이학박사학위논문

Highly efficient genome editing using RNA-guided base editors in mouse embryos

2018년 8월

서울대학교 대학원

화학부 생화학 전공

류 석 민

Abstract

Highly efficient genome editing using RNA-guided base editors in mouse embryos

Seuk-Min Ryu

Departments of Chemistry

The Graduate School

Seoul National University

The programmable nuclease, especially, CRISPR (clustered regularly interspaced short palindromic repeat) - Cas9 (CRISPR associated protein 9) has been significantly concerned as a powerful genome editing tool. The CRISPR-Cas9 system was dramatically developed and widely used to study gene functions and to generate model organisms in various animals and plants because it is highly specific and convenient to use. The applications were developed not only inducing double strand breaks (DSBs) effectively and precisely but using its specificity as a target-specific binding protein. Nickase Cas9 (nCas9) or dead Cas9 (dCas9) modified either one (D10A) or two (D10A/H840A) catalytic residues to lose its cleavage activity. Then, the modified protein was fused with other effector domains such as activators,

repressors, or deaminases and used for researches instead of knock-out by DSBs.

Point mutations are the major source of human genetic diseases rather than small insertions/deletions in the genome. To study human genetic disorders animal models are crucial elements. Using typical CRISPR-Cas9 system, inducing single-nucleotide substitutions is challenging because the most DNA DSBs are repaired by error-prone non-homologous end-joining rather than homologous recombination requiring a template donor. Among the recently developed Cas9 variants, base editors, composed of nCas9 or dCas9 and cytidine deaminases, substituting C · G to T · A or adenine deaminases, substituting A · T to G · C are powerful tools to induce point mutations effectively without a template donor in target specific manner.

I will describe inducing point mutations using cytidine deaminase in mouse embryos to generate mouse models because the base editing system had not been demonstrated yet in animals. I used Base Editor 3 to induce point mutations in two mouse genes, *Dmd* and *Tyr*. I generated a premature stop codon at the target sites efficiently and obtained homozygous mutant F0 mice with high frequencies.

Also, I will describe applying adenine base editors in mouse embryos. Also, I generated a mouse model targeting *Tyr* gene. However, the base editing systems have a limitation to design proper sgRNAs because of base editing windows where the editing efficiencies were the most effective between 14^{th} and 17^{th} position from protospacer adjacent motif sequence. With extended sgRNAs, I could increase base editing efficiencies in the position out of the canonical base editing windows.

Keywords: CRISPR-Cas9, Base editing, Base editor, Cytidine deaminase, Adenine deaminase, Mouse embryo editing

Student Number: 2013-22920

Table of Contents

List of Figures

List of Tables

Introduction

Target specific genome engineering is widely used to understand gene functions and non-coding elements. The programmable nucleases including zinc finger nucleases (ZFNs) (Bibikova et al., 2002; Kim et al., 1996; Lee et al., 2010; Urnov et al., 2005; Urnov et al., 2010), transcriptional activator-like effector nucleases (TALENs) (Boch et al., 2009; Christian et al., 2010; Kim et al., 2013; Miller et al., 2011; Moscou and Bogdanove, 2009) and clustered regularly interspaced repeat (CRISPR)-CRISPR associated (Cas) system (Cho et al., 2013; Cong et al., 2013; Jiang et al., 2013; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013b) are the key tools for genome engineering. Although those nucleases were derived from all the different origins the basic strategy of genome editing is the same. The programmable nucleases induce DNA double-strand breaks (DSBs) at the specific target sites. Then, DNA repair mechanisms are activated at the cleaved DNA loci.

When DSBs were generated the two major DNA repair pathways (Liang et al., 1998), non-homologous end joining (NHEJ) and homology-directed repair (HDR), are activated. NHEJ is a dominant repair pathway in higher eukaryotic cells or organisms (Shrivastav et al., 2008) and does not require homologous template donor DNAs. Because NHEJ is an error-prone pathway, it generates small insertions and deletions (indels) at the cleaved sites and causes frameshifts at the coding region and disturbs gene functions. This mechanism is used to study genetic information of targeted genes by knock-out (Kernstock et al., 2012; Kim et al., 2013; Steentoft et al., 2011). Also, another repair pathway, HDR is an error-free pathway requiring

donor DNA templates (Smithies et al., 1985). Even the pathway has an extremely low efficiency, it is usually used for gene correction, knock-in or inducing point mutations because it does not generate small indels (Cui et al., 2011; Soldner et al., 2011; Wefers et al., 2013; Yang et al., 2013).

The first two generation of programmable nucleases, ZFNs and TALENs, are fusion proteins consisted of DNA binding proteins such as zinc-finger proteins (ZFPs) or transcription activator-like effectors (TALEs) and nuclease domain of FokI type IIS restriction endonuclease (Bitinaite et al., 1998; Boch et al., 2009; Kim et al., 1996; Moscou and Bogdanove, 2009). Since the dimerization of FokI nuclease domains is required for the DNA double-strand cleavage, ZFNs and TALENs are delivered as a pair always. Also, because the DNA binding proteins have the specificity about the target sites. It is quite inconvenient and wastes time to change target sites since the new protein combination is necessary for the every the target site. In CRISPR-Cas system, however, Cas9 protein is a unique protein and only has nuclease activity inducing DSBs rather than target specificity. Only, CRISPR-RNA (crRNA) has the specificity about the genome locus (Horvath and Barrangou, 2010). Therefore, only crRNAs need to be changed depending on the target sites which is more convenient and less laborious to prepare (Cong et al., 2013; Ding et al., 2013) and has advantages for the multiple targeting. This characteristic could expand the genome engineering system applied to various research fields. Especially, CRISPR-Cas system is dramatically improved and enlarged because of its simplicity and effectiveness.

CRISPR-Cas system is originated from bacterial adaptive immune system

(Barrangou et al., 2007; Horvath and Barrangou, 2010; Sorek et al., 2013; Terns and Terns, 2011). When bacteria were infected by virus or bacteriophage, short DNA fragments from foreign DNA were integrated to the host genome at specific loci called CRISPR array. When the bacteriophage or virus infect the bacteria again, the integrated foreign DNA sequences are transcribed from the CRISPR array as pre-CRISPR RNA (pre-crRNA) and matured to process CRISPR RNA (crRNA). The matured crRNAs form a complex with Cas protein and lead the complex to the specific sites on the viral genome where have the complementary sequences with crRNAs. Then Cas protein cleaves the specific viral genome locus to protect from infection (Horvath and Barrangou, 2010). This characteristic inducing DSBs at the specific target sites is adopted in the field of genome engineering. Among the various bacteria species I borrowed type II CRISPR-Cas system composed of Cas9 protein, crRNA and trans-activating crRNA (tracrRNA) from *Streptococcus pyogenes* was mainly used.

Even CRISPR-Cas system was improved to induce DSBs more efficiently and specifically, various Cas9 variants were developed to use as a DNA binding protein in target specific manner. Nickase Cas9 (nCas9) or dead Cas9 (dCas9) modified either one (D10A) or two (D10A/H840A) catalytic residues to lose its cleavage activity. Then, the modified protein was fused with other effector domains such as activators, repressors, or deaminases. CRISPR-activator (CRISPRa) (Konermann et al., 2015; Maeder et al., 2013; Perez-Pinera et al., 2013) or CRISPR-repressor (CRISPRi) (Kiani et al., 2014; Radzisheuskaya et al., 2016; Rock et al., 2017) regulate transient gene expression level without the direct modification on the

genome. Recently developed Cas9 variants which is fused with cytidine or adenine deaminases called Base Editors (BEs) are able to induce point mutation efficiently without the template donor DNA.

The conventional method to introduce single-nucleotide substitutions uses HDR pathway requiring donor DNA templates with extremely low efficiencies. In 2016, several groups showed cytidine deaminases such as rAPOBEC1 (rat apolipoprotein B editing complex 1) (Komor et al., 2016) or AID (activation-induced deaminase) (Ma et al., 2016; Nishida et al., 2016) fused with nCas9 or dCas9 inducing C・G to T・A substitutions without DSBs and donor DNA templates. Even the cytidine deaminases fused with nCas9 or dCas9 are originated from different species the mechanism inducing C to T single-nucleotide substitution is basically same. When Cas9 protein and guide RNA (gRNA) complex binds to the target sites, the complex unwinds DNA double-strand and forms R-loops like wild-type Cas9 protein. Then, one strand which is called target strand interacts with guide RNA by Watson-Crick base paring and another strand, non-target strand is exposed as single-strand DNA. Cytidine in the non-target strands is deaminated and converted to uridine. During repair and replication process uridine which does not exist in DNA is converted to thymidine (Komor et al., 2016; Nishida et al., 2016).

Since uridine generated by deamination of cytidine in DNA is an error it is repaired by Uracil-DNA-Glycosylase (UDG). UDG excises uracil and generates apurinic/apymidinic sites (AP sites). Then, the AP site is filled with cytosine because of guanine of another DNA strand in most cases. In some cases, AP site is filled randomly with any bases during replications. However, when the UDG inhibitor

(UGI) is treated, uracil is not excised because the repair pathway is blocked and converted to thymine during replication. The efficiencies of C to T conversion is increased with UGI. Among the various version of BEs I used Base Editor 3 (BE3) consists of cytidine deaminase, nCas9 and UGI demonstrated the highest substitution efficiencies (Komor et al., 2016).

Using this new system, I generated mouse models (Kim et al., 2017b). Genetically engineered mouse is important to apply scientific researches to medicine or therapy. CRISPR-Cas system also affects the improvement of the conventional methods using ES-cells. The mutations are generated in ES-cells and performed positive or negative selection. Then ES-cells containing the desired mutations are injected into mouse blastocysts and researchers could obtain chimeric F0 mice are generated. After breeding for couple of generation, homozygous mutant mice could be obtained. This conventional methods take about 18 months to obtain homozygous mutant mice on average. CRISPR-Cas9 system can be delivered into mouse 1-cell stage zygotes and generate knock-out mice directly. It does not require any selection steps like the conventional methods. Also, F0 pups could be one of mosaic, heterozygous or homozygous mutant mice. Therefore, CRISPR-Cas system could reduce the time consuming to generate homozygous mutant mice. It usually takes 6 months on average. However, this improvement was especially applied to generate knock-out mouse models which use NHEJ repair pathways rather than HDR pathway requiring template DNA donors. Therefore, it is still remained as a limitation to generate mouse models efficiently.

Here I generated homozygous mutant mice inducing point mutations

efficiently using BE3 without template donors. I induced point mutations in two mouse genes, *Dmd* and *Tyr* genes generating a premature stop codon by C-to-T conversion at the target sites. BE3 mRNA and sgRNA or RNP complex were delivered by microinjection or electroporation, respectively. I could obtain blastocysts with high frequencies of single-nucleotide substitutions and homozygous mutant mice efficiently in both target genes.

Recently another type of base editor was developed. The new base editor called as adenine base editor (ABEs) composed of adenine deaminase and *Streptococcus pyogenes* Cas9 nickase enable adenine-to-guanine (A-to-G) single-nucleotide substitutions without DNA donor templates. The adenine deaminase is originated from *Escherichia coli*. TadA, a tRNA adenosine deaminase (Kim et al., 2006). Because the wild-type adenine deaminase did not work efficiently, the *E. coli.* TadA was engineered through the rational design and directed evolution. Unlike cytidine deaminases, rAPOBEC1 (Komor et al., 2016) or AID (Nishida et al., 2016), *E. coli.* TadA activates as a dimer (Losey et al., 2006). The latest version of ABEs, ABE 7.10, consists of wild-type TadA (TadA) and mutant TadA (TadA*) heterodimer (Gaudelli et al., 2017). The construct and the base editing mechanism, however, are very similar as CBEs. The deaminases are fused to N-terminus of Cas9 nickase and they deaminate adenosine in non-target strand exposed as single-strand DNA and convert adenine to inosine (A-to-I). Finally inosine is treated as guanine by polymerases. Also, the efficient base editing locus which are called as base editing windows are quite similar as CBEs.

Like my previous study with Base Editor 3 (BE3), ABEs can introduce

point mutations efficiently in mouse embryos, especially, A・T to G・C conversion without template donor DNAs because ABEs like CBEs do not induce double-strand break (DSBs) and do not rely on the endogenous DSB repair pathways such as HDR pathway. Also, they leave few unwanted small insertions or deletions (indels), induced by error-prone non-homologous end joining, at a target site. I used the latest version of ABEs, ABE 7.10 to generate a mouse model and approached for a therapeutic method to correct nonsense mutation in Duchenne muscular dystrophy (*Dmd*) gene (Ryu et al., 2018).

Base editing systems have their unique characteristics which is not observed in wild-type Cas9 protein. There are some specific positions where the base editing efficiencies are the highest. The specific location is called base editing window. Typically, the base editing window is known as positions 12 to 17 (suboptimal) or position 14 to 17 (optimal) of upstream of protospacer adjacent motif (PAM) sequence at a target protospacer site. However, the base editing window sometimes becomes a limitation to choose target sites. Base editing systems could not distinguish a specific cytosine or adenine where the base editing window. Therefore, to adjust this editing window, CBEs, namely, Base Editor 3 (BE3) or Target-AID, have been combined with truncated sgRNAs in cultured human cells (Kim et al., 2017c) to narrow down the window. On the other hand, the target cytosine or adenosine sometimes locates slightly away from the editing window because PAM sequence is also not appropriate. Therefore, a research to broaden the base editing window rather than narrowing down is also important. It would help to expand the range of applications. Even there was a case that tested extended sgRNAs

in *E. coli* with CBEs (Banno et al., 2018). I showed that the extended sgRNAs could enlarge the base editing scope of ABEs in mouse embryos and human cells.

Here I show that long gRNAs enable adenine editing at positions one or two bases upstream of the window that is accessible with standard single guide RNAs (sgRNAs). I demonstrate application of this technology in mouse embryos and adult mice. I introduced the Himalayan point mutation in the *Tyr* gene by microinjecting ABE mRNA and an extended gRNA into mouse embryos, obtaining *Tyr* mutant mice with an albino phenotype. Furthermore, I delivered the split ABE gene, using trans-splicing adeno-associated viral vectors, to muscle cells in a mouse model of Duchenne muscular dystrophy to correct a nonsense mutation in the *Dmd* gene, demonstrating the therapeutic potential of base editing in adult animals.

Materials and Methods

1. mRNA preparation

pCMV-BE3 was obtained from Addgene (cat. no. 73021). rAPOBEC1-XTEN and UGI were excised from the pCMV-BE3 vector and then placed into the pET-nCas9 (D10A)-NLS vector7 to produce pET-Hisx6-rAPOBEC1-XTEN-nCas9-UGI-NLS, which was used as the BE3 mRNA template. The mRNA template was prepared by PCR using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with primers (F: 5′-GGT GAT GTC GGC GAT ATA GG-3′, R: 5′-CCC CAA GGG GTT ATG CTA GT-3′). The pET_ABE7.10_nCas9 plasmid was generated based on ABE amino acid sequences from David Liu's study3. The codons were optimized for expression in human cells. BE3 and ABE7.10 mRNAs were synthesized using an *in vitro* RNA transcription kit (mMESSAGE mMACHINE T7 Ultra kit, Ambion) and purified with a MEGAclear kit (Ambion).

2. Preparation of ribonucleoproteins (RNPs)

Rosetta expression cells (EMD Milipore) were transformed with the pET28-Hisx6-rAPOBEC1-XTEN-nCas9 (D10A)-UGI-NLS (BE3) expression vector, and protein expression was induced for 12~14 h at 18 °C with 0.5mM isopropyl β-d-1-thiogalactopyranoside (IPTG). After induction, bacterial cells were harvested by centrifugation, and the pelleted cells were lysed by sonication in lysis buffer (50mM NaH2PO4 (pH 8.0), 300mM NaCl, 10mM imidazole, 1% Triton X-100, 1mM PMSF, 1mM DTT, and 1 mg/ml lysozyme). The cell lysate was then cleared by centrifugation at 5,251*g* for 30 min, and the soluble lysate was incubated

with Ni-NTA beads (Qiagen) for 1 h at 4 °C. Then, the Ni-NTA beads were washed with wash buffer (50mM NaH2PO4 (pH 8.0), 300mM NaCl, and 20mM imidazole) 3 times, and the BE3 protein was eluted with elution buffer (50mM Tris-HCl (pH 7.6), 150–500mM NaCl, 10–25% glycerol, and 0.2 M imidazole). The purified BE3 protein was exchanged and concentrated with storage buffer (20mM HEPES (pH 7.5), 150mM KCl, 1mM DTT, and 10% glycerol) using an Ultracel 100K cellulose column (Millipore). The purity of the protein was determined by SDS-PAGE. sgRNAs were prepared by *in vitro* transcription by T7 RNA polymerase as described in the previous study (Cho et al., 2013).

3. Animals

Experiments involving mice were approved by the Institutional Animal Care and Use Committee (IACUC) of Seoul National University. C57BL/6J and ICR mouse strains were used as embryo donors and foster mothers, respectively, and C57BL/6J, ICR, and *Dmd* knockout mice were maintained in an SPF (specific pathogen free) facility under a 12-h dark-light cycle.

4. Microinjection and electroporation of mouse zygotes

Microinjection and electroporation including superovulation and embryo collection were performed as described previously10. For microinjection, solutions containing complexes of BE3 and ABE7.10 mRNA (10 ng/μl) and sgRNA (100 ng/μl) were diluted in pyrocarbonate (DEPC)-treated injection buffer (0.25mM EDTA, 10mM Tris, pH 7.4)12 and injected into pronuclei using a Nikon ECLIPSE T*i* micromanipulator and a *FemtoJet 4i* microinjector (Eppendorf). For electroporation of one-cell stage embryos, the glass chamber of a NEPA 21 electroporator (NEPA

GENE Co. Ltd.) was filled with 100μl opti-MEM (Thermo Fisher Scientific) containing BE3-sgRNA RNP complexes (10μg/100μl and 6.5μg/100μl, respectively) (Hur et al., 2016). After RNP or mRNA delivery, embryos were cultured in microdrops of KSOM+AA (Millipore) at 37 ℃ for 4 days in a humidified atmosphere consisting of 5% CO2 in air. Two-cell-stage embryos were transplanted into the oviducts of 0.5-dpc (days post coitus) pseudo-pregnant foster mothers to obtain offspring. For *in vitro* analysis, I cultured microinjected zygotes for 4 days to obtain blastocysts. The sequences of *in vitr*o transcription of sgRNAs are listed in Table 1 and 7.

5. Genotyping

Genotyping was performed by targeted deep sequencing. Genomic DNA was extracted from blastocyst stage embryos or ear clips of newborn mice using a DNeasy Blood & Tissue Kit (QIAGEN Cat. No. 69504, Germany) for PCR genotyping and subjected to targeted deep sequencing and Sanger sequencing.

6. Targeted deep sequencing

Target sites were amplified from genomic DNA using Phusion polymerase (Thermo Fisher Scientific). The paired-end sequencing of PCR amplicons was performed using an Illumina MiniSeq. A list of primers used in this study is shown in Tables 1, 2 and 8.

