

# Adenine base editing of the *DUX4* polyadenylation signal for targeted genetic therapy in facioscapulohumeral muscular dystrophy

Darina Šikrová,<sup>1</sup> Vlad A. Cadar,<sup>2</sup> Yavuz Ariyurek,<sup>1,3</sup> Jeroen F.J. Laros,<sup>1,4,5</sup> Judit Balog,<sup>1</sup> and Silvere M. van der Maarel<sup>1</sup>

<sup>1</sup>Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands; <sup>2</sup>Leiden University, 2300 RA Leiden, the Netherlands; <sup>3</sup>Leiden Genome Technology Center, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands; <sup>4</sup>Department of Clinical Genetics, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands; <sup>5</sup>National Institute for Public Health and the Environment (RIVM), 3721 MA Bilthoven, the Netherlands

**Facioscapulohumeral muscular dystrophy (FSHD) is caused by chromatin relaxation of the D4Z4 repeat resulting in misexpression of the D4Z4-encoded *DUX4* gene in skeletal muscle. One of the key genetic requirements for the stable production of full-length *DUX4* mRNA in skeletal muscle is a functional polyadenylation signal (ATTAAG) in exon three of *DUX4* that is used in somatic cells. Base editors hold great promise to treat DNA lesions underlying genetic diseases through their ability to carry out specific and rapid nucleotide mutagenesis even in postmitotic cells such as skeletal muscle. In this study, we present a simple and straightforward strategy for mutagenesis of the somatic *DUX4* polyadenylation signal by adenine base editing in immortalized myoblasts derived from independent FSHD-affected individuals. We show that mutating this critical *cis*-regulatory element results in downregulation of *DUX4* mRNA and its direct transcriptional target genes. Our findings identify the somatic *DUX4* polyadenylation signal as a therapeutic target and represent the first step toward clinical application of the CRISPR-Cas9 base editing platform for FSHD gene therapy.**

## INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD) is a hereditary skeletal muscle disorder that typically becomes manifest around the second decade of life and progresses with high inter- and intra-familial variability.<sup>1–3</sup> It is believed that this variability in disease progression and severity can be partially explained by the underlying epigenetic mechanism of the disease, being a failure to establish and/or maintain a repressive chromatin structure of the D4Z4 macrosatellite repeat at 4q35 in somatic cells. This leads to a variegated expression of the D4Z4 repeat-encoded *DUX4* gene in muscle cells.<sup>4</sup> *DUX4* is a pioneer transcription factor that under physiological conditions is expressed in keratinocytes,<sup>5</sup> testes,<sup>4</sup> and thymus<sup>6</sup> and in cleavage stage embryos, where it drives zygotic genome activation.<sup>4,7–9</sup> When misexpressed in muscle cells, it disrupts, among others, the bona fide muscle transcriptome.<sup>10,11</sup>

The repressive chromatin environment of the D4Z4 locus in somatic cells is likely established by a repeat-mediated epigenetic silencing

mechanism that partly depends on the D4Z4 repeat unit copy number.<sup>12</sup> There are two genetically distinct but overlapping forms of FSHD: FSHD type 1 (FSHD1; OMIM: MIM158900) and FSHD type 2 (FSHD2; OMIM: MIM158901).<sup>13,14</sup> The more common form, FSHD1, is caused by a shortening of the D4Z4 repeat to a size of 1–10 units,<sup>15</sup> whereas in FSHD2 the repeat size is within the lower range of healthy individuals (9–20 D4Z4 units). In the latter case, *DUX4* derepression is caused by a malfunction of D4Z4 chromatin modifiers.<sup>16–18</sup> Most FSHD2 individuals can be explained by heterozygous mutations in the gene encoding for the Structural Maintenance of Chromosomes flexible Hinge Domain-Containing protein 1 (*SMCHD1*),<sup>17</sup> a protein involved in, among other pathways, epigenetic inactivation of the X chromosome in mammals.<sup>19–23</sup> A small number of *SMCHD1* mutation-negative FSHD2 families have been reported in which mutations in the genes encoding for the chromatin modifiers DNA Methyltransferase 3B (*DNMT3B*) or Ligand Dependent Nuclear Receptor Interacting Factor 1 (*LRF1*) were shown to cause D4Z4 chromatin relaxation and *DUX4* expression in skeletal muscle.<sup>16,18</sup>

In addition to D4Z4 chromatin relaxation, the genetic background of the 4q subtelomere is critically important for FSHD manifestation. There are two equally common variants of this subtelomere, termed 4qA and 4qB;<sup>24</sup> however, only the 4qA variant is associated with the disease.<sup>25,26</sup> This is due to a sequence difference immediately distal to the distal D4Z4 unit, where the 4qA allele contains an additional 260 bp sequence termed pLAM that creates the third exon of *DUX4* with a functional ATTAAG polyadenylation signal (PAS) in somatic cells. Such genetic prerequisite for developing FSHD is supported by the finding that a contraction of the highly homologous D4Z4 repeat on chromosome 10 (10q26) does not lead to FSHD despite the presence of the pLAM sequence. However, this sequence

Received 8 January 2021; accepted 26 May 2021;  
<https://doi.org/10.1016/j.omtn.2021.05.020>.

**Correspondence:** Darina Šikrová, Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands.

**E-mail:** [darinasikrova@gmail.com](mailto:darinasikrova@gmail.com)

**Correspondence:** Silvere M. van der Maarel, Department of Human Genetics, Leiden University Medical Center, 2333 ZC Leiden, the Netherlands.

**E-mail:** [S.M.van\\_der\\_Maarel@lumc.nl](mailto:S.M.van_der_Maarel@lumc.nl)



contains a single-nucleotide polymorphism (SNP) in the corresponding *DUX4* PAS sequence (ATTAAA → ATCAAA), which renders it non-functional.<sup>27</sup> The critical importance of this *DUX4* PAS sequence was recently corroborated with the identification of two chromosome 10q-linked FSHD families in which the distal end of the disease-associated contracted D4Z4 repeat on chromosome 10, including the pLAM sequence, originated from chromosome 4.<sup>28</sup> Likewise, 4qB chromosomes lack the pLAM sequence altogether, and, consequently, a D4Z4 repeat contraction on this genetic background does not lead to the development of FSHD.<sup>26</sup>

Previously, it has been shown by different approaches, including the application of antisense oligonucleotides, DNA nucleases and U7 small nuclear RNA (snRNA), that interference with the usage of the endogenous 4qA *DUX4* PAS in myogenic cells derived from FSHD patients results in transcriptional downregulation of *DUX4* and its target genes,<sup>29–33</sup> further emphasizing the necessity of the annotated 4qA *DUX4* PAS for proper 3' end processing of *DUX4* pre-mRNA and suggesting that interfering with its usage is sufficient to alleviate the FSHD expression signature in myogenic cells.

Currently, there is no cure for FSHD and because of the underlying genetic character of the disease, CRISPR/Cas9 genome editing could be a promising tool for its treatment. Unfortunately, because of the repetitive nature of the *DUX4* gene (every D4Z4 unit contains one copy of the *DUX4* open reading frame [ORF]), a straightforward Cas9 nuclease-mediated knockout strategy might lead to multiple breaks, trigger genomic instability, and result in cell death as has been shown for targeting multicopy genomic regions.<sup>34</sup> Therefore, a different approach is required. The novel RNA-programmable base editing system, which consists of a wild-type (WT) tRNA adenosine deaminase (TadA) and an artificially evolved version of TadA (TadA\*) fused as a dimer to the D10A nicking version of *Streptococcus pyogenes* Cas9 (nSpCas9), hereafter referred to as nSpABE, enables robust adenine to guanine substitution without reliance on homology-directed repair (HDR) or introduction of double-stranded DNA breaks.<sup>35</sup> Such editing system has already been shown to faithfully edit the desired nucleotides also in postmitotic cells such as neurons<sup>36</sup> or skeletal muscle cells.<sup>37,38</sup> In this study, we aimed to take advantage of this system to demonstrate that the 4qA *DUX4* PAS can be efficiently disrupted with this approach, resulting in downregulation of *DUX4* transcript levels in FSHD myogenic cells.

## RESULTS

### Validation of sgRNA targeting *DUX4* polyadenylation signal in HAP1 cells

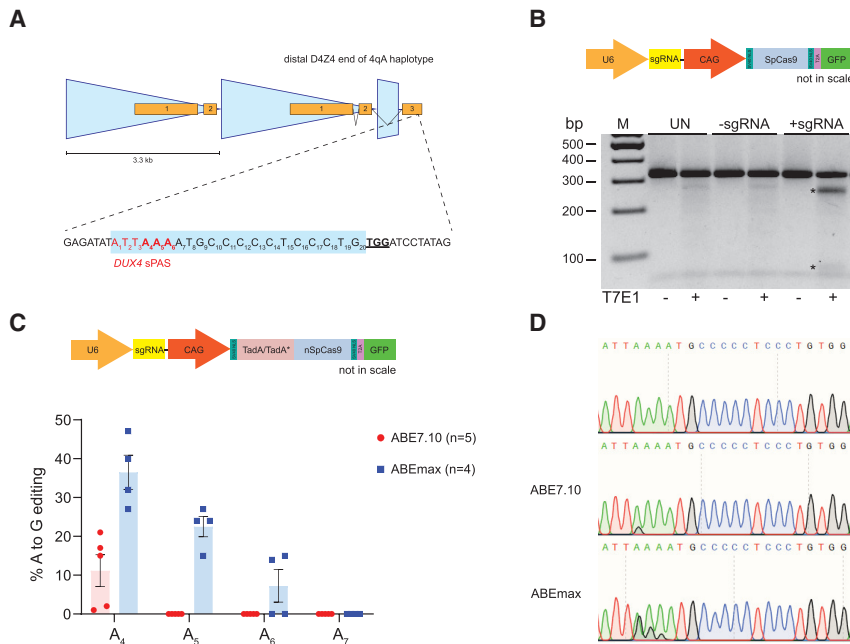
In myonuclei, the FSHD disease gene *DUX4* is transcribed from the distal unit of the D4Z4 repeat on the 4qA subtelomere, where its transcripts are stabilized by a PAS in exon 3. The adjacent SpCas9 protospacer adjacent motif (PAM) site (TGG) downstream of this PAS allows for the design of a single guide RNA (sgRNA) that places the last three adenines of the *DUX4* PAS (ATTAAA) in the activity window of nSpABE (Figure 1A). To test whether this sgRNA can effectively direct the Cas9 machinery to the locus of interest, we first per-

formed a T7E1 assay on HAP1 cells transfected with the sgRNA and a human codon-optimized SpCas9 nuclease. Despite having a repeat of 25 D4Z4 units on chromosome 4, which is most probably compacted into a dense chromatin structure perhaps hindering the interaction of the DNA with CRISPR/Cas9, we could clearly detect cleavage of the intended locus (Figure 1B). To evaluate A → G base editing of the *DUX4* PAS, we used a one-vector system for delivery of all adenine base editing components. HAP1 cells were individually transfected with two variants of the all-in-one vector in which the CAG promoter drives expression of the SpCas9 nickase fused to either the ABE7.10 or the ABEmax version of the adenine base editor, hereafter referred to as nSpABE7.10 and nSpABEmax, respectively (Figure 1C) and examined for A → G edits at the *DUX4* PAS site by Sanger sequencing. In nSpABE7.10-transfected cells, we could detect on average 11.2% ± 3.6% of A → G conversion for the adenine at position 4 of the protospacer (A<sub>4</sub>) as assessed by Sanger sequencing. We did not detect editing of adenines at positions 5 to 7 (A<sub>5-7</sub>) despite these adenines still fitting into the reported activity window of nSpABE7.10.<sup>35</sup> In nSpABEmax-transfected cells, we achieved more efficient adenine base editing at A<sub>4</sub> (36.5% ± 3.8%) as well as at downstream adenines A<sub>5</sub> (22.5% ± 2.25%) and A<sub>6</sub> (7.3% ± 3.6%), which is in agreement with a previous report that nSpABEmax is superior to nSpABE7.10 in terms of editing efficiency and processivity.<sup>40</sup>

Next, we assessed adenine editing of the *DUX4* PAS, using the ABEmax in combination with two other Cas9 orthologs, SaCas9 and CjCas9, since their cognate PAM sites, NNGRRT and NNNVRYM, respectively, are in the vicinity of the *DUX4* PAS such that adenines on the forward or reverse strand in the *DUX4* PAS could be amenable to adenine base editing (Figure S1A). We used the same all-in-one vector architecture as was used for nSpABEmax, including the same linker length, and the new constructs are hereafter referred to as nSaABEmax and nCjABEmax (Figure S1B). Surprisingly, both constructs failed to exert adenine base editing activity at the *DUX4* PAS in HAP1 cells based on evaluation by Sanger sequencing as was done for SpABE7.10 and SpABEmax (data not shown).

