

ORIGINAL ARTICLE

Mouse *Dux* is myotoxic and shares partial functional homology with its human paralog *DUX4*

Jocelyn O. Eidahl^{1,†}, Carlee R. Giesige^{1,2,†}, Jacqueline S. Domire¹,
Lindsay M. Wallace¹, Allison M. Fowler¹, Susan M. Guckes¹,
Sara E. Garwick-Coppens¹, Paul Labhart³ and Scott Q. Harper^{1,2,4,*}

¹Center for Gene Therapy, The Research Institute at Nationwide Children's Hospital, Columbus, OH, USA,

²Biomedical Sciences Graduate Program, The Ohio State University, Columbus, OH, USA, ³Active Motif,

Carlsbad, CA, USA and ⁴Department of Pediatrics, The Ohio State University, Columbus, OH, USA

*To whom correspondence should be addressed at: Scott Q. Harper, Center for Gene Therapy, The Research Institute at Nationwide Children's Hospital, 700 Children's Drive, Room WA3015, Columbus, OH 43205, USA. Tel: (614) 355-2893; Email: scott.harper@nationwidechildrens.org

Abstract

D4Z4 repeats are present in at least 11 different mammalian species, including humans and mice. Each repeat contains an open reading frame encoding a double homeodomain (*DUX*) family transcription factor. Aberrant expression of the D4Z4 ORF called *DUX4* is associated with the pathogenesis of Facioscapulohumeral muscular dystrophy (FSHD). *DUX4* is toxic to numerous cell types of different species, and over-expression caused dysmorphism and developmental arrest in frogs and zebrafish, embryonic lethality in transgenic mice, and lesions in mouse muscle. Because *DUX4* is a primate-specific gene, questions have been raised about the biological relevance of over-expressing it in non-primate models, as *DUX4* toxicity could be related to non-specific cellular stress induced by over-expressing a *DUX* family transcription factor in organisms that did not co-evolve its regulated transcriptional networks. We assessed toxic phenotypes of *DUX* family genes, including *DUX4*, *DUX1*, *DUX5*, *DUXA*, *DUX4-s*, *Dux-bl* and mouse *Dux*. We found that *DUX* proteins were not universally toxic, and only the mouse *Dux* gene caused similar toxic phenotypes as human *DUX4*. Using RNA-seq, we found that 80% of genes upregulated by *Dux* were similarly increased in *DUX4*-expressing cells. Moreover, 43% of *Dux*-responsive genes contained ChIP-seq binding sites for both *Dux* and *DUX4*, and both proteins had similar consensus binding site sequences. These results suggested *DUX4* and *Dux* may regulate some common pathways, and despite diverging from a common progenitor under different selective pressures for millions of years, the two genes maintain partial functional homology.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant disorder affecting 1 in 8,333 to 1 in 20,000 individuals worldwide (1,2). FSHD is characterized by weakness of facial, shoulder and limb muscles although presentation is often non-uniform.

Variability can be seen in the types of muscles affected, severity of weakness, age-at-onset, rates of progression and the involvement of asymmetrical or bilateral phenotypes (3,4). FSHD is associated with aberrant expression of the toxic *DUX4* gene, which encodes a double homeodomain transcription factor, embedded within an array of D4Z4 repeats located on the chromosome 4q subtelomere.

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Received: December 11, 2015. Revised: August 19, 2016. Accepted: August 25, 2016

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The prevailing FSHD pathogenesis model states that 4q D4Z4 repeats are normally located in a transcriptionally silent region of heterochromatin, but in FSHD the 4q subtelomere chromatin structure is relaxed, allowing for DUX4 transcription (5–11). This shift to an euchromatin-like state can occur through contraction in D4Z4 repeat copy number (FSHD1) or mutation of the structural maintenance of chromosomes flexible hinge domain 1 (SMCHD1) gene (FSHD2), which functions to silence the 4q D4Z4 array (7,12–14). Although chromatin loosening and DUX4 transcription are required for FSHD development, these events are not pathogenic unless they occur on a specific chromosome 4q variant, which harbours an untranslated exon containing a polyadenylation signal for the most telomeric DUX4 open reading frame (ORF) (7,15–18). As a result, DUX4 is absent or detected at extremely low levels in non-FSHD muscle and other adult tissues except testes, but more abundantly expressed and polyadenylated in FSHD muscles (11,18–20).

Over-expression strategies have been used to study the role of DUX4 in FSHD, and it is clear that DUX4 is toxic to muscle and non-muscle cells of numerous organisms, including humans, mice, zebrafish and frogs (21–29). The conserved toxic properties of over-expressed DUX4 are consistent with the hypothesis of its causation in FSHD, and seem to justify the methodology and the model systems used. On the other hand, the conservation of DUX4 toxicity across species also raised some concerns about the over-expression methodology. Specifically, D4Z4 repeats exist only in placental mammals, and within this group, close DUX4 homologs are present only in primates (30,31). Thus, organisms that lack DUX4-like genes would not evolve with selective pressure to maintain DUX4-regulated networks, thereby suggesting that the toxic phenotypes associated with DUX4 over-expression in species lacking D4Z4 repeats might not be due to activation of some conserved deleterious pathways, but instead could be related to non-specific effects of protein overload in general, or more specifically related to over-expression of a double homeodomain (DUX) family transcription factor family member, of which DUX4 is but one member.

To begin addressing these questions, in this study, we used identical methods to deliver and express seven DUX family members, including human DUX4, DUX1, DUX5, DUXA, DUX4-short (DUX4-s) and the mouse genes *Duxbl* and *Dux*, to human cells and mouse muscle, and assessed toxic phenotypes, or lack thereof, using several outcome measures. Among the genes tested, we found that human DUX4 and mouse *Dux* were similarly damaging while other homeodomain family members were non-toxic. These results were consistent with prior studies suggesting that *Dux* arises from D4Z4-like repeats in the mouse genome, and possesses DUX4-like toxic properties, including reducing C2C12 cell viability and/or myogenic differentiation, and causing morphological defects and death when expressed in *Xenopus laevis* embryos (32). Nevertheless, it was unclear if DUX4 and *Dux* function by activating similar toxic pathways. Thus, to further explore the mechanism of DUX4 and *Dux* toxicity, we performed ChIP-seq and RNA-seq to identify genomic regions commonly bound by the two proteins and define common gene expression changes. Our results suggest that DUX4 and mouse *Dux* share partial functional homology.

Results

DUX family transcription factors are not universally toxic in vitro and in vivo

DUX family transcription factors are classified based upon their two adjacent DNA binding homeodomains (HOX1 and HOX2).

DUX4 is the best-studied member of this group, largely because of its link to FSHD. Recently, DUX4 binding sites in the human genome were identified using ChIP-seq, some DUX4-activated genes have been validated, and structure-function studies showed that DUX4-induced toxicity was dependent upon DNA binding and transactivation activities (23,33,34). Although some studies of other DUX family members have been performed, the genes and biological pathways each controls are not well understood, and it is unclear if the toxicity associated with DUX4 over-expression is a universal property shared by the group. To address this initial question, in this study, we assessed the toxic properties of several human and mouse DUX family members in mouse cells and mouse muscle. These included mouse *Dux*, mouse *Duxbl*, human DUXA, human DUX1 and human DUX5, which share 33/56%, 45/63%, 45/57%, 81/95% and 81/95% amino acid similarity with their respective DUX4 double homeodomain regions (HOX1/HOX2), but lack significant alignment with DUX4 in the carboxy-terminal region located after the second homeodomain (Fig. 1). The one exception is a short stretch of sequence at the extreme carboxy-terminal end of DUX4 and mouse *Dux* (residues 411–424 of DUX4 and 661–674 of *Dux*), which are 78.6% similar (Fig. 1C). In addition, we tested DUX4-s, which is a smaller DUX4 isoform with identical HOX1 and HOX2 domains, but lacking the DUX4 transactivation domain.

We and others previously reported that DUX4 over-expression was toxic to immortalized mammalian cells, including human HEK293 fibroblasts, human TE671 rhabdomyosarcoma cells and mouse C2C12 myoblasts (21,23,24). We therefore used cell viability assays as an initial toxicity screen of DUX family members. Forty-eight hours after transfection of CMV expression plasmids containing the various DUX family members, we measured significant reduction of viability only in cells transfected with DUX4 and its mouse paralog *Dux* (mean reduction of 1.4×10^6 and 6.9×10^5 fold; p value < 0.05 , respectively), while cells treated with other DUX family members were not significantly affected. Furthermore, we mutagenized the first *Dux* DNA binding domain (HOX1) using an alanine substitution strategy similar to the one we previously reported for DUX4 (DUX4.HOX1 mutant), to generate an analogous *Dux.Hox1* mutant (23). We found that, like DUX4.HOX1, the *Dux.Hox1* DNA binding domain construct was non-toxic to C2C12 cells, suggesting DNA binding was essential for DUX4- and *Dux*-induced toxicity (Fig. 2A). Each protein was tagged at the carboxy-terminus with a V5-epitope. Western blots probed with HRP-coupled V5 antibodies confirmed expression of each DUX family protein in vitro. We noted that *Duxbl* was not expressed as highly as the other DUX family members, which could theoretically impact its potential toxicity. However, increasing *Duxbl* plasmid concentrations still failed to induce cell death (data not shown), and previous in vitro over-expression studies with *Duxbl* suggested it promoted cell proliferation and likewise did not stimulate cell death (Fig. 2B) (35). We performed identical experiments in HEK293s using several cell viability/cell death assays including quantifying caspase 3/7 activity and trypan blue exclusion counts. As we found previously, DUX4 consistently caused HEK293 cell death in every assay, while the impacts of mouse *Dux* on HEK293 cell viability were inconsistent and therefore not universally conclusive. No other DUX family member investigated showed toxicity in HEK293s (Supplementary Material, Fig. S1).