7. Whole genome sequencing

Whole genome sequencing used mouse gDNA, which was extracted from mouse ear and was performed at a sequencing depth of 30 × to 40 × using an Illumina

HiSeq X Ten sequencer (Macrogen, South Korea) about two samples, wild-type and BE3-treated sample of *Dmd* gene (D108). I mapped the sequencing data using Isaac aligner with a mouse reference genome (GRCm38/mm10). Variants were identified by Isaac Variant Caller (IVC) and the dbSNP142 was sorted out from the identified variants by SnpEff. To figure out the potential off-target sites, I picked out that C and G converted to the other bases among the remained variants. Then I excluded the common variants between wild type and the BE3-treated sample. The putative off-target sites were compared with the candidates from Cas-OFFinder considering mismatch up to 7-bp or 2-bp bulge + 5-bp mismatches. For analysis of ABE sample, I trimmed out SNVs in *Tyr* #4 with those in the WT control using the program 'Strelka' with the default 'eland' option. The resulting putative off-target sites were compared with homologous sites, identified using Cas-OFFinder, with up to seven mismatches or with up to five mismatches and a DNA or RNA bulge.

8. Immunofluorescent staining and imaging of tissue

For *Dmd* mutant mouse using BE3, *Tibialis anterior* (TA) muscle sections were immunostained with laminin or dystrophin antibodies. Laminin was detected with a 1:500 dilution of a rabbit polyclonal antibody (abcam, ab11575) followed by a 1:1,000 dilution of Alexa Fluor 568 anti-rabbit secondary antibody (Thermo Fisher Scientific). Dystrophin was detected with a 1:500 dilution of a rabbit polyclonal antibody (abcam, ab15277) followed by a 1:1,000 dilution of Alexa Fluor 488 anti-rabbit secondary antibody (Thermo Fisher Scientific). Fluorescence of sections was observed with a Leica DMI4000 B fluorescent microscope. For ABE7.10 treated mouse, *Tibialis anterior* (TA) muscles were excised from tendon to tendon, and

OCT-embedded samples were rapidly frozen in liquid nitrogen-cooled isopentane. To assess muscle pathology, 10-mm cryosections were prepared. Serial sections were immunostained with anti-dystrophin antibody (Abcam, 15277) or anti-nNOS antibody (Abcam, 76067) followed by Alexa Fluor 488 anti-rabbit secondary antibody (Thermo Fisher Scientific, A-11034). Laminin was double-stained with an anti-laminin antibody (Sigma, L0663) followed by Alexa Fluor 568 anti-rat secondary antibody (Thermo Fisher Scientific, A-11077). For quantification, the number of dystrophin-positive myofibers in the whole cross-sectional area was counted in 1,171 to 1,717 individual myofibers using Adobe Photoshop. The percentage of dystrophin-expressing fibers was calculated by dividing the number of dystrophin-positive fibers by the number of laminin-expressing fibers (a measure of the total number of fibers).

9. Cell culture and transfection conditions

HEK293T/17 cells (ATCC, CRL-11268) were maintained in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin (Welgene). HEK294T/17 cells (~1 x 10^5) were seeded on 48-well plates (Corning) and transfected at ~70% confluency with ABE 7.10-expressing plasmid (750 ng) and sgRNA-expressing plasmid (250ng) using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocol. Genomic DNA was isolated using a DNeasy Blood & Tissue Kit (Qiagen) at 96h after transfection. The sequences of oligonucleotides for sgRNA plasmids are listed in Table 7.

10. Production and titration of AAV vectors

HEK293T cells (ATCC, CRL-3216) were transfected with pAAV-ITR-ABE-NT-sgRNA or pAAV-ITR-ABE-CT, pAAV2/9 encoding AAV2rep and AAV9cap, and helper plasmid. HEK293T cells were cultured in DMEM with 2% FBS. Recombinant pseudotyped AAV vector stocks were generated using PEI coprecipitation with PEIpro (Polyplustransfection) and triple-transfection with plasmids at a molar ratio of 1:1:1 in HEK293T cells. After 72 h of incubation, cells were lysed and particles were purified by iodixanol (Sigma-Aldrich) step-gradient ultracentrifugation. The number of vector genomes was determined by quantitative PCR.

11. Intramuscular injection of AAV

Intramuscular delivery of AAV (1 × 1013 viral genome) in physiological saline (40 μl) was performed via longitudinal injection into TA muscles of 7-week-old male *Dmd* knockout mice anesthetized with 2–4% isoflurane. Muscles were injected using an ultra-fine insulin syringe with a 31G needle (BD). As a negative control, C57BL/6J and *Dmd* knockout (D108)5 mice were used. To confirm the injection target, I carefully checked the corresponding tendon reflexes.

12. Statistical analysis

No statistical methods were used to predetermine sample size for *in vitro* or *in vivo* experiments. All group results are expressed as mean ±s.e.m. unless stated otherwise. Comparisons between groups were made using the two-tailed Student's *t*-test. Statistical significance as compared with controls was denoted with $*P < 0.05$,

$^{**}P < 0.01$, $^{***}P < 0.001$ in the figures and figure legends. Statistical analysis was performed in Graph Pad PRISM 7.

Table 1. Sequences of oligos used for sgRNA and targetd deep sequencing.

Sequences of oligos used for sgRNA and targeted deep sequencing.

sgRNA sequence	
Dmd	AAGCCAGTTAAAAATTTGTAAGG
Tyr	ACCTCAGTTCCCCTTCAAAGGGG
sgRNA primer for T7 in vitro transcription (5'-3')	
Dmd-F	GAAAT AATACGACTCACTATAGAAGCCAGTTA AAAATTTGTAGTTTTAGAGCTAGAAATAGCAAG
Tyr-F	GAAATTAATACGACTCACTATAGACCTCAGTTCCCCTTCAAAGGTTTTAGAGCTAGAAATAGCAAG
R	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC
Targeted deep sequencing primer (5'-3')	
Dmd-1st PCR-F	GCTAGAGTATCAAACCAACATCATTAC
Dmd-1st PCR-R	TGCTTCCTATCTCACCCATCT
Dmd-2nd/adaptor PCR-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTACAACAATTGGAACAGATGAC
Dmd-2nd/adaptor PCR-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCTTCACTGTACAGAGCTCAATG
Tyr-1st PCR-F	TGTATTGCCTTCTGTGGAGTT
Tyr-1st PCR-R	GGTGTTGACCCATTGTTCATT T
Tyr-2nd/adaptor PCR-F	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAGTTTCCAGATCTCTGATGG
Tyr-2nd/adaptor PCR-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGCACTGGCAGGTCCTATTAT

Table 2. List of primers used for targetd deep sequencing at potential off-target sites of *Dmd* and *Tyr* sgRNAs.

	1st PCR		2nd PCR	
No.	**Forward (5' to 3')**	**Reverse (5' to 3')**	**Forward (5' to 3')**	**Reverse (5' to 3')**
Dmd-OT1	TTTTTGCTCCTTAC AAACAAGG	TTATGTGGCCTT GCTCATTG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTCAGAGGAGCTACAAGCAAGG	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCAGTGAGTCATTGCATCCATC
Dmd-OT2	ACCCCAAAATTCAC CAGAGA	TGAGGTCAGGGA TGGTGATT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTGAAAGGATACATGGGCTGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAATCCTGACAATCCATGAACA
Dmd-OT3	TTTGAGCCTCTGG GAACATT	TCAGCCTCTTCT CCACTTTTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTATCTATGCTGCCGATGATCC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCCAAATTGTATAACTGAAGCAG
Dmd-OT4	TGAAGTCCTAGAAA ACAAAAGCA	TGGCATTGCATT GAATCTGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAAAGGATAAACAGGCTGAGAAAAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTTGGGATATCTTTCCAATTTCTGA
Dmd-OT5	GCTTTCTAAAGCCT TTTTAGCTTTT	CACCTGCCAAGT GTGGTATG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTCTGGCAGTTTACCCAAGG	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTTGAAAGTGACAGCAATTCCATT
Dmd-OT6	CAACCCATATATATT TTGGCCAGT	TGCCAATTGCCC TTTCTATC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCACCTAATCAGTGGCCTTT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCATGAGCATGGGAAATCTT
Dmd-OT7	TTTCAGGGAAAGGT GACAGG	TGTGTGCAATAAA CCACAGTGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTTCAGGGAAAGGTGACAGG	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGCAGTTTCACGTCTGGGAGT
Dmd-OT8	CATCCAAAGTGGCT TGAACA	GTCTGGCCCAAA CAGAATGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCACCTCCCATAACCGTAAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCATGAATCTTCCCACAAGGAA
Dmd-OT9	GCACCTAGATTTTG GCCATC	AATGCCAAATGC ATTGAAGG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTAAAGGCATGCACAACCAC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAATGCCAAATGCATTGAAGG
Dmd-OT10	TGCAAGTTGTCTTC CGACTG	AGGCAAGGTGAG GACTCAGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCCAGACATGCACACACATA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCATTCTCACAGTTATCCCAAA
Dmd-OT11	CGAGAGGTTGAGA CCTGGAG	CTCCTGAGGGTA GGGAGCTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAACCACACTACTGCTCTCATGC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTACCCCTGAGAAATGAACACG
Dmd-OT12	GTTTGGCGTGGGA TATGACT	GCACATCTTCCAT GTGCTGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTACTCCAACCAACAGGACGAT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTTAGGCTGTGGTGATGGCTTT
Dmd-OT13	AGAAGAGGGCCAT GAGTCAA	GGTCTTAGCCTC CAGCCTCT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAAGATAACGCCATAGCTGCAC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCCTTCTTTTCCAGCTCCTT
Dmd-OT14	CCAGTCGTTAGGC CTGTGAG	GGGTGACCTTGA ATTCCTGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTCAGGAGCGCACTAGAACCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAGGCATGTACCATGACCACA
Dmd-OT15	GAGAACGAGTGCC AAAGGAG	CCCATGTATTTTC CCATTGC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTAACACATGGTTGGCTCCTT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCTAGTTATGAGGGAACAGTTGCT
WGS-*Dmd*-OT	TTGCGAAGGTTGTT GTTTCC	CCCATTGCATGA AGACAAAG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTGCGAAGGTTGTTGTTTCC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCAACAAAAGGAGATCAAAGGGAT
Tyr-OT1	GCATGAGCACACA CTGAAATATAA	AGATGCCTGCTC TGTCTTTC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAGTCCCAAGTACCCTACTAC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAAACAAGAGCTCACATTATTGG
Tyr-OT2	AGTGGTTGGCTTCT CTCTTATC	GCATGGTATGTA CTCCCTGTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCTTCTCTCTTATCCCACTCATATC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCCACAGACAGTAAGAAGTTCA
Tyr-OT3	CATGTGTCTTCCTG GCTATCTT	TCTGTGGCCTAG AGGAGTAAT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTACTTGTGCTATGCATTGGTAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTATCAAGAGAAGGCAGCACATAG
Tyr-OT4	TTGTCTTCCTGTGT CTGCCTTA	AGATTATGCCCAA GGGGTTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCCTATCCTATCCCCCTCTGC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGGGACTAAACCACCACCAGA
Tyr-OT5	TGGCCAGAAGACTA GGATGG	AGGGTTTGCAAC TCCATAGG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTCTTTTCCCAGTTTCCTTTCC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTTGTAGCAGAGAATGGCCTTG
Tyr-OT6	TCACACCAGCTTGC CATT	GAGAAAGGACCA AAGGAGTTGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTACACCAGCTTGCCATTCTT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGGGAGAGGTCCTTGATCCTAT
Tyr-OT7	CCAACCAGAACCA CCAGAAT	CTCTTCCTCTTC CTCTTCCTCT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCAGAATTCCCAGGGACTAAGC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCTCCCTCCTCCTGATTCTATGA
Tyr-OT8	CAGTTTCGGTAGCC TTGACTTA	CAATTGATAGTG GCTGCCTAGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCAGTGCGATAACCCTTCTTGT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGGATTGCGAAACAGTGGATCT
Tyr-OT9	GGGAAATAGTAAGT AACAAGGAGAA	GAGACTGGAACA GCAAACAC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCCACTTGTATGAGGGTGTTTCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAAATGCCCATGCAGCTCT
Tyr-OT10	TTTTGTTGTCAGCT GGCTTG	TCCAGGGATTTT GTGTTGGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCCCTATCCCATCCACTTTCC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTAAGCCATCAACAAAGGATGG
Tyr-OT11	CTTTCCCAGTGCCA CCTAAA	GAGCCTTACAAA TACAGAGATGGA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCTTTCCCAGTGCCACCTAAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGACTAAGCCATCAACCAAGGA
Tyr-OT12	CAAGGCCGGAGAG TTTACTAAG	CCTCCGCTAACA CAACATACA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCCATCTAAGTAGGAAGGCTAGA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCAGCAGGGAGTAGGATGTAAGTA
Tyr-OT13	TACTCTGCTGCAAG AGGATTTC	TGTGTGTGTGTG TGTGTGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGATGGATGCCTACCTGACAAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGAAGTAGACAGCATAGAGTACAGAAG
Tyr-OT14	AAAGGACTGAAGG AGTTGAAGG	ACGTCCAGGAAG TTCTCTTTATG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCAAAGAGCTCACAGGGACTAAA	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTCCTGCTTCTATGAGGGTGTTC
Tyr-OT15	GGCCCTGTCTATTT ACTAGAGTTG	GGATATAACTCAC AGACCTCAAGAA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGGCTCCACATTTCCATTCATTC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGGATATAACTCACAGACCTCAAGAA
Tyr-OT16	AGGAAGGAAGAAA CTGAAACCA	GTGAGGCAAACC CACAAGTA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAAAGAGGCCCAAGGATCAC	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTATACATCTGACCCACCTTGC
Tyr-OT17	ATTGTTGTGTCTTC TGCCCTA	GCTCTATTACCCA GTTCCTTCC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAGTCACAACTGTAGGCACATT	GTGACTGGAGTTCAGACGTGTGCTCTTCCG ATCTGTTCCTTCCTCTACCAGAAAGC

OT: Off-target, WGS: Whole genome sequencing

Results

1. Targeting dystrophin (*Dmd)* and tyrosinase (*Tyr)* genes by cytidine-deaminase-mediated base editing in mouse blastocysts

I used Base Editor 3 (BE3) (rAPOBEC1-nCas9-UGI (uracil glycosylase inhibitor)) (Komor et al., 2016) to induce point mutations in two mouse genes, *Dmd* and *Tyr*, which encode dystrophin and tyrosinase, respectively (Figure 1). I expected that a premature stop codon would be generated by a single C-to-T conversion at the target site in each gene. First, I carried out base editing in mouse embryos by microinjection of BE3-encoding mRNA and single guide RNAs (sgRNAs). Targeted point mutations were observed in 11 out of 15 (73%) or 10 out of 10 (100%) blastocysts at the target site in the *Dmd* or *Tyr* gene, respectively, with mutation frequencies that ranged from 16% to 100% (Figure 2 and Table 3, 4). Targeted deep sequencing showed that a C-to-T conversion was the major mutagenic pattern at the two target sites. C-to-A or C-to-G conversions were also observed in four *Dmd* (36%) and two *Tyr* (20%) mutant blastocysts.

I purified the recombinant rAPOBEC1-nCas9 (D10A)-UGI protein from *Escherichia coli*. and delivered the resulting BE3 ribonucleoproteins (RNPs), which consisted of the recombinant BE3 protein and *in vitro*-transcribed sgRNAs, into mouse embryos by electroporation (Hur et al., 2016). At day 4 after electroporation, BE3 RNPs efficiently induced mutations in blastocysts with a frequency of 81% (13

out of 16) or 85% (11 out of 13) at the *Dmd* or *Tyr* target site, respectively (Figure 3 and Table 3, 4).

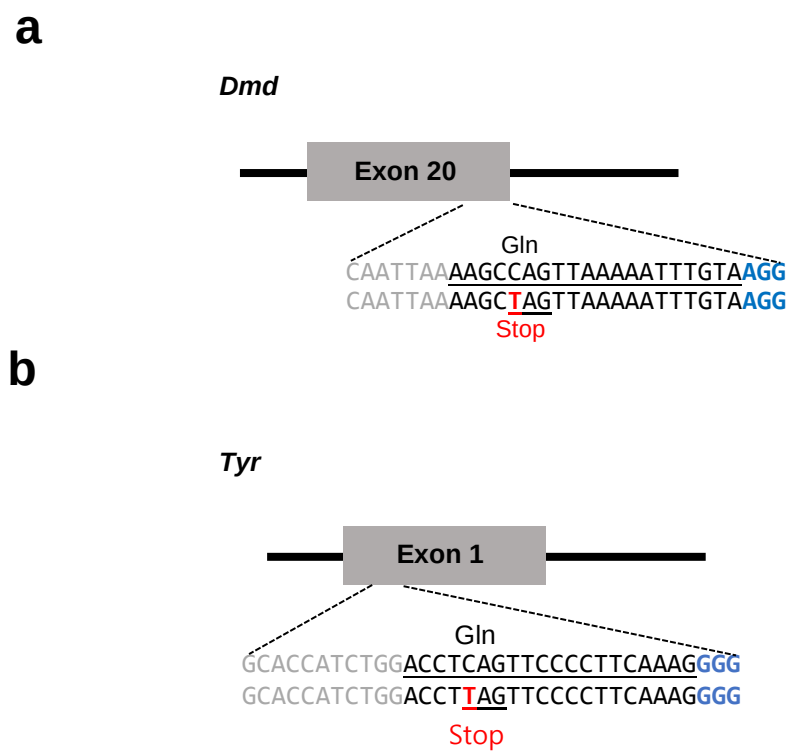


Figure 1. Scheme for the design of sgRNAs to induce premature stop codon. (a) The target sequence at the dystrophin (*Dmd*) locus. The PAM sequence and the sgRNA target sequence are shown in blue and black, respectively. The nucleotide substituted by cytidine-deaminase-mediated base editing is shown in red. (b) The target sequence at the *Tyr* locus. The PAM sequence and the sgRNA target sequence are shown in blue and black, respectively. The nucleotide substituted by BE3-mediated base editing is shown in red.