### Base editing of *DUX4* PAS in patient-derived immortalized FSHD1 and FSHD2 myoblasts

To explore the effect of the mutated PAS on *DUX4* steady-state transcript levels, we carried out base editing in FSHD patient-derived immortalized myoblasts, since HAP1 cells do not express *DUX4*. We used three different FSHD myogenic cell lines with different genetic characteristics, D4Z4 methylation status and *DUX4* expression levels (Figures S2A and S2B). We selected one FSHD2 cell line that has a heterozygous missense mutation in SMCHD1 (K204E) combined with an 11-unit-long 4qA D4Z4 repeat and two FSHD1 cell lines, one with a 3-unit-long 4qA D4Z4 repeat (FSHD1<sup>3U</sup>) and one with an 8-unit-long 4qA repeat (FSHD1<sup>8U</sup>). Shorter D4Z4 repeats are generally correlating with lower D4Z4 methylation levels,<sup>41</sup> a more severe FSHD phenotype and a worse prognosis,<sup>2</sup> whereas repeats in the upper size limit of FSHD1 typically show a higher incidence of familial non-penetrance and a milder disease presentation.<sup>3,42</sup> Furthermore, we chose cell lines heterozygous for 4qA and



**Figure 1. Adenine base editors can edit the *DUX4* PAS**

(A) Schematic representation of the distal end of the 4qA-derived D4Z4 macrosatellite repeat (each blue triangle represents one D4Z4 repeat unit) including the adjacent downstream sequence containing the polyadenylation signal of *DUX4* in exon 3 (*DUX4* exons are indicated by orange boxes) and zoom in on the sequence to be targeted by the adenine base editor. The sgRNA protospacer is outlined in the blue box, the PAM site for SpCas9 is underlined in bold and the *DUX4* PAS sequence (ATTAAA) is in red font with adenines that can be targeted by the adenine base editor in bold. (B) Schematic map of the pX458 vector for simultaneous sgRNA and SpCas9 nuclease expression (top). Result of the T7E1 assay performed on HAP1 cells that were transfected with the pX458 vector expressing the sgRNA targeting the *DUX4* PAS together with a plasmid encoding for puromycin resistance to select for transfected HAP1 cells (bottom). Untransfected cells (UN) or cells transfected with no sgRNA-containing vector (-sgRNA) served as negative control. Asterisks mark the T7E1 cleavage products. (C) Schematic map of the modified all-in-one pX458 vector encoding for the adenine base editors (top). Editing efficiency was assessed in HAP1 cells for the ABE7.10 and ABEmax versions of the adenine base editor. The A→G editing efficiency was calculated from Sanger sequencing tracks with EditR<sup>39</sup> for each adenine in the

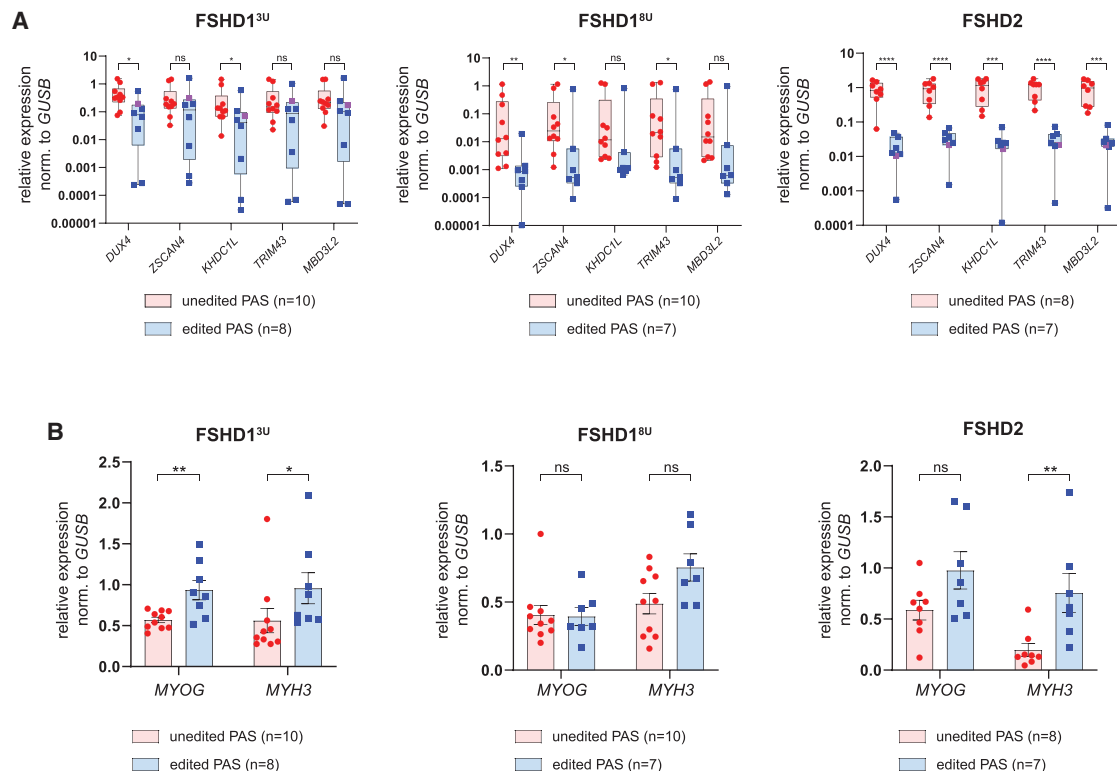
editing window. Graph shows mean  $\pm$  SEM of at least four independent biological replicates (dots). (D) Representative Sanger sequencing tracks for ABE7.10- or ABEmax-mediated editing of the *DUX4* PAS used for quantification.

4qB to facilitate unequivocal assignment of successful editing of the FSHD allele, except for FSHD1<sup>8U</sup>, which carries two variant alleles of 4qA (with the healthy allele being of the 4qA161L variant and the FSHD allele of the 4qA161S variant).<sup>43</sup> However, these two allelic variants of 4qA161 can be distinguished by the presence of a SNP (Figure S3A). Clonal cell cultures from all three cell lines were genotyped for the *DUX4* PAS after transfection with nSpABEmax and single-cell sorting of GFP<sup>+</sup> cells. Untransfected cells underwent the same sorting procedure to obtain clones with a WT PAS sequence to ensure the same experimental conditions and population doublings between compared groups. Successfully edited clones showed a plethora of A→G editing outcomes (Figure S3A). We also obtained one clone from the FSHD1<sup>3U</sup> and one clone from the FSHD2 cell line in which the editing attempt resulted in small deletions fully or partially encompassing the *DUX4* PAS (Figure S3A). *DUX4* steady-state mRNA levels were measured as well as those of four well-established *DUX4* target genes (*ZSCAN4*, *KHDC1L*, *TRIM43*, and *MBD3L2*)<sup>11,44</sup> serving as an indirect readout for *DUX4* transcription factor activity. The steady-state mRNA levels of *DUX4* and its target genes were reduced in all three cell lines upon editing of the *DUX4* PAS under proliferating (Figure S3B) as well as differentiating (Figure 2A) conditions. Since it has been shown that *DUX4* expression increases during myogenic differentiation,<sup>45</sup> we analyzed the expression of early (*MYOG*) as well as late (*MYH3*) myogenic markers by qRT-PCR to rule out the possibility that lower *DUX4* levels were due to reduced differentiation potential of edited clones (Figure 2B). On the contrary, edited clones showed equal if not slightly increased myogenic differentiation, which is in agreement with previous findings that *DUX4* in-

hibits myogenic differentiation, thereby lowering its levels would improve differentiation.<sup>10</sup> However, unedited clones showed a high variability in *DUX4* expression levels and those of its target genes ranging from 1 order of magnitude in the FSHD1<sup>3U</sup> and FSHD2 lines up to 3 orders of magnitude in clones derived from the FSHD1<sup>8U</sup> line. Such high expression variability thus makes it difficult to confidently determine the effect of *DUX4* downregulation conferred by base editing.

#### Reducing the clonal variability in *DUX4* expression

Since D4Z4 displays highly variable transcriptional activity between individuals<sup>46</sup> and across cells from the same individual (this study), a behavior that is also described for genomic loci known as metastable epialles<sup>47</sup> of which their epigenetic profile is stochastically established in early embryogenesis, we hypothesized that starting the editing from a monoclonal cell culture rather than a polyclonal culture may resolve a large part of inter-clonal variability in *DUX4* expression. This would facilitate a better comparison of *DUX4* levels between *DUX4* PAS pre-editing and post-editing clones in the absence of large expression variability at WT baseline. We therefore first tested the “mitotic stability” of *DUX4* expression by deriving new daughter clones from two clones showing different levels of *DUX4* expression (referred to as *DUX4*<sup>high</sup> and *DUX4*<sup>low</sup>) originating from the FSHD1<sup>8U</sup> line, as it showed the highest *DUX4* expression variability. Indeed, after resorting, new single-cell derived cultures exhibited more homogeneous *DUX4* and *DUX4* target gene (*ZSCAN4* and *MBD3L2*) expression levels comparable to the parental clone as measured by qRT-PCR (Figure S4).



**Figure 2. Adenine base editing of the *DUX4* PAS reduces expression of *DUX4* and its target genes in FSHD myogenic cells derived from polyclonal cultures**

(A) mRNA levels as assessed by qRT-PCR of *DUX4* and four *DUX4* target genes (*MBD3L2*, *ZSCAN4*, *TRIM43*, and *KHDC1L*) in PAS unedited versus edited clones derived from two FSHD1 and one FSHD2 cell lines differentiated into myotubes. Statistical significance was calculated with unpaired two-tailed t test (ns, non-significant; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ) on log<sub>2</sub> transformed expression values to correct for skewed distribution. Expression values normalized to *GUSB* as housekeeping gene are plotted. Line represents mean, and whiskers represent min and max value. Individual dots represent individual clones; the two violet clones carry a deletion affecting the *DUX4* PAS. (B) mRNA levels of two myogenic markers (*MYOG* and *MYH3*) for all unedited and edited clones of all three FSHD cell lines are plotted. Statistical significance was calculated with unpaired two-tailed t test (ns, non-significant; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). Bars represent mean  $\pm$  SEM, with individual clone expression values plotted as individual dots.

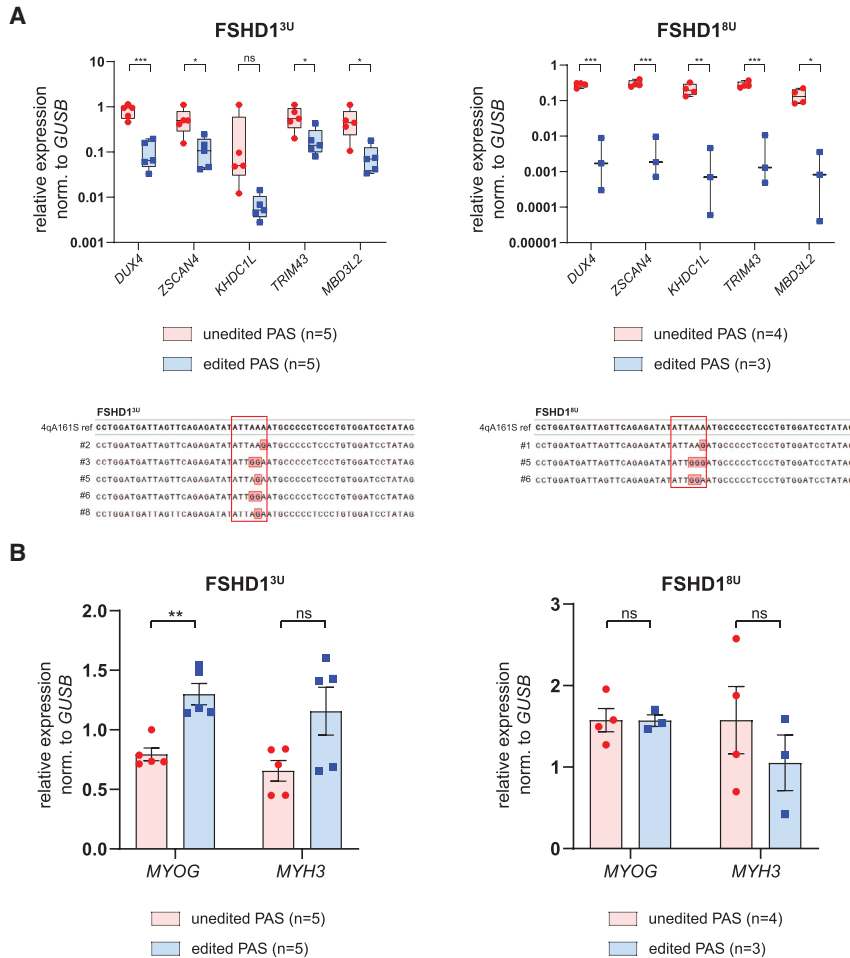
We selected one unedited *DUX4*<sup>high</sup> clone derived from either the FSHD1<sup>3U</sup> or the FSHD1<sup>8U</sup> cell line and repeated the editing procedure to obtain new *DUX4* PAS unedited and edited clones. As expected, deriving new unedited clones from a monoclonal culture resulted in lower *DUX4* expression variability between clones, with clones carrying an edited *DUX4* PAS showing significantly reduced *DUX4* steady-state mRNA levels as well as *DUX4* target gene levels (Figure 3A). Again, the reduced *DUX4* expression levels could not be attributed to a difference in myogenic differentiation, as shown by comparable expression of the two myogenic differentiation markers between edited and unedited clones (Figure 3B). Interestingly, editing the *DUX4* PAS seems to have a more negative impact on *DUX4* mRNA levels in FSHD1<sup>8U</sup> (~1,000-fold downregulation) than in cells from FSHD1<sup>3U</sup> line (~10-fold downregulation).

