcctgggtggatccggcagctacgccctgggtccgtatcag
TALE_DddA R	tcagattagttgagccgccagagcaaccgcctttggtcgggctcttcggg
TALE_UGI F	tctggcggctcaactaatctgagcgacatcattg
TALE_ UGI R	cagatccgaaaatggatatacaagctccc
mRNA F	gaaattaatacgactcactatagggagacccaagctggctagcaccgccaccatggccctgtcccgtgcggtttgtggcac
mRNA R	agggaagaaagcgaaaggag

Ⅲ. Results

1. Creation of non-toxic, full-length $DddA_{tox}$ variants by random mutagenesis

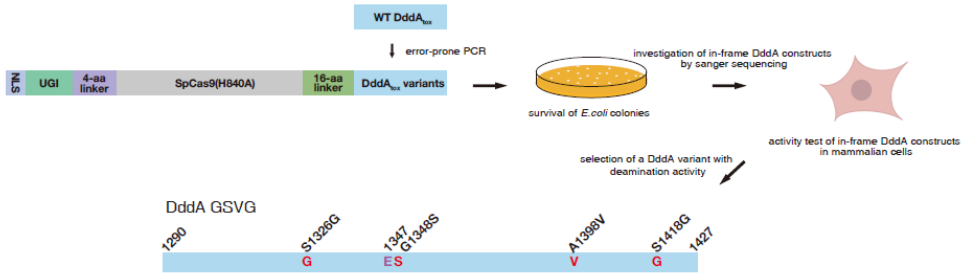
Full-length $DddA_{tox}$ could not able to be cloned to H840A nCas9 due to cytotoxicity. Therefore, I tried error-prone PCR to introduce random mutations in the DddAtox coding sequence and collected surviving colonies. Most of colonies had mutations that causes out-of-frame, but 23 clones had no frameshift (Table 1, Figure 1).

Through these experiments, a non-toxic, full-length $DddA_{tox}$ variant with four-point modifications: S1326G, G1348S, A1398V, S1418G (termed "GSVG") was able to be obtained, after measuring base editing efficiencies of 23 clones in human cells. The numbers used to indicate the position of the GSVG target window were obtained by counting backward and forward from the proto-spacer, toward the 5` upstream and toward the 3` downstream region, respectively. The base editing and indel frequencies were measured by targeted deep sequencing (Table 2, 3). This variant was also successfully fused to the C terminus or the N terminus of dCas9, D10A nCas9, H840A nCas9, and Cas9. In human cells, these fusion proteins, with the exception of those containing the wild-type Cas9, induced C-to-T conversions at various sites with efficiencies of up to 38% (Figure 2). As expected, fusion proteins containing Cas9 induced indels rather than base substitutions.

In order to pinpoint which mutations are critical in the GSVG variant, I

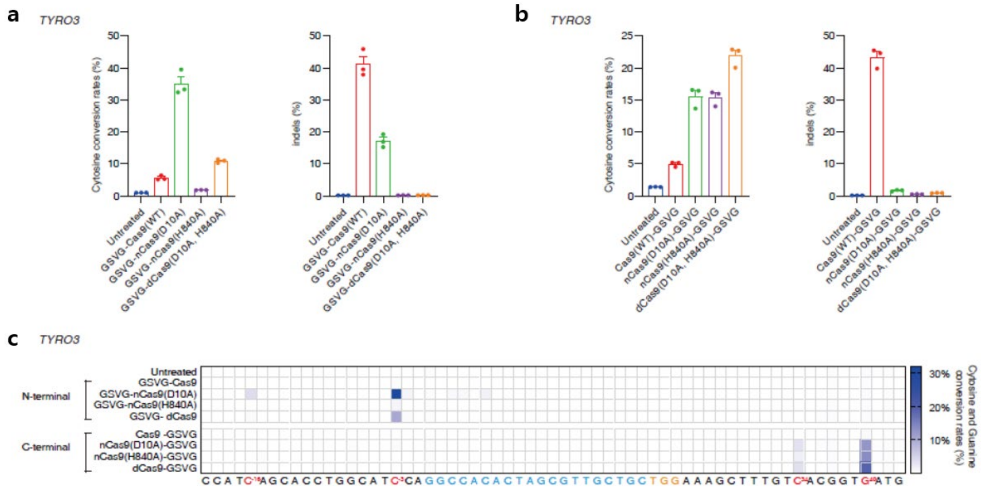
attempted to create four revertants, namely, SSVG, GGVG, GSAG, and GSVS (reverted residues are underlined), via site-directed mutagenesis. I was able to obtain SSVG, GSAG, and GSVS revertants but failed to obtain the GGVG variant fused to the C terminus of nCas9. It can be assumed that G1348 right next to E1347 which is the catalytic site plays the most important role (Figure 3).

Surprisingly, unlike previously known to be inactive in the split form, the E1347A $DddA_{tox}$ variant had been shown to be active at full-length (Figure 3). I was able to confirm the deaminase activity of the recombinant E1347A-D10A nCas9 fusion protein, expressed in and purified from *E. coli*, under cell-free *in vitro* conditions using a PCR amplicon containing the TYRO3 site (Figure 4).



(By Ji Min Lee in Seoul National University)

Figure 1. Schematic diagram of the screen for non-toxic DddA$_{tox}$ variants generated by error-prone PCR. The E1347 active site and replaced residues are shown in violet and red, respectively.



(By Young Geun Mok in Institute for Basic Science and Ji Min Lee in Seoul National University)

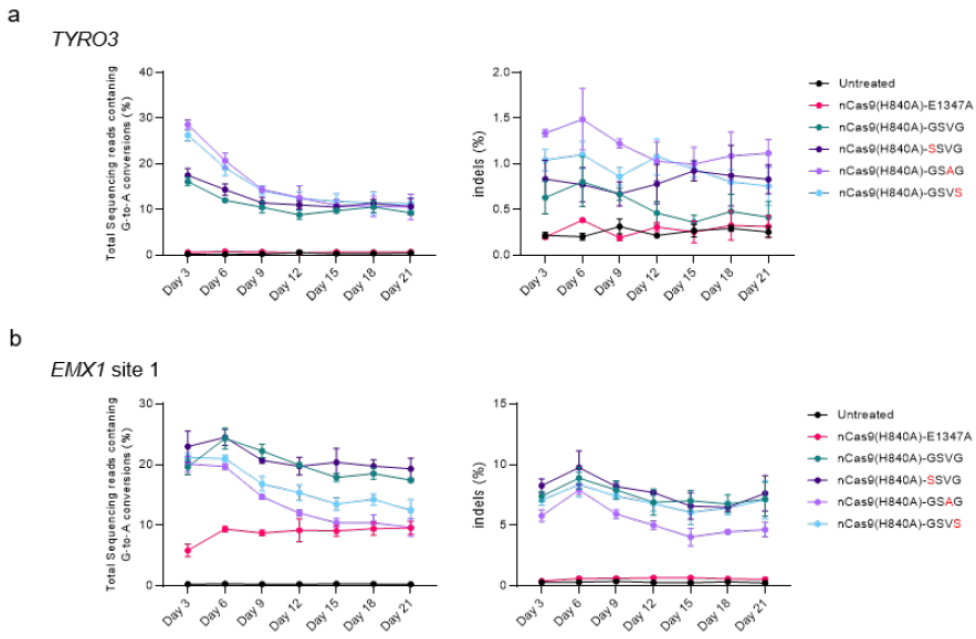
Figure 2. Characterization of the $DddA_{tox}$ GSVG variant derived from random mutagenesis. (a), (b) Editing and indel frequencies at the TYRO3 site induced by the GSVG variant fused to the N (a) and C (b) termini of Cas9, D10A nCas9, H840A nCas9, and D10A, H840A dCas9. (c) Heat map showing the frequencies of C-to-T substitutions at various positions in the TYRO3 site. The protospacer is shown in blue and the PAM in orange. Cytosines that underwent editing are shown in red. Means ± s.e.m. (a, b) and heat map colors (c) were determined from three independent experiments.

Table 2. PCR primers for amplifying target site sequence in genome.

	Forward primer(5` to 3`)	Reverse primer(5` to 3`)
TYRO3	ctgtcaacaaagtgctggcc	agttacacagggccttcgtg
Hman ND1	ggttcggttggtctctgcta	atggccaacctcctactcct
Hman ND4	gccattctcatccaaacc	ggttgagggataggaggag
Hman ND6	ggtttgtggggttttcttct	caaccaccaccccatcata
mouse ND5	cgcagctacaggaaaatcagc	atggtattcctgtgagggcg
Hman MTND4P12	ctaattctctttgaggagcatggttag	tatcacttccagccacctatttcc

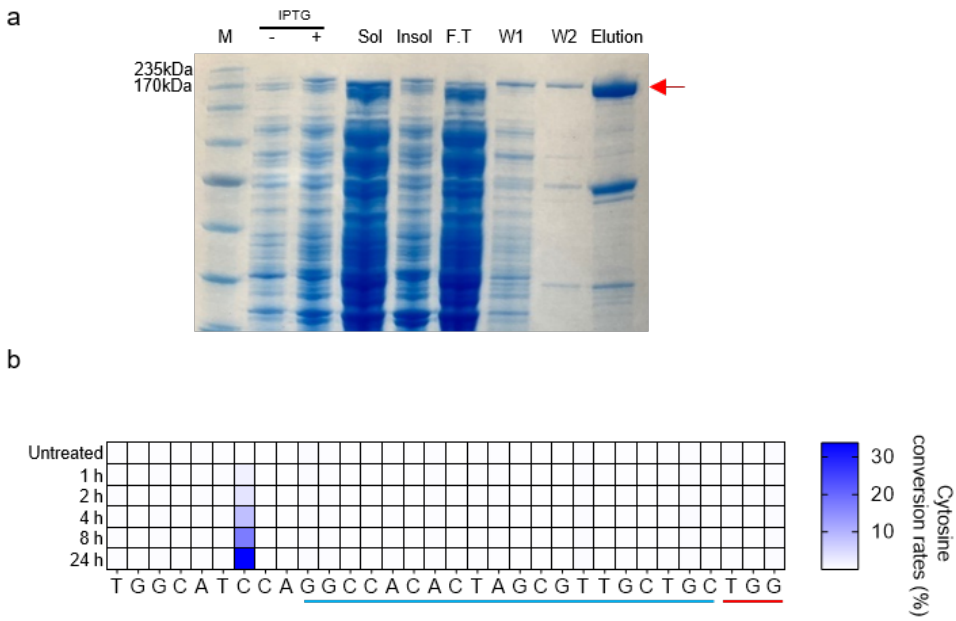
Table 3. PCR primer list for targeted deep sequencing.