a

***Dmd*, mRNA microinjection**

		Frequency (%)	
Wt	ACAGCAATTAAAAGCCAGTTAAAAATTTGTAAGG		
#64	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	94	
#65	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	53	
	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	38	
#66	ACAGCAATTAAAAGACAGTTAAAAATTTGTAAGG	55	
	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	34	
	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	6	
	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	3	
#67	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	16	
#68	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	97	
#72	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	71	
	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	25	
#79	ACAGCAATTAAAAGGTAGTTAAAAATTTGTAAGG	99	
#80	ACAGCAATTAAAAGTAAGTTAAAAATTTGTAAGG	39	
	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	20	
	ACAGCAA----------AAATTTGTAAGG	40	(-15 bp)
#87	ACAGCAATTAAAAGCAAGTTAAAAATTTGTAAGG	84	
	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	12	
#88	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	77	
	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	22	
#95	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	54	
	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	45	

b

***Tyr*, mRNA microinjection**

		Frequency (%)	
Wt	GCACCATCTGGACCTCAGTTCCCCTTCAAAGGGG		
#47	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	73	
#48	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	99	
#49	GCACCATCTGGACCTTAGTTTCCCTTCAAAGGGG	49	
	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	45	
#50	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	99	
#51	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	56	
	GCACCATCTGGACCTCAGTTCCCTTTCAAAGGGG	14	
	GCACCATCTGGACCTCA--------TCAAAGGGG	25	(-8 bp)
#52	GCACCATCTGGACCTTAGTTCCC-TTCAAAGGGG	100	(-1 bp)
#53	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	26	
	GCACCATCTGGACCTAAGTTCCCCTTCAAAGGGG	23	
#54	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	57	
#55	GCACCATCT----------G GGG	16	(-21 bp)
	GCACCATCTGGACCTAAGTTCCCCTTCAAAGGGG	15	
	GCACCATCTGGATCTTAGTTCCCCTTCAAAGGGG	14	
	GCACCATCTGGACCTTAGTTACCCTTCAAAGGGG	11	
#56	GCACCATCTGGACCTTAGTTCCCCTTCAAAGGGG	86	

Figure 2. Targeted mutagenesis in mouse embryos by microinjection of deaminase-nCas9 mRNA and sgRNAs. (a) and (b) Alignments of mutant sequences from blastocysts that developed after microinjection of BE3-encoding mRNA and sgRNAs into zygotes. The target sequence is underlined. The PAM sequence and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles. Wt, wild-type.

a

***Dmd*, RNP electroporation**

		Frequency (%)	
Wt	ACAGCAATTAAAAGCCAGTTAAAAATTTGTAAGG		
#17	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	85	
#18	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	66	
	ACAGCAATTAAAAGCAAGTTAAAAATTTGTAAGG	18	
#19	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	64	
	ACAGCAATTAAAAGCAAGTTAAAAATTTGTAAGG	24	
#20	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	57	
	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	30	
#21	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	100	
#22	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	96	
#23	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	42	
#24	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	51	
#25	ACAGCAATTAAAAGTTAGTTAAAAATTTGTAAGG	68	
	ACAGCAATTAAAAGTCAGTTAAAAATTTGTAAGG	8	
#26	ACAGCAATTAAAAGCCAGTT--------GTAAGG	98	(-8 bp)
#28	ACAGCAATTAAAAGTAAGTTAAAAATTTGTAAGG	100	
#31	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	77	
#32	ACAGCAATTAAAAGCTAGTTAAAAATTTGTAAGG	100	

b

***Tyr*, RNP electroporation**

		Frequency (%)	
Wt	GCACCATCTGGACCTCAGTTCCCC-TTCAAAGGGGTGG		
#83	GCACCATCTGGACCTAAGTTCCCC-TTCAAAGGGGTGG	25	
	GCACCATCTGGACCTCAGTTACCC-TTCAAAGGGGTGG	18	
	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	14	
#34	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	54	
	GCACCATCTGGACCTGAGTTCCCC-TTCAAAGGGGTGG	45	
#36	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	48	
#37	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	56	
#38	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	94	
#40	GCACCATCTGGACCTGAGTTCCCC-TTCAAAGGGGTGG	59	
	GC----------TTCAAAGGGGTGG	21	(-22 bp)
	GCACCATCTGGACCTCAGTTCCCC-TTCAAAGGGGTGG	14	
#41	GCACCATCTGGACCTTAGTTTCCC-TTCAAAGGGGTGG	44	
#42	G----------G	35	(-35 bp)
#44	GCACCATCTGGACCTTAGTTCCCCTTTTAAAGGGGTGG	25	(+1 bp)
	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	9	
	GCACCATCTGGACCTCAGTTCTCC-TTCAAAGGGGTGG	3	
#45	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	55	
#46	GCACCATCTGGACCTCAG----------TGG	76	(-16 bp)
	GCACCATCTGGACCTTAGTTCCCC-TTCAAAGGGGTGG	23	

Figure 3. Targeted mutagenesis in mouse embryos by electroporation of rAPOBEC1-nCas9 (D10A)-UGI RNPs. (a) and (b) Alignments of mutant sequences from blastocysts that developed after electroporation of BE3 RNP into zygotes. The target sequence is underlined. The PAM site and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles. Wt, wild-type.

Table 3. Summary of the numbers of embryos used and mutants obtained afted targeting the *Dmd* site.

Target gene	Methods		No. of examined embryos	No. of two-cell stage embryos (%)	No.of blastocysts (%)	No. of transferred embryos	No. of offspring (%)	Mutant ratio (%)	
								No. of mutants/total blastocysts or offspring	No. of nonsense mutants/total blastocysts or offspring
Dmd	mRNA+sgRNA	Microinjection	33	30 (91)[a]	15 (50)[b]	NA	NA	11/15 (73)	10/15 (67)
	mRNA+sgRNA	Microinjection	83	66 (80)[a]	NA	66	11 (17)[c]	5/9 (56)	4/9 (44)
	RNP	Electroporation	55	52 (95)[a]	16 (31)[b]	NA	NA	13/16 (81)	9/16 (56)

NA: not applicable.
[a]Calculated from the number of examined embryos.
[b]Calculated from the number of developed 2-cell stage embryos.
[c]Calculated from the number of transferred embryos.

Table 4. Summary of the numbers of embryos used and mutants obtained afted targeting the *Tyr* site.

Target gene	Methods		No. of examined embryos	No. of two-cell stage embryos (%)	No. of blastocysts (%)	No. of transferred embryos	No. of offspring (%)	Mutant ratio (%)	
								No. of mutants/total blastocysts or offspring	No.of nonsense mutants/total blastocysts or offspring
Tyr	mRNA+sgRNA	Microinjection	23	21 (91)[a]	10 (48)[b]	NA	NA	10/10 (100)	10/10 (100)
	RNP	Electroporation	53	47 (89)[a]	14 (30)[b]	NA	NA	11/13 (85)	9/13 (69)
	RNP	Electroporation	63	55 (87)[a]	NA	55	7 (13)[c]	7/7 (100)	4/7 (57)

NA: not applicable.
[a]Calculated from the number of examined embryos.
[b]Calculated from the number of developed 2-cell stage embryos.
[c]Calculated from the number of transferred embryos.

2. Generating substitutions in mouse models *via* Base Editor 3

a. Generating mouse models using Base Editor 3 targeting *Dmd* and *Tyr* genes

After microinjection of BE3 mRNA and sgRNA, I transplanted mouse embryos into surrogate mothers and obtained offspring with point mutations in the *Dmd* gene. Five out of nine mice carried mutations at the target site. Notably, three (D102, D103, and D108) out of the five mutant mice harbored one or two mutant alleles and lacked the wild-type allele. The other two mutant mice (D107 and D109) harbored the wild-type allele at a frequency of 10% in addition to a mutated allele. One mutant mouse (D109) had a 20-base-pair (bp) deletion rather than a point mutation (Figure 4 and 5), consistent with previous reports that Cas9 nickase can still induce indels (Cho et al., 2014; Mali et al., 2013a; Ran et al., 2013). One mutant F0 mouse (D108) with no wild-type allele carried a premature stop codon in the *Dmd* gene caused by a single C-to-T conversion (Fig. 1c). Immunostaining showed that this *Dmd* mutant mouse rarely expressed the dystrophin protein, consistent with its mutant genotype (Figure 6).

Also, I transplanted mouse embryos into surrogate mothers and obtained offspring with targeted mutations in the *Tyr* gene after electroporation of BE3 RNPs. Notably, all seven pups carried various mutations at the *Tyr* target site (Figure 7a). Targeted deep sequencing showed that two newborn mice (T113 and T114) were homozygous for a nonsense mutation at the target site; a stop codon was generated by a single C-to-T conversion (Figure 7b). These two newborn mice showed an

albino phenotype in the eye, consistent with their mutant genotype (Figure 7c). The other mutant mice harbored missense mutations (Q68E or Q68K) or were mosaic, carrying the wild-type allele. Taken together, these results show that base editing is highly efficient in mouse embryos. Note that I had previously demonstrated site-specific mutagenesis in mice using TALENs (Sung et al., 2013) and Cas9 (Sung et al., 2014) and Cpf1 (Hur et al., 2016) nucleases but rarely observed a 100% mutation frequency with no evidence of mosaicism in F0 animals.

Figure 4. Targeted deep sequencing of newborn pups targeting *Dmd* gene. Alignments of mutant sequences from newborn pups that developed after microinjection bf BE3-encoding mRNA and sgRNAs into zygotes. The target sequence is underlined. The PAM sequence and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles. Wt, wild-type.

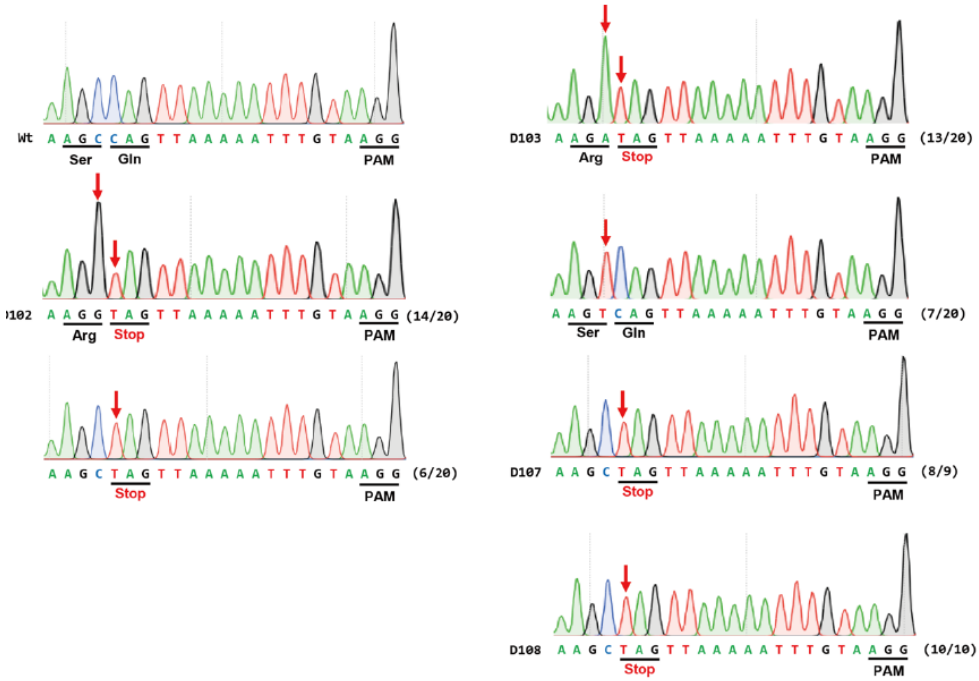


Figure 5. Sanger sequencing chromatograms of wild type and *Dmd* mutant mice.

Sanger sequencing chromatograms of DNA from wild-type and D108 mutant mice. The red arrow indicates the substituted nucleotide. The relevant codon identities at the target site are shown under the DNA sequence.

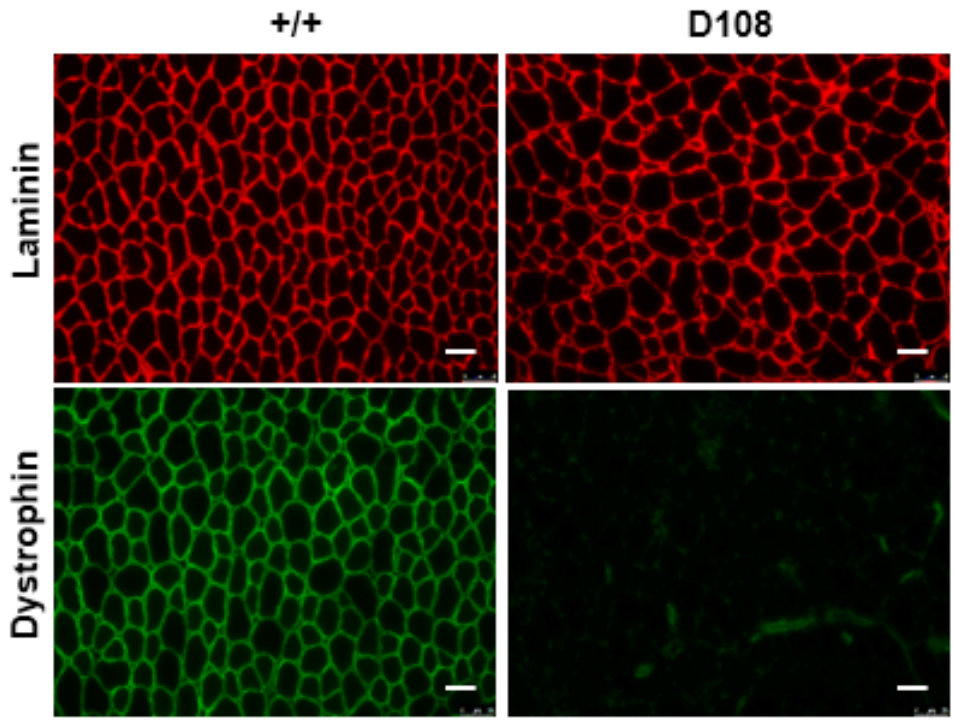
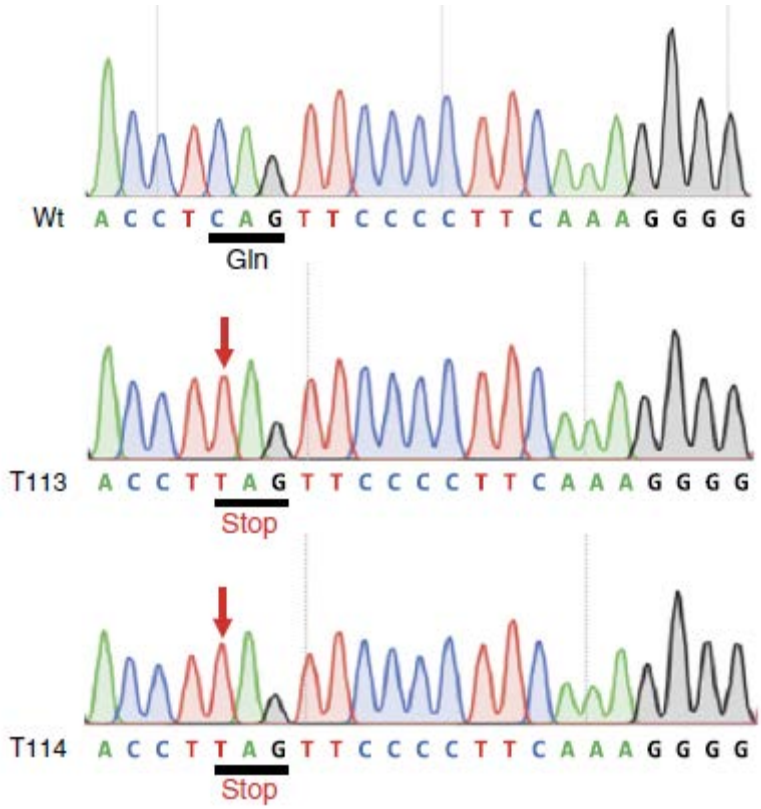


Figure 6. Dystrophin-deficient mutant mice generated by cytidine-deaminase-mediated base editing. Histological analysis of *tibialis anterior* (TA) muscles from wild-type and D108 mice. Laminin (control) and dystrophin are shown in red and green, respectively. All muscles were dissected from 4-week-old wild-type or *Dmd* mutant mice (D108) and frozen-in-liquid, nitrogen-cooled isopentane. Scale bars, 50um.

a

		Frequency (%)	
Wt	GTGGCACCATCTGGACCTCAGTTCCCCTTC--AAAGGGG		
T110	GTGGCACCATCTGGACCTTAGTTCCCCTTC--AAAGGGG	51	(Q68Stop)
	GTGGCACCATCTGGACCTGAGTTCCCCTTC--AAAGGGG	47	(Q68E)
T111	GTGGCACCATCTGGACCTTAGTTCCCCTTC--AAAGGGG	51	(Q68Stop)
	GTGGCACCATCTGGACCTGAGTTCCCCTTC--AAAGGGG	48	(Q68E)
T112	GTGGCACCATCTGGACCTCA----------GGG	49	(–14 bp)
	GTGGCACCATCTGGACCTAAGTTCCCCTTC--AAAGGGG	46	(Q68K)
	GTGGC----------TTC--AAAGGGG	5	(–22 bp)
T113	GTGGCACCATCTGGACCTTAGTTCCCCTTC--AAAGGGG	100	(Q68Stop)
T114	GTGGCACCATCTGGACCTTAGTTCCCCTTC--AAAGGGG	99	(Q68Stop)
T117	GTGGCACCATCTGGACCTGAGTTCCCCTTC--AAAGGGG	99	(Q68E)
T118	GTGGCACCATCTGGACCTCAGTTCCCCTTCAGAAAGGGG	41	(+2 bp)
	G----------GT----------GG	26	(–31 bp)

b



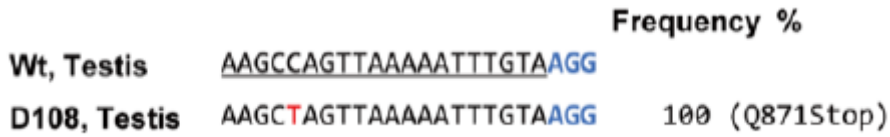
c

Figure 7. Generation of an albinism mouse model by cytidine-deaminase-mediated base editing. (a) Alignments of mutant sequences from newborn pups that developed after electroporation of the BE3 RNP into embryos. The target sequence is underlined. The PAM site and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles. Wt, wild-type. (b) Sanger sequencing chromatograms of DNA from wild-type and *Tyr* mutant mice. The red arrow indicates the substituted nucleotide. The relevant codon identities at the target site are shown under the DNA sequence. (c) *Tyr* mutant newborn pups that developed after electroporation of the BE3 RNP exhibited an albino phenotype in their eyes (back arrows, T113 and T114).

b. Germline transmission

I also confirmed germline transmission of the mutant allele to F1 offspring. I genotyped the testis of a *Dmd* homozygous mutant mouse (D108) by deep sequencing and only the mutant sequence was observed (Figure 8). Another homozygous mutant *Dmd* mutant mouse (D102) carrying two alleles (S870R, Q871Stop and Q871Stop) crossed to a wild-type C57BL/6J mouse. I obtained 10 pups and identified genotypes of the F1 offspring. All the mutant alleles observed in F0 mouse were transmitted to the germline (Figure 9). Because *Dmd* gene is inherited on X chromosome, male F1 pups showed 100 percent of mutant allele frequencies and female F1 pups showed about 50 percent of mutant allele frequencies.

a



b

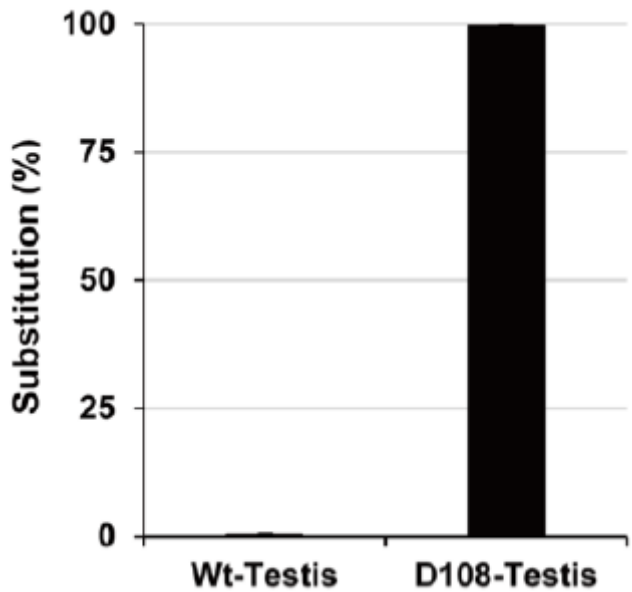


Figure 8. Germline transmission of the mutant *Dmd* mouse. (a) Alignments of mutant sequences from Wt (wild type) and D108 testis. The PAM site and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles. (b) Targeted deep sequencing analysis using gDNA of Wt (wild type) and D108 testis.