#### Editing of the *DUX4* PAS induces alternative pre-mRNA cleavage and polyadenylation

Previously, it was shown that hindering the *DUX4* PAS with phosphorodiamidate morpholino oligomers (PMOs) causes a redirection of the *DUX4* pre-mRNA cleavage site (CS) ~40 nt upstream of its ca-

nonical CS despite the absence of a recognizable alternative PAS motif in the upstream sequence.<sup>31</sup> Since base editing of the *DUX4* PAS does not completely abolish *DUX4* expression, we tested if the mutated PAS is still being used for *DUX4* transcript termination, albeit less efficiently, or if alternative PASs/CSs are being used. Using a semi-quantitative 3' rapid amplification of cDNA ends (3' RACE) to identify 3' UTR sequences of *DUX4* mRNAs from unedited and edited clones derived from all three FSHD immortalized cell lines from Figure 2A, we detected three different CSs 16–24 nt downstream of *DUX4* PAS in close proximity to each other in unedited cells (Figure 4A), as was previously described.<sup>31</sup> In edited clones, however, two different shifts in the CS occur, either proximally or distally to the canonical CS (Figure 4A & B, Table S3). Interestingly, the FSHD2 edited clones strictly used the proximal CS, the same one as reported by Marsollier et al.<sup>31</sup> after using PMOs against the *DUX4* PAS region, whereas the distal CS switch is predominant in the FSHD1 clones independent of their 4qA permissive allele size (Figure 4B). Moreover, opposite to the single proximal CS being used after PAS editing, the distal CS is not as deterministic, since we observed multiple different 3' ends in FSHD1 edited clones. Of note, the small





**Figure 3. Adenine base editing of the *DUX4* PAS reduces expression of *DUX4* and its target genes in FSHD myogenic cells derived from monoclonal cultures**

(A) mRNA levels as assessed by qRT-PCR of *DUX4* and four *DUX4* target genes (*MBD3L2*, *ZSCAN4*, *TRIM43*, and *KHDC1L*) in *DUX4* PAS unedited versus edited subclones derived from two clones with different FSHD1 cell line origins (top). Genotypes of edited clones aligned to the reference WT sequence with the *DUX4* PAS are highlighted in a red rectangle, and red colored bases denote mismatches (bottom). (B) mRNA levels of two myogenic markers (*MYOG* and *MYH3*) for unedited and edited clones from (A). Statistical significance was calculated with unpaired two-tailed t test (ns, non-significant; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ ). Bars represent mean  $\pm$  SEM, with individual clone expression values plotted as individual dots.

proportion of *DUX4* mRNAs using the canonical CS position in FSHD2 clones is coming from the single clone that carries a partial deletion of *DUX4* PAS. Despite the clear shift in the CS upon *DUX4* PAS editing, we could not detect a nearby PAS-like sequence ( $\pm 100$  nt from original PAS) which could explain the CS shifts. Overall these data show that *DUX4* PAS base editing prevents proper 3' end formation of the *DUX4* transcript.

#### Off-target analysis by targeted next generation sequencing

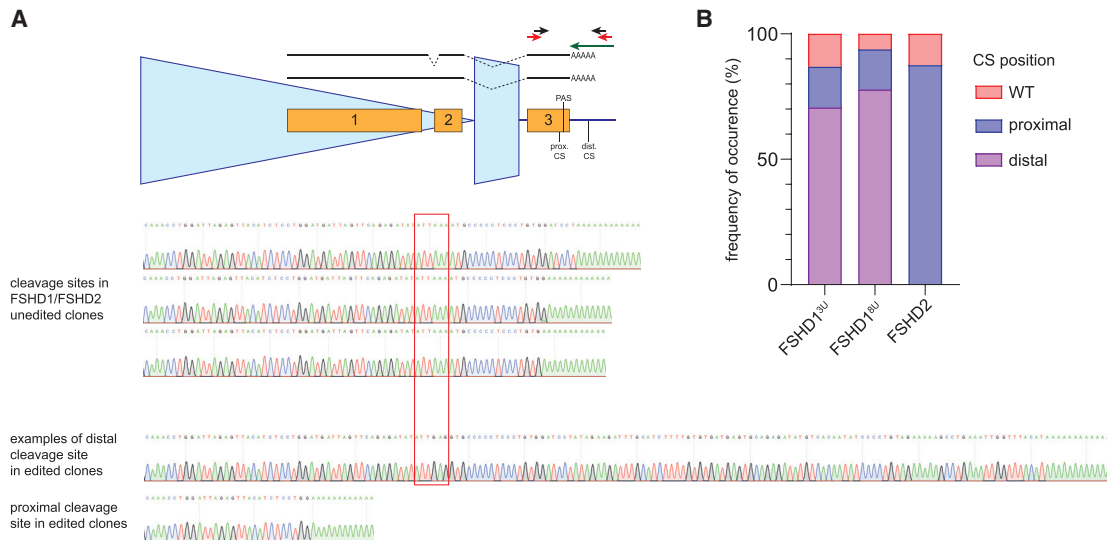
To explore potential off-target effects, we used the CRISPOR prediction tool<sup>48</sup> to identify genomic sites that have a sequence homology to the sgRNA used for targeting the *DUX4* PAS. This resulted in the identification of 227 potential off-target (OT) sites, of which none are predicted to target polyadenylation signals of other genes. Only 3 are predicted to target coding sequences, however, with low off-target scores due to the number and position of individual mismatches (Table S4). We further filtered predicted off-target sites by the following criteria: (1) having up to 4 mismatches outside of the PAM region and the seed region of the sgRNA, (2) containing at least one adenine in the editing window of nSpABEmax, and (3) representing a single copy locus. Based on these criteria, we performed targeted

next generation sequencing on 10 selected potential off-target sites in DNA samples obtained from HAP1 cells that were transfected with nSpABEmax with or without sgRNA targeting the *DUX4* PAS from Figures 1C and 1D (Figure 5A). At 7 out of 10 examined sites, deep sequencing did not reveal any appreciable increase in A  $\rightarrow$  G transitions within or near the editing window as compared to the control samples (Figure 5B). However, the nucleotide sequences of OT1 and OT10 contained a SNP in the HAP1 genome, producing an extra mismatch in the sgRNA protospacer (Figure 5A). Therefore, their off-target potential might be higher in genomes that do not contain this mismatch. At

three sites, OT2 (chr6: 13,331,126–13,331,148), OT5 (chr12: 2,444,719–2,444,741), and OT6 (chr2: 218,831,310–218,831,332), we detected editing efficiencies of 0.17%, 1.72%, and 0.43% of adenines within the editing window, respectively (Figures 5B and 5C). None of the three affected OT sites resides in coding regions. OT2 is in an intergenic region  $\sim 2$  kb upstream of the *TBCID7* gene, while OT5 and OT6 map to intron 3 of *CACNA1C* and intron 2 of *PRKAG3*, respectively. Both genes, *CACNA1C* and *PRKAG3*, are expressed in skeletal muscle according to the Human Protein Atlas,<sup>50</sup> but neither edit is predicted to affect the splicing of these genes when modeled with the Alamut software. In summary, these results show that sgRNA-dependent off-target DNA editing is likely rare.

#### DISCUSSION

So far, therapeutic attempts for FSHD have been mainly focused on oligonucleotide- or small molecule-based transient modulation of *DUX4* levels.<sup>51,52</sup> Three recent studies focused on gene therapy approaches that inhibit the production of full-length *DUX4* mRNA.<sup>32,33,53</sup> Two of these studies used CRISPR/Cas9 strategies, either employing a standard Cas9 nuclease to introduce deletions affecting the *DUX4* PAS by HDR with a provided template<sup>32</sup> or using



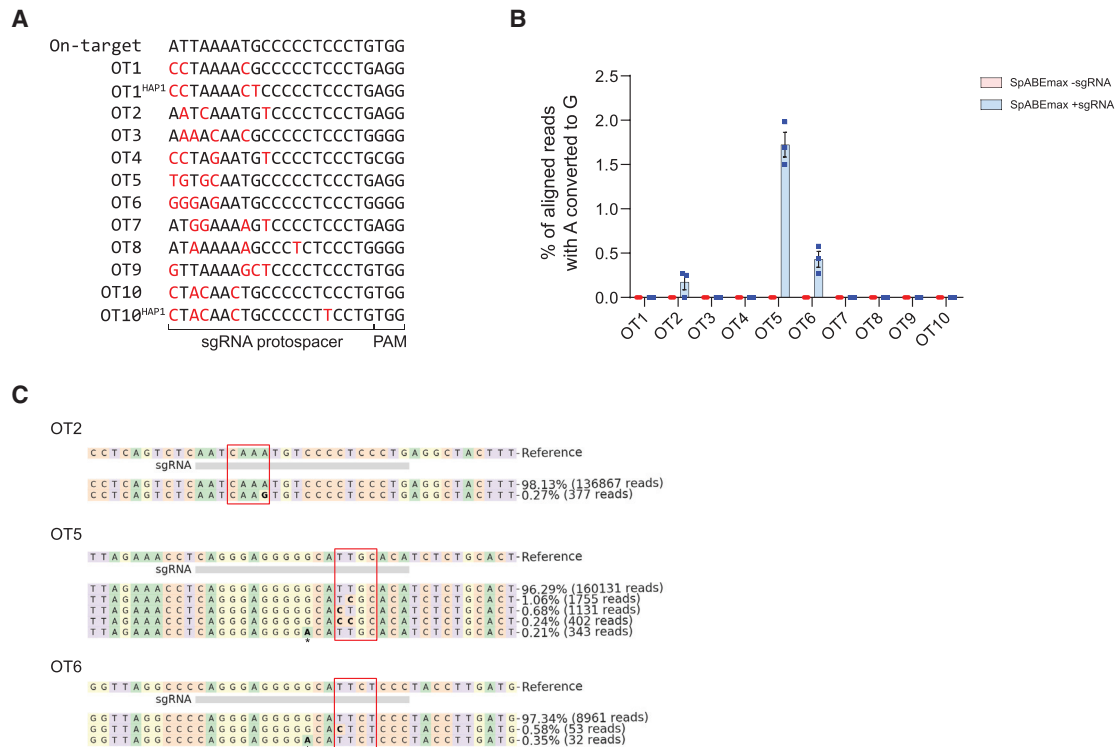
**Figure 4. Editing of the *DUX4* PAS induces alternative pre-mRNA cleavage site**

(A) Schematic of the terminal D4Z4 repeat unit with short ending (4A161S haplotype) showing the design of 3' RACE experiment to determine the cleavage and polyadenylation site of *DUX4* mRNA in the edited clones. Two known *DUX4* mRNA isoforms are depicted with splicing or retention of intron 1 (top). Arrows represent primers used for oligo-dT reverse transcription (green), first PCR (red), and second nested PCR (black). The identified proximal and distal cleavage sites, for which Sanger sequencing traces are provided, are marked. Sanger sequencing tracks (bottom) show representative examples of 3' ends of *DUX4* mRNA in *DUX4* PAS unedited and edited FSHD1/FSHD2 clones. The red rectangle outlines the *DUX4* PAS sequence. Three different CSs were identified in unedited clones (as reported previously<sup>31</sup>), while different shifts in CSs were identified in edited clones. One representative Sanger sequencing track is shown for each CS choice. (B) Barplots representing the frequency of occurrence of different CSs identified in *DUX4* PAS edited clones with respect to WT CSs from  $\geq 4$  clones for each condition.