We next tested the effects of DUX family member over-expression in vivo. To do this, we delivered two doses (high and low; 3×10^{10} and 3×10^9 DNase Resistant Particles, DRP) of serotype 6 adeno-associated viral vectors (AAV6) carrying CMV

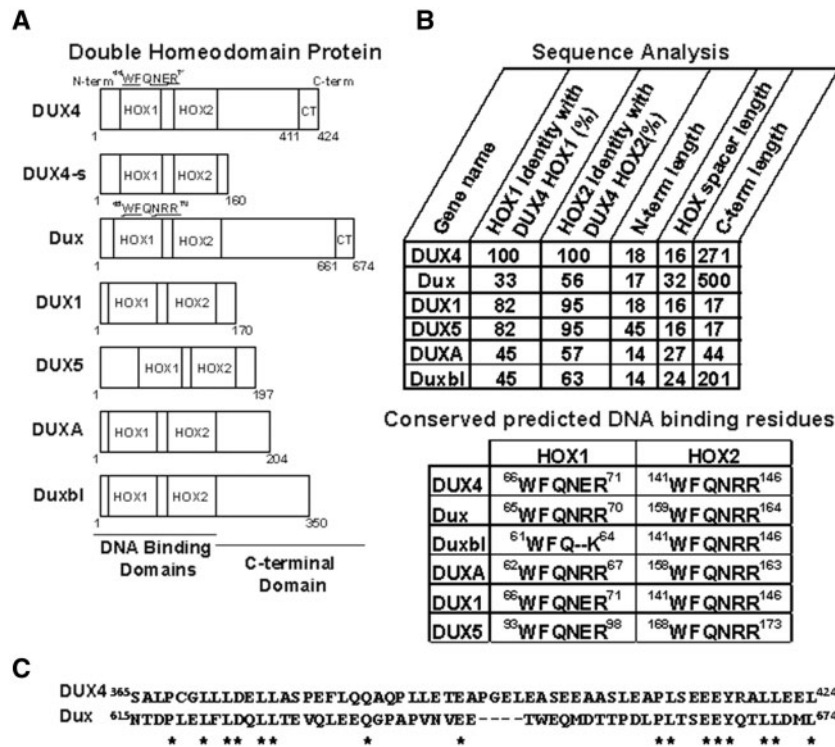


Figure 1 DUX family proteins used in this study. (A) Schematic of 7 DUX family proteins showing location of DNA binding homeodomains (HOX1 and HOX2) and variable carboxyl-terminal domains. Numbers indicate amino acid residues. In the DUX4.HOX1 and Dux.Hox1 DNA binding mutants, underlined residues were mutated to alanines. CT, conserved extreme C-terminus in DUX4 and Dux proteins, shown in (C). (B) (Top) Comparison of amino acid sequence homology in DUX family proteins. Columns 1 and 2 show amino acid identity in the HOX1 and HOX2 domains of indicated proteins, relative to the same domains in DUX4. 'N-term length' indicates the amino acid length from residue 1 to the beginning of the first homeodomain. 'C-term length' indicates the length from the end of the second homeodomain to C-terminus of the protein. 'HOX spacer length' indicates the number of amino acids between the first and second homeodomains. (Bottom) Alignment of conserved DNA binding residues within HOX1 and HOX2. In DUX4 and Dux, 5 of these were mutated to alanines to construct DUX4.HOX1 and Dux.Hox1, as shown in (A). (C) Alignment of extreme C-terminal residues of DUX4 (residues 365-424) and Dux (residues 615-674). Asterisks indicate residue identity. Note high identity in the last 14 amino acids.

promoter-driven DUX4, DUX4-s, Dux, Duxbl, DUXA, DUX1, DUX5 and our two DNA binding domain mutants (DUX4.HOX1 and Dux.Hox1) to mouse tibialis anterior (TA) muscle via intramuscular (IM) injections (Fig. 3A). As previously, each construct was V5-epitope tagged to enable detection with the same antibody. Using this same methodology, our lab previously showed that AAV.CMV.DUX4 vectors produced dose-dependent lesions that included histological evidence of mononuclear cell infiltration and myofibers with central nuclei, which is an indication of muscle degeneration and regeneration (23). Eight days, or 2 and 4 weeks after injection we harvested muscles and performed immunofluorescence and histological stains of fresh frozen cryosections. DAPI and V5 immunofluorescence verified protein expression and nuclear localization similar to DUX4 (Fig. 3B and C). Like our *in vitro* results in C2C12 cells, only DUX4- and Dux-expressing muscles showed histological evidence of damage, indicated by degenerating and regenerating myofibers, including the presence of myofibers with central nuclei, mononuclear cell infiltration and fibre size variability in hematoxylin and eosin (H&E)-stained cryosections (Fig. 3D). All other DUX family genes (DUX4-s, DUXA, DUX1, DUX5, Duxbl) lacked significant histopathological evidence of toxicity (Fig. 3B and C). We quantified histological changes at 4 weeks post-transduction and found that AAV.DUX4- and AAV.Dux-transduced fibres had significantly increased numbers of centralized nuclei ($P < 0.0001$, chi-square) (Fig. 3D). In addition, AAV.DUX4- and AAV.Dux-transduced fibres had reduced average fibre diameter size

(mean fibre diameter was decreased by 15% and 14.9% respectively compared to saline controls) (Fig. 3D). In comparison, DUX1, DUXA, Duxbl, DUX5, Dux.Hox1 and DUX4-s were similar to vehicle-treated animals (mean fibre size for each treatment was within 2% of saline-injected control fibre diameter mean) (Fig. 3D). Like DUX4, DNA binding was important for the toxic function of mouse Dux *in vivo*, as Dux.Hox1-injected mice were protected from Dux-induced pathology. Overall, these data suggested that DUX protein family members are not universally toxic, and only DUX4 and Dux caused muscle damage *in vivo*. Importantly, to ensure reproducibility and control for potential variability in AAV vector preparations, all *in vivo* assays were performed in two cohorts using two independent lots of AAV vectors for each construct.

Identification of human DUX4 and mouse Dux binding sites in human myoblasts

DUX4 and mouse Dux are transcription factors, but their normal biological roles within an organism and the mechanisms underlying their toxicity are unclear. We hypothesized that the toxic properties of the two proteins were directly related to the genes and pathways they activate, and their shared toxicity may be due to their ability to activate common gene targets. Thus, we sought to identify common DUX4- and Dux-controlled genes to provide clues about the mechanisms underlying their toxicity

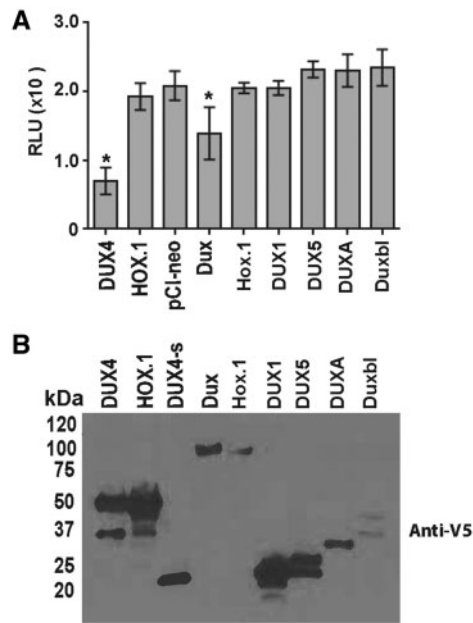


Figure 2 *In vitro* analysis of DUX family proteins. (A) Luminescence-based ATP assay. C2C12s were transfected with indicated plasmids and abundance of ATP was monitored 48 h later. Data are reported in relative luminescent units (RLU) with background removed (cells only), and each condition was performed in triplicate in two independent experiments. * indicates significant difference from "pCI-neo" empty vector control, $P < 0.05$, ANOVA. (B) Western blot. Protein extracts from cells transfected with expression plasmids encoding V5-epitope tagged DUX family ORFs were visualized using an HRP-coupled anti-V5 antibody. Predicted molecular weights of each protein: DUX4 and DUX4.HOX1, 52 kDa; DUX4s, 22 kDa; Dux and Dux.Hox1, 22 kDa; DUX1, 22 kDa; DUXA, 24 kDa; Duxbl, 38 kDa; DUX5, 22 kDa.

and/or normal function. To accomplish this, we transfected each gene into human myoblasts and performed chromatin immunoprecipitation followed by next generation sequencing (ChIP-Seq). We identified 4,009 (Dux) and 8,442 (DUX4) peaks (peak defined as the nucleotide start and end position at a genomic location) at a P-value threshold of $1e-6$ for Dux and DUX4, respectively, using the model-based analysis of ChIP-Seq (MACS) peak finding algorithm (Fig. 4A, Supplementary Material, Table S1) (36). We analysed the top 1,000 peak sequences using MEME/TOMTOM (<http://meme.nbcr.net/meme/>) to identify the consensus binding motifs for each protein (37). For DUX4, this was 5' – TAACCTAATCA – 3'; for mouse Dux, 5' – TGATTCAATCA – 3' (Fig. 4B, Supplementary Material, Table S2). We queried each motif to the JASPER database and found that our DUX4 consensus binding motif matched the DUX4 binding sequence identified in a previous ChIP-seq study (P-value $6.9e-7$) (33). This DUX4 sequence was the top match identified for Dux, as well (P-value $9.5e-6$).