	Forward primer(5' to 3')	Reverse primer(5' to 3')
TYRO3	acactctttccctacacgacgctcttccgatctctgtcaacaaagtgctggcc	gtgactggagttcagacgtgtgctcttccgatctcttgactcccatgcctcctg
Hman ND1	acactctttccctacacgacgctcttccgatctaagggtggagaggttaaaggag	gtgactggagttcagacgtgtgctcttccgatctccctggtcaacctcaaccta
Hman ND4	acactctttccctacacgacgctcttccgatctgacttcaaactctactcccactaatag	gtgactggagttcagacgtgtgctcttccgatctgttgtggtaaatatgtagagggag
Hman ND6	acactctttccctacacgacgctcttccgatctctctttcacccacagcacc	gtgactggagttcagacgtgtgctcttccgatctgattgttagcggtgtggtcg
mouse ND5	acactctttccctacacgacgctcttccgatctcctccacccatgactaccat	gtgactggagttcagacgtgtgctcttccgatctgggtgagagcacaaatagctg
Hman MTND4P12	acactctttccctacacgacgctcttccgatctggctgagcgagcctca	gtgactggagttcagacgtgtgctcttccgatctgacttctagcaagcctcactaatc



(By Young Geun Mok in Institute for Basic Science)

Figure 3. Time-dependence of editing and indel frequencies induced by $DddA_{tox}$ variants. E1347A, GSVG, SSVG, GSAG, and GSVS variants fused to the C terminus of H840A nCas9 had base editing activity at the TYRO3 site (position G_{40}) (a) and EMX1 site 2 (position G_{38}) (b). Means ± s.e.m. were determined from three independent experiments.



(By Jaesuk Lee in Seoul National University)

Figure 4. *In vitro* assay to measure the deaminase activity of E1347A-D10A nCas9. (a) Nickel agarose bead-based purification of His-tagged E1347A-D10A nCas9 protein from E. coli cell lysates was monitored using polyacrylamide gel electrophoresis. Coomassie blue was used to stain the gel. Lane 1, molecular weight indicators, with sizes of representative markers indicated to the left. Lane 2, sample from cells in which protein expression was not triggered by IPTG. Lane 3, Sample from cells in which protein expression was stimulated with IPTG. Lane 4, soluble fraction after sonication. Lane 5, insoluble fraction after sonication. Lane 6, flow-through fraction from the column. Lane 7, first wash fraction. Lane 8, second wash fraction. Lane 9, elution fraction. The red arrow indicates the E1347A-D10A nCas9 protein. (b) Time-dependent *in vitro* activity of E1347A-D10A nCas9 measured by targeted deep sequencing of a PCR amplicon containing the TYRO3 target. The

protospacer and PAM are underlined in blue and red, respectively. The colors in the heat map were determined from three independent experiments.

2. mtDNA editing by monomeric DdCBEs

a. On-target activity of mDdCBE

The results show that E1347A and GSVG variants can be used as monomeric $DddA_{tox}$ cytosine deaminases (mDdCBE). GSVG and E1347A variants were successfully fused to the C terminus of TALE arrays designed to bind to three mitochondrial genes, *ND4, ND6*, and *ND1*. The resulting mDdCBEs containing the GSVG variant achieved base editing at intended target nucleotide positions with high frequencies of up to 31% (*ND4*) (Figure 5a, b) 27% (*ND6*) (Fig. 5c, d), and 42% (*ND1*) (Fig. 5e, f), on par with the original split DdCBE pairs targeted to these sites (shown as L-1397N + R-1397C, indicating $DddA_{tox}$ split at G1397 fused to the left- or right-TALE array, and L-1333N + R-1333C, indicating $DddA_{tox}$ split at G1333 fused to the left- or right-TALE array in Figure 5). mDdCBEs containing E1347A also generated targeted C-to-T conversions, albeit less efficiently, with frequencies of up to 7.2% (*ND4*), 8.9% (*ND6*), and 13.7% (*ND1*) at these sites.

I tried to find out if mDdCBEs work in other mammalian cells. As a result, it was shown that mDdCBEs were highly efficient in mouse NIH3T3 cells, exhibiting editing frequencies of up to 32% at a target site in the *MT-ND5* gene (Figure 6).

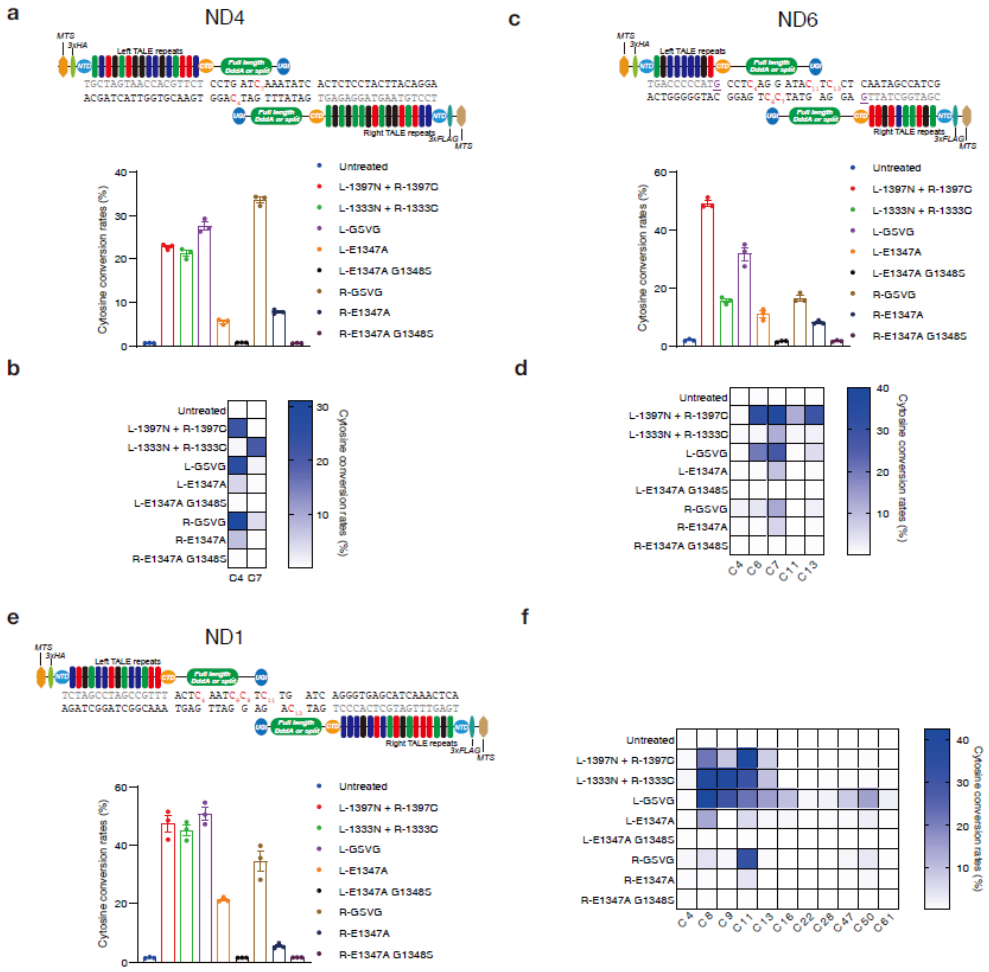


Figure 5. Base editing in the mitochondrial genome induced by mDdCBEs and DdCBEs. (a), (c), (e) Editing efficiencies of mDdCBEs and DdCBEs at the *ND4* (a), *ND6* (c), and *ND1* (e) sites. Target cytosines and TALE-binding sites are shown red and gray, respectively. The left and right TALE arrays are represented by L and R, respectively. Mismatches between the site recognized by the *ND6*-specific TALE array and the reference genome are underlined and shown in purple. (b), (d), (f) Heat maps showing the frequencies of C-to-T conversions at the indicated positions in the

ND4 (b), *ND6* (d), and *ND1* (f) sites. Means ± s.e.m. (a, c, e) and heat map colors (b, d, f) were determined from three independent experiments.