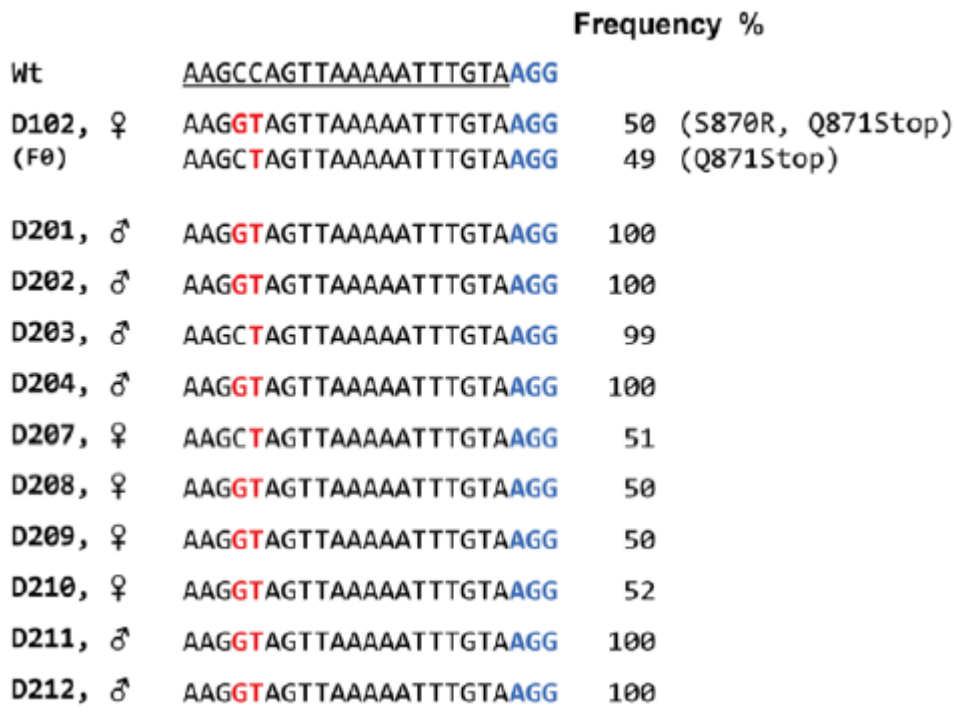


Figure 9. Mutant alleles of F1 *Dmd* mutant mice. Germline transmission of the mutant allele in pups (D201 to D204 and D207 to D212) obtained from the *Dmd* mutant mouse (D102). Deep sequencing was used to confirm the genotype. The PAM site and substitutions are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles.

c. Analysis of off-target effects

To assess BE3 off-target effects, I first identified, using Cas-OFFinder (Bae et al., 2014), potential off-target sites with up to 3-nucleotide mismatches in the mouse genome, and analyzed genomic DNA isolated from newborn mice via targeted deep sequencing. No off-target mutations were detectably induced at these sites (Figure 10, 11 and Table 5, 6). I next performed whole genome sequencing (WGS) to identify BE3 off-target mutations in the *Dmd* mutant mouse (D108) (Figure 12). Among 319,663 or 296,518 protospacer adjacent motif (PAM)-containing sites that differ from the on-target site by up to 7 or 5 mismatches, respectively, with 0 or up to 2 bulges, respectively, just a single site was identified as a potential off-target site. Targeted deep sequencing with genomic DNA isolated from six different organs, however, showed that this site was a false positive. At the BE3 on-target site, no wild-type allele was present in DNA samples obtained from various organs, confirming that the D108 mutant mouse was not a mosaic. Taken together, these results show that BE3 can be highly specific in vivo. Note, however, that WGS with a typical sequencing depth of 30–40 × cannot reveal rare off-target mutations in a population of cells or organisms (Kim et al., 2015).

Table 5. Potential off-target sites of *Dmd* in the mouse genome.

No.	Gene		Sequence	Chromosome	Position	Direction
Dmd-On	*Dmd*	Exon	AAGCCAGTTAAAAATTTGTAAGG	chrX	83781748	+
Dmd-OT1	*Atp11c*	Intron	AAGCCAGTTAgAAgTTTGTAAGG	chrX	60354718	+
Dmd-OT2	Intergenic region	-	AAGCaAGTgtAAAATTTGTATGG	chr3	123915964	-
Dmd-OT3	*Ptprd*	Intron	AAtaCAGTTAAtAATTTGTAAGG	chr4	78045958	+
Dmd-OT4	*Cfh*	Intron	AAGCaAGTaAAAcATTTGTATGG	chr1	139933967	+
Dmd-OT5	*Sos1*	Intron	ttGCCAGTTtAAAATTTGTAAGG	chr17	80462531	-
Dmd-OT6	Intergenic region	-	AAGCaAtTgAAAAATTTGTATGG	chr17	81921018	+
Dmd-OT7	*Grid2*	Intron	AAGCCAGaaAcAAATTTGTAGGG	chr6	64297703	-
Dmd-OT8	Intergenic region	-	AAGCtAGTggAAAATTTGTACGG	chr6	147759582	-
Dmd-OT9	Intergenic region	-	AActCAaTTAAAAATTTGTATGG	chr14	26539529	-
Dmd-OT10	*Spag9*	Intron	AAGCCAGagAAtAATTTGTAGGG	chr11	94052933	+
Dmd-OT11	*Tle1*	Intron	AAaCCAGTTAAAtATTTcTAAGG	chr4	72159600	+
Dmd-OT12	*RNf114b*	Intron	AAGCCAtTTAAAAATTTGagTGG	chr13	47235657	-
Dmd-OT13	*Tnfrsf21*	Intron	AAcCCAGTTAgAAATTTtTATGG	chr17	43052831	+
Dmd-OT14	*Aak1*	Intron	AAGaCAGaTAAAAATTTGgAGGG	chr6	86907351	-
Dmd-OT15	*Gpm6b*	3' UTR	AgGCCAGaTAAAAATTTGaAGGG	chrX	166385994	+
WGS-*Dmd*-OT	Intergenic region	-	cAGCCtGTTtAtccTTTGTgTAG	chr12	68958733	+

Table 6. Potential off-target sites of *Tyr* in the mouse genome.

Tyr-On	*Tyr*	Exon	ACCTCAGTTCCCCTTCAAAGGGG	chr7	87493130	-
Tyr-OT1	*Zmat4*	Intron	cCCTCAGTTCCaCTTCAgAGAGG	chr8	23975596	-
Tyr-OT2	Intergenic region	-	ACCTCAcTTgCCCTTCtAAGTGG	chr8	102059942	+
Tyr-OT3	*Il2*	Intron	ACCTCAGTcCCCCTTtAcAGAGG	chr3	37123254	+
Tyr-OT4	Intergenic region	-	ACCTCAGTTCCCCTaCActGGGG	chr3	9150768	-
Tyr-OT5	Intergenic region	-	ACCTCAGTTCCCCTaCActGGGG	chr3	81773804	-
Tyr-OT6	Intergenic region	-	tCCTCAGTTCCCCTTCActGGGG	chr7	11778250	+
Tyr-OT7	Intergenic region	-	cCCTCAGTTCCCCTaCAcAGAGG	chr4	47766122	-
Tyr-OT8	Intergenic region	-	ACCTCAGTTtCCCTTCcAgGAGG	chr4	54553337	-
Tyr-OT9	2810429I04Rik	Intron	cCCTCAGTTCCCCTTCActGGGG	chr13	3491261	+
Tyr-OT10	Intergenic region	-	cCCTCAGTTCCCCTaCAcAGGGG	chr13	74957186	+
Tyr-OT11	MGP_C57BL6NJ_G0001126	-	cCCTCAGTTCCCCTaCAcAGGGG	chr2	79031825	+
Tyr-OT12	*Rai14*	Intron	ACCTCAGTTtCCCcTCAAAaTGG	chr15	10596653	+
Tyr-OT13	Intergenic region	-	ACCTCAGTTgtCCTTCAAAcAGG	chr6	107384348	-
Tyr-OT14	D6Ertd474e	Intron	cCCTCAGTTCCCCTaCAAtGGGG	chr6	143247456	-
Tyr-OT15	Intergenic region	-	AaCTCtGTTCCCCTTCtAAGTGG	chr18	23151245	+
Tyr-OT16	Intergenic region	-	AtCTCAGTTtCCCTTCAcAGGGG	chr11	71708976	-
Tyr-OT17	*Foxk2*	Intron	ACCTtAGTTCCCtTTCAAAcTGG	chr11	121271649	+

OT: Off-target, WGS: Whole genome sequencing

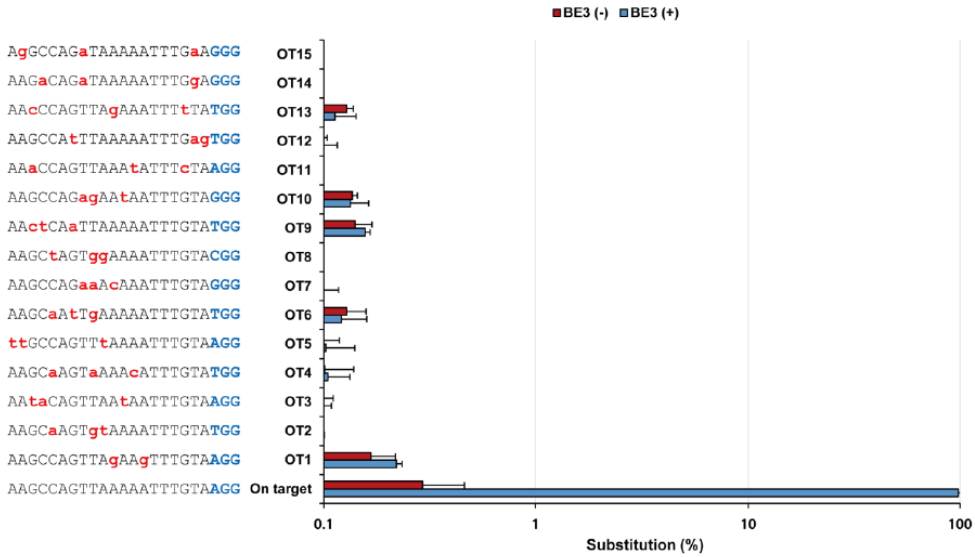


Figure 10. No off-target mutations were detectably induced at potential off-target sites in *Dmd* mutant mice. Targeted deep sequencing was used to measure base editing efficiencies at potential off-target sites in *Dmd* mutant mice (n=3). Mismatched nucleotides and PAM sequences are shown in red and in blue, respectively.

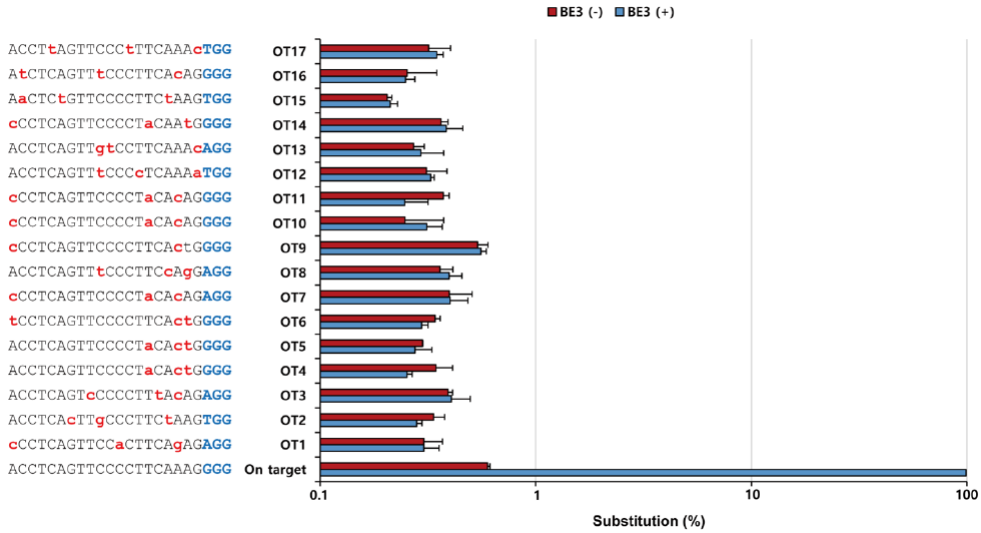


Figure 11. No off-target mutations were detectably induced at potential off-target sites in *Tyr* mutant mice. Targeted deep sequencing was used to measure base editing efficiencies at potential off-target sites in *Tyr* mutant mice (n=2). Mismatched nucleotides and PAM sequences are shown in red and in blue, respectively.

a

Genome wide sequencing analysis

	D108	Wt	Uniquely assigned to D108
total variants (SNPs + small indels)	118000	123963	
SNPs	36959	37648	12126
SNPs after excluding dbSNPs	21366	21734	8722
C -> T SNPs	3127	3160	1114
C -> A SNPs	1190	1208	448
C -> G SNPs	1303	1325	693
G -> A SNPs	3205	3183	1194
G -> T SNPs	1249	1307	452
G -> C SNPs	1343	1395	696
Possible positions matching SNPs after excluding dbSNPs			
On-target	1	0	1
Up to 2-bp bulge + 5-bp mismatch with NRG PAM (296518 sites)	10	9	0
Up to 7-bp mismatches with NRG PAM (319663 sites)	21	24	1

Wt: Wild type

b

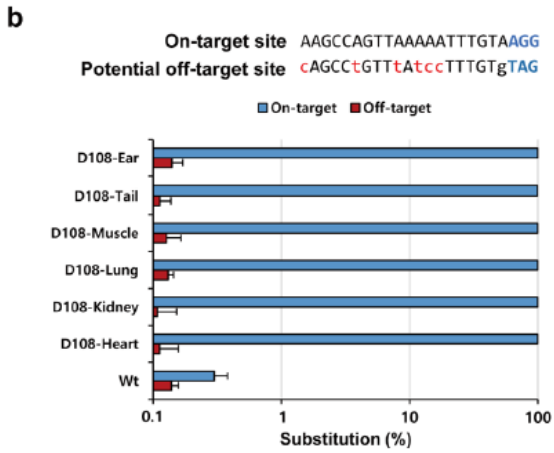


Figure 12. Whole genome sequencing of *Dmd* mutant and wild-type (Wt) mice. (a) Summary of whole genome sequencing analysis. A *Dmd* mutant mouse (D108) and a wild type mouse (Wt) were separately sequenced using Illumina Hiseq X10. Total variants (SNPs + small indels) were identified using ISACC. After filtering out naturally occurring variants in the SNP database (dbSNP) and excluding SNPs also found in the wild-type genome, I obtained 8,722 SNPs in the D108 genome. I further excluded base substitutions other than C to T or A or G conversions. I next compared the DNA sequences at the remaining SNP sites with the on-target sequence. Among 319,663 or 296,518 proto-spacer adjacent motif (PAM)-containing sites that differ from the on-target site by up to 7 or 5 mismatches, respectively, with 0 or up to 2 bulges, respectively, just a single site was identified as a potential off-target site. (b) The off-target candidate site was invalidated using targeted deep sequencing of genomic DNA isolated from various organ (heart, kidney, lung, muscle, tail and ear). Mismatched nucleotides and PAM sequences are shown in red and in blue, respectively.

3. Extended sgRNAs broaden the base editing window in HEK293T cells using ABEs

First, I transfected HEK293T cells with plasmids encoding ABE7.10 and a conventional GX19 sgRNA ("G" or "g" indicates a matched or mismatched guanine, respectively, at the 5′ end, whereas "X19" indicates a 19-mer RNA sequence complementary with the protospacer DNA sequence) targeted to the HEK2 site used in previous studies (Gaudelli et al., 2017; Kim et al., 2017a) or eight extended sgRNAs with additional nucleotides at the 5′ terminus and performed targeted deep sequencing to measure substitution frequencies (Figure 13 and 14a, b). As expected, the GX19 sgRNA was highly efficient with an editing frequency of 73% at position 16 but was only slightly active (<0.6%) at positions 18 and 19 upstream of the canonical base editing window. In contrast, extended sgRNAs with a few additional nucleotides at the 5′ terminus (GX20, gX21, gX22, and gX23) showed base editing activity at positions 18 and 19 with frequencies that ranged from 1.3% to 10%. sgRNAs with >23 base spacers were inactive at these positions and less active at positions even in the optimal editing window. Compared to the GX19 sgRNA, the GX20 sgRNA showed a 15-fold increase in the editing efficiency at positions outside of the canonical window (Figure 14a). Likewise, gX21, ggX20, gX22, and gX23 sgRNAs, targeted to four additional endogenous sites containing adenines at various positions, expanded the editing window, compared to gX19 and GX19 sgRNAs (Figure 14c-f). Taken together, our results show that sgRNAs with a few extra nucleotides, targeted to five endogenous sites (Table 9) in the human genome,

induced A-to-G conversions at positions upstream of the current editing window, demonstrating that it is feasible to expand the window of adenine base editing in human cells (Figure 15 and Table 10). These extended sgRNAs broaden the editing window by 25% or 50%, from positions 14–17 (four nucleotides) to positions 14–18 or 19 (five or six nucleotides). Extended sgRNAs combined with engineered S. pyogenes Cas9 variants, such as xCas9, which recognizes the NG PAM (Hu et al., 2018), may further expand the targetable range of base editing.

Table 7. Sequences of oligonucleotides.