Next, we mapped DUX4 and Dux enriched peaks from the ChIP-seq study to the human genome. Each peak was annotated to an associated gene if the interval location was within a set distance of 10,000 bp upstream or downstream of the gene transcriptional start site (TSS) (Supplementary Material, Fig. S2). A correlation analysis between all DUX4 and Dux peaks (percent of total peaks in each sample) revealed that the majority (96.2 and 92.0%, respectively) of DUX4- and Dux-bound sites were unique to each respective protein. We displayed the total merged peak regions as a scatter plot showing three separate populations, one corresponding to DUX4-specific peaks, a

second to Dux-specific peaks and third to peaks shared by both DUX4 and Dux (Fig. 4A, Supplementary Material, Fig. S3). The common peak population shared by the two proteins consisted of 319 binding sites (3.8% of all DUX4 sites; 8.0% of all Dux sites) annotated to 235 individual genes (Supplementary Material, Table S3). We next used "Find Individual Motif Occurrences" (FIMO) software to identify the consensus binding motifs for DUX4 and Dux within the set of 319 common ChIP-Seq binding sites (38). Of these sequences, 78% (250/319) contained consensus motifs for both DUX4 and Dux, while 18% contained only one type of consensus motif (29/319 DUX4 only; 27/319 Dux only) (Supplementary Materials, Fig. S4, Table S4). Thirteen of the 319 common binding sites (4%) contained no consensus binding sites for either protein. As an example, the previously validated DUX4 target gene, ZSCAN4, contained 6 DUX4 binding sites located at positions +474, +496, +513, +561, +583 and +600. Four of these six DUX4 consensus binding sites in ZSCAN4 also matched the Dux consensus binding motif (+474, +513, +561 and +600) (Supplementary Materials, Fig. S4, Table S4).

The presence of only the DUX4 consensus binding motif in genes bound by Dux in our ChIP-Seq experiment, and conversely, the presence of the Dux consensus binding motif in genes bound by DUX4, suggested both proteins were able to recognize either sequence. To confirm that DUX4 and Dux were capable of binding similar sequences, we co-transfected each DUX family protein into HEK293 cells along with a DUX4-responsive GFP reporter (39). We found that DUX4 and Dux, and not their DNA binding mutants, activated GFP expression while other DUX family proteins did not (Supplementary Material, Fig. S5) (39).

We also performed Ingenuity Pathway Analysis (IPA) to identify some broad categories of overlap between human DUX4- and mouse Dux-bound genes, as well as pathways unique to Dux alone (Supplementary Material, Table 5). Of the 235 genes annotated to the common binding sites, only 6 (ZSCAN4, GTF2F1, PALM, C19orf21/MISP, CLK1, PPIL3) were among the top 100 DUX4 and Dux annotated genes when sorted by ChIP-seq peak value (Fig. 4C). We performed a GO enrichment analysis on these 6 genes to identify their corresponding functions, which included transcriptional factor activity, kinase and isomerase activity, plasma membrane association and actin binding (Fig. 4C).

We next compared our annotated DUX4 binding sites in human myoblasts (referred to here as DUX4) with the raw data (FASTQ files) from the previously reported DUX4 ChIP-seq study (for distinction, we refer to this published dataset here as DUX4-G, where G = Geng) (33). We performed an identical analysis using our methodology, and identified 33,919 peaks for DUX4-G (whereas Geng, et al. reported 39,737 to 62,028 peaks using different analysis parameters). DUX4-G had ~4 times as many called peaks compared to our DUX4 sample, but 71% (6,020 of 8,442) were overlapping between the two samples (Fig. 4A). This represented 88% (3,385) of all genes in our DUX4 gene list, with overlapping network functions (Supplementary Material, Table 5). These results support good reproducibility between the two datasets. Similar to our DUX4 and Dux comparison, the correlation analysis between all DUX4-G and Dux peaks (percent of total peaks in a sample) revealed that the majority (98.8 and 90.0%, respectively) of DUX4-G- and Dux-controlled genes were unique to each respective protein and only 396 binding sites overlapped (1.2% of all DUX4-G sites; 9.9% of all Dux sites) in the ChIP-Seq experiments. The percentage of overlapping binding sites was greater between DUX4-G and Dux compared to DUX4 (9.9 versus 8.0%) likely due to the more abundant number of peaks detected by Geng et al. for DUX4.

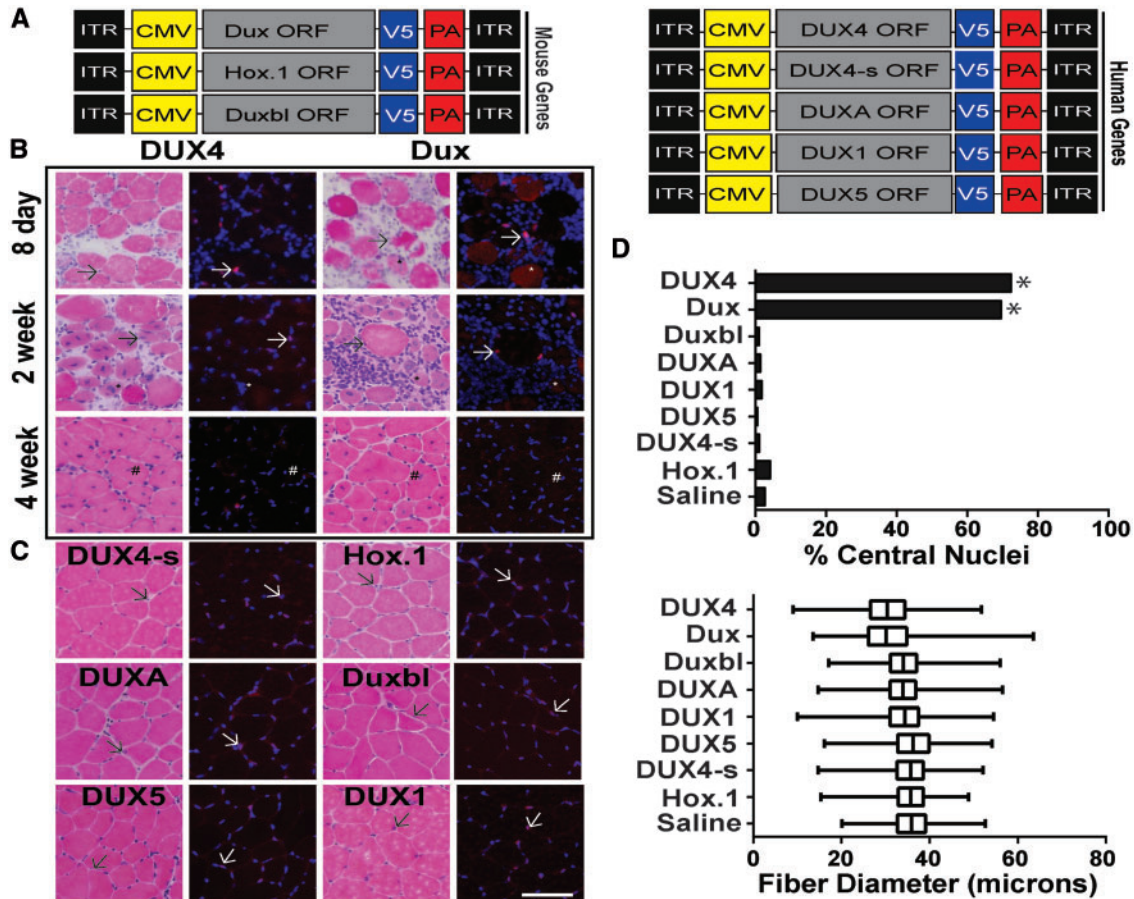


Figure 3 Dux and DUX4 cause myotoxicity *in vivo*. (A) Schematic of adeno-associated viral vector constructs (AAV6). Each DUX family ORF was driven by the cytomegalovirus (CMV) promoter and contained a carboxyl-terminal V5-epitope tag followed by an SV40 poly-adenylation signal (PA). Each construct was flanked by AAV2 inverted terminal repeats (ITR). (B) Photomicrographs of mouse muscle cryosections stained with hematoxylin and eosin (H&E) staining and anti-V5 immunofluorescence at indicated times following AAV.DUX4 or AAV.Dux delivery. (C) DUX4s, Dux.Hox1, DUXA, Duxbl, DUX5, and DUX1 are shown at 4 weeks post-injection. Arrows indicate positive V5-stained (red) nuclei. Blue, DAPI stain to identify nuclei. * denotes degenerating fibres positive for V5 staining and # show examples of regenerated myofibres with central nuclei. Scale bar = 100 μ m. (D) (Top) Percent central nuclei in muscle cryosections 4 weeks post-injection. Central nuclei from fibres of 5 independent animals were counted per treatment. * denote significantly increased numbers of centralized nuclei ($P < 0.0001$, chi-square) (Bottom) Distribution of fibre diameter (microns) as a percentage of total fibres counted during sampling at 4 weeks post-injection for each construct ($n = 5$ muscles per group; 5 representative x20 photomicrographs per section).