TTTTCCTACTGGTCCGATTCCACCCCCTC$_{10}$ACG ACTAATAATAACTTTATTTTAA
AAAAGGATGACCAGGCTAAGGTGGGGGAG$_{10}$TGC$_{13}$ TGATTATTATTGAAATAAAATT

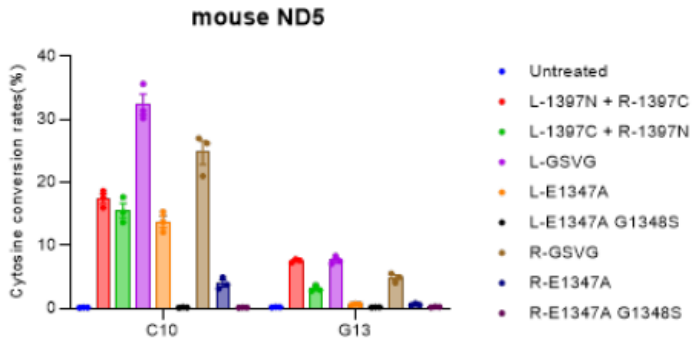
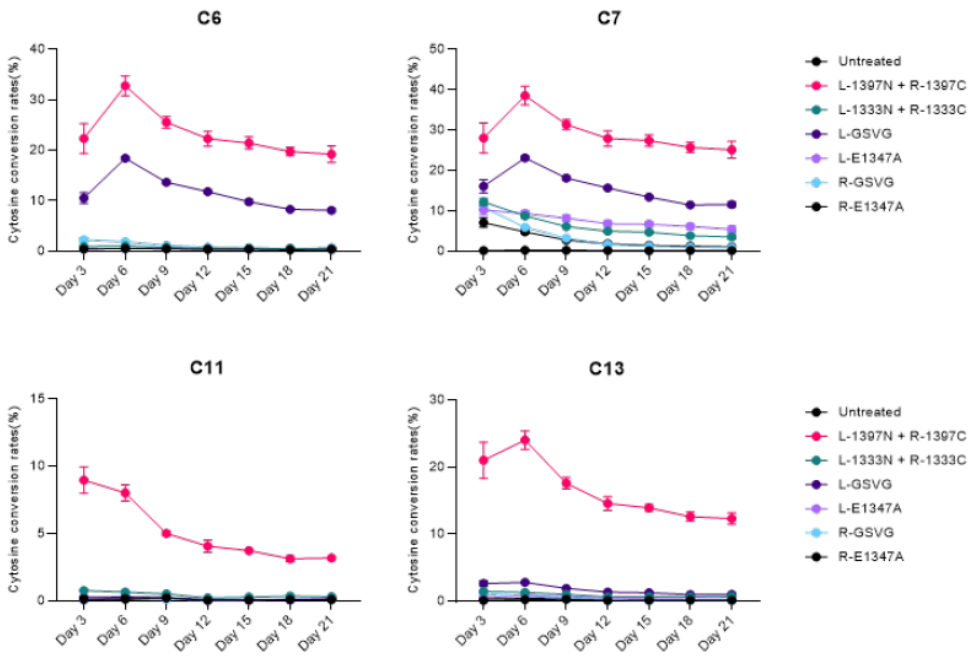


Figure 6. Mitochondrial base editing frequencies induced by the indicated DdCBEs and mDdCBEs in mouse NIH3T3 cells at the ND5 site. Target cytosines are shown in red and the left and right TALE-binding sites are shown in blue and green, respectively. The cytosine conversion rates were measured by targeted deep sequencing. Means ± s.e.m. were determined from three independent experiments.

b. Cytotoxicity and nuclear off-target analysis of mDdCBE

I next investigated whether mDdCBEs targeted to the mitochondria were cytotoxic or would induce off-target mutations in the nuclear genome. I found that mtDNA edits induced by *ND6*-specific mDdCBEs were maintained in HEK293T cells for at least 21 days (Figure 7). It suggests that mtDNA editing by mDdCBEs was well-tolerated.

mDdCBEs containing a mitochondrial targeting sequence (MTS) rather than a nuclear localization signal did not induce off-target mutations at a potential off-target site with a single-nucleotide mismatch in the nuclear genome (Figure 8).



(With Young Geun Mok in Institute for Basic Science)

Figure 7. Time-dependence of editing frequencies induced by the indicated DdCBEs and mDdCBEs in HEK293T cells. Editing efficiencies at each target cytosine in the *ND6* site are shown. The cytosine conversion rates were measured by targeted deep sequencing. Means ± s.e.m. were determined from three independent experiments.

a

mtDNA *ND4* on-target

TGCTAGTAACCACGTTCTCCTG$_{4}$ATC$_{7}$AAATATCACTCTCCTACTTACAGGA
ACGATCATTGGTGCAAGAGGAC$_{4}$TTG$_{7}$TTTATAGTGAGAGGATGAATGTCCT

Nuclear *MTNND4P12* pseudogene (chr5:134,926,846-134,926,895)

TGCTAGTAACCACATTCTCCTG$_{4}$ATC$_{7}$AAATATCACTCTCCTACTTACAGGA
ACGATCATTGGTGTAAGAGGAC$_{4}$TTG$_{7}$TTTATAGTGAGAGGATGAATGTCCT

b

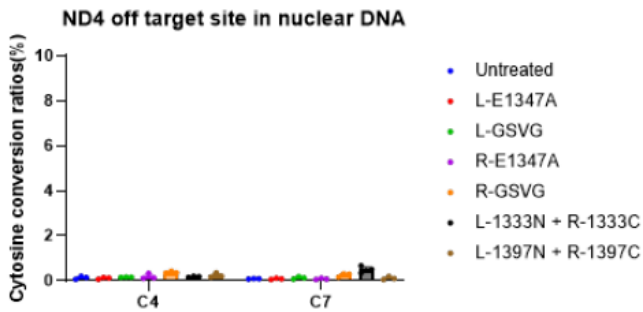


Figure 8. Off-target editing activity of mitochondrially-targeted DdCBEs and mDdCBEs in the nuclear DNA of HEK293T cells. (a) The on-target editing site in the mitochondrial ND4 gene and the corresponding nuclear DNA sequence with the highest homology. The nucleotide mismatch between the mtDNA and the nuclear MTNND4P12 pseudogene is shown in yellow. (b) Editing frequencies induced by the indicated DdCBEs and mDdCBEs at the off-target site in MTNND4P12. Target cytosines are shown in red and the left and right TALE-binding sites are shown blue and green, respectively. The cytosine conversion rates were measured by targeted deep sequencing. Means ± s.e.m. were determined from three independent experiments.

3. Advantages of mDdCBEs over conventional dimeric DdCBEs

a. Targeting single TALE binding site

Since the dimeric DdCBE requires two binding sites, the target TC motif is limited compared to the monomeric DdCBE. TALE-binding sites typically have preference to a thymine at the 5' and 3' termini. Although it is possible to ignore the need for a thymine at the 5' terminus or to use an engineered TALE N-terminal domain that recognizes all four bases (Lamb *et al*., 2013), it is unknown whether the resulting TALE proteins used in a DdCBE pair would be as efficient and specific as conventional TALE proteins recognizing a thymine at the 5' terminus. By *in silico* analysis, all TC motifs which dimeric DdCBE cannot adhere efficiently in human mitochondrial DNA were organized and analyzed (Figure 9a, Table 4). As shown, 8.4% of TC motifs in human mtDNA can potentially be edited by mDdCBE but not DdCBE.

As an example, I chose the *MT-TC* gene encoding tRNA-Cys: Various single-nucleotide substitutions in this gene are associated with myopathy or hearing loss17. I was able to design mDdCBEs but not dimeric DdCBEs to target a site in this gene, where there is no thymine within a stretch of 39 bp downstream of a potentially editable TC motif (Figure 9b). The mDdCBE containing the GSVG variant targeted to this site achieved C-to-T conversions at a frequency of 16%, demonstrating the advantage of mDdCBEs over conventional DdCBEs (Figure 9c, d).

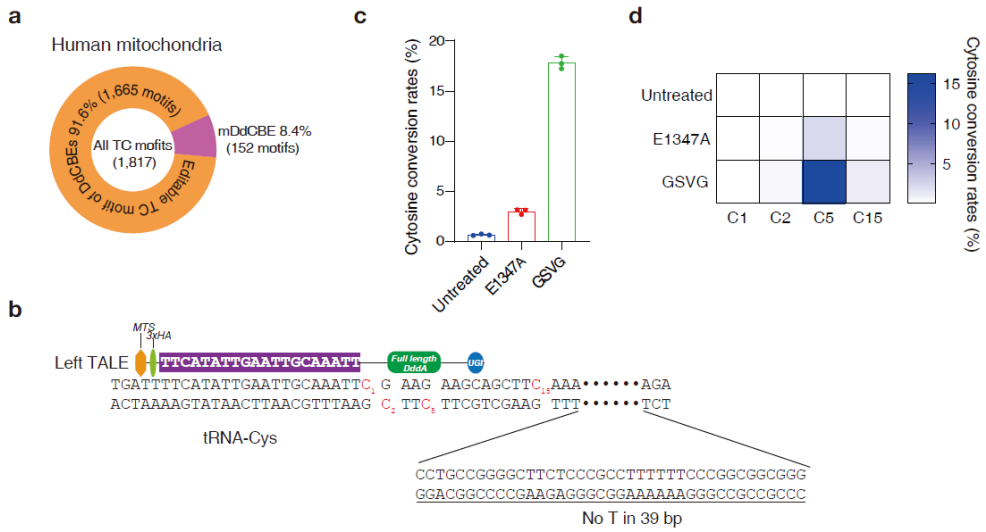


Figure 9. mDdCBE allows base editing in mtDNA at sites with a single TALE binding site. (a) Among all editable TC motifs in human mtDNA, 8.4% can potentially be edited by mDdCBE but not DdCBE. (b) Schematic of a mDdCBE target site in the *MT-TC* gene in human mtDNA. Target cytosines are shown in red. (c) Editing frequencies induced by mDdCBEs at the *MT-TC* site. (d) Heat map showing the frequencies of C-to-T substitutions at various positions in the MT-TC site. Means ± s.e.m. (c) and colors in the heat map (d) were determined from three independent experiments.