	Forward (5' to 3')	Reverse (5' to 3')
HEK2_gX30	caccgAAGGAAACTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAGTTTCCTTc
HEK2_gX27	caccgGAAACTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAGTTTCc
HEK2_gX25	caccgAACTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAGTTc
HEK2_gX24	caccgACTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAGTc
HEK2_gX23	caccgCTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAGc
HEK2_gX22	caccgTGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCAc
HEK2_gX21	caccgGGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCCc
HEK2_gX20	caccgGAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTCc
HEK2_gX19	caccgAACACAAAGCATAGACTGC	aaacGCAGTCTATGCTTTGTGTTc
Site 18_gX30	caccgACACACAGACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTCTGTGTGTc
Site 18_gX27	caccgCACAGACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTCTGTGc
Site 18_gX25	caccgCAGACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTCTGc
Site 18_gX24	caccgAGACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTCTc
Site 18_gX23	caccgGACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTCc
Site 18_gX22	caccgACACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGTc
Site 18_gX21	caccgCACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTGc
Site 18_gX20	caccgACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTc
Site 18_gX19	caccgCACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGc
Site 18_ggX20	caccggACACACACACTTAGAATCTG	aaacCAGATTCTAAGTGTGTGTGTcc
Site 19_gX30	caccgCACACAGACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGTCTGTGTGc
Site 19_gX27	caccgACAGACACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGTCTGTc
Site 19_gX25	caccgAGACACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGTCTc
Site 19_gX24	caccgGACACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGTCc
Site 19_gX23	caccgACACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGTc
Site 19_gX22	caccgCACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTGc
Site 19_gX21	caccgACACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGTc
Site 19_gX20	caccgCACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGc
Site 19_gX19	caccgACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTc
Site 19_ggX20	caccggCACACACACACTTAGAATCTGT	aaacACAGATTCTAAGTGTGTGTGcc
HBB-E2_gX30	caccgAGGCTGCCTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAGGCAGCCTc
HBB-E2_gX27	caccgCTGCCTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAGGCAGc
HBB-E2_gX25	caccgGCCTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAGGCc
HBB-E2_gX24	caccgCCTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAGGc
HBB-E2_gX23	caccgCTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAGc
HBB-E2_gX22	caccgTATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATAc
HBB-E2_gX21	caccgATCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGATc
HBB-E2_gX20	caccgTCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGAc
HBB-E2_gX19	caccgCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGc
HBB-E2_ggX20	caccggTCAGAAAGTGGTGGCTGGTG	aaacCACCAGCCACCACTTTCTGAcc
HBB-E3_gX30	caccgTAATATCCCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGGGATATTAc
HBB-E3_gX27	caccgTATCCCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGGGATAc
HBB-E3_gX25	caccgTCCCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGGGAc
HBB-E3_gX24	caccgCCCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGGGc
HBB-E3_gX23	caccgCCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGGc
HBB-E3_gX22	caccgCCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGGc
HBB-E3_gX21	caccgCCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGGc
HBB-E3_gX20	caccgCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGc
HBB-E3_gX19	caccgAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTc
HBB-E3_ggX20	caccggCAGTTTAGTAGTTGGACTTA	aaacTAAGTCCAACTACTAAACTGcc
Dmd-gX20	caccg AACTAGCTTTTAATTGCTGT	aaacACAGCAATTAAAAGCTAGTTc
Tyr-gX19_F	GAAATTAATACGACTCACTATAGCATAACAGAGACTCTTACAGTTTTAGAGCTAGAAATAGCAAG	
Tyr-GX20_F	GAAATTAATACGACTCACTATAGCCATAACAGAGACTCTTACAGTTTTAGAGCTAGAAATAGCAAG	
Tyr-GX21_F	GAAATTAATACGACTCACTATAGGCCATAACAGAGACTCTTACAGTTTTAGAGCTAGAAATAGCAAG	
Universal_R	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTA AAAC	

Table 8. List of primers used for targeted deep sequencing.

	1st PCR		2nd PCR	
No.	**Forward (5' to 3')**	**Reverse (5' to 3')**	**Forward (5' to 3')**	**Reverse (5' to 3')**
HEK2	AACCAGT GTCAGGG AGCTGT	ATCCACA GCAACAC CCTCTC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAGGACGTCTGCCCAATATGT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTAGCCCCATCTGTCAAACTGT
Site18	CCCACGG TCACTTCT TGTTT	GGGGTAA GGACAGA CGTTCA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCCTTGTAAGTGTTGCTGTCC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTCTCGCCTGCAGAAAGGTAT
Site19	CCCACGG TCACTTCT TGTTT	GGGGTAA GGACAGA CGTTCA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCCTTGTAAGTGTTGCTGTCC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTCTCGCCTGCAGAAAGGTAT
HBB-E2	GTTTGCA GCCTCAC CTTCTT	ACAATCC AGCTACC ATTCTGC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTATCTTCCTCCCACAGCTCC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCAAGGCCCTTCATAATATCCCC
HBB-E3	GTTTGCA GCCTCAC CTTCTT	ACAATCC AGCTACC ATTCTGC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTATGCACTGACCTCCCACATT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCCCTGGCCCACAAGTATCA
Tyr-On	TGAATGA CCTTCTTC AGGACA	TGTTTTTA ACTGCCC CCAAA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGAACAATGGCTGCGAAGG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTGGTTACCCACATTGCATTCC
Tyr-OT1	TTACTCGC ATGCAAC CATGT	GGCCTGG TCATCATC ACTTC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTGTGGAATTCAGGTCTGCAC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTGTGCTCACAGCACCAGAGAA
Tyr-OT2	CATGGCG TCTCTTCA TAGCA	GGATTCC CTGGAAC TGGAAT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCCCAACAAACTGTGGGAAAT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTAGAGCATCGAGCCATTTTTC
Tyr-OT3	TTGTTGG AAAGAGC AAAGCA	TGTTTAG GGATGAC TGATTGG A	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTGTTGGAAAGAGCAAAGCA	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTGAAGGTTACCCGCTACATTG
Tyr-OT4	CTCTGCC AGGTGGC TTTTAC	CTTGCCTC AAATGCA CTGAA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCTGCCAGGTGGCTTTTACAT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTCTGCCTTGGTTTCTGTGTG
Tyr-OT5	TCTTGATC CTGCCAAT GATG	CGGAGGA CCCTGGA TTTTAT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTACACCACATCCATCCACCTT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTGTCTTTCTCCCTGGTCTCC
Tyr-OT6	GGAAATC CTGGAGA GCATCA	CATCTATG GCGATTG GGAGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGGAAATGCTCCTGCCACATA	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTGTTTTGCTGGGTTGACACCT
Tyr-OT7	AGGCACA AGGAGTC TGCATT	AGACAGG GGAGGTG GGTAGT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCGGAGATGTGAGACCAGAGC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTGGGTAGTTTAAGGCCCAGA
Tyr-OT8	TCCTCAG GCTCAGA ACCAGT	TCAGCCA CAGGAGG AACTTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTACCAGGCAAGCTCATTCATC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCGTGCTTCTCTACTTCCCAAG
Tyr-OT9	AGCATTG GCAGCAG GAATAA	ATGCCTG GCTCCTT GTAATG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTTCTGTTTTCCAGTGTTTTAGTCA	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTTGCTGCATTCAGATCCACAT
Dmd-On	CTGCCAAT TGCTGAG TGAGA	CTGACAA AATCAAG TAGCCAT GT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCTGCCAATTGCTGAGTGAGA	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCAATTTGCCA AGAAGCACCT
Dmd-OT1	GCGAGGA AGAGGGA AAGAAT	CATTGCAT GGTCAGA GCAAC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCGAGGAAGAGGGAAAGAAT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTAACAGCCAAGGTCAGCAAAG
Dmd-OT2	GAAGAAC GTGGCTG AAGGAG	AGTGGAG TGCACAG CAGATG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCAAAGTGCCCCTGGTATGAC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTACCTCTTCTTCGCCCCTTTA
Dmd-OT3	ATGGACC CAACACT GAGGAG	ATATTCTG CTCCACC GCAAC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGGCTGCCTCTTTGGCTATC	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCAAACTGTCCCACACTCCAA
Dmd-OT4	GCATGGC AGAAGTG TGCTAA	ACATAGC CGAGTGG AGCCTA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTGTTGTTCAGGAAGCAAACACT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCATAGCCGAGTGGAGCCTAT
Dmd-OT5	GAGGCTA AGGAGCC TCAGGT	CGGGATC TTG CTCTGAA CTC	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGAAACCCAGAAATGGATGGA	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCGGGATCTTGCTCTGAACTC
Dmd-OT6	TCCTCCTG ATTCCCAT TCTG	GATGCATC TCTGGCA ATTCA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTAGCTGCCGGTAAACAGTGAG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTGTGGGAAACGAAGGGAAAGT
Dmd-OT7	GGAGGAG GGGGAAA TACAAG	GAACATG GGGGAGC AGATAA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCAAGAAGATTGTTCGGGTTG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCCTATGTACATGCCCCACAC
Dmd-	GTGTCCA	CTTCCAA	ACACTCTTTCCCTACACGACGCTCTTCCG	GTGACTGGAGTTCAGACGTGTGCTCTTC

OT8	AGCTCCC ATGTCT	AAC CTCAGCC AAA	ATCTGGCTCAAAAGAAAAGCAGCA	CGATCTTCCACCGCTATGGGTTTTTA
Dmd-OT9	CATATGCC ATCAGTC GTGCT	GGCTGTAT TG GGCATCT GTT	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGTGGGCTCACGGAATGTAAG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTCCCGCACCCATGTTATAGAT
Dmd-OT10	GGTCATTG GTTTCCA GAGCA	CAATTTGC CA AGAAGCA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTTGCCAATTGCTGAGTGAGAG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTGCCAAGAAGCACCTTTTGAT
Dmd-OT11	GTAACGC GATATGCC TGGTT	TGGCTCAT AC CATCATCC AA	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTCCAAGCCAGCTTTAGTCAGG	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT GCAGAGGCCACTAGGGTATTT
Dmd-OT12	GGCAGCT TCTGAGT GAGCTT	GGTAGAG CACTCGC CTAACG	ACACTCTTTCCCTACACGACGCTCTTCCG ATCTGCTGAGGGTCTTCGTGTTCT	GTGACTGGAGTTCAGACGTGTGCTCTTC CGATCT GGTAGAGCACTCGCCTAACG

OT: Off-target

Table 9. ABE target sites in HEK293T cells.

No.	Sequence	Chromosome	Position	Direction	Reference
HEK2	GAACACAAAGCATAGACTGCGGG	chr5	87944780 - 87944799	+	3, 8
Site18	ACACACACACTTAGAATCTGTGG	chr1	184974900 - 184974919	+	3
Site19	CACACACACTTAGAATCTGTGGG	chr1	184974901 - 184974920	+	3
HBB-E2	TCAGAAAGTGGTGGCTGGTGTGG	chr11	5225630 - 5225649	-	
HBB-E3	CAGTTTAGTAGTTGGACTTAGGG	chr11	5225531 - 5225550	+	

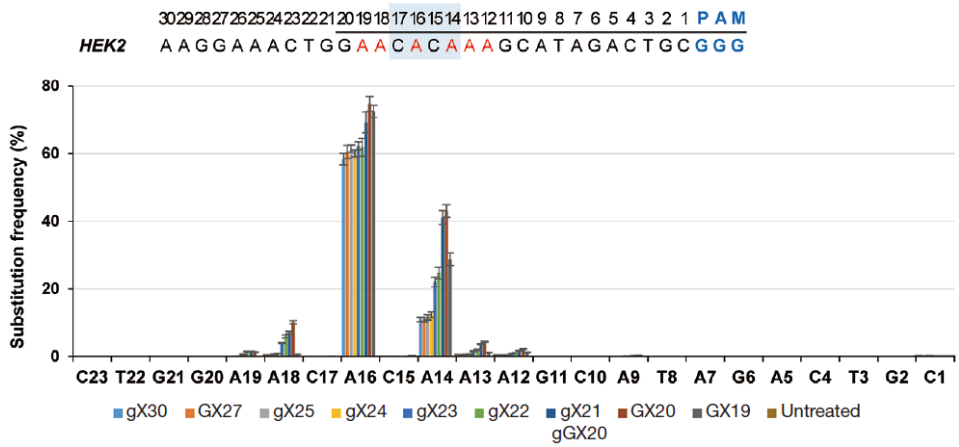
(By Kayeong Lim in Seoul National University)

Table 10. Relative editing efficiencies at each base position at 5 endogenous sites in the human genome.

	Position of target A												
sgRNA	A23	A22	A21	A20	A19	A18	A17	A16	A15	A14	A13	A12	A11
gX30	0.000	0.000	0.000	0.000	0.000	0.000	*** 0.561	0.981	1.000	0.570	0.077	0.028	0.000
gX27	0.000	0.000	0.000	0.000	0.000	0.000	*** 0.560	0.987	1.000	0.581	0.080	0.027	0.000
gX25	0.000	0.000	0.000	0.000	0.000	0.007	*** 0.571	0.980	1.000	0.585	0.079	0.028	0.000
gX24	0.000	0.000	0.000	0.000	0.000	0.037	*** 0.561	0.970	1.000	0.536	0.077	0.032	0.000
gX23	0.000	0.000	0.000	0.000	0.000	** 0.076	*** 0.565	0.947	0.970	0.649	0.092	0.032	0.000
gX22	0.000	0.000	0.000	0.006	* 0.165	*** 0.109	*** 0.584	0.915	0.922	0.692	0.142	0.040	0.000
gX21	0.000	0.000	0.000	* 0.027	*** 0.063	*** 0.187	** 0.540	0.934	0.940	0.774	0.358	0.065	0.000
ggX20	0.000	0.000	0.000	0.000	** 0.024	*** 0.101	*** 0.557	0.886	0.906	0.816	0.405	0.119	0.000
gX20	0.000	0.000	0.000	0.000	** 0.017	*** 0.101	*** 0.662	0.929	0.952	0.827	0.407	0.105	0.000
gX19	0.000	0.000	0.000	0.000	0.000	0.015	0.406	0.965	0.955	0.718	0.136	0.055	0.000

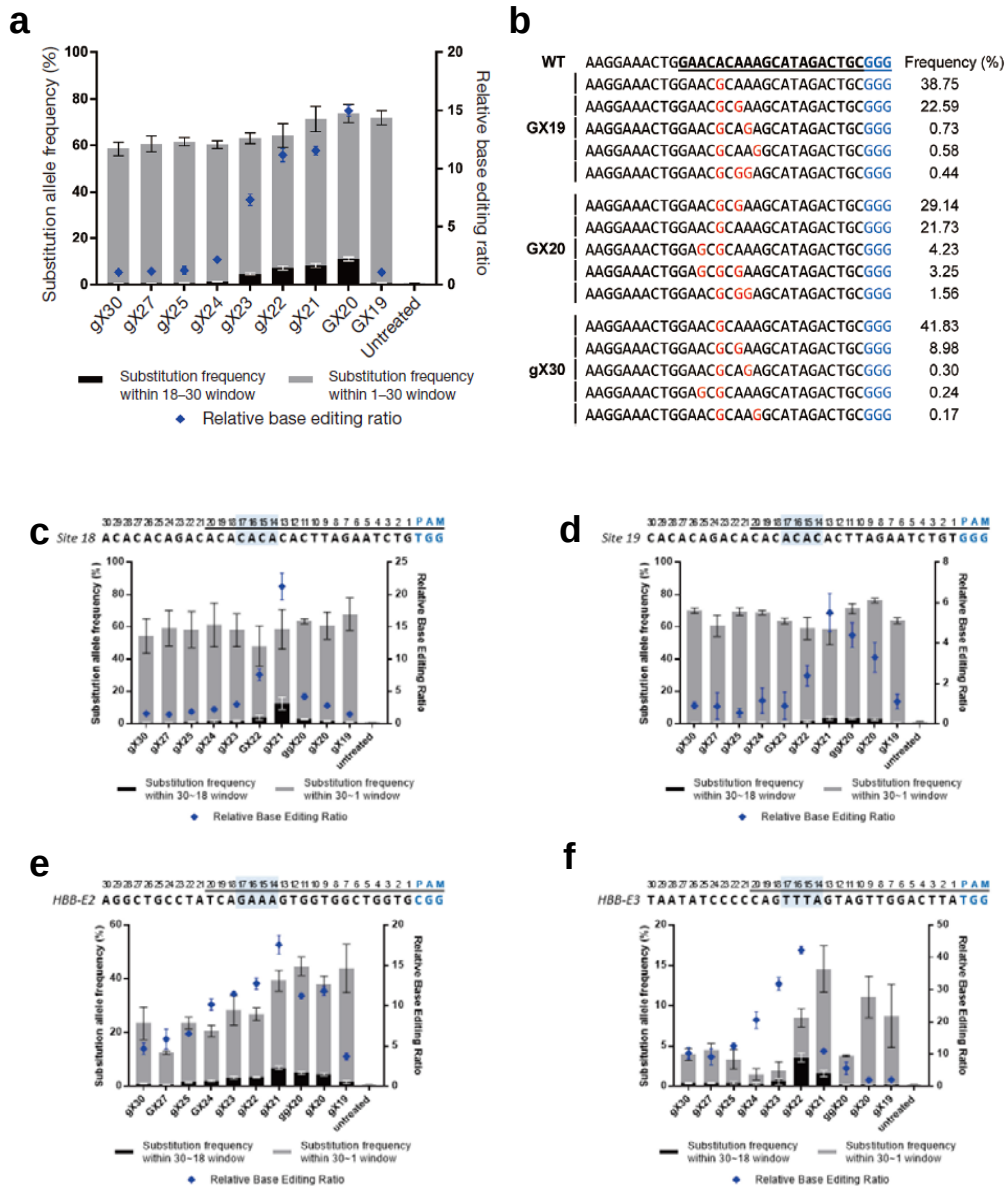
Two-tailed Student's t-test, *P < 0.05, **P < 0.01, ***P < 0.001.