In an attempt to correlate ChIP-seq binding with gene activation, we initially used a candidate approach to validate some potential gene targets of DUX4 and Dux by qPCR. We first focused on expression of ZSCAN4, PRAMEF12, GTF2F1, SFRS8 (aka SFRS2B), NFYA, CWC15 and RFPL1 because these genes were bound by DUX4 and Dux in our ChIP-seq experiments and previously identified as DUX4-upregulated targets (33). We also measured levels of ANKRD1 (which had previously been shown to be significantly downregulated upon DUX4 over-expression), and TRAF6, which was bound by Dux but not DUX4 in our ChIP-seq study. To do this, we transfected human myoblasts with expression plasmids containing DUX4, DUX4.HOX1, DUX4-s, Dux, Dux.Hox1 or GFP, harvested RNA, and performed quantitative RT-PCR using primers/probes for each respective human gene. We found that ZSCAN4 and PRAMEF12, which have been used as DUX4-responsive biomarkers, were absent or very low in normal human myoblasts expressing GFP, DUX4.HOX1, DUX4-s, and Dux.Hox1, but significantly upregulated 34,772-fold and 5,790-fold in cells expressing DUX4, and 13,338-fold and 4,665-fold in cells expressing Dux, respectively (Fig. 4D, Supplementary Material, Table S6A). In contrast, GTF2F1, RFPL1, SFRS8, NFYA and

CWC15 were significantly upregulated only in DUX4 expressing cells (4.7-fold, 1685-fold, 5.9-fold, 5.1-fold and 5.4-fold, respectively) but not in Dux expressing cells despite the presence of Dux binding sites within each gene (Fig. 4D, Supplementary Material, Table S6A). Interestingly, ANKRD1 was significantly downregulated by DUX4, which was consistent with a prior study, but significantly upregulated in Dux expressing cells (33). Finally, TRAF6 was not significantly changed in response to any treatment (Fig. 4D, Supplementary Material, Table S6A).

In a second candidate gene approach to validate our ChIP-seq data with qPCR, we investigated potential gene expression changes in the P53 pathway. We focused on the P53 pathway because we and others previously showed that DUX4 toxicity operated at least partly through activation of P53-associated cell death (21–23,40). A total of 33 genes in the P53 pathway had binding sites for DUX4, Dux, or both in human myoblasts (Supplementary Material, Table 6B). Using similar methodology to the experiment in Figure 4, we performed Taqman qPCR assays on 29 of these genes containing DUX4 or Dux binding sites. None were significantly changed following Dux over-expression, and only 2 were increased upon DUX4 treatment (SIRT1, BRIP1), while one was decreased (NBA1).

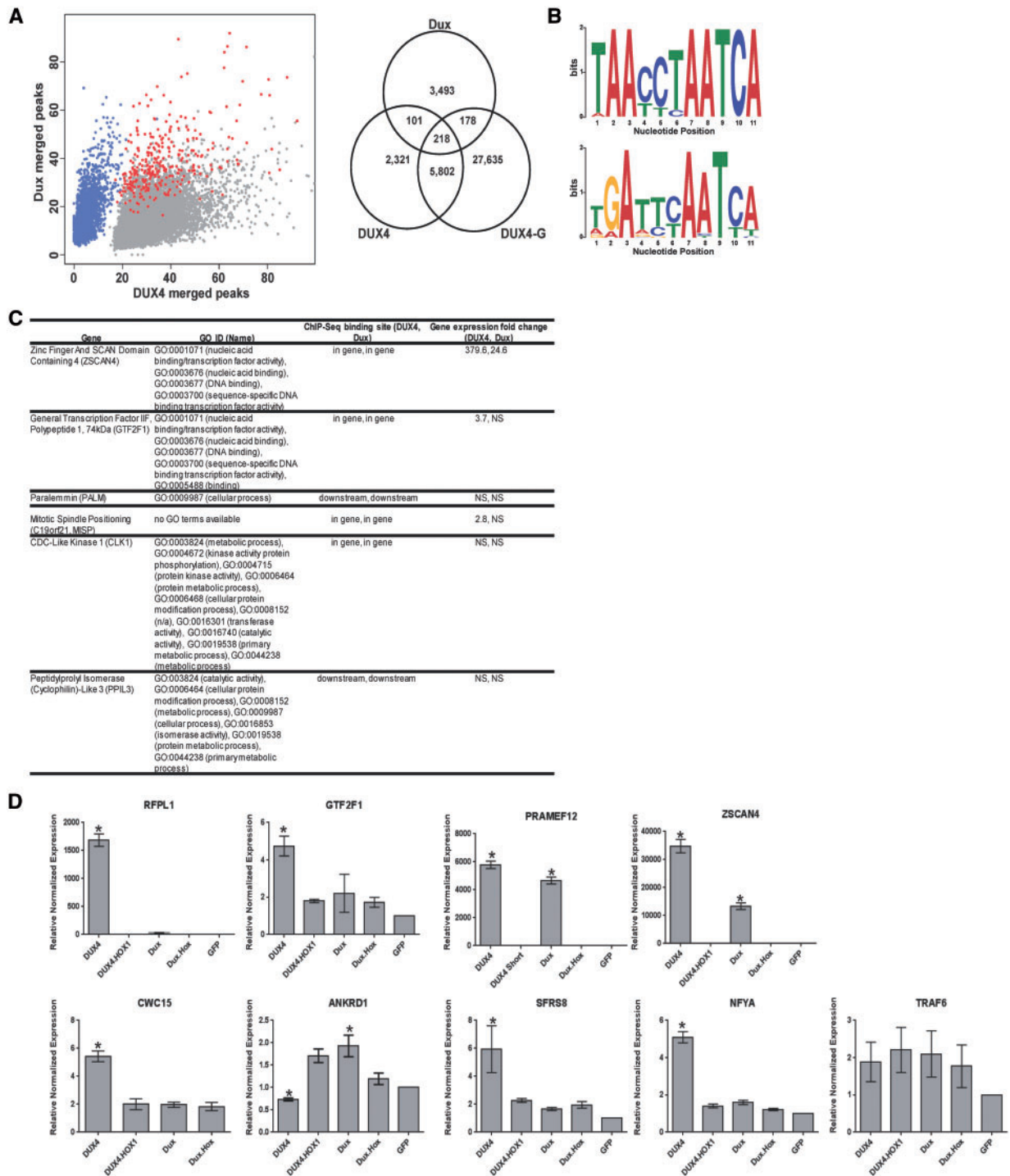


Figure 4 Characterization of DUX4 and Dux DNA binding sites in myoblasts. (A) (Left) Scatter plot of merged peak regions for DUX4 and Dux in human myoblasts. Gray dots correspond to DUX4 specific peaks, blue dots correspond to Dux specific peaks and red dots correspond to peaks in common between DUX4 and Dux. The XY axes denote the number of alignments (“tag”) within each of the peak regions. (Right) Venn diagram showing the overlap of peak regions for DUX4, Dux and DUX4-G in human myoblasts. DUX4-G refers to a published dataset of DUX4 ChIP-seq data by Geng, et al (33). Additional readout of the merged peak regions (heatmap and density plot) to visualize the unique peak populations for DUX4 and Dux can be found in [Supplementary Figure 3](#). (B) Conserved sequence motifs identified by MEME (<http://meme.nbcr.net/meme/>) in the DUX4 (top) and Dux (bottom) top 1,000 ChIP-Seq peak sequences (37). Nucleotides (measured in bits) of the identified consensus motif are displayed in a sequence logo representation (62). X-axis corresponds to the 11 nucleotide positions ([Supplementary Material, Table 2](#)). (C) Table listing genes annotated to DUX4 and Dux overlapping DNA binding sites and their corresponding Gene Ontology IDs. The peak values associated with these DNA binding sites are within the top 100 values for both DUX4 and Dux samples. Columns indicate genomic binding sites identified by ChIP-Seq and gene expression fold change identified by RNA-Seq. (D) Twenty-four hours after transfection of DUX4, DUX4.HOX1, Dux, Dux.Hox1, and GFP in human myoblasts, quantitative RT-PCR was performed to measure expression of the indicated genes containing ChIP-seq binding sites for DUX4 and Dux. Expression was normalized to GFP transfected cells with human RPL13A used as the reference gene, with means \pm SEM shown. * indicates significant difference from GFP control, $P < 0.05$, ANOVA. Each individual assay was performed in triplicate from $n = 3$ samples for each condition. In addition, ZSCAN4 and PRAMEF12 experiments were performed three independent times; RFPL1, GTF2F1 and SFRS8 experiments were performed two independent times; and ANKRD1, CWC15, TRAF6, SFRS8 and NFYA expression levels were determined in one experiment. The graphs here show representative data from one experiment each.

DUX4 and Dux upregulate common genes in human myoblasts

We hypothesized that DUX4 and Dux could potentially activate common genes and pathways despite the low abundance of common binding sites from our ChIP-seq experiments. Our candidate gene/QPCR approach for ChIP-seq validation proved laborious, and demonstrated that the presence of DUX4 or Dux ChIP-seq binding sites within or near a gene did not necessarily predict gene expression changes. We therefore turned to a high throughput method and performed RNA-Seq to identify unique and common gene expression changes in human myoblasts overexpressing either DUX4 or Dux, compared to pCI-Neo control-transfected cells. Compared to controls, 226 transcripts were significantly upregulated in human myoblasts expressing mouse Dux (>2-fold change; range 2- to 25-fold increase), while 3,244 transcripts were significantly increased in cells expressing human DUX4 (>2-fold increase; range 2- to 547- fold increase). This represented 14X more changes in human myoblasts expressing human DUX4 compared to mouse Dux. Importantly, DUX4 and Dux protein levels were relatively equivalent (Fig. 2). In addition, 104 transcripts were downregulated upon DUX4 expression (>2-fold decrease; range 2- to 62-fold decrease), but we found no significant decreases in any transcript upon Dux transfection. Of the 226 upregulated genes associated with mouse Dux expression, 181 (or 80%) were also upregulated in DUX4 expressing myoblasts (Fig. 5A, Supplementary Materials, Tables S7 and S8). IPA analysis identified some broad functional categories for shared genes upregulated by both DUX4 and Dux, including embryonic and organismal development, cell signaling, lipid metabolism, cell-mediated immune response, and cardiovascular system development (Supplementary Materials, Table S9).