Table 4. ***In silico*** **analysis of target sites that contain only a single TALE-binding sequence in human mitochondria.**

chrom	5' binding seq	spacer seq	3' binding seq	L len	S len	R len	TC position
MT	TCACAGGTCTATCACCCT			18			36
MT	TATCACCCTATTAACCACT	CACGGGAGCTCTCCATGCATTTGGT	ATTTTCGTCTGGGGGGTA	19	25	18	41
MT	TAGCATTGCGAGACGCT	GGAGCCGGAGC	ACCCTATGTCGCAGTA	17	11	16	107
MT	TGGAGCCGGAGCACCCT	ATGTCGC	AGTATCTGTCTTTGA	17	7	15	120
MT	TGGAGCCGGAGCACCCT	ATGTCGCAGTATCTGTCTTTG	ATTCCTGCCTCATCCTA	17	21	17	128
MT	TTTCCACACAGACAT			15			295
MT	TCATAACAAAAAATTT	CCACCAAACCCCCCCTCCCCCGCTTCTGGCC	ACAGCACTTAAACA	16	31	14	311
MT			ACCCTAACACCAGCCTA			17	366
MT			AATTTTATCTTTTGGCGGTA			20	389
MT	TAACACCAGCCTAACCAGAT	TTCA	AATTTTATCTTTTGGCGGTA	20	4	20	394
MT			ACCAACCAAACCCCA			15	545
MT			ACCCCCCACAGTTTA			15	564
MT			AAAGCAATACACTGA			15	594
MT			AACCTTTAGCAATA			14	822
MT			AATAGAAGCCGGCGTA			16	922
MT	TTAAATCAACAAAACTGCT	CGCCAGAACACTACGAGCC	ACAGCTTAAAACTCA	19	19	15	1147
MT	TACGAGCCACAGCT	TAAAACTCA	AAGGACCTGGCGGTGCTTCA	14	9	20	1165
MT	TACGAGCCACAGCT	TAAAACTCAAAGGACCTGGCGGTGCTTC	ATATCCCTCTAGAGGA	14	28	16	1170
MT	TAAGAGTAGAGTGCTTAGT	TGAACAGGGCCCTGAAGCGCGT	ACACACCGCCCGTCA	19	22	15	1469
MT			ACCCTCCTCAAGTA			14	1491
MT			AGCTACCTAAGAACA			15	1913
MT			ACCTAAGAACAGCTA			15	1916
MT			ACAGCTAAAAGAGCA			15	1927
MT	TTGTAAATTTAACTGTTAGT	CCAAAGAGGAACAGCTCTTTGG	ACACTAGGAAAAAA	20	22	14	2116
MT			AATCTTACCCCGCCTGTTTA			20	2476
MT	TAAAAAAAGTAAAAGGAACT	CGGCAAATCTT	ACCCCGCCTGTTTA	20	11	14	2484
MT			ATGGAGCTTTAATTTA			16	2724
MT			AATTTATTAATGCA			14	2733
MT	TCCTAAACTACCAAACCT	GCATTAAAAATTTCGGTTGGGGCG	ACCTCGGAGCAGAA	18	24	14	2809
MT			AGCAGTACATGCTA			14	2842

(*Continued*)

MT			ATTCTAGAGTCCATA			15	2948
MT	TCTGAGTTCAGACCGGAGT			19			3093
MT			AGAGCCCGGTAATCGCA			17	3235
MT			AATCGCATAAAACTTA			16	3242
MT	TCTACCATCACCCT	CTACATCACCGCCCCGACCTTAGCTCTC	ACCATCGCTCTTCTA	14	28	15	3531
MT	TACATCACCGCCCCGACCT	TAGCTCTC	ACCATCGCTCTTCTA	19	8	15	3541
MT			ATTACTCCTGCCATCA			16	3822
MT			ACCTTGCCGAAGGGGA			16	3900
MT			AAGGGGAGTCCGAA			14	3901
MT			AGGGGAGTCCGAACTA			16	3910
MT			AACTAGTCTCAGGCTTCA			18	3916
MT			AACTAGTCTCAGGCTTCA			18	3920
MT	TTGCCGAAGGGGAGT	CCGA	ACTAGTCTCAGGCTTCA	15	4	17	3922
MT	TTGCCGAAGGGGAGT	CCGAACTAGTCTC	AGGCTTCAACATCGA	15	13	15	3930
MT	TTGCCGAAGGGGAGT	CCGAACTAGTCTC	AGGCTTCAACATCGA	15	13	15	3932
MT	TCCGAACTAGTCTCAGGCT	TC	AACATCGAATACGCCGCA	19	2	18	3939
MT			ATAATCCTTCTAATA			15	4669
MT	TTCTGAGTCCCAGAGGT	TACCCAAGGCACCCCTCTGAC	ATCCGGCCTGCTTCTTCTCA	17	21	20	4839
MT	TTACCCAAGGCACCCCT	CTGAC	ATCCGGCCTGCTTCTTCTCA	17	5	20	4841
MT	TTACCCAAGGCACCCCT	CTGACATCCGGCCTGCTTCTTCTC	ACATGACAAAAACTA	17	24	15	4846
MT	TAGCCCCCATCTCAAT	CATATACCAAATCTCTCCCTCACT	AAACGTAAGCCTTCTCCTCA	16	24	20	4904
MT	TACTACTATCTCGCACCT	GAAACAAGCTAACATGACT	AACACCCTTAATTCCA	18	19	16	5185
MT	TAACATGACTAACACCCT	TAATTCCATCC	ACCCTCCTCTCCCTA	18	11	15	5206
MT	TCATCATCCCCACCAT	CATAGCCACCATC	ACCCTCCTTAACCTCTA	16	13	17	5324
MT	TCATCATCCCCACCATCAT	AGCCACCATCACCCTCCTT	AACCTCTACTTCTA	19	19	14	5330
MT	TACAAAACCCACCCCAT	TCCTCCCCACACTCATCGCCCTT	ACCACGCTACTCCTA	17	23	15	5459
MT	TTAACAGCTAAGCACCCT	AATC	AACTGGCTTCAATCTA	18	4	16	5713
MT			AAGCCCCGGCAGGTTTGA			18	5758
MT			AGCCCCGGCAGGTTTGA			17	5760
MT			AGCTGCTTCTTCGA			14	5777
MT			AATTTGCAATTCAA			14	5787

(*Continued*)

MT			ATTGGAACACTATA			14	5943
MT			ACAACGTTATCGTCA			15	6062
MT			ACAGTCTACCCTCCCTTA			18	6280
MT			ACCCTCCCTTAGCA			14	6287
MT			AGCAGGGAACTACTCCCA			18	6294
MT	TCCCTTAGCAGGGAACT	ACTCCCACCCTGGAGCCTCCGTAG	ACCTAACCATCTTCTCCTTA	17	24	20	6322
MT	TCCCTTAGCAGGGAACTACT	CCCACCCTGGAGCCTCCGTAG	ACCTAACCATCTTCTCCTTA	20	21	20	6328
MT	TACTCCCACCCTGGAGCCT	CCGTAGACCTAACCATCTTCTCCTTACACCT	AGCAGGTGTCTCCTCTA	19	31	17	6344
MT	TGCCATAACCCAAT	ACCAAACGCCCCTCTTCGTCTG	ATCCGTCCTAATCA	14	22	14	6452
MT			ACCCCATTCTATACCA			16	6580
MT			ATTCTATACCAACA			14	6582
MT			ATACCAACACCTATTCTGA			19	6591
MT	TACACCCTAGACCAAACCT	ACGCCAAAATCC	ATTTCACTATCATA	19	12	14	7139
MT	TCGGACTACCCCGATGCAT	ACACCACATGAAACATCCT	ATCATCTGTAGGCTCA	19	19	16	7259
MT			ACCCCCCAAAGCTGGTTTCA			20	7462
MT	TAGACAAAAAAGGAAGGAAT	CGA	ACCCCCCAAAGCTGGTTTCA	20	3	20	7463
MT	TCGAACCCCCCAAAGCT	GGTTTCA	AGCCAACCCCATGGCCTCCA	17	7	20	7483
MT	TTTCATGATCACGCCCTCAT	AATCATTTTCCTTATCTGCTTCCTAGTCCTGT	ATGCCCTTTTCCTA	20	32	14	7687
MT	TAATTACATCACAAGACGT	CTTGCACTC	ATGAGCTGTCCCCA	19	9	14	8070
MT	TCTAAACCAAACCACT	TTCACCGCTACACGACCGGGGGTAT	ACTACGGTCAATGCTCTGA	16	25	19	8148
MT	TTTTAAGTTAAAGATT	AAGAGAACCAACACCTCTTT	ACAGTGAAATGCCCCA	16	20	16	8354
MT			ATTCATTGCCCCCA			14	8539
MT	TGAACGAAAATCTGT	TCGCTTC	ATTCATTGCCCCCA	15	7	14	8544
MT	TAGGCCTACCCGCCGCAGT	ACTGATC	ATTCTATTTCCCCCTCTA	19	7	18	8592
MT	TAGGCCTACCCGCCGCAGT	ACTGATC	ATTCTATTTCCCCCTCTA	19	7	18	8595
MT	TATCCTTAATCATTTTTAT	TGCCACAACTAACCTCCTCGG	ACTCCTGCCTCACTCA	19	21	16	8778
MT	TCACATGCCTATCATATAGT	AAAACCCAGCCCATGACCCCT	AACAGGGGCCCTCTCA	20	21	16	9253
MT			AGCCCTCCTAATGA			14	9272
MT			AAAAAGGCCTTCGA			14	9431
MT			ACCCCGCTAAATCCCCTA			18	9569
MT			AAGCACTGCTTATTA			15	9687