Cas9 coupled to a transcriptional inhibitor domain to repress *DUX4* expression<sup>53</sup>. The third study used custom U7 nuclear RNAs (snRNAs) to mask important regulatory features of *DUX4* mRNA maturation such as splice sites and the *DUX4* PAS.<sup>33</sup> In this study, we demonstrate the use of a CRISPR/Cas9-based genome editing application to directly modify the *DUX4* locus while avoiding DNA double-strand breaks. We show that by using an adenine base editor we can target and disable one of the important genetic prerequisites for FSHD manifestation, the *DUX4* somatic polyadenylation signal. We were able to successfully edit the *DUX4* PAS with SpCas9-based base editors nSpABE7.10 and nSpABEmax, with the latter showing higher editing efficiency, which is in agreement with previous reports.<sup>54</sup> Fusing ABEmax to two other Cas9 orthologs, namely SaCas9 and CjCas9, has previously been shown to also result in adenine editing activity.<sup>54–56</sup> However, we did not observe adenine to guanine conversion at the *DUX4* PAS when using such fusion proteins in HAP1 cells as determined by Sanger sequencing. The T7E1 assay nevertheless did show evidence for recruitment of the SaCas9 nuclease to the *DUX4* PAS site (Figure S1C), suggesting that the complex can be recruited to the *DUX4* PAS but that the nSaABEmax fusion protein is likely not efficient at this site. Previously, a lower editing efficiency has been reported for nSaABEmax as compared to nSpABEmax,<sup>54</sup> which could explain our findings. Recently, a new version of the adenine base editor, termed ABE8e, was described.<sup>55</sup> When paired with a variety of Cas effectors, including SaCas9, it demonstrated a further enhanced editing efficiency. Therefore, coupling ABE8e to SaCas9 might result in successful adenine base ed-

iting of the *DUX4* PAS. In addition, such a fusion construct would be more favorable compared to the SpCas9 construct because of its smaller size, which could facilitate the use of the adeno-associated viral (AAV) system for its *in vivo* delivery and testing. Alternatively, an AAV split system could be used for *in vivo* delivery of SpABEmax or SpABE8e. Indeed, such an approach has been already tested for delivering base editors to a range of tissues,<sup>37,57</sup> reaching 20% editing efficiency in skeletal muscle tissue.<sup>57</sup> Since published strategies were aiming at whole body delivery and were not optimized for skeletal muscle targeting or expression, further optimization by using a tissue-specific promoter and a muscle-trophic AAV serotype might increase the editing efficiencies in the skeletal muscle. On the other hand, the failure to detect editing of the *DUX4* PAS with nCjABEmax might be attributed to a suboptimal nearby PAM sequence (5'-AATCATC-3') that was predicted for the targeting. We identified this PAM site based on the PAM consensus sequence (5'-NNNVRYM-3') reported by Yamada et al.<sup>58</sup> Another study by Kim et al.<sup>59</sup> reported a slightly different PAM consensus sequence (5'-NNNNRYAC-3') for CjCas9 targeting that is more refined and differs from the sequence that we used for deriving our sgRNA. Moreover, such a fusion construct has not been characterized in depth yet; therefore there is no knowledge about its precise editing window or its efficiency.

As anticipated, editing of the *DUX4* PAS in immortalized myogenic lines obtained from different FSHD-affected individuals resulted in lower *DUX4* mRNA levels and lower *DUX4* transcription factor activity as indirectly measured by the steady-state mRNA levels of its



**Figure 5. sgRNA-dependent off-target analysis in HAP1 cells**

(A) DNA sequences of 10 predicted off-target sites identified by CRISPOR.<sup>48</sup> Nucleotide mismatches compared to the *DUX4* PAS target sequence are highlighted with red font. Two off-target sites (OT1 and OT10) carried an extra mismatch in HAP1 cells as compared to the reference sequence obtained from GRCh38. (B) Editing frequencies at predicted off-target sites were assessed in HAP1 cells that were transfected with nSpABEmax and either with or without *DUX4* PAS targeting sgRNA. The A→G editing efficiency was assessed by amplicon next generation sequencing and analyzed with CRISPResso2.<sup>49</sup> Graph shows mean ± SEM of 3 independent biological replicates. (C) Representative allele frequencies of three off-target sites (OT2, OT5, and OT6) with the highest editing outcome are shown. OT5 and OT6 sequences are shown in forward orientation, while sgRNA targets the reverse complement strand. The editing windows are highlighted in the red box. Only allele frequencies of at least 0.1% were considered. The mutation rate in the G homopolymer (marked by asterisk) preceding the editing window was not included in the editing frequency calculation plotted in (B) since it occurred also in the control samples and was more likely introduced either during PCR steps or Illumina sequencing itself rather than in an sgRNA-dependent fashion.

target genes. We could not determine if editing more adenines at once or if editing an adenine at a particular position in the *DUX4* PAS motif results in a more profound *DUX4* downregulation, since multiple clones with the same editing outcome would be required to confidently assess this. Nevertheless, we show that even a single adenine substitution is sufficient to negatively impact proper 3' end processing of the *DUX4* transcript. To our surprise, mutating the *DUX4* PAS in this manner does not completely abolish the production of polyadenylated *DUX4* transcripts as opposed to the situation on chromosome 10, which might suggest the presence of other *cis* modifiers acting as regulators of *DUX4* expression than just the previously recognized SNP in 4q/10q *DUX4* PAS motif. These *cis* factors are likely in linkage disequilibrium with the *DUX4* PAS considering the exclusive linkage of FSHD with the presence of a *DUX4* PAS. Interestingly, in two independent FSHD1 clonal cell lines we observed different steady-state *DUX4* mRNA level reduction upon editing (Figure 3A). Since we cannot correlate this outcome to the initial *DUX4* expression levels, to the nucleotide edit at the *DUX4* PAS, or to the methylation levels at the targeted region, this outcome may be a

reflection of its suspected role as metastable epiallele, as the chromatin environment has also been suggested to influence PAS usage efficiency.<sup>60,61</sup> Such individualistic response will require further studies to elucidate its mechanism and to be able to predict the benefit of this approach for FSHD patients.

In addition, the study by Joubert et al. reported the use of either paired transcription activator-like effector nucleases (TALENs) or paired CRISPR/Cas9 nucleases to excise the *DUX4* PAS sequence with the aim of incorporating a mir-1 sequence by HDR in immortalized myoblasts.<sup>32</sup> This approach yielded only 2 successfully edited clones out of 227 (0.8%). In contrast, with our approach we achieved 30/163 successfully edited immortalized myoblast clones (nearly 20%) across five different experiments including three different FSHD cell lines (Table S5). Nevertheless, despite the limited number of successfully edited clones in the Joubert study, they also observed reduced, but not abolished, *DUX4* and *DUX4* target gene levels and a switch in the *DUX4* mRNA cleavage and polyadenylation site, which corroborates our findings. The increased editing efficiency in our study could

be explained by the fact that adenine base editors act independently of the HDR pathway, a pathway that is only available in S and G2 phases of the cell cycle. This cell cycle-independent feature of the ABE system makes it a viable candidate for its future *in vivo* translatability. The main bottleneck for adenine editing efficiency may therefore very well be the optimal delivery of editing components to skeletal muscle tissue.

One of the main concerns for the use of genome editing platforms is their potential off-target effect. Adenine base editors have been shown to suffer from sgRNA-dependent off-target DNA editing, albeit to a lesser extent than cytosine base editors.<sup>62</sup> In this study, we detected at least three sites that were edited in an sgRNA-dependent fashion but to a much lesser extent than the intended site. We observed ~23-fold more efficient editing at the A<sub>4</sub> position of the on-target site, i.e., 40% as assessed by Sanger sequencing in gDNA samples that were used also for the inspection of off-target editing in HAP1 cells, as compared to the most efficiently edited off-target site (OT5, 1.7%) as assessed by Illumina short read sequencing. Additionally, off-target editing of cellular RNAs by adenine base editors has been reported.<sup>63</sup> However, we have not explored this particular side effect of nSpABEmax. In any case, both DNA and RNA off-target activity of adenine base editors can be minimized by making use of further engineered adenine deaminases<sup>55,64,65</sup> linked to higher-fidelity Cas9 versions<sup>66–68</sup> and modified sgRNAs<sup>69</sup> and by reducing exposure time and/or effector molecule concentrations by employing different delivery strategies such as in the form of ribonucleoprotein particles.<sup>55,70</sup> The specificity of the adenine base editing approach for *DUX4* PAS targeting should therefore be carefully evaluated to ensure safety in case of its therapeutic application.

Base editors have been already used to achieve efficient gene silencing by targeting *cis*-regulatory elements important for proper gene expression by either introducing in-frame stop codons,<sup>71,72</sup> mutating a start codon,<sup>73</sup> or disrupting splice sites.<sup>74,75</sup> Since deviations from the canonical PAS hexamers generally reduce their cleavage and polyadenylation efficiency,<sup>76</sup> we explored how many polyadenylation signals genome-wide would be amenable for such an editing approach. We focused on the two most widely used hexameric motifs, namely AATAAA and ATTAAA, as they constitute ~80% of all identified polyadenylation signals (Figure S5A). These PAS motifs can be disrupted with adenine base editors either by modifying any of the adenines of the last three nucleotide positions of the PAS motif on the coding strand or alternatively by targeting the adenine on the non-coding strand that pairs with the middle thymine on the coding strand, leading to its substitution with a cytosine (Figure S5B). Based on these criteria, we established that ~25% of all PASs with either AATAAA or ATTAAA motifs are editable with nSpABEmax (Figure S5C). However, it should be pointed out that weakening the core PAS motif might not always lead to the expected transcriptional downregulation, since other *cis* auxiliary elements are known to influence the efficiency of PAS usage.<sup>77</sup> Moreover, alternative polyadenylation is widespread for genes that contain multiple functional PASs;<sup>78</sup> therefore invalidating only one of them might not be suffi-

cient to achieve an overall desired level of silencing. Rather, since alternative polyadenylation is tissue specific and globally regulated, PAS editing might represent a more refined tool for gene editing in some conditions. Therefore, the utility of this approach requires locus-specific validation. Nevertheless, due to challenging gene structure, *DUX4* represents an excellent candidate for adenine base editing-mediated mutagenesis of its PAS as a means for its expression interference.

## MATERIALS AND METHODS

### Cloning

To create the all-in-one base editing vector pX458-ABE7.10, overlapping PCR products of the TadA dimer from pCMV-ABE7.10 (Addgene #102919), nCas9-SV40 NLS from pX335 (Addgene #42335), and T2A-GFP from pX458 were cloned in pX458 using the AgeI and EcoRI restriction sites. The pX458-ABEmax vector was created by cutting out the TadA dimer together with the N-terminal domain of Cas9 from pX458-ABE7.10 using the AgeI and ApaI sites and replacing it with the PCR-amplified TadA dimer missing the N-terminal domain of Cas9 from the pCMV-ABEmax-GFP vector (Addgene #112101). The pX601-SaABEmax vector was cloned by first creating a new insert consisting of the TadA dimer linked to the N-terminal domain of SaCas9. This was achieved by overlapping PCR amplifications on pCMV-ABEmax (for the TadA dimer) and pX601 (for the SaCas9 domain) during which a D10A mutation was introduced into SaCas9. The resulting PCR product was cloned in pX601 using the XbaI and HindIII sites. The pX601-CjABEmax was created by first mutating the KpnI site upstream of the CAG promoter in the pX601-SaABEmax vector by replacing it with the same PCR fragment containing a KpnI mutation and cloned using XbaI and AgeI. Next, the SaABEmax-T2A-GFP-bGH insert was replaced by CjABEmax-T2A-GFP-bGH, which was produced by overlapping PCRs on pX601-SaABEmax for TadA dimer, pX404 (Addgene #68338) for CjCas9 (D8A mutation was introduced during this PCR step), and pX601-SaABEmax for T2A-GFP-bGH PAS. The final insert was cloned into pX601-SaABEmax via the AgeI and KpnI sites. Further, the SaCas9 sgRNA expression cassette was replaced with an CjCas9 sgRNA expression cassette. The CjCas9 sgRNA expression cassette was assembled by overlapping PCRs on pX601 to amplify the U6 promoter sequence and on the pU6-Cj-sgRNA plasmid (Addgene #89753) to amplify the sgRNA scaffold. The resulting insert was cloned into the pX601-CjABEmax plasmid created in the previous step via the KpnI and NotI sites. All sgRNAs were cloned into their target vector according to the Zhang lab's protocol.<sup>79</sup> For the pX458 vector (Addgene #48138) and its adenine base editor derivatives (SpABE7.10 and SpABEmax), the BbsI sites were used, and for the pX601 vector's derivatives (SaABEmax and CjABEmax) the BsaI sites were used. For optimal transcription from the U6 promoter, an extra G nucleotide was added to the 5' end of the sgRNA in case the sequence did not start with one already. All constructs were verified by Sanger sequencing. All primers used are listed in Table S1. The following restriction enzymes were used for cloning: AgeI-HF (New England Biolabs, ##R3552), EcoRI (Thermo Fisher Scientific, #ER0271), ApaI (New England Biolabs, #R0114), HindIII (New



England Biolabs, #R0104), KpnI-HF (New England Biolabs, #R3142), NotI-HF (New England Biolabs, #R3189), BbsI (New England Biolabs, #R3539), and BsaI (Thermo Fisher Scientific, #ER0291).