We next compared our ChIP-seq and RNA-Seq results to identify upregulated gene targets common to both DUX4 and Dux. For DUX4, 316 of the 3,244 upregulated genes (9.7%) contained ChIP-seq binding sites. The same analysis for Dux revealed that 33 of 226 upregulated genes had Dux ChIP-seq binding sites (15%). Of these, 14 were in common between DUX4 and Dux, including the DUX4/FSHD biomarkers PRAMEF12 and ZSCAN4, and ERBB4, which can promote apoptosis upstream of P53 (Fig. 5B, Supplementary Materials, Table S3) (41,42). In addition to ERBB4, we also found that DUX4 and Dux both significantly upregulated GADD45G, which could contribute to P53 pathway activation by activating p38 and JNK pathways (43,44), although neither protein had binding sites located in or near the GADD45G gene, suggesting an indirect activation.

A recent report showed that DUX4 required at least two binding sites to induce gene expression of reporter plasmids (45). To address this, we generated a list of genes upregulated by DUX4 containing ChIP-seq binding sites, and a similar list for Dux. We then used FIMO software to identify consensus-binding sequences within the ChIP-Seq merged peaks corresponding to genes from these lists (Supplementary Materials, Table S4). We sorted the gene lists by fold change and number of consensus binding sites found within the peak sequences (Supplementary Material, Table S10). We did not detect an obvious correlation between fold change upregulation and quantity of consensus binding sites within the peak sequence.

Discussion

Several systems for studying DUX4 expression have emerged in recent years, but in general they can be divided into two

categories: those in which DUX4 arises from an endogenous D4Z4 locus, and DUX4 over-expression models (23,25–27,46–48). These “endogenous” models hold advantages as they allow the study of DUX4 from its natural cis-acting control elements, and recapitulate epigenetic phenotypes associated with FSHD. In contrast, they also present some difficulties, as DUX4 expression in these systems may be absent, very low, or alternatively, elevated but only in a small percentage of myonuclei (0.01 – 0.1%) (11,25,40,46–49). As a result, cytotoxic or myopathic phenotypes caused by DUX4 expression in “endogenous” systems may be completely lacking, very subtle, and/or inconsistent. Alternatively, over-expression methods permit more uniform and detectable DUX4 expression across a population of cells, in whole muscle groups, or throughout an organism. In addition, since elevated levels of DUX4 are considered a key component of FSHD pathogenesis, there is logic to modeling FSHD through DUX4 over-expression. Using this approach, numerous groups have shown that DUX4 is toxic to muscle and non-muscle cells of several vertebrate species, including zebrafish, frogs, mice and humans (20–22,24–28,32,33,46,50). For example, we found that DUX4 caused dose-dependent muscle lesions and weakness in adult mice transduced with AAV.DUX4 vectors (23). In contrast, a DUX4 DNA binding domain mutant (DUX4.HOX1) was benign when over-expressed in human cells, mouse muscle and developing zebrafish, suggesting that DUX4 toxicity was specifically related to its ability to bind DNA and activate downstream deleterious pathways (23). Taken together, these data could be interpreted in two opposing ways. On one hand, the conserved toxic properties of DUX4, as well as the absence of toxicity from the DUX4.HOX1 mutant, support the hypothesis of DUX4 causation in FSHD, and seemed to justify the use of over-expression methodology in human and non-human model systems alike. Moreover, these data suggested DUX4 toxicity was associated with its transcription factor function, and unrelated to non-specific protein overload within target cells, which has been reported for otherwise benign proteins such as GFP (51). On the other hand, DUX4 is a primate-specific gene, so a case could be made that the conservation of DUX4 toxicity in zebrafish, frogs, mice and humans would be unlikely to occur through similar mechanisms, and therefore suggested that data obtained from DUX4 over-expression studies should be cautiously interpreted for its specificity and relevance to FSHD. This interpretation of the current data is based on the notion that animals without DUX4 would lack selective pressure to maintain DUX4 binding sites within critical pathways, and would therefore not co-evolve DUX4-regulated networks.

If the toxic effects of DUX4 were related to its ability to trans-activate-specific genes, it is difficult to envision that DUX4 would control the same genes in primates as it would in non-primates, which lack DUX4. Nevertheless, over-expressing DUX4 would certainly impact gene expression in each of these organisms; the DUX4 consensus binding site is an 11 nucleotide (nt) motif occurring with a random frequency of 1 in 3,815,000, meaning that the zebrafish, mouse and primate genomes would contain about 393, 708 and 786 DUX4 consensus sites based upon chance alone. Although not all sites would impact gene expression - since transcription factors and their binding sites may be regulated by transcriptional co-factors, chromatin status and position near a promoter or enhancer region - we considered the possibility that the conserved toxic properties of DUX4 could be related to non-specific, promiscuous transcription caused by over-expressing a DUX family transcription factor in general. Our results did not support this hypothesis, as only mouse Dux showed toxic phenotypes similar to DUX4 (Figs. 2 and 3). Together, these data demonstrated that DUX family members

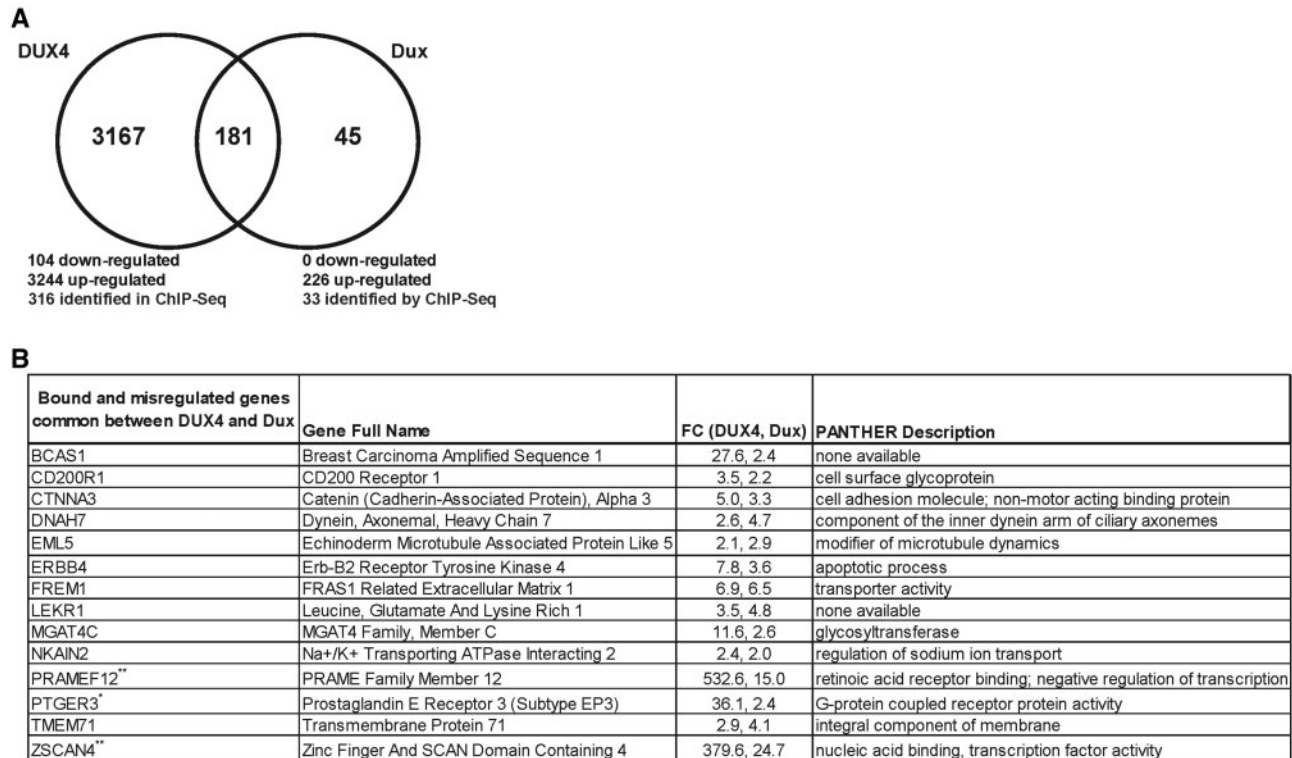


Figure 5 Common expression changes in DUX4- and Dux-expressing myoblasts (A) Venn diagram of genes up- or down-regulated in human myoblasts transfected with DUX4 and Dux identified by RNA-seq. (B) Compilation of ChIP-seq and RNA-seq data. The fourteen genes listed were significantly upregulated by both DUX4 and Dux and contained ChIP-seq binding sites for both proteins within 10,000 bp upstream or downstream of the corresponding gene's TSS. FC, fold-change. PANTHER description (<http://pantherdb.org/>). Asterisks indicate genes with identical binding sites in our ChIP-seq experiments. One asterisk (*) indicates the presence of one consensus binding site sequence, and two asterisks (**) indicates the presence of more than one consensus binding site sequences.

were not universally toxic, and suggested that human DUX4 and mouse Dux could be functionally homologous. If true, this has positive implications for studying DUX4 and FSHD in mice but does not explain DUX4 toxicity in zebrafish or frogs, since they do not contain D4Z4 repeats or express DUX4-like proteins.