(*Continued*)

MT	TTCACCATTTCCGACGGCAT	CTACGGCTC	AACATTTTTTGTAGCCA	20	9	17	9790
MT	TTTTTTGTAGCCACAGGCT	TCCACGGACTTC	ACGTCATTATTGGCTCA	19	12	17	9820
MT	TTTTTTGTAGCCACAGGCT	TCCACGGACTTC	ACGTCATTATTGGCTCA	19	12	17	9825
MT	TTTTTTGTAGCCACAGGCT	TCCACGGACTTCACGTCATT	ATTGGCTCAACTTTCCTCA	19	20	19	9830
MT	TAATATTTCACTTT	ACATCCAAACATC	ACTTTGGCTTCGAA	14	13	14	9900
MT	TACCACAACTCAACGGCT	ACATAGAAAAATCCACCCCTT	ACGAGTGCGGCTTCGA	18	21	16	10159
MT	TAGAAAAATCCACCCCT			17			10170
MT	TAGAAAAATCCACCCCT			17			10181
MT	TAGAAAAATCCACCCCT			17			10182
MT	TTACGAGTGCGGCT			14			10192
MT	TGCGGCTTCGACCCTAT	ATCCCCCGCCCGCGTCCCTTTCTCC	ATAAAATTCTTCTTA	17	25	15	10205
MT			ATCCCTCTACTATTTTTTA			19	10867
MT			ACTATTTTTTAACCA			15	10870
MT	TTCACAGCCACAGAACT	AATC	ATATTTTATATCTTCTTCGA	17	4	20	11107
MT			ACGCAGGCACATACTTCCTA			20	11188
MT	TATGACTCCCTAAAGCCCAT	GTCGAAGCCCCCATCGCTGGGTC	AATAGTACTTGCCGCA	20	23	16	11434
MT	TTACCCCCCACTATTAACCT	ACTGGGAGAACTCTCTGTGCT	AGTAACCACGTTCTCCTGA	20	21	19	11896
MT			ATGTTCATACACCTA			15	12067
MT			AAGCTTCAAACTAGA			15	12564
MT			AAGCTTCAAACTAGA			15	12566
MT			AAACTAGACTACTTCTCCA			19	12576
MT			AGCCCTACTCCACTCA			16	13065
MT	TTAGGCGCTATCACCACT	CTGTTCGCAGCAGTCTGCGCCCTT	ACACAAAATGACATCA	18	24	16	13206
MT	TCAACCAACCACACCT	AGCATTCCTGCACATCTGT	ACCCACGCCTTCTTCA	16	19	16	13314
MT	TCAACCAACCACACCT	AGCATTCCTGCACATCTGT	ACCCACGCCTTCTTCA	16	19	16	13323
MT	TTCGAAAAATAGGAGGACT	ACTC	AAAACCATACCTCTCA	19	4	16	13425
MT	TTCGAAAAATAGGAGGACT	ACTCAAAACCATACCTCTC	ACTTCAACCTCCCTCA	19	19	16	13438
MT	TCCCTCACCATTGGCAGCCT	AGCATTAGCAGGA	ATACCTTTCCTCACA	20	13	15	13481
MT	TCGAAACCGCAAACATAT	CATACACAAACGCCTGAGCCCT	ATCTATTACTCTCA	18	22	14	13556
MT	TATCATACACAAACGCCT	GAGCCCTATCT	ATTACTCTCATCGCTA	18	11	16	13565
MT	TACACAAACGCCTGAGCCCT	ATCTATTACTCTC	ATCGCTACCTCCCTGA	20	13	16	13573

(*Continued*)

MT	TGGCAGCCGGAAGCCT	ATTCGC	AGGATTTCTCATTA	16	6	14	13725
MT	TGGCAGCCGGAAGCCT	ATTCGCAGGATTTCTCATTACT	AACAACATTTCCCCCGCA	16	22	18	13730
MT	TGGCAGCCGGAAGCCT	ATTCGCAGGATTTCTCATTACT	AACAACATTTCCCCCGCA	16	22	18	13735
MT	TGGCAGCCGGAAGCCT	ATTCGCAGGATTTCTCATTACT	AACAACATTTCCCCCGCA	16	22	18	13737
MT	TTTCCCCCGCATCCCCCT	TCCAAACAACAATCCCCCTCT	ACCTAAAACTCACA	18	21	14	13782
MT	TCGGATTCTACCCTAGCAT	CACACACCGCACAATCCCCT	ATCTAGGCCTTCTTA	19	20	15	13947
MT	TCACACACCGCACAAT	CCCCTATCT	AGGCCTTCTTACGA	16	9	14	13954
MT	TCAACCCAAAAAGGCAT			17			14100
MT			ATACTCTTTCACCCA			15	14346
MT			ACCTCCATCGCTAA			14	14372
MT	TCCATCGCTAACCCCACT	AAAACACTCACCAAGACCTC	AACCCCTGACCCCCA	18	20	15	14410
MT	TAACCCCACTAAAACACT	CACCAAGACCTC	AACCCCTGACCCCCA	18	12	15	14415
MT			AATACTAAACCCCCA			15	14578
MT			AACAGAAACAAAGCA			15	14650
MT	TCCAACATCTCCGCATGAT	GAAACTTCGGCTCACTCCTTGGCGCCTGCCTG	ATCCTCCAAATCACCA	19	32	16	14854
MT	TACTCAGTAGACAGT	CCCACCCTCACACGATTCTTTACCTTTC	ACTTCATCTTGCCCTTCA	15	28	18	15276
MT	TAAAATCACCTTCCACCCT			19			15428
MT	TTCCACCCTTACTACACAAT			20			15436
MT	TACACAATCAAAGACGCCCT	CGGCTTACTTCTCTTCCTTCTCTCCTT	AATGACATTAACACTA	20	27	16	15446
MT	TACACAATCAAAGACGCCCT	CGGCTTACTTCTCTTCCTTCTCTCCTT	AATGACATTAACACTA	20	27	16	15448
MT			AAGCCCGAATGATA			14	15550
MT	TGATATTTCCTATTCGCCT			19			15591
MT	TGATATTTCCTATTCGCCT	ACACAATTCTCCGATCCGTCCCTAACAAACT	AGGAGGCGTCCTTGCCCTA	19	31	19	15594
MT	TTAGCACCCAAAGCTAAGAT			20			16019
MT	TTTACCGTACATAGCACAT	TACAGTCAAATCCCTTCTCGTCCCC	ATGGATGACCCCCCTCA	19	25	17	16353
MT	TCCCTTGACCACCATCCT			18			16425
MT	TCCTCCGTGAAATCAAT			17			16434
MT	TCAATATCCCGCACAAGAGT	GCTACTCTCCTCGCTCCGGGCCC	ATAACACTTGGGGGTA	20	23	16	16444
MT	TAGCCCACACGTTCCCCT	TAAATAAGACATCACGATGG	ATCACAGGTCTATCA	18	20	15	16566

(*L len: Left TALE binding site length, S len: spacer length, R len: Right TALE binding site length)

b. AAV-mediated base editing in mtDNA by mDdCBE

mDdCBE can be delivered via a single AAV vector with a ~4.7-kb cargo capacity, but dimeric DdCBE cannot (Figure 10a). I produced AAV2 particles encoding mDdCBEs targeted to the *ND4* and *ND1* sites and transduced HEK293T cells with variable viral doses. At day 6 post-infection, base editing frequencies reached as high as 99.1% at the *ND4* site and 59.8% at the *ND1* site with high multiplicity of infection (Figure 10b-e). This result suggests that nearly homoplasmic (> 99%) mutations can be induced in mtDNA using AAV-mediated DdCBE delivery.