### Cell culture and transfection

Detailed information about cell lines used in this study can be found in Table S2. The HAP1 cell line was maintained in IMDM-GlutaMAX (Thermo Fisher Scientific, #31980) supplemented with 10% fetal bovine serum (FBS) (Gibco, #10270106) and 1% (v/v) penicillin-streptomycin (Gibco, #15140). Immortalized myoblast cell lines 073iMB (FSHD1<sup>8U</sup>) and 200iMB (FSHD2) were a kind gift from Prof. S. Tapscott, Fred Hutchinson Cancer Research Center. The 2402iMB line (FSHD1<sup>3U</sup>) was obtained by immortalizing primary myoblasts, which were a kind gift of Prof. R. Tawil from the University of Rochester, by stable integration of hTERT and CDK4 retroviruses as described previously.<sup>80</sup> All initial primary myoblast lines originated from the Fields Center for FSHD and Neuromuscular Research at the University of Rochester Medical Center and were obtained following the informed consent after the study had been approved by the relevant institutional review board. All myogenic lines were maintained in Ham's F-10 Nutrient Mix (Gibco, #31550) supplemented with 20% (v/v) FBS, 1% (v/v) penicillin-streptomycin, 10 ng/mL fibroblast growth factor (FGF)-b (PromoKine, #C-60240), and 1  $\mu$ M dexamethasone (Sigma-Aldrich, #D2915). Myogenic differentiation was achieved by switching myoblasts at 100% confluency to DMEM (Gibco, #31966021) supplemented with 2% (v/v) KnockOut serum replacement (Gibco, #10828028). All cell lines were maintained at 37°C and 5% CO<sub>2</sub> and were tested for *Mycoplasma* contamination with the MycoAlert Mycoplasma detection kit (Lonza, #LT07-318) according to the vendor's instructions. One day prior to transfection, 2  $\times$  10<sup>5</sup> HAP1 cells were seeded in a 12-well plate. Transfection was performed with 1.5  $\mu$ g of the base editing vector and 0.5  $\mu$ g of a vector containing puromycin resistance cassette (AA19\_pLKO.1-puro.U6.sgRNA.BveI-stuffer plasmid, a kind gift from Prof. M.A.F.V. Gonçalves, Leiden University Medical Center) using Lipofectamine 3000 (Thermo Fisher Scientific, #L3000008) according to the manufacturer's instructions. The next day, the medium was replaced with medium containing 0.5  $\mu$ g/mL of puromycin, and cells were selected for 48 h, after which the medium was replaced again with non-puromycin medium and cells were grown for an additional 72 h, after which they were harvested for subsequent analysis. For myoblast experiments, 3  $\times$  10<sup>5</sup> myoblasts were seeded in a 6-well plate, and the following day cells were transfected with 2  $\mu$ g of plasmid DNA using Lipofectamine 3000 (Thermo Fisher Scientific, #L3000008) according to the manufacturer's instructions. Medium was changed the next day, and cells were harvested for further analysis 72 h after transfection.

### T7E1 cleavage assay

CRISPR/Cas9-induced indels at the targeted locus were examined with the T7E1 cleavage assay. Three days after transfection, cells were harvested in lysis buffer for genomic DNA (100 mM Tris-HCl pH 8.0, 50 mM EDTA pH 8.0, 2% [w/v] SDS) and DNA was extracted by protein precipitation by addition of saturated salt to the solution

and subsequent isopropanol precipitation. The target locus was amplified by PCR using DreamTaq (Thermo Fisher Scientific, #EP0701) with the following cycling conditions: 95°C for 5 min followed by 35 cycles of 95°C for 25 s, 67°C for 25 s, and 72°C for 20 s, with a final extension step at 72°C for 5 min. Resulting PCR products were subjected to reannealing in a thermal cycler with the following conditions: 95°C for 5 min followed by cooling down from 95°C to 85°C at 2°C/s and from 85°C to 25°C at 0.1°C/s. After reannealing, 10  $\mu$ L of PCR product was incubated with T7E1 enzyme (New England Biolabs, #E3321) according to the manufacturer's instructions. Resulting products were resolved on a 2% Tris-borate-EDTA (TBE) agarose gel with ethidium bromide.

### Fluorescence-activated cell sorting

Cells were trypsinized, collected in their respective culturing media, and spun down, and the cell pellet was resuspended in fluorescence-activated cell sorting (FACS) buffer (10% [v/v] FBS in PBS). Cells were sorted with a BD FACS Aria III cell sorter according to GFP fluorescence, and collected cells were used for further analysis or expansion.

### DUX4 PAS genotyping and quantification of base editing efficiency

Exon 3 of *DUX4* containing the PAS was amplified from genomic DNA by PCR as described in the T7E1 cleavage assay. The product's purity was first assessed by an electrophoretic separation on a 2% TBE agarose gel and then extracted from the gel with the NucleoSpin Gel and PCR Clean-up kit (Bioké, #740609) and submitted for Sanger sequencing with the forward primer used in the PCR. Base editing efficiency in the initial test in HAP1 cells was assessed by Sanger sequencing and estimated with Edit-R<sup>39</sup> (online tool available at <http://baseeditr.com/>).

### RNA isolation, cDNA synthesis, and qRT-PCR

Cells were harvested in QIAzol lysis reagent (QIAGEN, #79306), and RNA was isolated with the RNeasy mini kit (QIAGEN, #74101) with DNase I treatment according to the manufacturer's protocol. Oligo-dT-primed cDNA was synthesized from 2  $\mu$ g of input RNA with the Minus First Strand cDNA synthesis kit (Thermo Fisher Scientific, #K1621). Gene expression was measured with the CFX384 system (Bio-Rad) in technical triplicates using iQ SYBR Green Supermix (Bio-Rad, #1708887). qRT-PCR primers are listed in Table S1. *GUSB* was used as a housekeeping gene.

### 3' RACE

The 3' RACE was carried out as reported previously<sup>31</sup> with minor modifications. The cDNA synthesis was carried out with the Minus First Strand cDNA synthesis kit with modified oligo-dT primer: 5'-GCTGTCAACGATACGCTACGTAACGGCATGACAGTGTTTTTTTTTTTTTTTTTTTTTT-3'. The first PCR was performed using 2  $\mu$ L of cDNA as template in a final volume of 20  $\mu$ L with AccuPrime Taq high-fidelity DNA polymerase (Thermo Fisher Scientific, #1236086) with previously published forward and reverse primers and according to established PCR cycling conditions.<sup>31</sup> Nested PCR

was performed using 2  $\mu$ L of primary PCR product with AccuPrime *Taq* high-fidelity DNA polymerase with previously published forward and reverse primers and according to established PCR cycling conditions.<sup>31</sup> Final PCR products were purified from 2% TBE agarose gel and subcloned into the TOPO-TA vector (Thermo Fisher Scientific, #450641). At least 6–8 individual bacterial colonies were screened to determine the *DUX4* mRNA 3' ends.

#### Methylation analysis of *DUX4* exon 3 (FasPAS region) by bisulfite PCR followed by TOPO-TA subcloning

500 ng of genomic DNA was converted with the EZ DNA Methylation-Lightning kit (Zymo Research, #D5030) according to the manufacturer's protocol. The FasPAS region was amplified from converted DNA with previously published primers (Table S1) using high-fidelity AccuPrime *Taq* DNA polymerase (Thermo Fisher Scientific, #12346086) with the following PCR program: 95°C for 4 min followed by 35 cycles of 95°C for 4 min, 58°C for 20 s, and 72°C for 40 s, followed by a final extension step at 72°C for 5 min. PCR products were purified by electrophoresis and isolated from gel with the NucleoSpin Gel & PCR Clean-up kit (Bioké, #740609) followed by subcloning into the TOPO-TA vector. Plasmid DNA from individual bacterial colonies was sent for Sanger sequencing using the M13R primer, and methylation levels were assessed with BiQ Analyzer software. Methylation lollipop plots were produced with the online QUMA tool (<http://quma.cdb.riken.jp/top/index.html>).

#### sgRNA-dependent off-target analysis using targeted next generation sequencing

Potential off-target sites were predicted by CRISPOR (<http://crispor.tefor.net/crispor.py>).<sup>48</sup> Ten predicted off-target sites were chosen based on the MIT specificity score and uniqueness of the region for specific amplification. Genomic regions of interest were amplified with specific primers containing appropriate Illumina forward and reverse adaptor sequences (Table S1). For the first PCR, 100 ng of genomic DNA was used as starting material in a 25  $\mu$ L reaction further containing 0.4  $\mu$ M of forward and reverse primer and 12.5  $\mu$ L of 2 $\times$  KAPA HiFi HotStart ReadyMix (Kapa Biosystems, #KK2601). PCR reactions were carried out as follows: 95°C for 3 min followed by 27 cycles of 98°C for 20 s, 64°C for 15 s, and 72°C for 15 s, with a final extension step at 72°C for 3 min. This first PCR product was purified with AMPure beads (Beckman Coulter, #A63881) with a 0.8 PCR-to-beads ratio according to the manufacturer's instructions, and DNA was eluted in 10  $\mu$ L of EB buffer. A subsequent barcoding PCR was performed in a total volume of 25  $\mu$ L using 3  $\mu$ L of purified first PCR product, 2  $\mu$ L of Illumina barcoding primer mix, and 12.5  $\mu$ L of 2 $\times$  KAPA HiFi HotStart ReadyMix. The barcoding PCR was carried out as follows: 95°C for 3 min followed by 7 cycles of 98°C for 20 s, 60°C for 20 s, and 72°C for 20 s, with a final extension step at 72°C for 3 min. PCR products were purified with AMPure beads in a 0.8 PCR-to-beads ratio according to manufacturer's instructions, and DNA was eluted to 10  $\mu$ L of EB buffer. The concentration of the final purified amplicons was measured with Qubit, and all amplicons were pooled in equimolar ratio and sequenced on an Illumina MiSeq instrument. Paired-end

reads were evaluated for mutations by alignment to the provided predicted off-target sequence using CRISPResso2<sup>49</sup> (CRISPRessoBatch\_batch\_settings 'my\_tab\_separated\_batchfile'–amplicon\_seq 'my\_reference\_sequence'–base\_edit -g 'my\_sgrna\_sequence' -wc -10 -w 20). The effect of intronic mutations on gene splicing was predicted with Alamut Visual software (Interactive Biosoftware, Rouen, France, version 2.15).

#### Genome-wide detection of editable polyadenylation signals

In order to find all editable polyadenylation signals in the genome with an AATAAA or ATTAAA motif, we constructed a regular expression that combines the polyadenylation signal motif sequence with a PAM site for SpCas9 (5'-NGG-3') at appropriate distance from the targeted base so that it falls into the reported activity window of nSpABEmax.<sup>54</sup> This regular expression was used to find all matching patterns in the human reference genome GRCh38 (<https://hgdownload.cse.ucsc.edu/goldenPath/hg38/bigZips/>). A similar approach was used to find all occurrences on the reverse complement strand. The results of this search were intersected with a list of known polyadenylation signals ([ftp://ftp.ebi.ac.uk/pub/databases/genecode/Gencode\\_human/release\\_35/genecode.v35.polyAs.gff3.gz](ftp://ftp.ebi.ac.uk/pub/databases/genecode/Gencode_human/release_35/genecode.v35.polyAs.gff3.gz)) to obtain the final list of editable polyadenylation signals. We used the 'famotif2bed' subcommand of the Fastools (<https://fastools.readthedocs.io/en/latest/>) package (version 1.0.2) for finding patterns in a reference sequence using regular expressions. All genome arithmetic was done using bedtools (<https://bedtools.readthedocs.io/en/latest/>) (version 2.27.1). The full procedure is available online (<https://github.com/jfjlaros/motif-edit>) under the MIT Open Source license.

#### Statistical methods

GraphPad Prism software v.8.4.2 was used for calculation of statistics. Sample sizes were not pre-determined prior to experiments, and a concrete statistical test is stated in the respective figure legend.