D4Z4 repeats have been identified in at least 11 different placental mammal species, including humans and rodents, with the greatest conservation occurring in the DUX4 or DUX4-like ORF (30,31,52). Indeed, non-primate and human D4Z4 sequences have almost no nucleotide similarity outside the ORF. Nevertheless, evolutionary studies suggest close DUX4 homologs exist only in primates, while the mouse Dux gene is considered a distantly related paralog that emerged independently of primate DUX4 by retrotransposition from a common progenitor (DUXC), which has since been lost in primates and rodents (30,31,52). Although the two paralogs would have been subject to different selective pressures over millions of years, the shared toxic phenotypes could have resulted from maintenance of some primordial role of their DUXC progenitor. Unfortunately, we did not directly test the potential toxicity of DUXC in our system. The DUXC ORF has been suggested to exist in dogs and cows, and although we were able to isolate DUXC clones from the dog, we were never able to identify complete open reading frames to determine if DUXC was toxic like Dux and DUX4 (Supplementary Material, Fig. S6). However, we performed ChIP-seq and RNA-seq to explore the hypothesis that DUX4 and Dux could have maintained some conserved gene regulation function. We found that DUX4 and Dux have similar, but not identical, consensus binding motifs and share only a small percentage of binding sites within the human myoblast genome (4 – 8% of all sites bound by DUX4 and Dux,

respectively). Our overall analysis of ChIP-Seq binding sites reveals a unique population is bound by Dux (Fig. 4A). Moreover, DUX4 had 2X as many binding sites as Dux, and activated 14X more genes in the same myoblast cells. Considering the genes are paralogous and evolved in different species, it is perhaps not surprising that the human cells were more responsive to the human DUX4 gene than to mouse Dux. Nevertheless, although a relatively small number of genes were activated by mouse Dux, 80% of these were also activated by DUX4. Thus, combined, our *in vitro* and *in vivo* toxicity experiments, and ChIP-seq and gene expression data, suggested that DUX4 and Dux may share some partial functional homology. We identified 14 genes that are bound and upregulated by both DUX4 and Dux, including ERBB4, which can promote cell death through P53 signalling (Fig. 5B) (41,42). Another potentially important candidate gene involved in P53-mediated cell death was GADD45G, which was significantly increased in both DUX4- and Dux-expressing cells, but did not appear to be a direct transcriptional target of either protein (Supplementary Material, Table 7B) (43,44). We are interested in P53 signalling because we previously showed that DUX4 overexpression in human HEK293 cells and adult mouse muscle caused cytotoxicity or myopathy that was dependent on its ability to bind DNA and indirectly activate the p53 signalling pathway (23). Several other groups reported similar findings in different systems (21,22,28,40). In future studies, we will explore the potential role of ERBB4 and GADD45G alone, or in combination, as well as other genes commonly regulated by DUX4 and Dux, as part of the mechanism underlying DUX4- and Dux-associated toxicity. However, we also note that our previous investigation of the P53 pathway utilized intact, whole mouse muscles as a model

system, and it is therefore possible that DUX4 and Dux could regulate different genes in myoblasts versus myotubes or whole muscle. Thus, it may be prudent to explore other cell types, including differentiated human muscle cells.

Regarding another potential mechanism of toxicity, one hypothesis involves DUX4 activation of the DEFEB103B gene, which was proposed to modulate innate immunity and inhibit muscle differentiation. Although 88% of the genes changed in our DUX4 ChIP-seq were also present in the prior report that identified DEFEB103B as a potentially important gene in FSHD pathogenesis, in our study DEFEB103B was not bound by DUX4 or Dux, nor was its expression significantly changed. In contrast, several previously reported DUX4 target genes were upregulated in both DUX4- and Dux-treated cells, including ZSCAN4, PRAMEF1 and TRIM43, thereby further supporting the validity of our overall methods (Fig. 4D, Supplementary Material, Table S7).

By comparing the structure of the various DUX family members, we gained additional insight in the domains mediating DUX4 and Dux toxicity. First, we showed that conserved residues within the first DNA binding domain (residues 65-70 in mouse Dux) were critical for Dux-associated toxicity, which was an identical finding to what we reported for DUX4. Importantly, these conserved residues are also present in both the HOX1 and HOX2 homeodomains of all DUX family members analysed here, including non-toxic DUX1 and DUX5, which have 81-95% conservation with the DUX4 homeodomains (Fig. 1B). These results suggested that the high homology between the DNA binding domains is not predictive of their toxicity. Indeed, among the DUX family genes tested, DUX4 and Dux had the least amount of homology within their homeodomains but were the only ones with toxic properties. The fact that DUX4 and Dux had similar consensus binding motifs but few overlapping binding sites could suggest that other transcriptional co-factors play a role in the binding specificity and function (i.e. toxicity) of the various DUX family proteins. We hypothesize that the variable carboxy-terminal regions of the respective DUX family proteins could play a role in mediating protein interactions required for efficient transactivation function (Fig. 1A and B). Indeed, overall the DUX4 and Dux carboxy-terminal domains are divergent, but their extreme C-terminal residues show high conservation indicating a common functional role including through stimulating the toxic phenotypes reported here (Fig. 1C) (31). None of the other non-toxic DUX family members contain this extreme C-terminal motif, including the benign DUX4-s isoform that has identical homeodomains as DUX4 but lacks this carboxy-terminal motif (Figs. 2 and 3). In addition, the interactions of DUX protein carboxy-terminal domains could result in their tethering to particular chromatin regions, which is necessary for temporal and spatial patterning (53–55). We now know that DUX4 recruits the histone acetyltransferase, p300, to chromatin thereby enhancing transcriptional activation of target genes, including ZSCAN4 (56). By triggering an increase in acetylation to particular regions in the chromatin, DUX4 is hypothesized to function as a pioneer transcription factor (56). Conservation between DUX4 and Dux within their corresponding carboxy-terminal motifs suggest interactions with co-activators, like p300, could remain conserved as well. Further studies should be designed to address whether Dux is capable of facilitating H3K27 acetylation at particular chromatin regions and in turn promote gene expression, as recent reports have described for DUX4 (56). Finally, we note that excluded from our list of DUX family members is DUX4c, which is identical to DUX4 in the double homeodomain region, but divergent in the C-terminus. DUX4c shares some common protein binding partners with

DUX4 (57), but its over-expression causes muscle differentiation defects and not cell death (58,59). Future work should include domain-switching experiments to determine which C-terminal residues are important for the toxic function of DUX4 and Dux, as well as to identify proteins that could interact with these motifs.

In conclusion, we reported here that DUX family members are not universally toxic, and only mouse *Dux* caused toxic phenotypes similar to those seen in cells and tissues over-expressing DUX4. Like DUX4, Dux-associated toxicity required residues important for DNA binding in HOX1. Finally, mouse Dux and human DUX4 showed largely divergent patterns of potential gene regulation based upon their ChIP-seq binding sites, but 80% of the genes upregulated by Dux were also upregulated by DUX4 in human myoblasts. These results suggest that DUX4 and Dux may share some partial functional overlap.

Materials and Methods

Protein alignment

Protein alignments of all Dux proteins were compared using EMBOSS Needle (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). The following UniProt/Swiss-Prot identifications were used DUX4_HUMAN (Accession Q9UBX2), A1JVI8_MOUSE (Accession A1JVI8), Q7TNE6_MOUSE (Accession Q7TNE6), DUXA_HUMAN (A6NLW8), DUX1_HUMAN (O43812) and DUX5_HUMAN (Q96PT3) for the amino acid sequence alignments of DUX4, Dux, Duxbl, DUXA, DUX1 and DUX5, respectively.

Cell culture

Human embryonic kidney cells (HEK293) were cultured in DMEM supplemented with 10% fetal bovine serum, L-glutamine and penicillin/streptomycin at 37 °C in 5% CO₂. Human immortalized myoblasts (WS236, 15V, bicep, unaffected control cells) (19) were cultured in media containing DMEM supplemented with 16% Medium 199, 15% fetal bovine serum, 30 ng/ml zinc sulphate, 1.4 µg/ml vitamin B12, 55 ng/ml dexamethasone, 2.5 ng/ml human growth factor, 10 ng/ml fibroblast growth factor and 20 mM HEPES. Cells were maintained as myoblasts and were not differentiated for any of the described experiments.

Chromatin immunoprecipitation and sequencing (ChIP-Seq)

Human immortalized myoblasts were transfected with plasmids encoding C-terminal epitope tagged DUX4 and Dux using the Human Dermal Fibroblast Nucleofector kit (Amaxa). Cell fixation was performed 24 h post-transfection by the addition of 1% formaldehyde for 15 min and quenched with 0.125 M glycine. Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and DNA sheared to an average length of 300-500 bp. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. An aliquot of chromatin (30 µg) was precleared with protein A agarose beads (Invitrogen). Genomic DNA regions of interest were isolated using 4 µg of antibody against V5-tag (Abcam ab15828). Complexes were washed, eluted from the beads with SDS buffer then subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at

65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation.

Illumina sequencing libraries were prepared from the ChIP and Input DNA by the standard consecutive enzymatic steps of end-polishing, dA-addition and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on HiSeq 2500. Sequences (50 nt reads) were aligned to the human genome (hg19) using the BWA algorithm (60). Duplicate reads were removed and uniquely mapped reads (mapping quality ≥ 25) were used for further analysis. Alignments were extended *in silico* at the 3'-ends to a length of 200 bp, which is the average genomic fragment length in the size-selected library, and assigned to 32-nt bins along the genome. The resulting histograms were stored in BAR and bigWig files. Peak locations for DUX4 and Dux were determined using the MACS algorithm (v1.4.2) with a cutoff of P-value = $1e-6$ (36). Signal maps and peak locations were used as input data to Active Motif's (<http://www.activemotif.com/>) proprietary analysis program, which created Excel tables containing detailed information on sample comparison, peak metrics, peak locations and gene annotations. The data discussed in this publication have been deposited in the NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE75791 (61).