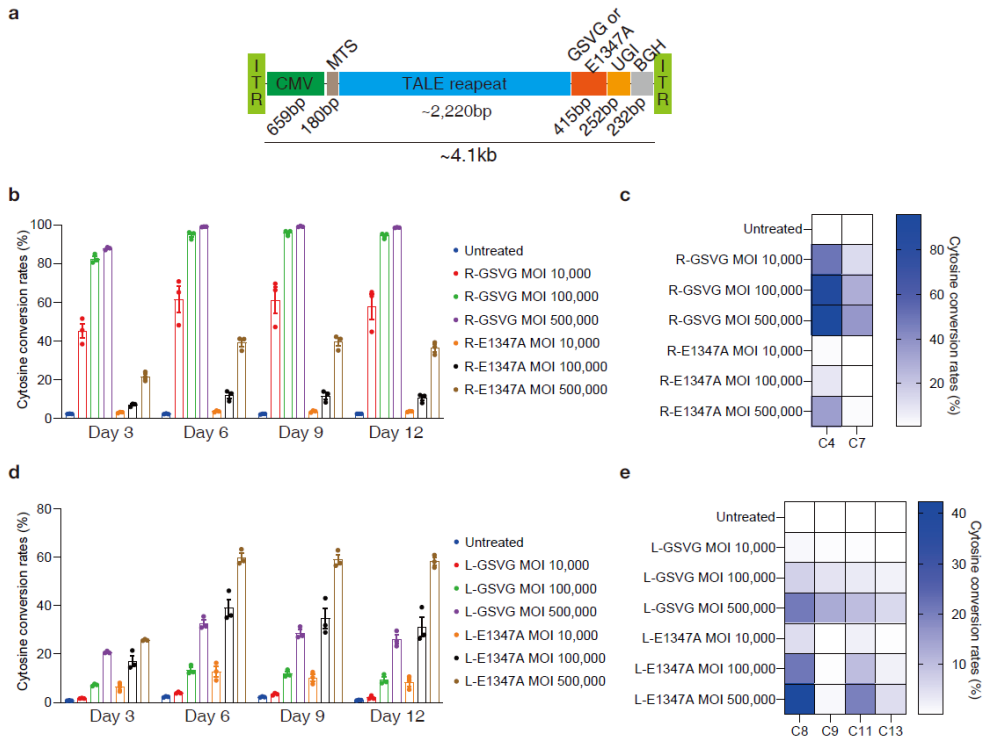


Figure 10. AAV-mediated base editing in mtDNA in HEK293T cells. (a) Schematic showing the AAV vector encoding mDdCBE. Time-dependence of editing frequencies induced by AAV2-encoded mDdCBEs at the *ND4* (b) and *ND1* (d) sites. Heat maps showing the frequencies of C-to-T substitutions at various positions in the *ND4* (c) and *ND1* (e) sites on Day 12. The cytosine conversion rates were measured by targeted deep sequencing. Means ± s.e.m. (b and d) and colors in the heat maps (c and e) were determined from three independent experiments.

4. Mitochondrial genome-wide target specificity

I assessed the mitochondrial genome-wide target specificity of split DdCBEs and mDdCBEs by performing high-throughput sequencing of DNA samples isolated from cells transfected with the *ND1*- and *ND6*-specific split DdCBEs or mDdCBEs containing the E1347A or GSVG variant (Figure 11a, b). At the same time, I also analyzed four TALE-free split or full-length $DddA_{tox}$ constructs to test whether they would induce random mutations across the mitochondrial genome. A mitochondrial-specific PCR primer set was used to amplify the whole mitochondrial DNA (Table 5). None of these TALE-free $DddA_{tox}$ constructs were mutagenic, compared with the negative control: Average frequencies of mitochondrial genome-wide C-to-T editing induced by these constructs ranged from 0.018% to 0.019%, not much different from that obtained with the negative control (0.019%). The split DdCBE pairs targeted to the two sites, however, showed off-target C-to-T editing with average frequencies that ranged from 0.031% to 0.039% (*ND1*) or 0.14% to 0.19% (*ND6*). The mDdCBEs targeted to the same sites also were mutagenic, inducing off-target C-to-T editing with average frequencies that ranged from 0.041% to 0.21% (*ND1*) or 0.12% to 0.23% (*ND6*) (Figure 11a, b). This result suggests that mDdCBE target specificities are not necessarily poorer than those of split DdCBE pairs.

Off-target activities of DdCBEs and ZFDs can be reduced or eliminated by delivering such constructs as *in vitro* transcripts rather than plasmids encoding them (Lim *et al.*, 2022). I investigated whether *in vitro* transcripts encoding mDdCBEs can also reduce off-target editing frequencies. *ND6*-targeted mDdCBEs, associated

with relatively high average frequencies of mitochondrial genome-wide off-target C-to-T editing, were transfected via variable amounts of mDdCBE-encoding mRNA into HEK293T cells. As expected, high editing frequencies of up to 20% were obtained with increasing doses of up to 800 ng mRNA (Figure 11c). Importantly, the use of 800 ng mRNA was as efficient as transfection of plasmid DNA in terms of on-target editing frequencies and reduced off-target C-to-T editing by 3.7 fold. Thus, the average frequency of mtDNA-wide off-target editing with 800 ng mRNA was 0.058% (Figure 11d), whereas that obtained with the mDdCBE plasmid was 0.21% (Figure 11b). Use of 200 ng mRNA achieved base editing with modest frequencies of up to 10% without inducing mtDNA-wide off-target editing, compared with that in the untreated control (Figure 11c, d). These results show that the amount of mDdCBE-encoding mRNA can be titrated to reduce or avoid off-target editing in human mtDNA.

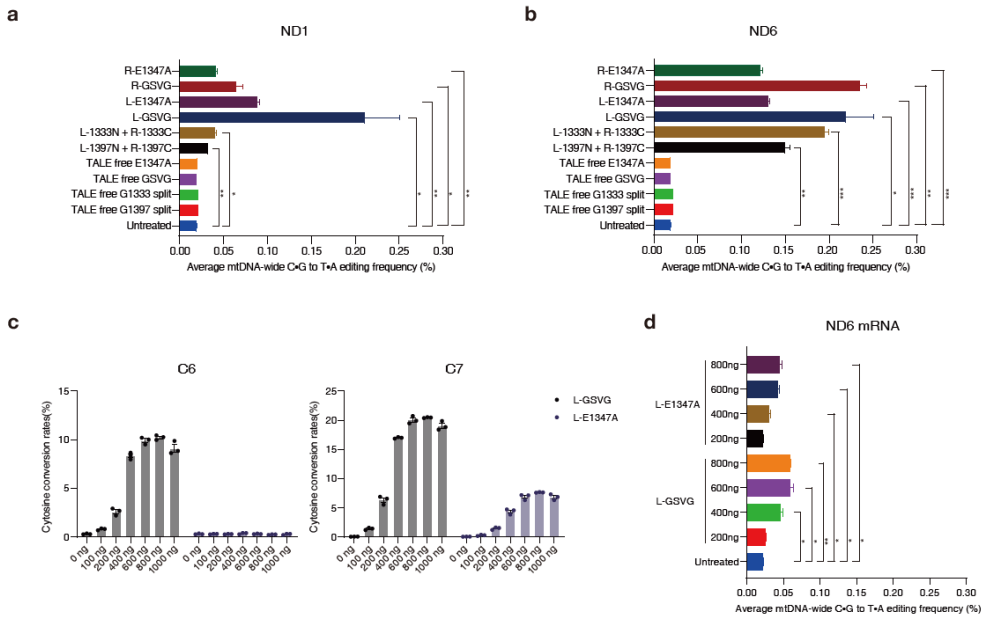


Figure 11. Mitochondrial genome-wide analysis of off-target editing by DdCBEs and mDdCBEs. (a), (b) Average frequencies of off-target C•G-to-T•A editing in mtDNA. HEK293T cells were transfected with plasmids encoding DdCBEs, mDdCBEs, TALE-free MTS-split $DddA_{tox}$-UGI, and TALE-free MTS-non-split DddA–UGI targeted to the *ND1* (a) and *ND6* (b) sites. (c) Editing frequencies at the C6 and C7 positions in the *ND6* site following transfection of various concentrations of mDdCBE-encoding mRNAs into HEK293T cells. (d) Average frequencies of off-target C•G-to-T•A editing in mtDNA from HEK293T cells transfected with mDdCBE-encoding mRNA. Means ± s.e.m. were determined from two (a, b, d) and three (c) independent experiments.

Table 5. Primer list for whole mitochondria genome amplification.

Name	Sequence 5` to 3`
Mito 9397-1892 F	aaagcacataccaaggccac
Mito 9397-1892 R	ttggctctccttgcaaagtt
Mito 15195-9796 F	tatccgccatcccatacatt
Mito 15195-9796 R	aatgttgagccgtagatgcc
Mito 2478-10858 F	gcaaatcttaccccgcctg
Mito 2478-10858 R	aattaggctgtgggtggttg
Mito 10653-2688 F	gccatactagtctttgccgc
Mito 10653-2688 R	ggcaggtcaatttcactgg

Ⅳ. Discussion

In this study, I presented full-length, non-toxic $DddA_{tox}$ variants, which can be fused to D10A or H840A nCas9 to make novel CRISPR RNA-guided cytosine base editors with altered editing windows or to custom-designed TALE DNA-binding proteins to create mDdCBEs enabling mtDNA editing in human and mouse cells. In addition, I demonstrated the deaminase activity of E1347A $DddA_{tox}$ variant by *in vitro* assay and cell experiments.

There were previous studies that engineering mtDNA by TALE (Bacman *et al.*, 2018; Bacman *et al.*, 2013). Although, because of the lack of DNA repair system in mitochondria, their methodology had limitations; mitoTALEN can only be applied by removing mutant mtDNAs (Kazak *et al*., 2012; Moretton *et al*., 2017; Shokolenko *et al*., 2013). In spite of miroTALEN, mDdCBEs can make *de novo* disease model and treat patients with no or few normal mtDNA in mitochondria population.