#### Data availability

The sequencing data generated for the off-target editing evaluation are available at BioProject: PRJNA732823.

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtn.2021.05.020>.

#### ACKNOWLEDGMENTS

We thank members of the van der Maarel group for all their helpful suggestions. This research was supported by funds from Friends of FSH Research, from the Prinses Beatrix Spierfonds (W.OP14-01 and W.OR21-04), and from Spieren voor Spieren. D.Š., J.B., and S.M.M. are members of the European Reference Network for Rare Neuromuscular Diseases [ERN EURO-NMD].

#### AUTHOR CONTRIBUTIONS

D.Š. designed and performed experiments, analyzed results, and wrote the manuscript. V.A.C. cloned and tested SpABEmax, SaABEmax, and CjABEmax in HAP1 cells. Y.A. and D.Š. performed and

analyzed deep sequencing for off-target regions. J.F.J.L. and D.Š. performed the genome-scale analysis of editable polyadenylation signals. J.B. immortalized 2,402 primary myoblasts and edited the manuscript. S.M.M. provided feedback and co-wrote the manuscript. All authors contributed to manuscript review.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES

- Wohlgemuth, M., Lemmers, R.J., Jonker, M., van der Kooij, E., Horlings, C.G., van Engelen, B.G., van der Maarel, S.M., Padberg, G.W., and Voermans, N.C. (2018). A family-based study into penetrance in facioscapulohumeral muscular dystrophy type 1. *Neurology* 91, e444–e454.
- Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., and Upadhyaya, M. (1995). Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum. Mol. Genet.* 4, 951–958.
- Statland, J.M., Donlin-Smith, C.M., Tapscott, S.J., Lemmers, R.J.L.F., van der Maarel, S.M., and Tawil, R. (2015). Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. *Neurology* 85, 2147–2150.
- Snider, L., Geng, L.N., Lemmers, R.J.L.F., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., and Miller, D.G. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* 6, e1001181.
- Gannon, O.M., Merida de Long, L., and Saunders, N.A. (2016). DUX4 Is Derepressed in Late-Differentiating Keratinocytes in Conjunction with Loss of H3K9me3 Epigenetic Repression. *J. Invest. Dermatol.* 136, 1299–1302.
- Das, S., and Chadwick, B.P. (2016). Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS One* 11, e0160022.
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* 49, 941–945.
- Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934.
- Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet.* 9, e1003947.
- Bosnakovski, D., Gearhart, M.D., Toso, E.A., Ener, E.T., Choi, S.H., and Kyba, M. (2018). Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci. Rep.* 8, 16957.
- Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., and Tapscott, S.J. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38–51.
- Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2014). Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet. Muscle* 4, 12.
- Sacconi, S., Briand-Suleau, A., Gros, M., Baudoin, C., Lemmers, R.J.L.F., Rondeau, S., Lagha, N., Nigumann, P., Cambieri, C., Puma, A., et al. (2019). FSHD1 and FSHD2 form a disease continuum. *Neurology* 92, e2273–e2285.
- Sacconi, S., Lemmers, R.J.L.F., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am. J. Hum. Genet.* 93, 744–751.
- van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037–2042.
- van den Boogaard, M.L., Lemmers, R.J.L.F., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F.P., Kriek, M., et al. (2016). Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am. J. Hum. Genet.* 98, 1020–1029.
- Lemmers, R.J.L.F., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W.E., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370–1374.
- Hamanaka, K., Šikrová, D., Mitsuhashi, S., Masuda, H., Sekiguchi, Y., Sugiyama, A., Shibuya, K., Lemmers, R.J.L.F., Goossens, R., Ogawa, M., et al. (2020). Homozygous nonsense variant in *LRIF1* associated with facioscapulohumeral muscular dystrophy. *Neurology* 94, e2441–e2447.
- Blewitt, M.E., Gendrel, A.-V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40, 663–669.
- Gendrel, A.-V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). SmcHD1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* 23, 265–279.
- Gendrel, A.-V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Grealley, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smcHD1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* 33, 3150–3165.
- Nozawa, R.-S., Nagao, K., Igami, K.-T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBIX1 pathway. *Nat. Struct. Mol. Biol.* 20, 566–573.
- Brideau, N.J., Coker, H., Gendrel, A.-V., Siebert, C.A., Bezstarosti, K., Demmers, J., Poot, R.A., Nesterova, T.B., and Brockdorff, N. (2015). Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. *Mol. Cell. Biol.* 35, 4053–4068.
- van Geel, M., Dickson, M.C., Beck, A.F., Bolland, D.J., Frants, R.R., van der Maarel, S.M., de Jong, P.J., and Hewitt, J.E. (2002). Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* 79, 210–217.
- Lemmers, R.J.L.F., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 884–894.
- Lemmers, R.J.L.F., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.-J.B., Frants, R.R., and van der Maarel, S.M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* 32, 235–236.
- Lemmers, R.J.L.F., van der Vliet, P.J., Klooster, R., Sacconi, S., Camaño, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650–1653.
- Lemmers, R.J.L.F., van der Vliet, P.J., Blatnik, A., Balog, J., Zidar, J., Henderson, D., Goselink, R., Tapscott, S.J., Voermans, N.C., Tawil, R., et al. (2021). Chromosome 10q-linked FSHD identifies DUX4 as principal disease gene. *J. Med. Genet.* Published online January 12, 2021. <https://doi.org/10.1136/jmedgenet-2020-107041>.
- Chen, J.C., King, O.D., Zhang, Y., Clayton, N.P., Spencer, C., Wentworth, B.M., Emerson, C.P., Jr., and Wagner, K.R. (2016). Morpholino-mediated Knockdown of DUX4 Toward Facioscapulohumeral Muscular Dystrophy Therapeutics. *Mol. Ther.* 24, 1405–1411.
- Anseau, E., Vanderplanck, C., Wauters, A., Harper, S.Q., Coppée, F., and Belayew, A. (2017). Antisense Oligonucleotides Used to Target the DUX4 mRNA as Therapeutic Approaches in Facioscapulohumeral Muscular Dystrophy (FSHD). *Genes (Basel)* 8, 93.

31. Marsollier, A.-C., Ciszewski, L., Mariot, V., Popplewell, L., Voit, T., Dickson, G., and Dumonceaux, J. (2016). Antisense targeting of 3' end elements involved in *DUX4* mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: a new gene-silencing approach. *Hum. Mol. Genet.* 25, 1468–1478.
32. Joubert, R., Mariot, V., Charpentier, M., Concordet, J.P., and Dumonceaux, J. (2020). Gene Editing Targeting the *DUX4* Polyadenylation Signal: A Therapy for FSHD? *J. Pers. Med.* 11, 7.
33. Rashnonejad, A., Amini-Chermahini, G., Taylor, N.K., Wein, N., and Harper, S.Q. (2020). Designed U7 snRNAs inhibit *DUX4* expression and improve FSHD-associated outcomes in *DUX4* overexpressing cells and FSHD patient myotubes. *Mol. Ther. Nucleic Acids* 23, 476–486.
34. Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 6, 914–929.
35. 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.
36. Lim, C.K.W., Gapinske, M., Brooks, A.K., Woods, W.S., Powell, J.E., Zeballos, C.M.A., Winter, J., Perez-Pinera, P., and Gaj, T. (2020). Treatment of a Mouse Model of ALS by In Vivo Base Editing. *Mol. Ther.* 28, 1177–1189.
37. Levy, J.M., Yeh, W.H., Pendse, N., Davis, J.R., Hennessey, E., Butcher, R., Koblan, L.W., Comander, J., Liu, Q., and Liu, D.R. (2020). Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.* 4, 97–110.
38. Ryu, S.-M., Koo, T., Kim, K., Lim, K., Baek, G., Kim, S.-T., Kim, H.S., Kim, D., Lee, H., Chung, E., and Kim, J.-S. (2018). Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* 36, 536–539.
39. Kluesner, M.G., Nedveck, D.A., Lahr, W.S., Garbe, J.R., Abrahamte, J.E., Webber, B.R., and Moriarity, B.S. (2018). EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* 1, 239–250.
40. Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., and Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* 36, 843–846.
41. Lemmers, R.J.L.F., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2015). Inter-individual differences in CpG methylation at *D4Z4* correlate with clinical variability in FSHD1 and FSHD2. *Hum. Mol. Genet.* 24, 659–669.
42. Ricci, G., Mele, F., Govi, M., Ruggiero, L., Sera, F., Vercelli, L., Bettio, C., Santoro, L., Mongini, T., Villa, L., et al. (2020). Large genotype-phenotype study in carriers of *D4Z4* borderline alleles provides guidance for facioscapulohumeral muscular dystrophy diagnosis. *Sci. Rep.* 10, 21648.
43. Lemmers, R.J., van der Vliet, P.J., Balog, J., Goeman, J.J., Arindrarto, W., Krom, Y.D., Straasheijm, K.R., Debipersad, R.D., Özel, G., Sowden, J., et al. (2018). Deep characterization of a common *D4Z4* variant identifies biallelic *DUX4* expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* 26, 94–106.
44. Yao, Z., Snider, L., Balog, J., Lemmers, R.J.L.F., Van Der Maarel, S.M., Tawil, R., and Tapscoff, S.J. (2014). *DUX4*-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum. Mol. Genet.* 23, 5342–5352.
45. Balog, J., Thijssen, P.E., Shadle, S., Straasheijm, K.R., van der Vliet, P.J., Krom, Y.D., van den Boogaard, M.L., de Jong, A., F Lemmers, R.J., Tawil, R., et al. (2015). Increased *DUX4* expression during muscle differentiation correlates with decreased SMCHD1 protein levels at *D4Z4*. *Epigenetics* 10, 1133–1142.
46. Tawil, R. (2008). Facioscapulohumeral muscular dystrophy. *Neurotherapeutics* 5, 601–606.
47. Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., and Whitelaw, E. (2002). Metastable epialleles in mammals. *Trends Genet.* 18, 348–351.
48. Concordet, J.P., and Haessler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46 (W1), W242–W245.
49. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* 37, 224–226.
50. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.
51. Cohen, J., DeSimone, A., Lek, M., and Lek, A. (2021). Therapeutic Approaches in Facioscapulohumeral Muscular Dystrophy. *Trends Mol. Med.* 27, 123–137.
52. Bouwman, L.F., van der Maarel, S.M., and de Greef, J.C. (2020). The prospects of targeting *DUX4* in facioscapulohumeral muscular dystrophy. *Curr. Opin. Neurol.* 33, 635–640.
53. Himeda, C.L., Jones, T.I., and Jones, P.L. (2020). Targeted epigenetic repression by CRISPR/dSaCas9 suppresses pathogenic *DUX4-f1* expression in FSHD. *Mol. Ther. Methods Clin. Dev.* 20, 298–311.
54. Huang, T.P., Zhao, K.T., Miller, S.M., Gaudelli, N.M., Oakes, B.L., Fellmann, C., Savage, D.F., and Liu, D.R. (2019). Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* 37, 626–631.
55. Richter, M.F., Zhao, K.T., Eton, E., Lapinaite, A., Newby, G.A., Thuronyi, B.W., Wilson, C., Koblan, L.W., Zeng, J., Bauer, D.E., et al. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 38, 883–891.
56. Li, X., Qian, X., Wang, B., Xia, Y., Zheng, Y., Du, L., Xu, D., Xing, D., DePinho, R.A., and Lu, Z. (2020). Programmable base editing of mutated TERT promoter inhibits brain tumour growth. *Nat. Cell Biol.* 22, 282–288.
57. Koblan, L.W., Erdos, M.R., Wilson, C., Cabral, W.A., Levy, J.M., Xiong, Z.M., Tavarez, U.L., Davison, L.M., Gete, Y.G., Mao, X., et al. (2021). In vivo base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature* 589, 608–614.
58. Yamada, M., Watanabe, Y., Gootenberg, J.S., Hirano, H., Ran, F.A., Nakane, T., Ishitani, R., Zhang, F., Nishimasu, H., and Nureki, O. (2017). Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol. Cell* 65, 1109–1121.e3.
59. Kim, E., Koo, T., Park, S.W., Kim, D., Kim, K., Cho, H.Y., Song, D.W., Lee, K.J., Jung, M.H., Kim, S., et al. (2017). In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.* 8, 14500.
60. Zhang, J., Zhang, Y.Z., Jiang, J., and Duan, C.G. (2020). The Crosstalk Between Epigenetic Mechanisms and Alternative RNA Processing Regulation. *Front. Genet.* 11, 998.
61. Nanavaty, V., Abrash, E.W., Hong, C., Park, S., Fink, E.E., Li, Z., Sweet, T.J., Bhasin, J.M., Singuri, S., Lee, B.H., et al. (2020). DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. *Mol. Cell* 78, 752–764.e6.
62. Xin, H., Wan, T., and Ping, Y. (2019). Off-Targeting of Base Editors: BE3 but not ABE induces substantial off-target single nucleotide variants. *Signal Transduct. Target. Ther.* 4, 9. <https://doi.org/10.1038/s41392-019-0044-y>.
63. Grünwald, J., Zhou, R., Garcia, S.P., Iyer, S., Lareau, C.A., Aryee, M.J., and Joung, J.K. (2019). Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* 569, 433–437.
64. Grünwald, J., Zhou, R., Iyer, S., Lareau, C.A., Garcia, S.P., Aryee, M.J., and Joung, J.K. (2019). CRISPR DNA base editors with reduced RNA off-target and self-editing activities. *Nat. Biotechnol.* 37, 1041–1048.
65. Zhou, C., Sun, Y., Yan, R., Liu, Y., Zuo, E., Gu, C., Han, L., Wei, Y., Hu, X., Zeng, R., et al. (2019). Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* 571, 275–278.
66. Lee, J.K., Jeong, E., Lee, J., Jung, M., Shin, E., Kim, Y.-H., Lee, K., Jung, I., Kim, D., Kim, S., and Kim, J.S. (2018). Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* 9, 3048.
67. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529, 490–495.
68. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351, 84–88.