Our transfection conditions differed from the Geng *et al.* report for the following reasons. We transfected unaffected control immortalized myoblasts using plasmids encoding C-terminal epitope tagged DUX4 or Dux using a nucleofector kit whereas their group transduced unaffected control primary myoblasts using a lentivirus expressing DUX4. Prior to ChIP-Seq, we immunoprecipitated DUX4 using an antibody targeting the C-terminal V5 tag, whereas Geng *et al.* immunoprecipitated DUX4 using polyclonal rabbit anti-sera targeting DUX4. In both cases, cells were harvested 24 h post-transfection or -transduction.

RNA sequencing (RNA-seq)

Human immortalized myoblasts were transfected identical to cells described above for ChIP-Seq. Following assessment of the quality of total RNA using the Agilent 2100 bioanalyzer and RNA Nano Chip kit (Agilent Technologies, CA), RNA was DNase treated and 1.5 μg was subjected to rRNA with Ribo-Zero™ rRNA removal kit for human/mouse/rat (illumina). To generate directional signal in RNA seq data, libraries were constructed from first strand cDNA using ScriptSeq™ v2 RNA-Seq library preparation kit (Epicentre Biotechnologies, WI). Briefly, 50 ng of rRNA-depleted RNA was fragmented and reverse transcribed using random primers containing a 5' tagging sequence, followed by 3' end tagging with a terminal-tagging oligo to yield di-tagged, single-stranded cDNA. Following purification by a magnetic-bead based approach, the di-tagged cDNA was amplified by limit-cycle PCR using primer pairs that anneal to tagging sequences and add adaptor sequences required for sequencing cluster generation. Amplified RNA-seq libraries were purified using AMPure XP System (Beckman Coulter). Quality of libraries were determined via Agilent 2200 TapeStation using High Sensitivity D1000 screen tape, and quantified by Qubit fluorometer with dsDNA BR assay (Invitrogen by Thermo Fisher Scientific). Paired-end 150 bp sequence reads were generated using the Illumina HiSeq 4000 platform. The data discussed in this publication have been deposited in the NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE85935 (61).

Real-time polymerase chain reaction

Human myoblasts (500,000 cells, $n = 3$) were transfected with 6 μg of AAV.DUX4.V5, AAV.Dux.V5, AAV.DUX1.V5, AAV.DUX5.V5, AAV.DUXA.V5, AAV.Duxbl.V5, AAV.Dux.Hox1.V5 or AAV.eGFP plasmid using the NHDF Nucleofector Kit (Lonza). Cells were harvested in TRIzol RNA isolation reagent (Life Technologies) 18–20 h post-transfection. RNA was isolated, DNase treated and reverse transcribed into cDNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). TaqMan gene expression assays (Applied Biosystems) were used to quantify human RPL13A (Hs01494366_g1), PRAMEF12 (Hs04193637_mH), ZSCAN4 (Hs00537549_m1), NFYA (Hs00953589_m1), SRSF8 (Hs00259455_s1), CWC15 (Hs00982912_m1), GTF2F1 (Hs00157845_m1), RFPL1 (Hs00798084_s1), ANKRD1 (Hs00173317_m1) and TRAF6 (Hs00939745_g1). Efficiencies were comparable among all probes. RPL13A was used as the reference gene. The normalized expression ($\Delta\Delta\text{Cq}$) was calculated relative to control GFP transfected cells.

Western blot protein visualization

HEK293 cells (900,000 cells) were transfected with 1.6 μg of AAV.DUX4.V5, AAV.Dux.V5, AAV.DUX1.V5, AAV.DUX5.V5, AAV.DUXA.V5, AAV.Duxbl.V5, AAV.Dux.Hox1.V5 plasmid using Lipofectamine 2000 reagent (Thermo Scientific). Cells were harvested in M-PER Mammalian Protein Extraction Reagent (Thermo Scientific) 24 h post-transfection. Total protein (15 μg) was quantified using the DC Protein Assay (Bio-Rad) was analysed, respectively, by 12% SDS-polyacrylamide gel. Protein was visualized by western blot using an anti-V5 antibody (Invitrogen; R961-25).

Luminescent cell viability assay

C2C12 cells (32,650 cells/well) were transfected ($n = 3$) with 400 ng AAV.DUX4.V5, AAV.DUX4.HOX1, AAV.Dux.V5, AAV.DUX1.V5, AAV.DUX5.V5, AAV.DUXA.V5, AAV.Duxbl.V5, AAV.Dux.Hox.V5 or pCI-neo plasmid using Lipofectamine 2000 (Thermo Scientific) and plated simultaneously on 96-well plates. Abundance of ATP was monitored using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). Luminescence, proportional to the abundance of ATP, was measured 48 h post-transfection using a luminescent plate reader (GloMax microplate luminometer, Promega) after 30 min. Data were reported as mean luminescence with the background (cells only) value removed.

DUX4-activated reporter in cells

HEK293 cells (1×10^6 cells/well) were cotransfected with 600 ng AAV.DUX4.V5, AAV.DUX4.HOX1, AAV.Dux.V5, AAV.Dux.Hox, or pCI-neo and 600 ng pLenti.DUX4-activated GFP, previously reported in (39), reporter plasmids using Lipofectamine 2000 (Thermo Scientific) and plated simultaneously on 6-well plates. GFP expression was visually monitored 24 h post-transfection with a fluorescent stereo microscope (Leica M165 FC microscope., Leica Microsystems).

Adeno-associated virus injections

Six- to eight-week-old C57BL/6 females received $3e9$ or $3e10$ DNase resistant particles (DRP) of adeno-associated virus (AAV) 6 bilaterally via direct 50 μl intramuscular injection into the

anterior (TA), and harvested 2 or 4 weeks post-injection ($n = 10$ muscles per group each time point at (DRP) units). *In vivo* transduction was determined by V5 immunofluorescence staining as described below. To ensure repeatability of data, two different AAV lots were produced per vector, such that two cohorts of animals were used with $n = 5$ per cohort per dose.

Histological analysis

Tibialis anterior (TA) muscles were dissected from intramuscularly injected mice at indicated times post-injection for histological analysis ($n = 10$ muscles per group each time point at $3e9$ or $3e10$ DRP units). Ten micron cryosections were H&E stained as previously described (23). Fibre diameter and central nuclei quantifications were determined from muscles 4 weeks post-injection ($n = 5$ muscles per group; 5 representative $\times 20$ photomicrographs per section), using cellSens Standard Software (Olympus, Tokyo, Japan). Fibre size diameter histograms represent percentage of total fibres analysed.

Immunohistochemistry

Gene expression and subcellular localization of all Dux proteins were visualized using V5 immunofluorescence as previously described (23).

Statistical analysis

All statistical analyses (Luminescent Cell Viability Assay, *in vivo* histological analysis, RT-PCR) were performed in GraphPad Prism 5 (GraphPad Software, La Jolla, CA) using indicated statistical tests.

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank members of the Harper laboratory for assistance and support, Dr. Yi-Wen Chen for mouse Dux cDNA, Dr. Alexandra Belayew for DUX1 plasmid, Dr. Daniel G. Miller for the DUX4-responsive reporter plasmid, TRINCH Viral Vector Core Facility members for assistance with AAV production and the University of Massachusetts Wellstone Muscular Dystrophy Program for providing the human immortalized myoblasts used in our studies. ChIP-seq analysis was performed under contract by ActiveMotif.

Conflict of Interest statement. None declared.

Funding

This work was supported by grants from the Chris Carrino Foundation, the National Institutes of Health [5R01AR062123-04] to S.Q.H.; Wellstone Muscular Dystrophy Centers at Nationwide Children's Hospital [U54HD066409 to C.R.G.] and University of Massachusetts [U54HD060848-09 to S.Q.H.]; the National Center for Advancing Translational Sciences Award [TL1TR001069 to C.R.G.]; FSH Society [FSHS-MGBF-020] to J.O.E.; OSU/NCH Center for Muscle Health and Neuromuscular Disorders [5T32NS077984-02 to J.O.E.] and [82132515 to C.R.G.]. The content is solely the responsibility of the authors and does not necessarily represent the

official views of the National Center for Advancing Translational Sciences or the National Institutes of Health. Funding to pay the Open Access publication charges for this article was provided by the National Institutes of Health [5R01AR062123-04].