Interestingly, the split DdCBE pairs and mDdCBEs often produced different editing patterns. For example, the DdCBE pair (shown as L-1333N + R-1333C in Figure 5b) specific to the *ND4* gene was poorly active at the C4 position with an editing frequency of 0.8%, whereas two mDdCBEs containing the GSVG variant (shown as L-GSVG (Left TALE fused to the GSVG variant) and R-GSVG (Right TALE fused to the GSVG variant) in Figure 5b) were highly active at this position with editing frequencies of 26% and 31% (Figure 5b). I also noted that the *ND1* site-specific R-GSVG was highly selective, inducing C-to-T conversions

primarily at the C11 position, whereas two split DdCBE pairs targeted to the same site and L-GSVG were poorly discriminatory, inducing base edits at multiple positions in the editing window (Figure 5f). These results show that dimeric DdCBEs and mDdCBEs can create different mutation patterns and suggest that the two forms can be complementary to each other to induce diverse mutations at a given target site.

I examined the positions of C-to-T edits induced by mDdCBEs containing either the GSVG variant or the E1347A variant. I plotted the cytosine-editing frequencies of a total of 9 mDdCBEs at each nucleotide position downstream of a TALE-binding sequence to define an editing window for mDdCBEs (Figure 12). Positions 4 to 11 were more frequently converted than those immediately adjacent to or far downstream of the TALE-binding site. Thus, the editing window for mDdCBEs can be loosely defined as spanning nucleotides 4~11 downstream of a TALE-binding site.

mDdCBE-encoding genes, unlike those encoding dimeric DdCBEs, can be packaged into AAV and other viral vectors with a limited cargo space, facilitating *in vivo* studies and gene therapy. Importantly, the ND4-specific mDdCBE delivered via AAV induced C-to-T edits at a frequency of up to 99.1%, suggesting that nearly homoplasmic mutations can be achieved in mtDNA without drug selection. This suggests the potential of treating diseases that most of mitochondria population needs to be corrected via AAV encoding mDdCBE.

Furthermore, mDdCBEs can edit sites for which only a single TALE protein can be designed. I also found that mDdCBEs often yield mutation patterns different

from those obtained with dimeric DdCBEs. Scrutinizing other organelles by *in silico* analysis, there would be a lot more single TALE binding sites. For example, the length of *Arabidopsis thaliana* chloroplast genome is 154,478 bp, about 10 times larger than the human mitochondria genome built of 16,569 bp (Sato *et al*., 1999; Taanman, 1999). Through this, it can be estimated that there are much more single TALE binding sites in *Arabidopsis thaliana* chloroplasts than in human mitochondria.

I found, however, that certain, but not all, mDdCBEs were more prone to induce off-target editing in mtDNA, compared with dimeric DdCBEs. Fortunately, I was able to reduce or eliminate mitochondrial genome-wide off-target effects by using mDdCBE-encoding mRNAs rather than plasmid DNA. Furthermore, off-target effect can be reduced by the improvement of spatiotemporal controllability. There are the methods to control CRISPR-Cas9 function by small molecule activation, small molecule inhibition, bioresponsive delivery carriers and optical/ thermal/ultrasonic/magnetic activation of protein (Davis *et al*., 2015; Dow *et al*., 2015; Furuhata *et al*., 2017; Jain *et al*., 2016; Maji *et al*., 2017; Nguyen *et al*., 2016; Polstein and Gersbach, 2015; Shahbazi *et al*., 2019; Zhou and Deiters, 2016; Zhuo *et al*., 2021). By applying these techniques to mDdCBEs, it will be possible to make them specifically expressed only in the desired cells or to reduce the unwanted residual deaminase effect.

I believe that, together with dimeric DdCBEs and ZFDs for C-to-T editing and transcriptional activator-like effector nucleases (TALEDs) for A-to-G editing

(Cho *et al*., 2022), mDdCBEs can broaden the scope of organellar genome editing. And also, targeting sites is broadened by DdCBEs with evolved $DddA_{tox}$ to increase average editing efficiencies at AC and CC targets from less than 10% for canonical DdCBE to 15–30% and up to 50% (Mok *et al*., 2022). Combining new $DddA_{tox}$ variants and mDdCBE will lead to higher activity and expending targeting scope.

This study suggests further lines of investigation to improve base editing outcomes, including the modification of the editing window of $DddA_{tox}$ and characterization of the factors that distinguish well-edited and poorly edited sites. Non-toxic, full-length $DddA_{tox}$ was obtained by random mutagenesis with no evolution pressure to deaminase activity or deliberate cytotoxicity. Therefore, engineering $DddA_{tox}$ via directed evolution can illuminate new variants with altered sequence context and substrate preferences or higher activity with low toxicity.

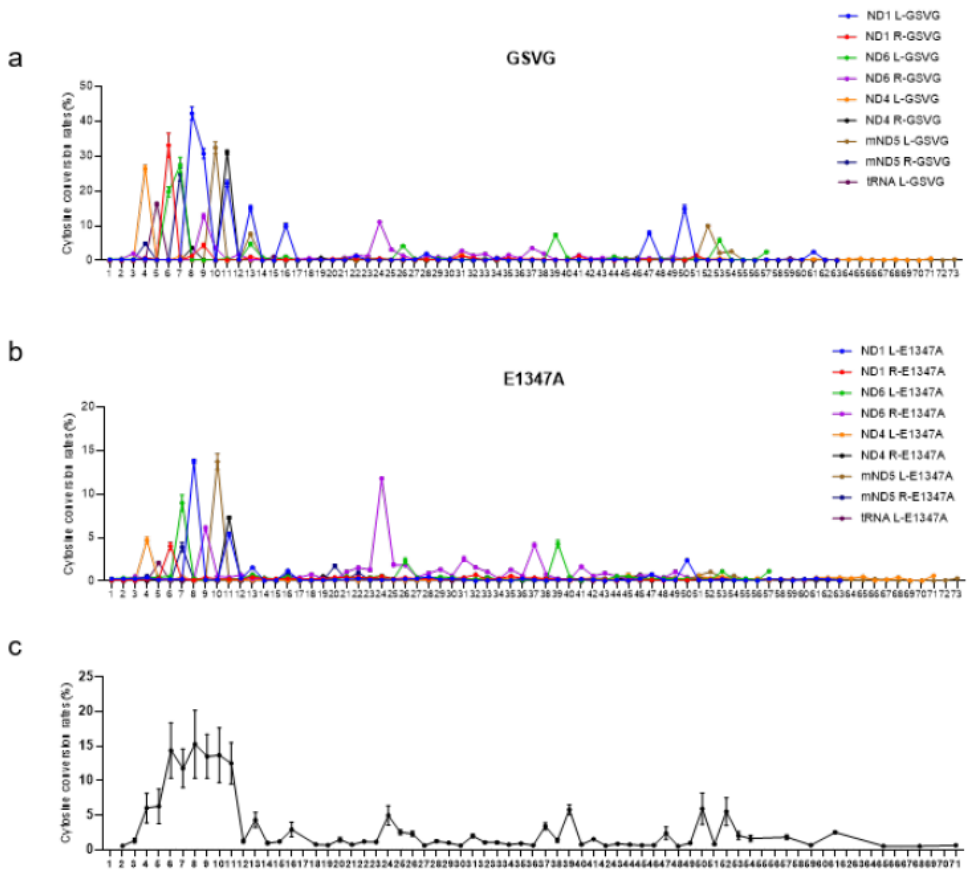


Figure 12. Defining the mDdCBE editing window. Editing frequencies at each nucleotide position downstream of the TALE-binding sequence were obtained at 9 sites following treatment with mDdCBEs containing either the GSVG variant (a) or the E1347A variant (b). Summary of the 18 sets of results (c).

Ⅴ. Reference

Archibald, J.M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. Curr Biol *25*, R911-921.

Bacman, S.R., Kauppila, J.H.K., Pereira, C.V., Nissanka, N., Miranda, M., Pinto, M., Williams, S.L., Larsson, N.G., Stewart, J.B., and Moraes, C.T. (2018). MitoTALEN reduces mutant mtDNA load and restores tRNA(Ala) levels in a mouse model of heteroplasmic mtDNA mutation. Nat Med *24*, 1696-1700.

Bacman, S.R., Williams, S.L., Pinto, M., Peralta, S., and Moraes, C.T. (2013). Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. Nat Med *19*, 1111-1113.

Cho, S.I., Lee, S., Mok, Y.G., Lim, K., Lee, J., Lee, J.M., Chung, E., and Kim, J.S. (2022). Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. Cell *185*, 1764-1776 e1712.

Coulthurst, S. (2019). The Type VI secretion system: a versatile bacterial weapon. Microbiology (Reading) *165*, 503-515.

Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A., and Liu, D.R. (2015). Small molecule-triggered Cas9 protein with improved genome-editing specificity. Nat Chem Biol *11*, 316-318.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat Biotechnol *33*, 390-394.

Duchen, M.R. (2004). Mitochondria in health and disease: perspectives on a new

mitochondrial biology. Mol Aspects Med *25*, 365-451.

Furuhata, Y., Nihongaki, Y., Sato, M., and Yoshimoto, K. (2017). Control of Adipogenic Differentiation in Mesenchymal Stem Cells via Endogenous Gene Activation Using CRISPR-Cas9. ACS Synth Biol *6*, 2191-2197.

Gammage, P.A., Rorbach, J., Vincent, A.I., Rebar, E.J., and Minczuk, M. (2014). Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. EMBO Mol Med *6*, 458-466.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., and Liu, D.R. (2017). Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. Nature *551*, 464-471.