(By Kayeong Lim in Seoul National University)



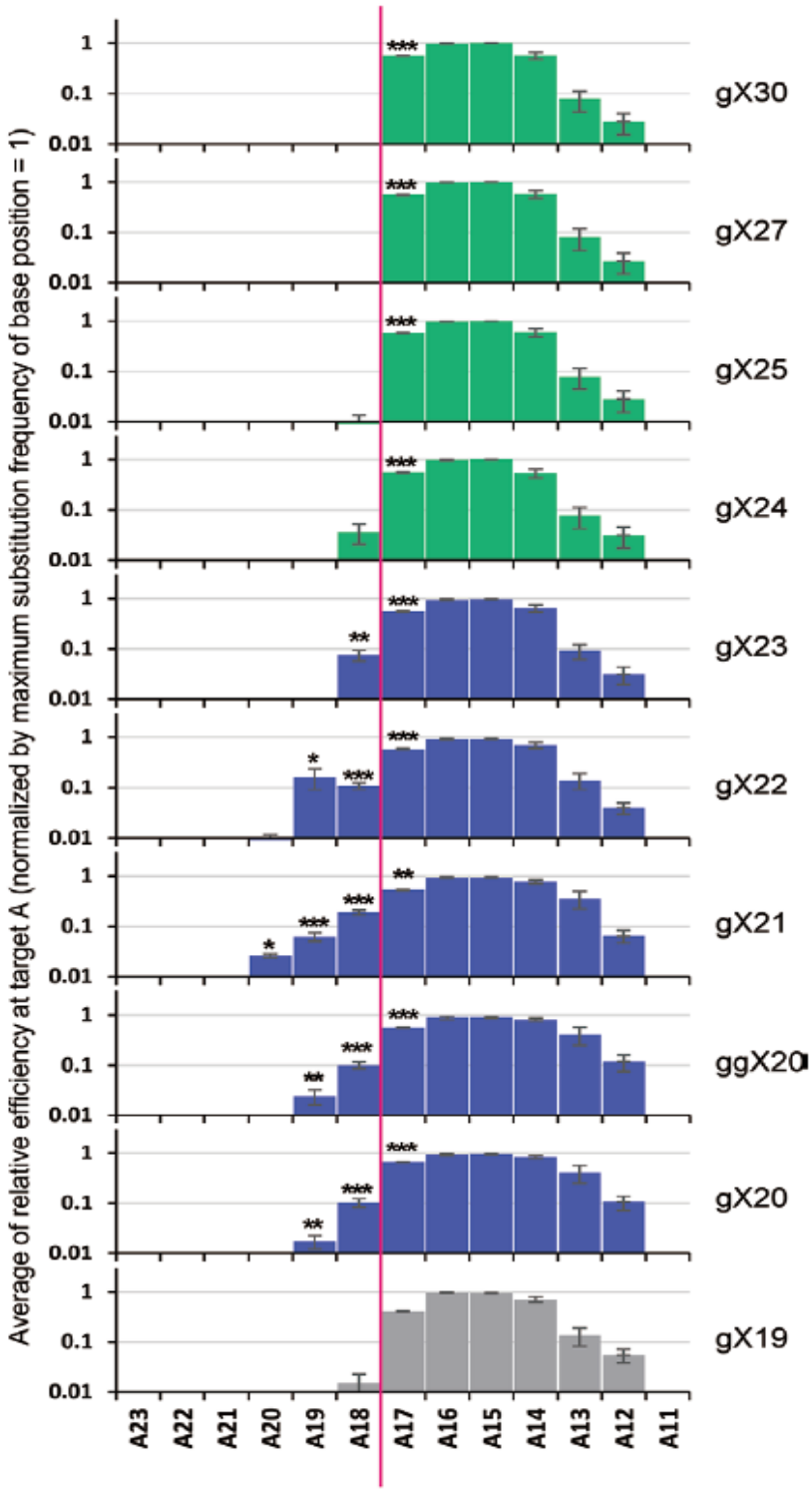
(By Kayeong Lim in Seoul National University)

Figure 13. Substitution frequencies of ABE targeting *HEK2* in HEK293T cells. Extended sgRNAs with up to a 30-nucleotide spacer were tested in parallel with the conventional GX19 sgRNA at the *HEK2* target site in HEK293T cells. Substitution frequencies were measured at each base position by targeted deep sequencing.



(By Kayeong Lim in Seoul National University)

Figure 14. Activities of ABEs using extended sgRNAs in HEK293T cells. Base editing efficiencies of ABEs with extended sgRNAs at (a) *HEK2*, (c) Site 18, (d) Site 19, (e) the *HBB-E2* site, and (f) the *HBB-E3* sites. (Left Y axis) Grey or black bars indicate the percentage of alleles that have a base-substitution mutation at the target adenine within two windows [positions 1-30 or 18-30 (upstream from PAM)], respectively. (Right Y axis) Blue dots indicate the relative base editing ratio, which was calculated by dividing the substitution allele frequency within the window of positions 18-30 by that within the window of positions 1-30. The PAM is shown in blue. Data are presented as mean ± s.e.m. (n = 3 biologically independent samples). (b) Sequences of the most frequent mutant alleles at the *HEK2* site. Substituted nucleotides are shown in red. The PAM is shown in blue.



(By Kayeong Lim in Seoul National University)

Figure 15. Relative substitution efficiencies at each sgRNA position. Relative efficiencies at each base position (the substitution frequency at each position normalized to the maximum substitution frequency) were averaged over five tested sites (*HEK2*, Site 18, Site 19, *HBB-E2, HBB-E3*). The red line indicates the boundary of the conventional activity window. Data are presented as mean ± s.e.m. (n = 3 biologically independent samples). Two-tailed Student's t-test, *P < 0.05, ** P < 0.01, *** P < 0.001.

4. Generating mouse models using ABEs targeting *Tyr* gene

a. Inducing Himalayan mutation by extended sgRNAs in blastocysts

I next sought to create the Himalayan mutation (Ch/Ch) in the *Tyr* gene in mice using ABEs. This mutant allele is caused by a naturally occurring A-to-G single-nucleotide variation in the mouse *Tyr* gene, resulting in a histidine-to-arginine amino acid change (H420R) in the tyrosinase protein, an essential enzyme for pigmentation in animals. The Himalayan mouse is characterized by a partial albinism in adults but is indistinguishable from albinos at birth (Kwon et al., 1989). The target adenine in the *Tyr* gene is located at position 18, one nucleotide upstream of the canonical base editing window (Figure 16), providing us with the right conditions for a proof-of-principle experiment in animals with extended sgRNAs. I delivered each of three sgRNAs (gX19, GX20, and GX21) with ABE7.10 mRNA into mouse embryos by microinjection and analyzed genomic DNA isolated from the resulting blastocysts. The three sgRNAs were almost equally efficient with average editing efficiencies of 78%, 95%, and 78%, respectively (Figure 17a). The Himalayan allele, however, was detected in just one out of eight embryos (13%) with a frequency of 10%, when the gX19 sgRNA was injected (Figure 17b and 18). Thus, the overall frequency of the Himalayan allele obtained with the gX19 sgRNA was 1.3%. Notably, the mutant allele was created in 5 of 16 (31%) and 8 of 12 (67%) embryos with frequencies of up to 47% and 94%, respectively, when GX20 and GX21 sgRNAs, respectively, were used. As a result, the overall frequencies of the

Himalayan allele obtained with these extended sgRNAs were 8.5% (GX20) and 20% (GX21) (Figure 19).

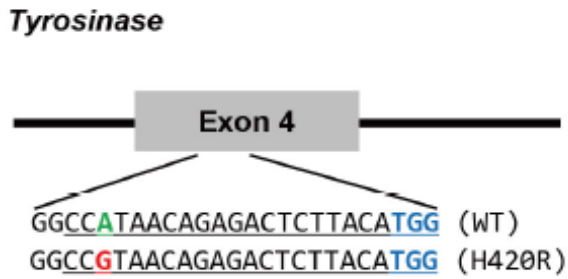
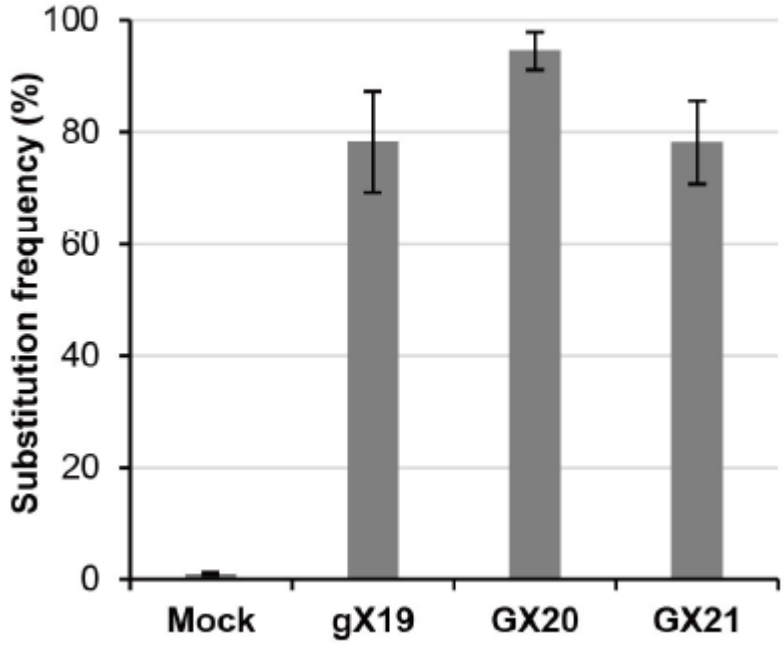


Figure 16. Scheme for the design of sgRNA to induce Himalayan mutation. The target sequence at the Tyrosinase (*Tyr*) locus. The PAM sequence is shown in blue. The sgRNA target sequence is underlined. The targeted adenine in the wild-type sequence and the expected change in the sequence are shown in green and red, respectively.

a


Substitution frequency (%)
100
80
60
40
20
0
Mock
gX19
GX20
GX21

b

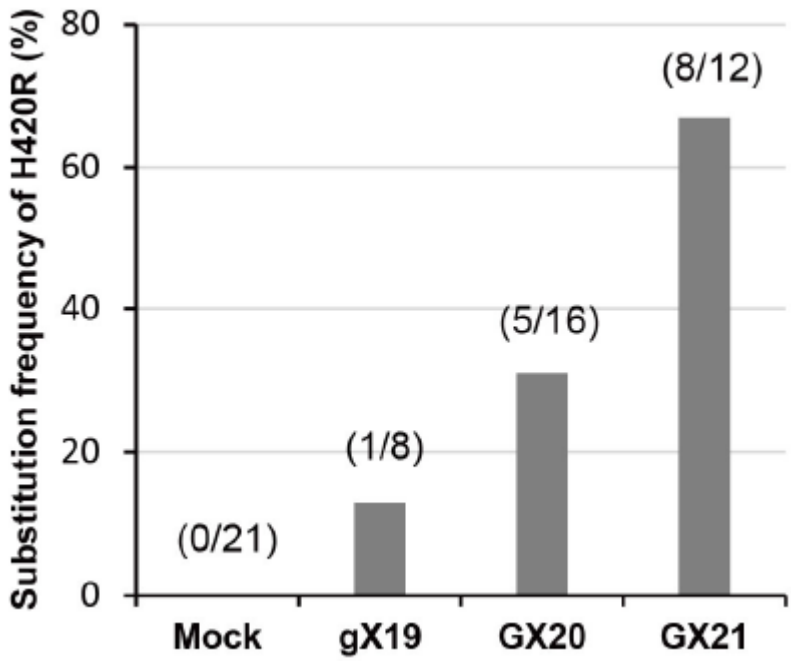

Substitution frequency of H420R (%)
80
60
40
20
0
(0/21)
(1/8)
(5/16)
(8/12)
Mock
gX19
GX20
GX21

Figure 17. Total substitution frequencies and Himalayan mutation frequencies in mouse embryos. (a) Point mutation efficiencies associated with sgRNAs of different lengths (gX19, GX20, and GX21). Data are presented as mean ± s.e.m. (*n* = 8, 16 and 12 blastocysts, respectively). (b) Frequencies of blastocysts that carry the H402R mutation (Himalayan allele).

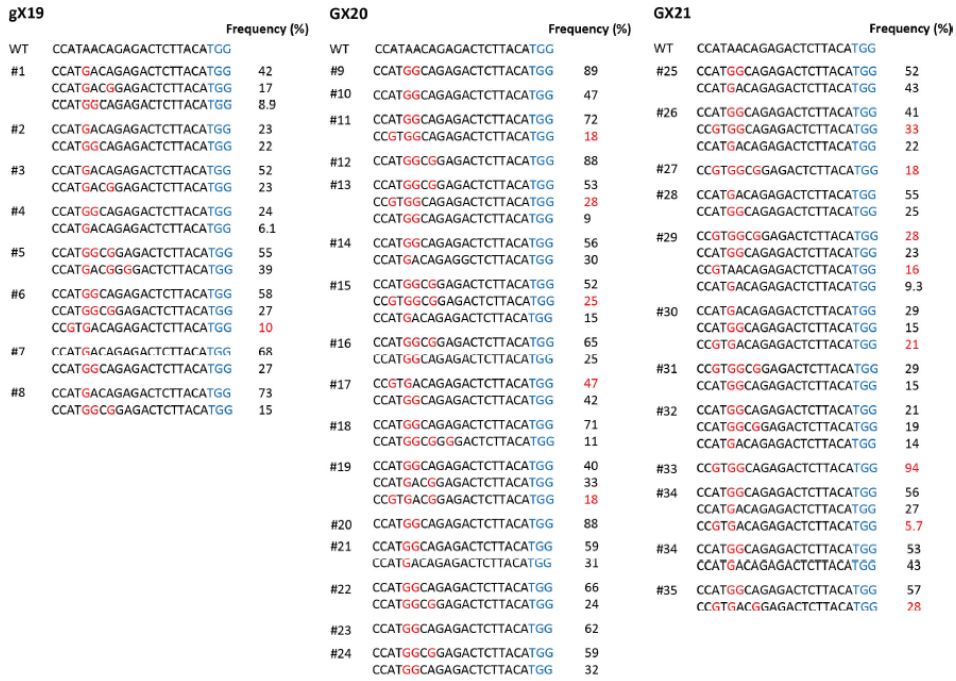


Figure 18. Adenine editing efficiencies in mouse embryos. Alignments of mutant sequences from blastocysts after microinjection of ABE mRNA and *Tyr*-targeting sgRNAs of different lengths. The PAM and substitutions are shown in blue and red, respectively. The numbers in the column on the right indicate the frequency of each allele. WT, wild-type. H402R mutant allele frequencies are shown in red.

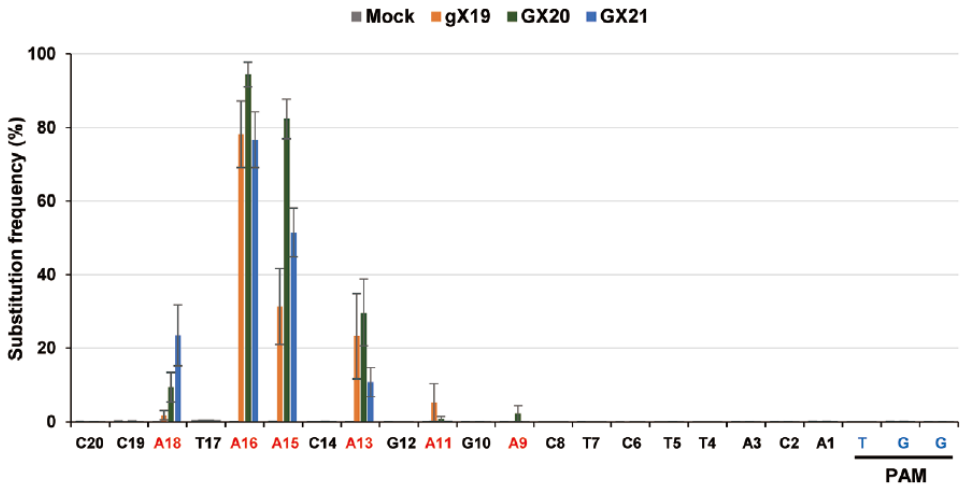


Figure 19. Substitution frequencies of ABE targeting *Tyr* gene in mouse embryos. Substitution frequencies at each position in the targeted locus. Results obtained with gX19, GX20, and GX21 sgRNAs are shown in yellow, green, and blue, respectively. Data are presented as mean ± s.e.m. (*n* = 8, 16, and 12 blastocysts, respectively)

b. Generating Himalayan mutant mouse models

I chose the GX21 sgRNA for further experiments and transplanted the resulting ABE-injected embryos into surrogate mothers to obtain several offspring carrying point mutations at the *Tyr* target site. Seven out of nine F0 mice (78%) harbored base substitutions, with frequencies that ranged from 12% to 100% (Figure 20). One F0 mouse showed an albino phenotype in the eye even at birth (Figure 21). Targeted amplicon sequencing revealed that this mouse was not genetically mosaic at the target site, carrying two Himalayan alleles. All of the other mutant mice were mosaic with at least three alleles or two non-Himalayan-mutant alleles. Altogether, I observed base editing in 43 out of 45 (96%) embryos or pups with frequencies that ranged from 18% to 100% (Figure 18 and Table 11).

Figure 20. Sequences of *Tyr* mutations in newborn pups. Alignments of mutant sequences from newborn pups. The altered nucleotides and the PAM site are shown in red and blue, respectively. The targeted position is shown in green.

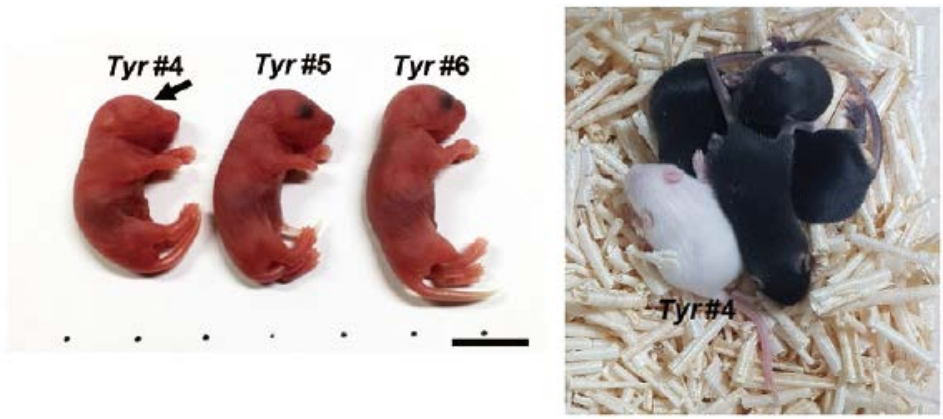


Figure 21. Phenotype of *Tyr* mutations in newborn pups. *Tyr* mutant F0 pups that developed after microinjection of the ABE mRNA; one pup exhibited an albino phenotype in its eyes (*Tyr* #4, black arrow) and coat color.

Table 11. Relative editing efficiencies at each base position at 5 endogenous sites in the human genome.

Target gene	No. of examined embroys	No. of two-cell-stage embryos (%)	No. of blastocysts (%)	No. of transferred embryos	No. of offspring (%)	Mutant ratio (%)		
						No. of mutants/total blastocysts or offspring (%)	No. of missense mutants/total blastocysts or offspring (%)	No. of Himalayan mutants/mutant blastocysts or offspring (%)
Buffer injection	41	39 (95)[a]	24 (62)[b]	NA	NA	0/21 (0)	0/21 (0)	0/21 (0)
Tyr (gX19)	15	13 (87)[a]	8 (62)[b]	NA	NA	8/8 (100)	8/8 (100)	1/8 (13)
Tyr (GX20)	24	21 (88)[a]	16 (76)[b]	NA	NA	16/16 (100)	16/16 (100)	5/16 (31)
Tyr (GX21)	18	16 (89)[a]	12 (75)[b]	46	9 (20)[c]	19/21 (90)	19/21 (90)	13/19 (68)

NA: not applicable
a: Calculated from the number of examined embryos
b: Calculated from the number of developed 2-cell stage embryos
c: Calculated from the number of transferred embryos

c. Germline transmission

I also verified germline transmission of mutant alleles to F1 offspring (Figure 22). I obtained 7 out of 9 mutant mice carrying A-to-G substitution induced by ABEs with from 87% to 100% efficiencies. *Tyr* #1, *Tyr* #2 and *Tyr* #3 female mutant mice crossed to a wild-type C57BL/6J male mouse. I obtained 10 pups and identified genotypes of the F1 offspring. I confirmed all the mutant alleles observed in F0 mouse were transmitted to the germline by targeted deep sequencing (Figure 22). Even the mutant alleles occupied the least portion as a result of a mosaicism are transmitted to next generation.