69. Ryan, D.E., Taussig, D., Steinfeld, I., Phadnis, S.M., Lunstad, B.D., Singh, M., Vuong, X., Okochi, K.D., McCaffrey, R., Olesiak, M., et al. (2018). Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs. *Nucleic Acids Res.* *46*, 792–803.
70. Rees, H.A., Komor, A.C., Yeh, W.-H., Caetano-Lopes, J., Warman, M., Edge, A.S.B., and Liu, D.R. (2017). Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* *8*, 15790.
71. Billon, P., Bryant, E.E., Joseph, S.A., Nambiar, T.S., Hayward, S.B., Rothstein, R., and Ciccia, A. (2017). CRISPR-Mediated Base Editing Enables Efficient Disruption of Eukaryotic Genes through Induction of STOP Codons. *Mol. Cell* *67*, 1068–1079.e4.
72. Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., and Adli, M. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* *14*, 710–712.
73. Wang, X., Liu, Z., Li, G., Dang, L., Huang, S., He, L., Ma, Y., Li, C., Liu, M., Yang, G., et al. (2020). Efficient Gene Silencing by Adenine Base Editor-Mediated Start Codon Mutation. *Mol. Ther.* *28*, 431–440.
74. Li, Z., Xiong, X., Wang, F., Liang, J., and Li, J.F. (2019). Gene disruption through base editing-induced messenger RNA missplicing in plants. *New Phytol.* *222*, 1139–1148.
75. Gapinske, M., Luu, A., Winter, J., Woods, W.S., Kostan, K.A., Shiva, N., Song, J.S., and Perez-Pinera, P. (2018). CRISPR-SKIP: programmable gene splicing with single base editors. *Genome Biol.* *19*, 107.
76. Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* *18*, 5799–5805.
77. Nunes, N.M., Li, W., Tian, B., and Furger, A. (2010). A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.* *29*, 1523–1536.
78. Beadoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res.* *10*, 1001–1010.
79. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* *8*, 2281–2308.
80. Zhu, C.H., Mouly, V., Cooper, R.N., Mamchaoui, K., Bigot, A., Shay, J.W., Di Santo, J.P., Butler-Browne, G.S., and Wright, W.E. (2007). Cellular senescence in human myoblasts is overcome by human telomerase reverse transcriptase and cyclin-dependent kinase 4: consequences in aging muscle and therapeutic strategies for muscular dystrophies. *Aging Cell* *6*, 515–523.

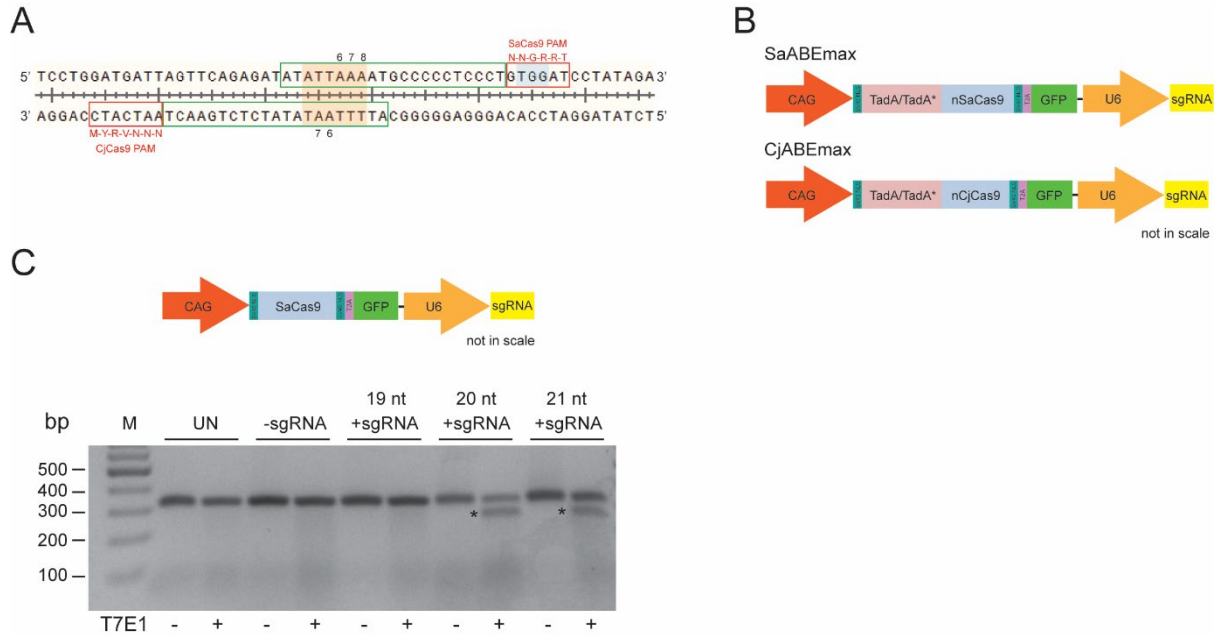
OMTN, Volume 25

## Supplemental information

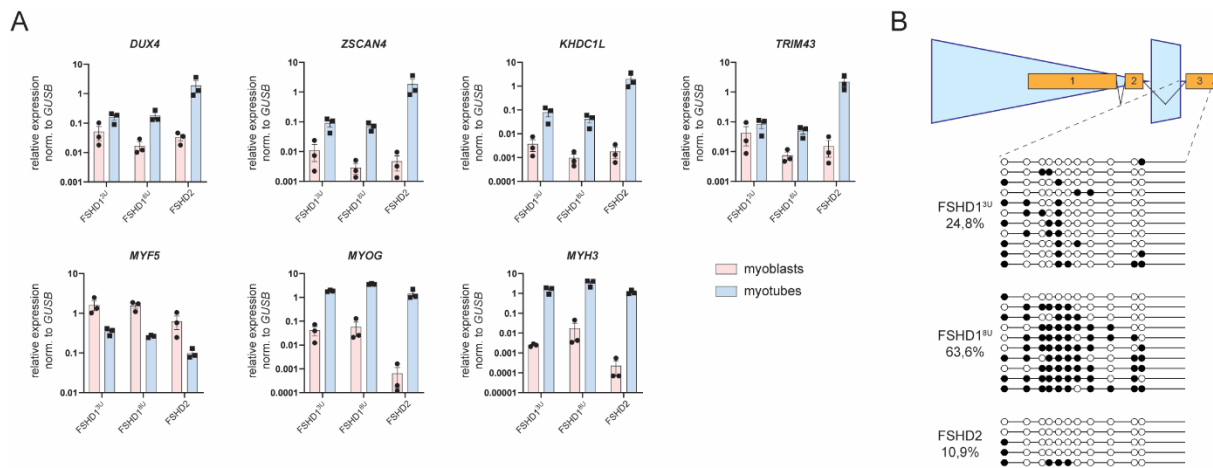
### **Adenine base editing of the *DUX4* polyadenylation signal for targeted genetic therapy in facioscapulohumeral muscular dystrophy**

**Darina Šikrová, Vlad A. Cadar, Yavuz Ariyurek, Jeroen F.J. Laros, Judit Balog, and Silvère M. van der Maarel**

## Supplemental Materials



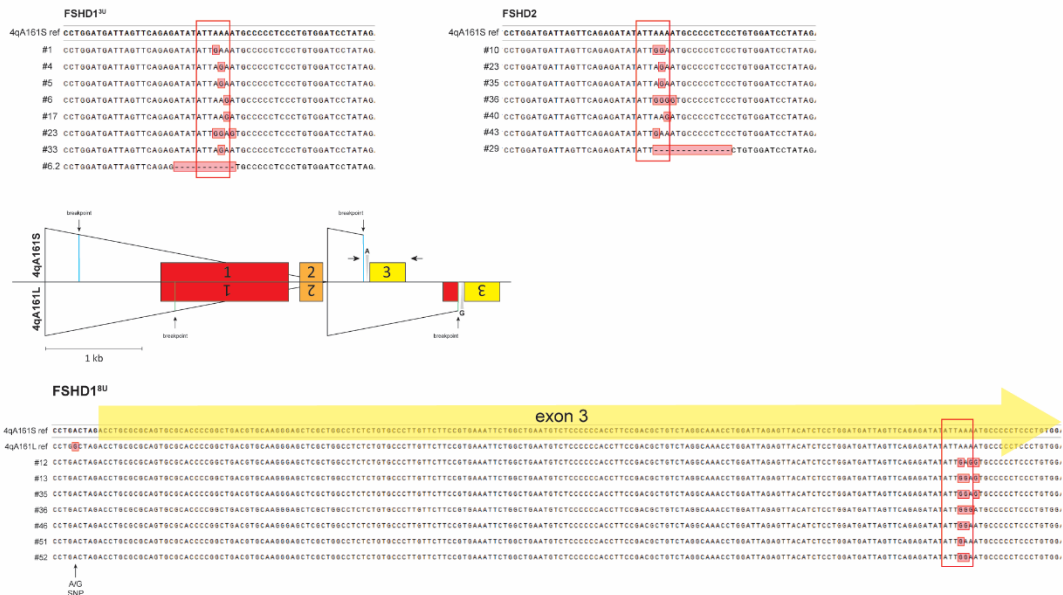
**Suppl. Figure 1. *DUX4* PAS is not editable either by nSaABEmax or by nCjABEmax. A)** DNA sequence surrounding of *DUX4* PAS (highlighted in the orange box). Cognate PAM sites for CjCas9 and SaCas9 are outlined in red rectangles, while sgRNA protospacer regions are outlined in green rectangles. The PAM site for SpCas9 is highlighted in the blue box. Adenines within the *DUX4* PAS amenable for base editing are numbered from the beginning of the sgRNA protospacer. **B)** Schematic maps of modified all-in-one vectors coding for nSaABEmax (top) and nCjABEmax (bottom). **C)** Schematic map of the pX601 vector for simultaneous sgRNA and SaCas9 nuclease expression (top). Result of the T7E1 assay performed on HAP1 cells which were transfected with a pX601 vector expressing the SaCas9 nuclease and sgRNAs of different length (19 nt-, 20 nt- or 21 nt-long) targeting the *DUX4* PAS (bottom). Untransfected cells (UN) or cells transfected with no sgRNA containing vector (-sgRNA) served as negative control. Asterisks mark the T7E1 cleavage products.