Gorman, G.S., Chinnery, P.F., DiMauro, S., Hirano, M., Koga, Y., McFarland, R., Suomalainen, A., Thorburn, D.R., Zeviani, M., and Turnbull, D.M. (2016). Mitochondrial diseases. Nat Rev Dis Primers *2*, 16080.

Guo, J., Chen, X., Liu, Z., Sun, H., Zhou, Y., Dai, Y., Ma, Y., He, L., Qian, X., Wang, J., *et al*. (2022). DdCBE mediates efficient and inheritable modifications in mouse mitochondrial genome. Mol Ther Nucleic Acids *27*, 73-80.

Hagstrom, E., Freyer, C., Battersby, B.J., Stewart, J.B., and Larsson, N.G. (2014). No recombination of mtDNA after heteroplasmy for 50 generations in the mouse maternal germline. Nucleic Acids Res *42*, 1111-1116.

Hood, R.D., Singh, P., Hsu, F., Guvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L., *et al*. (2010). A type VI secretion system of Pseudomonas aeruginosa targets a toxin to bacteria. Cell

Host Microbe *7*, 25-37.

Jain, P.K., Ramanan, V., Schepers, A.G., Dalvie, N.S., Panda, A., Fleming, H.E., and Bhatia, S.N. (2016). Development of Light-Activated CRISPR Using Guide RNAs with Photocleavable Protectors. Angew Chem Int Ed Engl *55*, 12440-12444.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477-1481.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., *et al*. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Jo, A., Ham, S., Lee, G.H., Lee, Y.I., Kim, S., Lee, Y.S., Shin, J.H., and Lee, Y. (2015). Efficient Mitochondrial Genome Editing by CRISPR/Cas9. Biomed Res Int *2015*, 305716.

Joung, J.K., and Sander, J.D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol *14*, 49-55. 10.1038/nrm3486.

Kang, B.C., Bae, S.J., Lee, S., Lee, J.S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., and Kim, J.S. (2021). Chloroplast and mitochondrial DNA editing in plants. Nat Plants *7*, 899-905.

Kazak, L., Reyes, A., and Holt, I.J. (2012). Minimizing the damage: repair pathways

keep mitochondrial DNA intact. Nat Rev Mol Cell Biol *13*, 659-671.

Kim, Y., Kweon, J., Kim, A., Chon, J.K., Yoo, J.Y., Kim, H.J., Kim, S., Lee, C., Jeong, E., Chung, E., *et al*. (2013). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol *31*, 251-258.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20-36.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *533*, 420-424.

Lamb, B.M., Mercer, A.C., and Barbas, C.F., 3rd (2013). Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. Nucleic Acids Res *41*, 9779-9785.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., *et al*. (2016). ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res *44*, D862-868.

Lee, H., Lee, S., Baek, G., Kim, A., Kang, B.C., Seo, H., and Kim, J.S. (2021). Mitochondrial DNA editing in mice with DddA-TALE fusion deaminases. Nat Commun *12*, 1190.

Lim, K., Cho, S.I., and Kim, J.S. (2022). Nuclear and mitochondrial DNA editing in human cells with zinc finger deaminases. Nat Commun *13*, 366.

Maji, B., Moore, C.L., Zetsche, B., Volz, S.E., Zhang, F., Shoulders, M.D., and Choudhary, A. (2017). Multidimensional chemical control of CRISPR-Cas9. Nat Chem Biol *13*, 9-11.

Mok, B.Y., de Moraes, M.H., Zeng, J., Bosch, D.E., Kotrys, A.V., Raguram, A., Hsu, F., Radey, M.C., Peterson, S.B., Mootha, V.K., *et al*. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature *583*, 631-637.

Mok, B.Y., Kotrys, A.V., Raguram, A., Huang, T.P., Mootha, V.K., and Liu, D.R. (2022). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nat Biotechnol.

Moretton, A., Morel, F., Macao, B., Lachaume, P., Ishak, L., Lefebvre, M., Garreau-Balandier, I., Vernet, P., Falkenberg, M., and Farge, G. (2017). Selective mitochondrial DNA degradation following double-strand breaks. PLoS One *12*, e0176795.

Nguyen, D.P., Miyaoka, Y., Gilbert, L.A., Mayerl, S.J., Lee, B.H., Weissman, J.S., Conklin, B.R., and Wells, J.A. (2016). Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. Nat Commun *7*, 12009.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., *et al*. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science *353*.

Picardi, E., D'Erchia, A.M., Montalvo, A., and Pesole, G. (2015). Using REDItools to Detect RNA Editing Events in NGS Datasets. Curr Protoc Bioinformatics *49*, 12 12 11-12 12 15.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system

for control of endogenous gene activation. Nat Chem Biol *11*, 198-200.

Ratnaike, T.E., Greene, D., Wei, W., Sanchis-Juan, A., Schon, K.R., van den Ameele, J., Raymond, L., Horvath, R., Turro, E., and Chinnery, P.F. (2021). MitoPhen database: a human phenotype ontology-based approach to identify mitochondrial DNA diseases. Nucleic Acids Res *49*, 9686-9695.

Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., and Tabata, S. (1999). Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res *6*, 283-290.

Schon, E.A., DiMauro, S., and Hirano, M. (2012). Human mitochondrial DNA: roles of inherited and somatic mutations. Nat Rev Genet *13*, 878-890.

Shahbazi, R., Sghia-Hughes, G., Reid, J.L., Kubek, S., Haworth, K.G., Humbert, O., Kiem, H.P., and Adair, J.E. (2019). Targeted homology-directed repair in blood stem and progenitor cells with CRISPR nanoformulations. Nat Mater *18*, 1124-1132.

Shokolenko, I.N., Wilson, G.L., and Alexeyev, M.F. (2013). Persistent damage induces mitochondrial DNA degradation. DNA Repair (Amst) *12*, 488-499.

Silva-Pinheiro, P., Nash, P.A., Van Haute, L., Mutti, C.D., Turner, K., and Minczuk, M. (2022). In vivo mitochondrial base editing via adeno-associated viral delivery to mouse post-mitotic tissue. Nat Commun *13*, 750.

Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta *1410*, 103-123.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nat Rev Genet *11*, 636-

646.

Wanrooij, P.H., Uhler, J.P., Simonsson, T., Falkenberg, M., and Gustafsson, C.M. (2010). G-quadruplex structures in RNA stimulate mitochondrial transcription termination and primer formation. Proc Natl Acad Sci U S A *107*, 16072-16077.

Zhou, W., and Deiters, A. (2016). Conditional Control of CRISPR/Cas9 Function. Angew Chem Int Ed Engl *55*, 5394-5399.

Zhuo, C., Zhang, J., Lee, J.H., Jiao, J., Cheng, D., Liu, L., Kim, H.W., Tao, Y., and Li, M. (2021). Spatiotemporal control of CRISPR/Cas9 gene editing. Signal Transduct Target Ther *6*, 238.

국문초록

지난 몇 십 년 간, 프로그램 가능한 DNA 결합 단백질을 이용한 유전자 공학 기술은 놀라운 발전을 이루었다. TALE은 *Xanthomonas*속 식물 병원성 박테리아에서 유래한, 숙주 식물세포 유전자의 전사를 변화시키는 기능을 가진 단백질이다. 일반적으로 10-30번의 반복을 보이는 TALE 배열을 통해 DNA 서열을 인식하고 결합할 수 있다. CRISPR-Cas 시스템은 RNA에 의해 유도되는 DNA 절단 기능을 하지고 있으며, 이를 통해 박테리아에서 외부 DNA 침입을 방어하는 역할을 가지고 있다. 이러한 프로그램 가능한 DNA 결합 단백질은 강력하고 다양하게 응용 가능한 도구로서 생물학 연구에 혁명을 불러일으켰다.

프로그램가능한 핵산분해효소는 비상동성 말단 결합과 상동 재조합에 의존하는 이중가닥 절단을 유도하여 표적 유전자 서열을 바꾼다. 그렇기에 이 방법은 해당 부위에서의 DNA 복구 메커니즘에 따라 제한될 수 있다. *Burkholderia cenocepacia*의 독에서 유래한 $DddA_{tox}$는 이중가닥 DNA에 사이티딘의 탈아미노화를 촉매한다. 세포독성을 피하기 위해 기존 $DddA_{tox}$는 비활성 상태의 두 조각으로 나누어 TALE과 합친 DddA 유래 사이토신 염기교정인자 (DdCBEs)의 형태로 쓰인다.

이 연구에서는 프로그램가능한 DNA 결합 단백질을 응용하여 핵 DNA와 미토콘드리아 DNA를 연구할 수 있음을 보일 것이다. 제1장에

서는 CRISPR-Cas9 시스템과 체세포 핵 이식을 이용하여 디스트로핀 돌연변이 개를 생산할 수 있다는 것을 보일 것이다. 또한 생성된 디스트로핀 돌연변이 개가 실제 질병 모델에 적합한 표현형을 가지고 있다는 것을 보인다.

제2장에서는 기존 이량체 DdCBE를 독성이 없는 단량체 DdCBE (monomeric DdCBE, mDdCBE) 로 개량하고 그 특성을 밝힐 것이다. mDdCBE를 미토콘드리아 DNA 연구에 쓸 수 있음을 보이고 표적 특이성에 대해 검증하였다. 또한 하나의 TALE만 붙을 수 있는 표적 서열과 AAV 실험을 통해 mDdCBE의 장점을 입증하였다.

학 번: 2013-20282