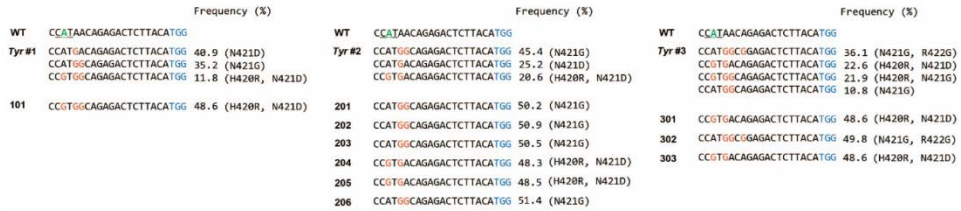


Figure 22. Germline transmission of *Tyr* mutant alleles. Germline transmission of mutant alleles to F1 pups (101, 201-206, and 301-303) from F0 *Tyr* mutant mice (*Tyr* #1, *Tyr* #2 and *Tyr* #3) was confirmed using targeted deep sequencing. The PAM site and altered nucleotides are shown in blue and red, respectively. The column on the right indicates frequencies of mutant alleles.

d. Analysis of off-target effects

To assess off-target effects, I first used Cas-OFFinder (Bae et al., 2014) to identify potential off-target sites with up to 3-nucleotide mismatches in the mouse genome (Table 12). No off-target mutations were detectably induced at these sites in three *Tyr* mutant mice (Figure 23). Furthermore, I performed whole genome sequencing (WGS) to assess off-target effects in the *Tyr* mutant mouse (*Tyr* #4) carrying two Himalayan alleles (Figure 24). I identified 207,848 candidate sites that differed from the on-target site by up to seven mismatches. Additionally, I further identified 1,030,669 candidate sites that differed from the on-target site by up to five mismatches with an RNA or DNA bulge. No off-target base editing was observed at these potential off-target sites, revealing the high specificity of the *Tyr*-targeted ABE.

Table 12. Potential off-target sites in the mouse genome.

No.	Gene		Sequence	Chromosome	Position	Direction
On	*Tyr*	-	CCATAACAGAGACTCTTACATGG	chr7	87438023	-
OT1	Intergenic region		ttATAACAGAGACTCTTACtGGG	chr8	127915201	-
OT2	*Rsrc1*	Intron	tCcaAACAGAGACTCTTACAAGG	chr3	67007324	-
OT3	Intergenic region		CgATtACAGAGACTCcTACAAGG	chr5	79532301	+
OT4	*Adgrb3*	intron	CCATAACtGAGACTCTgcCAAGG	chr1	25463979	-
OT5	Intergenic region		CaATAACAGAGgtTCTTACAAGG	chr1	113881388	-
OT6	Intergenic region		tCtTAACAGAGACTCTTAaAAGG	chrX	65801977	+
OT7	Intergenic region		aCATAgCAGgGACTCTTACACGG	Chr18	52601483	-
OT8	*Ikzf1*	Intron	CCAcAACAcAGACTCTTAaAAGG	Chr11	11758527	-
OT9	Intergenic region		CgATcACAGAaACTCTTACAAGG	chr11	42283167	+

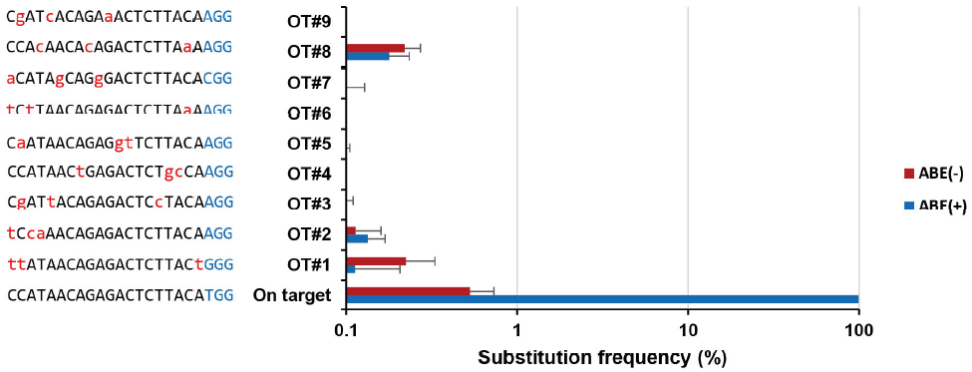


Figure 23. No off-target mutations at candidate sites in *Tyr* mutant mice. Potential off-target sites with up to 3 mismatches, relative to the wild-type sequence, were identified using Cas-OFFinder. Substitution frequencies at these potential off-target sites were measured using targeted deep sequencing. PAM sequences and mismatched nucleotides are shown in blue and red, respectively. Data are presented as mean ± s.e.m. (n = 9 animals).

	Tyr #4
Number of all unique SNVs after Strelka analysis	15627
Number of passed unique SNVs after Strelka analysis	2547
Number of SNVs at potential off-target sites (207,848) with up to 7 mismatches	0
Number of SNVs at potential off-target sites (1,030,669) with up to 5 mismatches and a DNA or RNA bulge	0

Figure 24. Whole genome sequencing to assess off-target effects in the *Tyr* mutant mouse. Genomes from the *Tyr* mutant mouse (*Tyr* #4) and a wild-type (WT) control mouse were sequenced using Illumina HiSeq X10. We identified unique single nucleotide variants (SNVs) in *Tyr* #4 by trimming out those in the WT control using the program 'Strelka' with the default 'eland' option. None of these SNVs other than the variations at the on-target site were found at potential off-target sites, identified using Cas-OFFinder, with up to 7 mismatches (207,848 sites) or with up to 5 mismatches and a DNA or RNA bulge (1,030,669 sites).

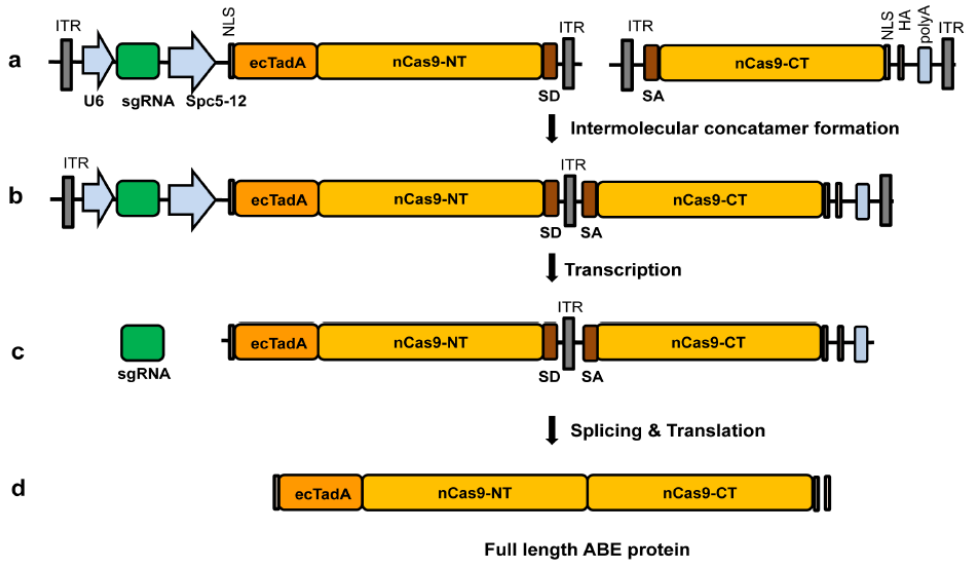
5. Therapeutic base editing with ABEs in a mouse model of Duchenne muscular dystrophy

a. Delivered using trans-splicing adeno-associated virus

I next set out to demonstrate therapeutic base editing in a mouse model of Duchenne muscular dystrophy with a premature stop codon (Kim et al., 2017b) in the *Dmd* gene, using ABE7.10 targeted to the *Dmd* nonsense mutation (Figure 25). Because the gene encoding ABE7.10, plus its sgRNA (6.1 kbp in size including promoters), cannot be packaged in a single adeno-associated virus (AAV) vector with a packaging limit of ~4.7 kbp, I employed a dual trans-splicing adeno-associated virus (tsAAV) vector system (Lai et al., 2005; Sun et al., 2000) to deliver the ABE gene in two parts into skeletal muscles. In a co-transduced cell, the two AAV vectors, each encoding the N-terminal or C-terminal half of ABE7.10, were heterodimerized efficiently through recombination of the identical inverted terminal repeat (ITR) sequences in each vector (Figure 26). Because the two vectors also contained a splicing donor or acceptor signal, the full-length ABE7.10 protein was expressed after mRNA maturation and splicing.

(By Taeyoung Koo in Institute for Basic Science)

Figure 25. Scheme for the design of sgRNA for correction of Duchenne muscular dystrophy. The ABE target sequence in exon 20 of the *Dmd* gene containing a nonsense mutation. The PAM sequence is shown in blue. The protospacer sequence is underlined. The mutated nucleotide in the *Dmd* knockout mouse and the guanine nucleotide corrected by ABE are shown in red and green, respectively.

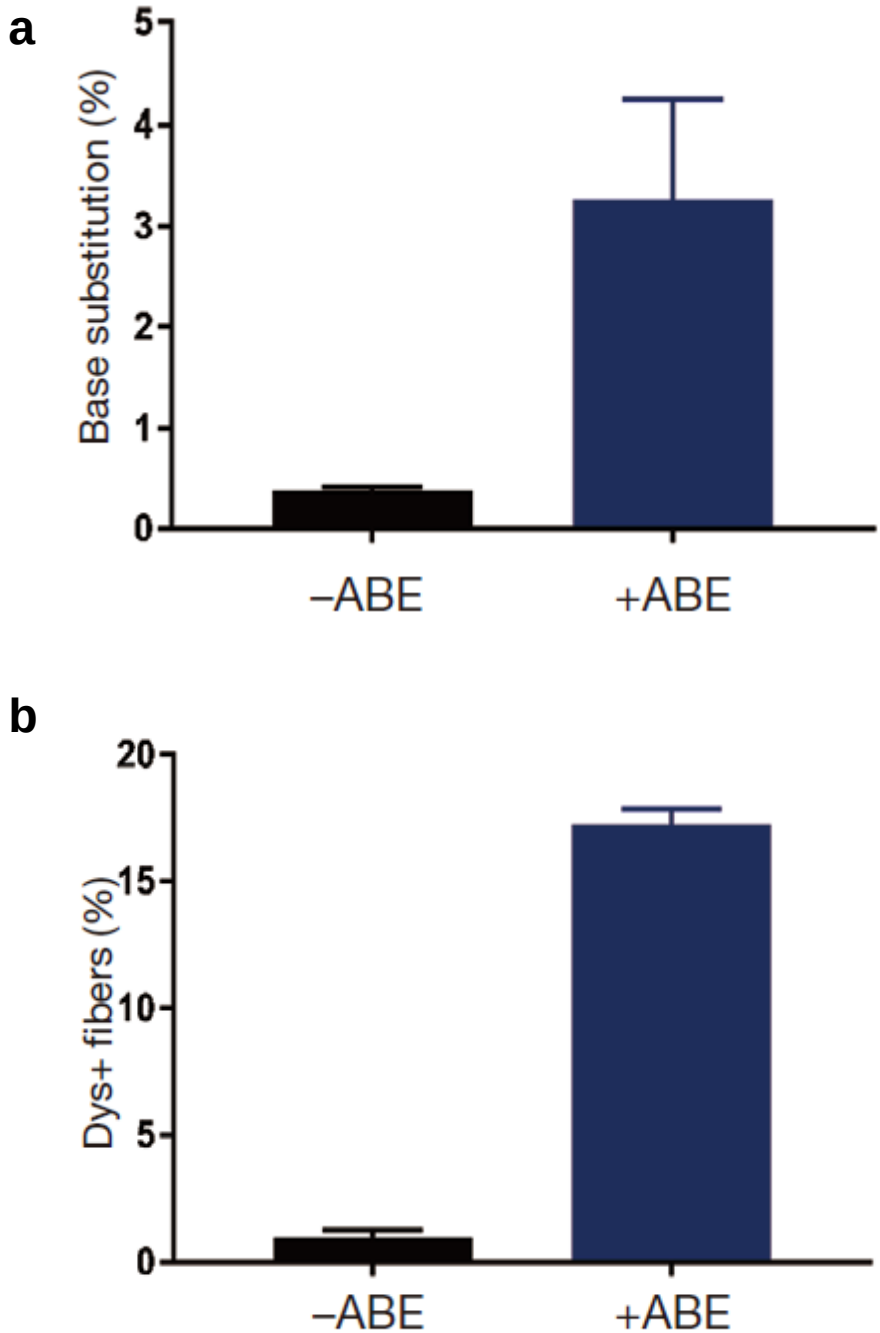


(By Taeyoung Koo in Institute for Basic Science)

Figure 26. Schematic diagram of trans-splicing AAV vector encoding ABE. The first vector carries a U6 promoter driven sgRNA, Spc512 promoter, ecTadA fused to 5'-end of nCas9-CT, followed by BLS, HA tag, and the bGH poly A signal (a) The tow split AAV-ABE vectors are coinfected into the target cells. (b) The two viral vectors are rejoined at the ITRs by recombination and led to heterodimer formation. (c) Target specific sgRNA and Pre-mRNA of ABE are made. (d) The intron along with the ITR by splicing is removed and ABE protein is made. SD, splicing donor; SA, splicing acceptor; NLS, nuclear localization signal; ITR, inverted terminal repeat.

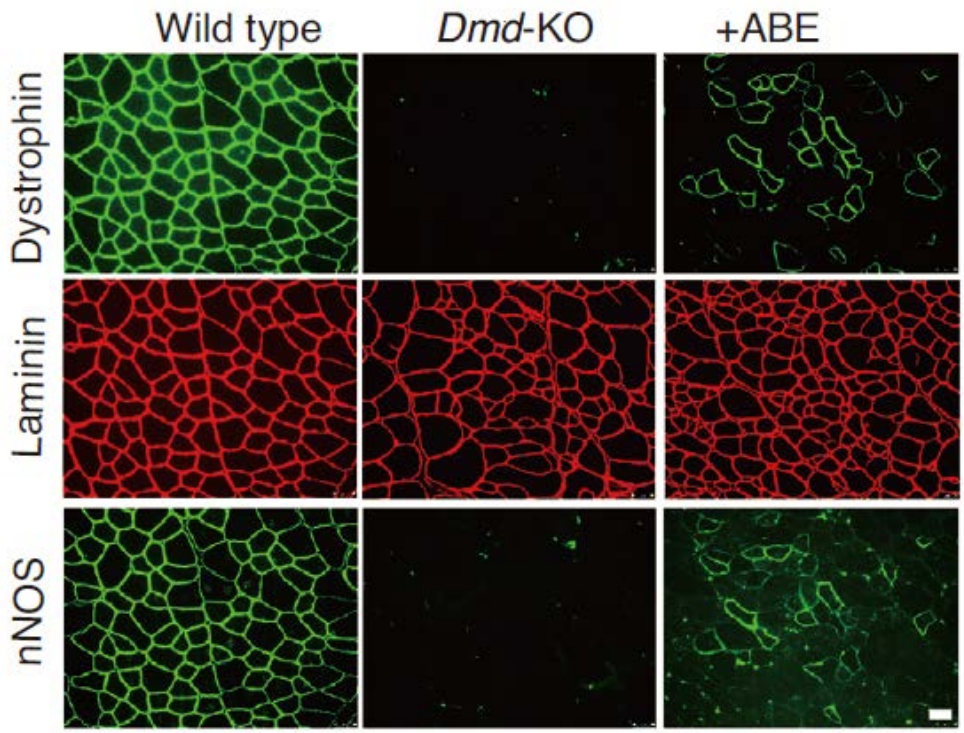
b. Rescue of a mouse model of Duchenne muscular dystrophy

Co-delivery of two trans-splicing viral vectors (tsAAV:ABE) *via* intramuscular administration into the tibialis anterior muscle in *Dmd* knockout mice induced precise A-to-G base substitutions, converting the premature stop codon to the glutamine codon, with a frequency of 3.3 ± 0.9%, 8 weeks post-injection, as assessed by targeted deep sequencing (Figure 27a). Unwanted indels were not induced at the target site by the tsAAV:ABE treatment, demonstrating an advantage of base editing over CRISPR–Cas9-mediated *Dmd* gene correction in mice (Long et al., 2016; Nelson et al., 2016; Tabebordbar et al., 2016). Notably, the *Dmd*-targeted tsAAV:ABE restored dystrophin expression in 17 ± 1% of myofibers (Figure 27b and 28). Note that ~4% of normal dystrophin expression is sufficient to improve muscle function (Li et al., 2010; van Putten et al., 2013). The tsAAV:ABE treatment also resulted in sarcolemmal localization of neuronal nitric oxide synthase (nNOS), a critical regulator of skeletal muscle exercise performance, suggesting that dystrophin produced in base-edited muscle cells functionally interacted with the dystrophinassociated protein complex (Figure 28). Taken together, our results demonstrate that ABEs can be used to make disease models with single-nucleotide substitutions in mice and that extended sgRNAs with a few additional nucleotides at the 5′ end can expand the window of base editing in vitro and in vivo. I also propose that ABEs, delivered via trans-splicing AAVs, may enable therapeutic base editing in vivo to correct point mutations causing genetic disorders.



(By Taeyoung Koo in Institute for Basic Science)

Figure 27. Therapeutic base editing with ABE in a mouse model of Duchenne muscular dystrophy. (a) Base editing frequencies at the *Dmd* target site were measured by deep sequencing. The error bar indicates s.e.m. (*n* = 3 animals). (b) Quantification of dystrophin positive (dys+) fibers in cross sections of TA muscles. Data are presented as mean ± s.e.m. (n = 3 animals)



(By Taeyoung Koo in Institute for Basic Science)

Figure 28. Histological analysis of *tibialis anterior* (TA) muscle. *Tibialis anterior* (TA) muscle from the wild-type, *Dmd* knockout, and ABE treated (+ABE) *Dmd* knockout mice 8 weeks after intramuscular delivery of tsAAV:ABE. Dystrophin and nNOS colocalized to the sarcolemma as shown in serial sections. Scale bar, 50 um.

c. Analysis of off-target effects

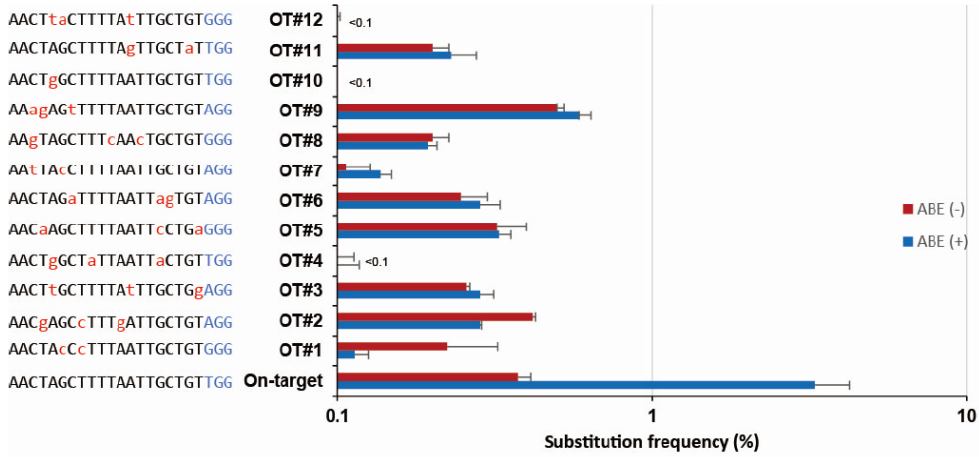
I also assessed off-target effects for the sgRNA targeting the nonsense mutation of *Dmd* gene. Like my previous experiment, off-target analysis of Himalayan mutant mouse, I identified potential off-target sites with up to 3-nucleotide mismatches in the mouse genome by Cas-OFFinder (Table 13) and obtained 12 candidates sites. Genomic DNA was isolated from ABE treated mouse and targeted deep sequencing was performed using Illumina Miniseq to figure out the point mutation efficiencies at the off-target candidates. Furthermore, no off-target mutations were detectably induced at homologous sites with up to three mismatches (Figure 29).