References

- INSERM, French Ministry of Health. (2008), In *Orphanet Report*. Orphanet, Vol. 2008.
- Deenen, J.C., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J., Bakker, E., Weinreich, S.S., Verbeek, A.L. and van Engelen, B.G. (2014) Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology*, **83**, 1056–1059.
- Flanigan, K.M. and Harper, S.Q. (2013) John Wiley & Sons, L. (ed.), In *Muscle Disease*, in press, pp. 288–297.
- Tawil, R. and Van Der Maarel, S.M. (2006) Facioscapulohumeral muscular dystrophy. *Muscle & Nerve*, **34**, 1–15.
- Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.*, **2**, 26–30.
- Wijmenga, C., Frants, R.R., Brouwer, O.F., Moerer, P., Weber, J.L. and Padberg, G.W. (1990) Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet*, **336**, 651–653.
- Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, 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.
- Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., et al. (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.*, **3**, 1287–1295.
- Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., et al. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene*, **236**, 25–32.
- de Greef, J.C., Lemmers, R.J., van Engelen, B.G., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R. and van der Maarel, S.M. (2009) Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mut.*, **30**, 1449–1459.
- Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genetics*, **6**, e1001181.
- van der Maarel, S.M., Tawil, R. and Tapscott, S.J. (2011) Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends in Mol. Med.*, **17**, 252–258.
- Sacconi, S., Lemmers, R.J., 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.
- Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., 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.
15. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.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.
 16. Thomas, N.S., Wiseman, K., Spurlock, G., MacDonald, M., Ustek, D. and Upadhyaya, M. (2007) A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus located on a 4qA-defined 4qter subtelomere. *J. Med. Genet.*, **44**, 215–218.
 17. Lemmers, R.J., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E. and van der Maarel, S.M. (2004) Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.*, **75**, 1124–1130.
 18. Dixit, M., Anseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Matteotti, C., van Acker, A.M., Leo, O., et al. (2007) DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl Acad. Sci. U S A*, **104**, 18157–18162.
 19. Jones, T.I., Chen, J.C., Rahimov, F., Homma, S., Arashiro, P., Beermann, M.L., King, O.D., Miller, J.B., Kunkel, L.M., Emerson, C.P., Jr., et al. (2012) Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.*, **21**, 4419–4430.
 20. Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R., et al. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.*, **18**, 2414–2430.
 21. Kowaljow, V., Marcowycz, A., Anseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.*, **17**, 611–623.
 22. Vanderplanck, C., Anseau, E., Charron, S., Stricwant, N., Tassin, A., Laoudj-Chenivresse, D., Wilton, S.D., Coppee, F. and Belayew, A. (2011) The FSHD atrophic myotube phenotype is caused by DUX4 expression. *PLoS One*, **6**, e26820.
 23. Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J. and Harper, S.Q. (2011) DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.*, **69**, 540–552.
 24. Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppee, F., et al. (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.*, **27**, 2766–2779.
 25. Dandapat, A., Bosnakovski, D., Hartweck, L.M., Arpke, R.W., Baltgalvis, K.A., Vang, D., Baik, J., Darabi, R., Perlingeiro, R.C., Hamra, F.K., et al. (2014) Dominant lethal pathologies in male mice engineered to contain an X-linked DUX4 transgene. *Cell Reports*, **8**, 1484–1496.
 26. Wuebbles, R.D., Long, S.W., Hanel, M.L. and Jones, P.L. (2010) Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.*, **3**, 386–400.
 27. Mitsuhashi, H., Mitsuhashi, S., Lynn-Jones, T., Kawahara, G. and Kunkel, L.M. (2013) Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.*, **22**, 568–577.
 28. Sharma, V., Harafuji, N., Belayew, A. and Chen, Y.W. (2013) DUX4 differentially regulates transcriptomes of human rhabdomyosarcoma and mouse C2C12 cells. *PLoS One*, **8**, e64691.
 29. Dandapat, A., Hartweck, L.M., Bosnakovski, D. and Kyba, M. (2013) Expression of the human FSHD-linked DUX4 gene induces neurogenesis during differentiation of murine embryonic stem cells. *Stem Cells Dev.*, **22**, 2440–2448.
 30. Leidenroth, A., Clapp, J., Mitchell, L.M., Coneyworth, D., Dearden, F.L., Iannuzzi, L. and Hewitt, J.E. (2012) Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma*, **121**, 489–497.
 31. Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A. and Hewitt, J.E. (2007) Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.*, **81**, 264–279.
 32. Bosnakovski, D., Daughters, R.S., Xu, Z., Slack, J.M. and Kyba, M. (2009) Biphasic myopathic phenotype of mouse DUX, an ORF within conserved FSHD-related repeats. *PLoS One*, **4**, e7003.
 33. 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., et al. (2012) DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell*, **22**, 38–51.
 34. Krom, Y.D., Dumonceaux, J., Mamchaoui, K., den Hamer, B., Mariot, V., Negroni, E., Geng, L.N., Martin, N., Tawil, R., Tapscott, S.J., et al. (2012) Generation of isogenic D4Z4 contracted and noncontracted immortal muscle cell clones from a mosaic patient: a cellular model for FSHD. *Am. J. Pathol.*, **181**, 1387–1401.
 35. Wu, S.L., Li, G.Z., Chou, C.Y., Tsai, M.S., Chen, Y.P., Li, C.J., Liou, G.G., Chang, W.W., Chen, S.L. and Wang, S.H. (2014) Double homeobox gene, *Duxbl*, promotes myoblast proliferation and abolishes myoblast differentiation by blocking MyoD transactivation. *Cell Tissue Res.*, **358**, 551–566.
 36. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
 37. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202. (2008)
 38. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
 39. Rickard, A.M., Petek, L.M. and Miller, D.G. (2015) Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.*, **24**, 5901–5914.
 40. Block, G.J., Narayanan, D., Amell, A.M., Petek, L.M., Davidson, K.C., Bird, T.D., Tawil, R., Moon, R.T. and Miller, D.G. (2013) Wnt/beta-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.*, **22**, 4661–4672.
 41. Icli, B., Bharti, A., Pentassuglia, L., Peng, X. and Sawyer, D.B. (2012) ErbB4 localization to cardiac myocyte nuclei, and its role in myocyte DNA damage response. *Biochem. Biophys. Res. Commun.*, **418**, 116–121.

42. Arasada, R.R. and Carpenter, G. (2005) Secretase-dependent tyrosine phosphorylation of Mdm2 by the ErbB-4 intracellular domain fragment. *J. Biol. Chem.*, **280**, 30783–30787.
43. Liebermann, D.A. and Hoffman, B. (2008) Gadd45 in stress signaling. *J. Mol. Signal.*, **3**, 15.
44. Lu, B., Yu, H., Chow, C., Li, B., Zheng, W., Davis, R.J. and Flavell, R.A. (2001) GADD45gamma mediates the activation of the p38 and JNK MAP kinase pathways and cytokine production in effector TH1 cells. *Immunity*, **14**, 583–590.
45. Zhang, Y., Lee, J.K., Toso, E.A., Lee, J.S., Choi, S.H., Slattery, M., Aihara, H. and Kyba, M. (2016) DNA-binding sequence specificity of DUX4. *Skeletal Muscle*, **6**, 8.
46. Krom, Y.D., Thijssen, P.E., Young, J.M., den Hamer, B., Balog, J., Yao, Z., Maves, L., Snider, L., Knopp, P., Zammit, P.S., et al. (2013) Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. *PLoS Genetics*, **9**, e1003415.
47. Ferreboeuf, M., Mariot, V., Bessières, B., Vasiljevic, A., Attié-Bitach, T., Collardeau, S., Morere, J., Roche, S., Magdinier, F., Robin-Ducellier, J., et al. (2014) DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.*, **23**, 171–181.
48. Winokur, S.T., Barrett, K., Martin, J.H., Forrester, J.R., Simon, M., Tawil, R., Chung, S.A., Masny, P.S. and Figlewicz, D.A. (2003) Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul Disord.*, **13**, 322–333.
49. Tassin, A., Laoudj-Chenivresse, D., Vanderplanck, C., Barro, M., Charron, S., Ansseau, E., Chen, Y.W., Mercier, J., Coppee, F. and Belayew, A. (2013) DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? *J. Cell. Mol. Med.*, **17**, 76–89.
50. Xu, H., Wang, Z., Jin, S., Hao, H., Zheng, L., Zhou, B., Zhang, W., Lv, H. and Yuan, Y. (2014) Dux4 induces cell cycle arrest at G1 phase through upregulation of p21 expression. *Biochem. Biophys. Res. Commun.*, **446**, 235–240.
51. Wallace, L.M., Moreo, A., Clark, K.R. and Harper, S.Q. (2013) Dose-dependent Toxicity of Humanized Renilla reniformis GFP (hrGFP) Limits Its Utility as a Reporter Gene in Mouse Muscle. *Mol. Ther. Nucleic Acids*, **2**, e86.
52. Leidenroth, A. and Hewitt, J.E. (2010) A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol. Biol.*, **10**, 364.
53. Lian, P., Liu, L.A., Shi, Y., Bu, Y. and Wei, D. (2010) Tethered-hopping model for protein-DNA binding and unbinding based on Sox2-Oct1-Hoxb1 ternary complex simulations. *Biophys. J.*, **98**, 1285–1293.
54. Sharkey, M., Graba, Y. and Scott, M.P. (1997) Hox genes in evolution: protein surfaces and paralog groups. *Trends in Genet.*, **13**, 145–151.
55. Ladam, F. and Sagerstrom, C.G. (2014) Hox regulation of transcription: more complex(es). *Dev. Dyn.*, **243**, 4–15.
56. Choi, S.H., Gearhart, M.D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N. and Kyba, M. (2016) DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.*, **44**, 5161–5173.
57. Ansseau, E., Eidahl, J.O., Lancelot, C., Tassin, A., Matteotti, C., Yip, C., Liu, J., Leroy, B., Hubeau, C., Gerbaux, C., et al. (2016) Homologous transcription factors DUX4 and DUX4c associate with cytoplasmic proteins during muscle differentiation. *PLoS One*, **11**, e0146893.
58. Ansseau, E., Laoudj-Chenivresse, D., Marcowycz, A., Tassin, A., Vanderplanck, C., Sauvage, S., Barro, M., Mahieu, I., Leroy, A., Leclercq, I., et al. (2009) DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. *PLoS One*, **4**, e7482.
59. Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R. and Kyba, M. (2008) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp. Neurol.*, **214**, 87–96.
60. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
61. Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
62. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.