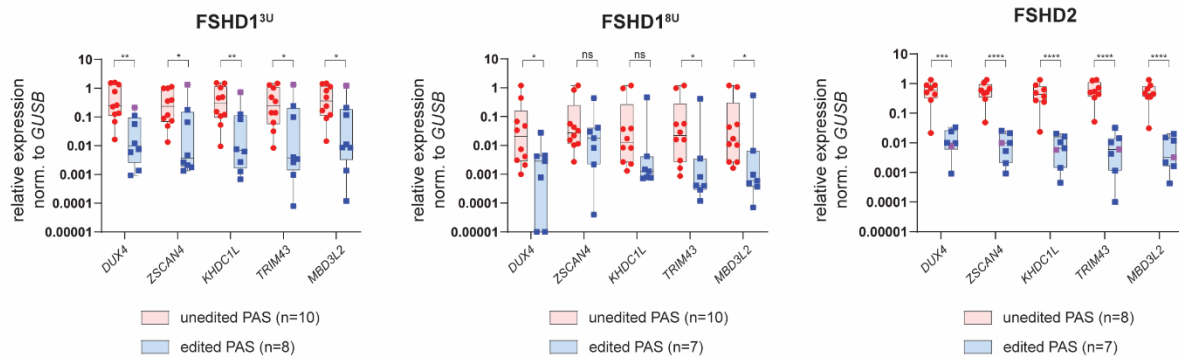
**Suppl. Figure 2. Expression and methylation profiles of FSHD immortalized myoblasts used for editing.** **A)** mRNA expression levels of *DUX4* and four *DUX4* target genes in 3 model FSHD cell lines used for base editing experiments at myoblast and myotubes stage. Expression of myogenic markers (*MYOG* and *MYH3*) is provided to show successful myogenic differentiation. *GUSB* was used as a housekeeping gene. Bars represent mean  $\pm$ SEM. Cells were grown three independent times and analysed for their gene expression. **B)** CpG methylation level of the FasPAS region encompassing exon 3 of *DUX4* in the three parental FSHD immortalized myoblast lines used for base editing. Individual rows represent a single molecule, empty circles denote unmethylated cytosines in a CpG context, while full circles denote methylated cytosines in a CpG context. Average methylation of the region (in %) is provided below the name of each sample. Note, that in case of FSHD1<sup>8U</sup>, both alleles (contracted and non-contracted allele) are amplified in bisulfite PCR.



A



B

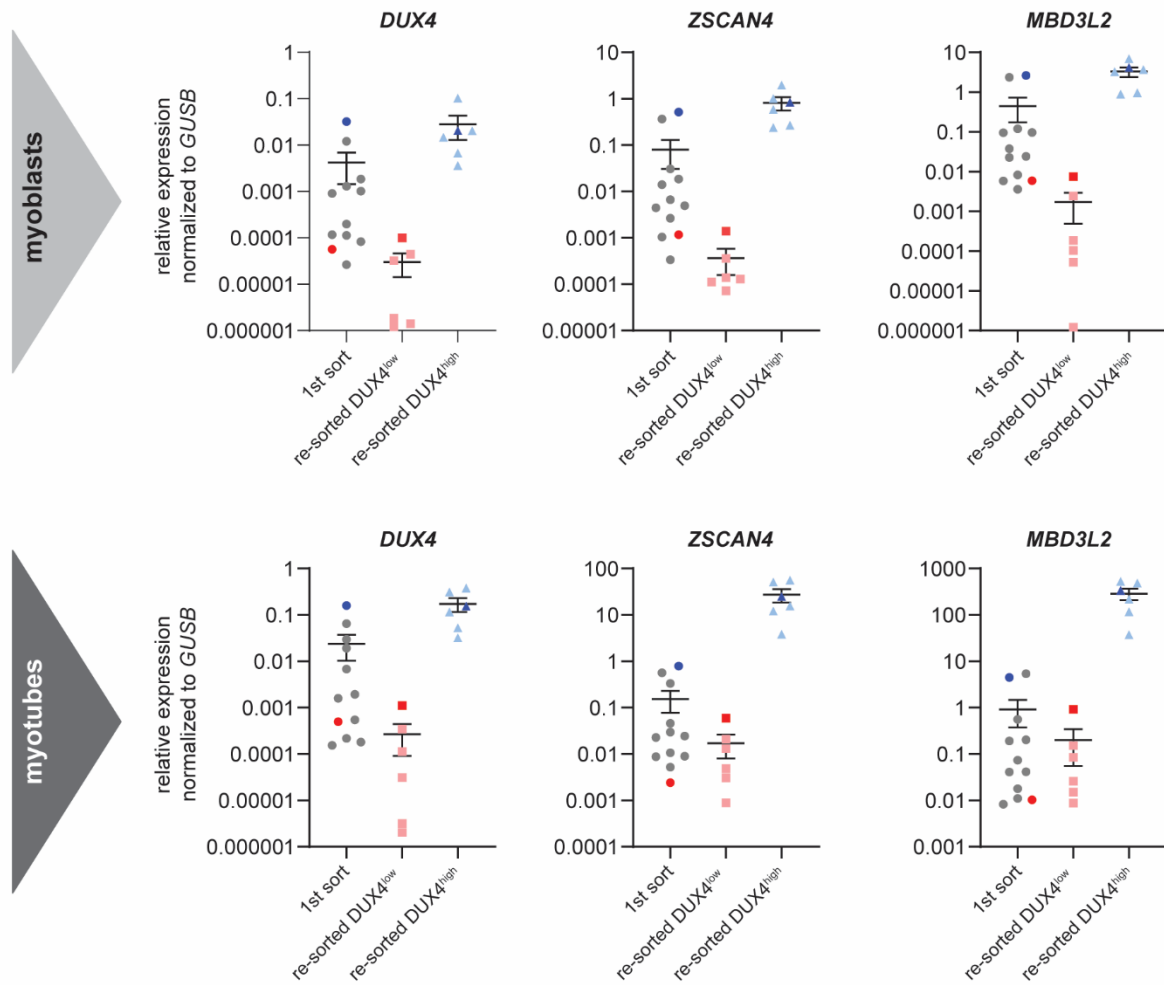


**Suppl. Figure 3. Genotypes of successfully edited clones and their expression data in proliferating myoblasts.**

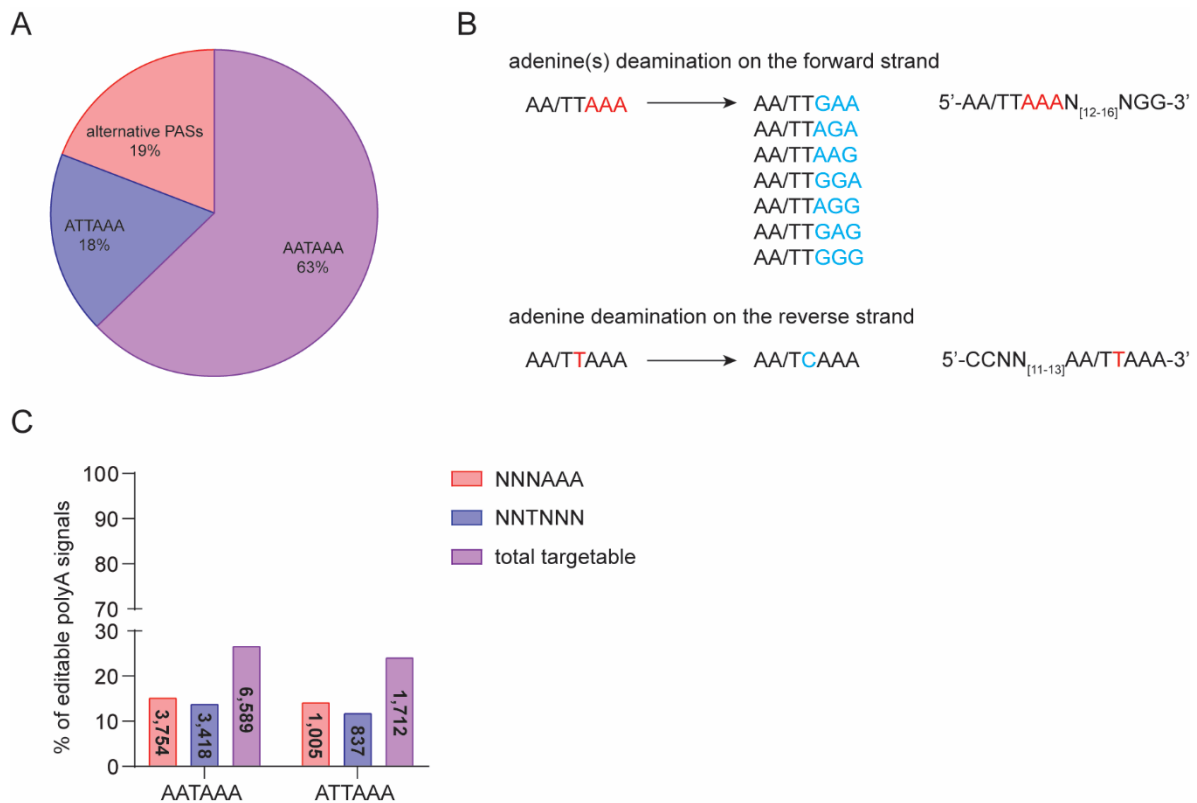
**A)** Genotypes of successfully *DUX4* edited clones obtained from three independent FSHD lines (top left: FSHD1<sup>3U</sup>, top right: FSHD2 and bottom: FSHD1<sup>8U</sup>) aligned to the WT reference sequence. The *DUX4* PAS sequence is highlighted in a red rectangle and red colored bases denote mismatches. Mirror schematic of the 4qA161S and 4qA161L D4Z4 haplotype termini is provided to show the genotyping approach for the FSHD1<sup>8U</sup> cell line. Red box represents exon 1, orange box represents exon 2 and yellow box represents exon 3 which corresponds to the yellow arrow highlighting the exon 3 sequence in the genotyping tracks. In the 4qA161L haplotype, a small 5' part of exon 1 (red box) precedes exon 3 due to a different breakpoint. A specific forward primer was used to selectively amplify the 4qA161S allele, which was confirmed by the presence of the SNP (A instead of G) in the Sanger sequencing tracks (marked by arrow) for all the genotyped clones. Reference

sequence for both, 4qA161S and 4qA161L allele is provided. B) mRNA levels as assessed by RT-qPCR of *DUX4* and four *DUX4* target genes (*MBD3L2*, *ZSCAN4*, *TRIM43* and *KHDC1L*) in PAS unedited vs edited clones derived from two FSHD1 and one FSHD2 cell lines during proliferation. Statistical significance was calculated with unpaired two-tailed t-test (ns: non-significant, \*: <0.05, \*\*: <0.01, \*\*\*: <0.001, \*\*\*\*: <0.0001) on log2 transformed expression values to correct for skewed distribution. Expression values normalized to *GUSB* as house-keeping gene are plotted. Line represent mean and whiskers represent min and max value. Individual data points represent individual clones, two violet clones carry a deletion over *DUX4* PAS.

A



**Suppl. Figure 4. *DUX4* expression signature is clonally stable. A)** mRNA levels as assessed by RT-qPCR of *DUX4* and its two target genes (*ZSCAN4* and *MBD3L2*) were measured in clonal lines established from FSHD1<sup>8U</sup> immortalized myoblasts and 5 new daughter clones derived from a parental clone with either high *DUX4* expression (dark blue colour) or low *DUX4* expression (dark red colour). Daughter clones are marked by light blue or light red colour. Expression data for both myoblasts (top) and myotubes (bottom) are provided.



**Suppl. Figure 5. A) Identification of polyadenylation signals targetable by nSpABEmax in the human genome. A)** Genome-wide prevalence of most common polyadenylation signal hexamers based on all annotated polyadenylation signals in Gencode. **B)** Representation of two different approaches of targeting polyadenylation signals by nSpABEmax either on the coding or non-coding strand and their possible outcomes. Targeted positions are in red and expected modified bases are in blue. **C)** Percentage of annotated polyadenylation signals in the GRCh38 human genome with the most prevalent motifs (AATAAA and ATTAAA) whose adenines are targetable by nSpABEmax either on the coding (red) or non-coding (blue) strand. The number within each bar represents the actual number of targetable polyadenylation signals.



**Supplemental Table 1.** Oligos used in the study. Table is available online as a separate excel sheet.

**Supplemental Table 2.** Detailed information about cell lines used in the study.

Cell Line (ID)	Cell Type	Clinical Status and <i>SMCHD1</i> Mutation Status	Sex	4q allele #1	4q allele #2
HAP1	derived from chronic myelogenous leukemia (KBM-7)	NA	M	25U 4qA161S	NA
2402 (FSHD3U)	immortalized myoblasts	FSHD1	M	3U 4qA161S	16U 4qB163
073 (FSHD8U)	immortalized myoblasts	FSHD1	M	7U 4qA161S	36U 4qA161L
200 (FSHD2)	immortalized myoblasts	FSHD2 (SMCHD1 p.Lys204Glu)	M	11U 4qA161S	39U 4qB168

**Supplemental Table 3.** Results of 3' RACE experiment. Calculation table is available online as a separate excel sheet.

**Supplemental Table 4.** Table of predicted off-target sites by nSpABEmax. Table is available online as a separate excel sheet.

**Supplemental Table 5.** Summary of efficiencies of deriving edited clones in immortalized myoblasts.

	# of post-FACS grown-out clones	# of edited clones	% efficiency (100*edited/all grown-out)
FSHD1 <sup>3U</sup> poly	40	8	20
FSHD1 <sup>8U</sup> poly	52	7	13,46153846
FSHD2 poly	46	7	15,2173913
FSHD1 <sup>3U</sup> mono	8	5	62,5
FSHD1 <sup>8U</sup> mono	17	3	17,64705882
<b>total</b>	163	30	18,40490798