Table 13. Base editing frequencies at potential off-target.

				Base substitution (%)*	
Target		**Location**	**Target sequence (5' to 3')**	**(-) RGEN**	**(+) RGEN**
Dmd exon 20	On-target	chrX	AACTAGCTTTTAATTGCTGT**TGG**	0.4%	3.3%
	Dmd-OT1	chr8	AACTA**c**C**c**TTTAATTGCTGT**GGG**	0.2%	0.1%
	Dmd-OT2	chr8	AAC**g**AGC**c**TTT**g**ATTGCTGT**AGG**	0.4%	0.3%
	Dmd-OT3	chr4	AACT**t**GCTTTTA**t**TTGCTG**g**AGG	0.3%	0.3%
	Dmd-OT4	chr4	AACT**g**GCT**a**TTAATT**a**CTGT**TGG**	0.1%	0.1%
	Dmd-OT5	chr13	AAC**a**AGCTTTTAATT**c**CTG**a**GGG	0.3%	0.3%
	Dmd-OT6	chr19	AACTAG**a**TTTTAATT**ag**TGT**AGG**	0.2%	0.3%
	Dmd-OT7	chr14	AA**t**TA**c**CTTTTAATTGCTGT**AGG**	0.1%	0.1%
	Dmd-OT8	chr14	AA**g**TAGCTTT**c**AA**c**TGCTGT**GGG**	0.2%	0.2%
	Dmd-OT9	chr9	AA**ag**AG**t**TTTTAATTGCTGT**AGG**	0.5%	0.6%
	Dmd-OT10	chrX	AACT**g**GCTTTTAATTGCTGT**TGG**	0.0%	0.0%
	Dmd-OT11	chr18	AACTAGCTTTTA**g**TTGCT**a**T**TGG**	0.2%	0.2%
	Dmd-OT12	chr11	AACT**ta**CTTTTA**t**TTGCTGT**GGG**	0.1%	0.1%

*Base editing frequenceis (%)

(By Taeyoung Koo in Institute for Basic Science)



(By Taeyoung Koo in Institute for Basic Science)

Figure 29. Base editing specificity of ABE in *Dmd* knockout mouse. Base editing frequencies at potential off-target sites identified by Cas-OFFinder were measured using targeted deep sequencing of genomic DNA isolated from muscles 8 weeks after injection of tsAAV: ABE. Mismatched nucleotides and the PAM sequence are shown in red and blue, respectively. OT; off-target site. Data are presented as mean ± s.e.m. (n = 3 animals)

Discussion

In this study, I applied new Cas9 variants, Base Editors composed of a cytidine or adenine deaminase fused to Cas9 nickase convert cytidine to thymine (C-to-T) or adenine to guanine (A-to-G), respectively, leading to single base-pair substitutions to generate mouse models. The conventional method to induce point mutation requires DNA template donors because it relies on the DNA repair pathway called homologous-directed repair (HDR) with an extremely low efficiencies. However, base editing systems do not need template donor DNA because these do not induce double-strand breaks and do not use the endogenous DNA repair pathways. Base editors deaminate cytosine or adenine in non-target strand which is exhibited as single-strand DNA. Because this systems had not been tested yet in mouse embryos to generate mouse models, I suggested that generation of mouse models with base editing systems would broaden the scope of genome editing to study of human genetic diseases.

Generating animal models is an important research to apply scientific researches to human for the therapy of genetic disorders. Especially, mouse model is the most widely used system to understand the characteristics of the diseases and to study the therapeutic methods of the disorders. Therefore, generating the accurate mouse models that mimic the genotype and phenotype of the human genetic disorders precisely is crucial. The major source of human genetic disease is single-nucleotide substitution rather than small insertions/deletions (indels) or chromosomal rearrangements in the genome. However, inducing point mutations to generate mouse model was a limitation because of the extremely low efficiencies. By CRISPR-Cas9 system the efficiencies of generating knock-out mouse models

carrying small indels are significantly improved but not for carrying point mutations. About 60 percent of the human genetic diseases caused by C-to-T (48%) or A-to-G (14%) substitutions, respectively. Theoretically, all together with cytosine and adenine base editors can be used to study about 60 percent of human genetic disorders for the mechanisms, characteristics and therapeutic methods. Therefore, here I suggested that the efficient mouse model generation with ABEs would have the synergy effects with my previous study with BE3 to understand the human genetic disorders caused by single-nucleotide substitutions and to develop the therapeutic methods by the precise generation of mouse disease models carrying the same substitutions like human.

I designed to induce the premature stop codon or Himalayan mutation in *Tyr* and *Dmd*, which encode tyrosinase and dystrophin or *Tyr* gene using BE3 or ABE 7.10, respectively. My results of BE3 showed that microinjection or electroporation of BE3 mRNA or RNPs resulted in efficient and precise base editing in mouse embryos. The expected mutations were observed in 11 out of 15 (73%) 10 out of 10 (100%) blastocysts at the target sites with high mutation frequencies, from 16% to 100%. After delivery of BE3 mRNA and sgRNAs targeting *Dmd* using microinjection and RNPs targeting *Tyr* using electroporation, I transplanted edited mouse embryos into surrogated mothers and obtained offspring. For *Dmd* gene I obtained 5 mutant pups out of 9 pups. Among the 5 mutant pups, one mutant F0 mouse carried expected homozygous mutant alleles by C-to-T conversion. Immunostaining showed that this *Dmd* mutant mouse rarely expressed the dystrophin protein. Also, for *Tyr* gene I obtained seven F0 pups. All seven pups carried various mutations at the *Tyr* target site by BE3. Among them two newborn

mice showed an albino phenotype in their eyes and carried homozygous nonsense mutation in *Tyr* gene.

Also, I used ABE 7.10 to induce Himalayan mutation in *Tyr* gene. However, the expected target was slightly away from the base editing window. I needed to increase editing efficiencies where out of the window. Therefore, I used extended sgRNAs to increase base editing efficiencies. The extended sgRNAs with a few additional nucleotide at the 5' end can expand the window of base editing in HEK293T cells. Also, I applied this strategy to mouse embryo editing. I designed 3 types of sgRNAs, gX19, GX20 and GX21, and delivered into mouse embryos with ABE 7.10 mRNA by microinjection. The three sgRNAs were almost equally efficient with average editing efficiencies of 78%, 95% and 78%, respectively. However, the number of blastocysts carrying the Himalayan mutant allele were increased when I used a longer sgRNAs from 13% to 67%. Therefore, I chose GX21 sgRNA to generate Himalayan mutant mice and transplanted edited embryos into surrogate mothers to obtain offspring as same as my previous experiment using BE3. Seven out of nine pups harbored base substitutions with frequencies that ranged 12% to 100%. One F0 mouse showed albino phenotype in the eyes and coat color. Targeted deep sequencing revealed the mouse carrying the homozygous Himalayan alleles.

I confirmed the germline transmission and off-target effects for the generated mice by BE3 and ABEs. The mutant alleles from F0 mice were well transmitted to next generation. To assess the off-target effects, I identified potential off-target sites using Cas-OFFinder up to 3-nucleotide mismatches in the mouse genome and analyzed genomic DNA isolated from newborn mice *via* targeted deep

sequencing. Also, I performed whole genome sequencing (WGS) to identify more off-target candidates which contain more mismatches than analyzed by targeted deep sequencing. No off-targets were detectably induced at these sites. Taken together, our results demonstrate that BE3 and ABEs can be used to make disease models with single-nucleotide substitutions in mice and that extended sgRNAs with a few additional nucleotides at the 5′ end can expand the window of base editing *in vitro* and *in vivo*.

Finally, I also proposed that ABEs, delivered via trans-splicing AAVs, may enable therapeutic base editing *in vivo* to correct point mutations causing genetic disorders. Also, I could observed that the wrecked expression of *Dmd* gene was rescued and expressed dystrophy protein was functional. I showed that the generation of mouse models carrying point mutations using base editing systems, CBEs and ABEs, efficiently and used the systems as a therapeutic methods. The mouse disease models to study human genetic diseases are usually generated as knock-out models rather than point mutations. These results demonstrate that RNA-programmable deaminases can be used for making various animal models with single amino-acid substitutions or nonsense mutations and for correction of the genetic defects in human embryos in the future.

References

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.

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in Escherichia coli. Nat Microbiol *3*, 423-429.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709-1712.

Bibikova, M., Golic, M., Golic, K.G., and Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics *161*, 1169-1175.

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.

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.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided

endonucleases and nickases. Genome Res *24*, 132-141.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., and Voytas, D.F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics *186*, 757-761.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., *et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cui, X., Ji, D., Fisher, D.A., Wu, Y., Briner, D.M., and Weinstein, E.J. (2011). Targeted integration in rat and mouse embryos with zinc-finger nucleases. Nat Biotechnol *29*, 64-67.

Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A., and Musunuru, K. (2013). Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. Cell Stem Cell *12*, 393-394.

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.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167-170.

Hu, J.H., Miller, S.M., Geurts, M.H., Tang, W., Chen, L., Sun, N., Zeina, C.M., Gao, X., Rees, H.A., Lin, Z., *et al.* (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature *556*, 57-63.

Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.M., Lee, Y.S., and Kim, J.S. (2016). Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. Nat Biotechnol *34*, 807-808.

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.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Kernstock, S., Davydova, E., Jakobsson, M., Moen, A., Pettersen, S., Maelandsmo, G.M., Egge-Jacobsen, W., and Falnes, P.O. (2012). Lysine methylation of VCP by a member of a novel human protein methyltransferase family. Nat Commun *3*, 1038.

Kiani, S., Beal, J., Ebrahimkhani, M.R., Huh, J., Hall, R.N., Xie, Z., Li, Y., and Weiss, R. (2014). CRISPR transcriptional repression devices and layered circuits in mammalian cells. Nat Methods *11*, 723-726.

Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I., and Kim, J.S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat Methods *12*, 237-243, 231 p following 243.

Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M., and Kim, J.S. (2017a). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat Biotechnol *35*, 475-480.

Kim, J., Malashkevich, V., Roday, S., Lisbin, M., Schramm, V.L., and Almo, S.C. (2006). Structural and kinetic characterization of Escherichia coli TadA, the wobble-specific tRNA deaminase. Biochemistry *45*, 6407-6416.

Kim, K., Ryu, S.M., Kim, S.T., Baek, G., Kim, D., Lim, K., Chung, E., Kim, S., and Kim, J.S. (2017b). Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol *35*, 435-437.

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.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017c). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol *35*, 371-376.

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.

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.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., *et al.* (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583-588.

Kwon, B.S., Halaban, R., and Chintamaneni, C. (1989). Molecular basis of mouse Himalayan mutation. Biochem Biophys Res Commun *161*, 252-260.

Lai, Y., Yue, Y., Liu, M., Ghosh, A., Engelhardt, J.F., Chamberlain, J.S., and Duan, D. (2005). Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. Nat Biotechnol *23*, 1435-1439.

Lee, H.J., Kim, E., and Kim, J.S. (2010). Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res *20*, 81-89.

Li, D., Yue, Y., and Duan, D. (2010). Marginal level dystrophin expression improves clinical outcome in a strain of dystrophin/utrophin double knockout mice. PLoS One *5*, e15286.

Liang, F., Han, M., Romanienko, P.J., and Jasin, M. (1998). Homology-directed repair is a major double-strand break repair pathway in mammalian cells. Proc Natl Acad Sci U S A *95*, 5172-5177.

Long, C., Amoasii, L., Mireault, A.A., McAnally, J.R., Li, H., Sanchez-Ortiz, E., Bhattacharyya, S., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. Science *351*, 400-403.

Losey, H.C., Ruthenburg, A.J., and Verdine, G.L. (2006). Crystal structure of Staphylococcus aureus tRNA adenosine deaminase TadA in complex with RNA. Nat Struct Mol Biol *13*, 153-159.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods *13*, 1029-1035.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat Methods *10*, 977-979.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013a). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat

Biotechnol *31*, 833-838.

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

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.

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

Nelson, C.E., Hakim, C.H., Ousterout, D.G., Thakore, P.I., Moreb, E.A., Castellanos Rivera, R.M., Madhavan, S., Pan, X., Ran, F.A., Yan, W.X., *et al.* (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. Science *351*, 403-407.

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*.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., *et al.* (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods *10*, 973-976.

Radzisheuskaya, A., Shlyueva, D., Muller, I., and Helin, K. (2016). Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res *44*, e141.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., *et al.* (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380-1389.

Rock, J.M., Hopkins, F.F., Chavez, A., Diallo, M., Chase, M.R., Gerrick, E.R., Pritchard, J.R., Church, G.M., Rubin, E.J., Sassetti, C.M., *et al.* (2017). Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. Nat Microbiol *2*, 16274.

Ryu, S.M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.T., Kim, H.S., Kim, D.E., Lee, H., Chung, E., *et al.* (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. Nat Biotechnol.

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.

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A., and Kucherlapati, R.S. (1985). Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature *317*, 230-234.

Soldner, F., Laganiere, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S., *et al.* (2011). Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell *146*, 318-331.

Sorek, R., Lawrence, C.M., and Wiedenheft, B. (2013). CRISPR-mediated adaptive immune systems in bacteria and archaea. Annu Rev Biochem *82*, 237-266.

Steentoft, C., Vakhrushev, S.Y., Vester-Christensen, M.B., Schjoldager, K.T., Kong, Y., Bennett, E.P., Mandel, U., Wandall, H., Levery, S.B., and Clausen,

H. (2011). Mining the O-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. Nat Methods *8*, 977-982.

Sun, L., Li, J., and Xiao, X. (2000). Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. Nat Med *6*, 599-602.

Sung, Y.H., Baek, I.J., Kim, D.H., Jeon, J., Lee, J., Lee, K., Jeong, D., Kim, J.S., and Lee, H.W. (2013). Knockout mice created by TALEN-mediated gene targeting. Nat Biotechnol *31*, 23-24.

Sung, Y.H., Kim, J.M., Kim, H.T., Lee, J., Jeon, J., Jin, Y., Choi, J.H., Ban, Y.H., Ha, S.J., Kim, C.H., *et al.* (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res *24*, 125-131.

Tabebordbar, M., Zhu, K., Cheng, J.K.W., Chew, W.L., Widrick, J.J., Yan, W.X., Maesner, C., Wu, E.Y., Xiao, R., Ran, F.A., *et al.* (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. Science *351*, 407-411.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. Curr Opin Microbiol *14*, 321-327.

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.

van Putten, M., Hulsker, M., Young, C., Nadarajah, V.D., Heemskerk, H., van der Weerd, L., t Hoen, P.A., van Ommen, G.J., and Aartsma-Rus, A.M. (2013).

Low dystrophin levels increase survival and improve muscle pathology and function in dystrophin/utrophin double-knockout mice. FASEB J *27*, 2484-2495.

Wefers, B., Meyer, M., Ortiz, O., Hrabe de Angelis, M., Hansen, J., Wurst, W., and Kuhn, R. (2013). Direct production of mouse disease models by embryo microinjection of TALENs and oligodeoxynucleotides. Proc Natl Acad Sci U S A *110*, 3782-3787.

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., and Jaenisch, R. (2013). One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. Cell *154*, 1370-1379.

국문 초록

ZFN, TALEN, 및 CRISPR-Cas9 등의 유전자 가위 기술을 이용한 유전체 교정연구는 다양한 연구 분야에서 널리 사용되고 있으며, 특히 세균내의 면역체계에서 유래한 CRISPR-Cas9는 ZFN, TALEN에 비해 사용하기가 쉬운 특징으로 인해 급속한 발전을 이루었다.

이러한 기술의 발전은 다양한 Cas9 단백질의 이형을 만들어내었고 이들중에는 Cas9 단백질의 DNA를 절단하는 활성을 없애고 표적 유전자에 정확하게 결합하는 기능만 남긴 후 다른 단백질에서 유래한 effector domain을 접합시켜 접합된 단백질의 특성을 표적 특이적으로 사용할 수 있는 이형들이 연구되어 왔다. 이 중 탈아미노효소와 결합한 염기교정 유전자 가위는 특정 유전자의 염기 하나를 치환하는 것이 가능하다. 이는 가장 최근에 나온 Cas9 단백질의 이형으로써 많은 연구자들로부터 주목받고 있으며 시토신을 티민으로 바꾸는 염기교정 가위와 아데닌을 구아닌으로 치환할 수 있는 염기교정 가위, 두 종류가 있다.

이번 연구에서 두 종류의 염기교정 유전자 가위를 이용하여 돌연변이를 가진 생쥐를 만들 수 있음을 처음으로 보였으며 이를 이용하여 다양한 형질전환 실험 동물을 만들 수 있는 가능성을 보여주었다. *Tyr* 유전자에 특정 염기를 교체하여 백색증이 있는 생쥐를 만들었을 뿐만 아니라 *Dmd* 유전자를 망가뜨려 근위축증이 발병한 생쥐를 만들었다. 그리고 이렇게 만들어진 근위축증을 가진 성체 생쥐에

아데노 부속 바이러스를 이용하여 아데닌 염기교정 유전자 가위를 전달해서 돌연변이 염기를 정상으로 치환하는 연구를 진행하였고 교정된 효과까지 확인함으로써 유전질환의 원인인 돌연변이를 정밀하게 교정할 수 있는 가능성을 제시한다.

뿐만 아니라 염기교정 유전자 가위의 특정 염기의 치환 능력은 특정 범위 안에서만 작동하는데 제한된 범위가 한계점으로 작동할 때가 있기 때문에 이 범위를 확장시키는 연구도 진행하였다. 결과적으로 연장된 guide RNA를 사용하여 염기교정 유전자 가위의 작동 범위를 넓힐 수 있었다. 이러한 결과를 바탕으로 새로운 종류의 유전자 가위를 동물 수준에서 처음으로 적용하여 더욱 다양한 종류의 형질전환 실험동물을 효과적으로 제작할 수 있음을 보여주었을 뿐만 아니라 유전질환의 치료방법으로도 사용할 수 있음을 입증하였다.

학 번 : 2013-22920