KLF18 is a necessary component of the DUX4-initiated transcriptional network and a candidate locus for phenotypic diversity

Danielle C. Hamm,[1] Aidan B. O'Donnell,[1] Sean R. Bennett,[1] Richard J.L.F. Lemmers,[2] Patrick J. van der Vliet,[2] Silvère M. van der Maarel,[2] and Stephen J. Tapscott[1,3,4]

[1]Human Biology Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA; [2]Department of Human Genetics, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands; [3]Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA; [4]Department of Neurology, University of Washington, Seattle, Washington 98195, USA

DUX4 initiates the first wave of zygotic genome activation (ZGA) in the human embryo, and its misexpression in skeletal muscle causes facioscapulohumeral muscular dystrophy (FSHD). However, other factors that help regulate ZGA-like transcription in human development and disease are still not fully understood. Here we identify Krüppel-like factor 18 (KLF18) as a component of a DUX4 feed-forward network that is necessary for expression of *KLF17* and a subset of DUX4-regulated totipotency-associated genes. We mapped the genome-wide binding profile of KLF18 downstream from DUX4 and showed that its activity is influenced by DUX4 and chromatin accessibility. We found that in contrast to the rodent ortholog, human KLF18 has a predicted β-solenoid structure composed of a variable number of tandem repeats (VNTR) with multiple different structural variants in the human population. Expression of different KLF18 variants in combination with DUX4 showed similar transcriptional activity, whereas assessment of *KLF18* structural variants in individuals with FSHD1 showed inconsistent association with disease severity that requires further study as a modifier locus. Together, these findings establish the polymorphic transcription factor KLF18 as a critical component of a DUX4-initiated feed-forward network and a candidate locus for human diversity.

[*Keywords*: DUX4; KLF17; KLF18; β-solenoid; facioscapulohumeral dystrophy; variable number of tandem repeats; zygotic genome activation]

Supplemental material is available for this article.

Received August 10, 2025; revised version accepted January 16, 2026.

Double homeobox 4 (*DUX4*) was initially identified as the causal gene of facioscapulohumeral muscular dystrophy (FSHD), a genetic disorder marked by progressive muscle weakness (Lemmers et al. 2010; Snider et al. 2010; Tawil et al. 2014). Since its identification as a functional retrogene, DUX4 has been recognized as a key transcriptional regulator in the early human embryo, where it plays a role in establishing totipotency (De Iaco et al. 2017; Hendrickson et al. 2017). After brief expression in the embryo, the *DUX4* locus is silenced in most somatic cells through epigenetic mechanisms (van Overveld et al. 2003; Zeng et al. 2009; Snider et al. 2010; Das and Chadwick 2016). In FSHD, DUX4 is aberrantly expressed in skeletal muscle cells due to loss of epigenetic silencing of the D4Z4 repeat array and the presence of a permissive haplotype on chromosome 4 (4qA) (for review, see Arends et al. 2025). FSHD has two main types, FSHD1 (~95% of cases) and FSHD2 (~5% of cases), which share clinical features but differ in their genetic basis. FSHD1 is caused by contraction of the D4Z4 repeat array to one to 10 units on a permissive 4qA allele, resulting in the loss of epigenetic silencing (Wijmenga et al. 1992; van Deutekom et al. 1993; Lemmers et al. 2010). FSHD2 is caused by mutations in chromatin repressors (mostly *SMCHD1* but also *DNMT3B* and *LRIF1*) leading to epigenetic derepression of a normal-sized D4Z4 array on a permissive 4qA allele (Lemmers et al. 2012; van den Boogaard et al. 2016; Hamanaka et al. 2020). DUX4 expression in FSHD skeletal muscle induces a gene expression pattern in muscle cells similar to that seen in the very early stages of embryonic development (Geng et al. 2012; Yao et al. 2014;

Corresponding author: stapscot@fredhutch.org
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.353253.125. Freely available online through the *Genes & Development* Open Access option.

Jagannathan et al. 2016; De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017; Zheng et al. 2025), disrupting myogenesis and leading to FSHD disease pathology (Kowaljow et al. 2007; Bosnakovski et al. 2008, 2018; Wang et al. 2019, 2021; Wong et al. 2020, 2024; van den Heuvel et al. 2022). Transcriptionally active DUX4 was also recently detected at low levels in many different cancers (Chew et al. 2019; Pineda and Bradley 2024), where it reactivates a totipotent gene signature (Smith et al. 2023).

During early embryonic development, transient DUX4 expression is restricted to the human 4 cell cleavage stage embryo followed by enrichment of other transcription factors during the 8 cell and morula stages (Hendrickson et al. 2017), corresponding to sequential activation of the minor and major waves of zygotic genome activation (ZGA). Rare populations of cultured human embryonic stem cells (hESCs) have been shown to fluctuate in and out of an 8 cell-like cell (8CLC) state and largely recapitulate the naturally occurring totipotent program observed during ZGA in the early embryo (Mazid et al. 2022; Taubenschmid-Stowers et al. 2022). Recent characterization of these in vitro cultured 8CLCs resembling the early human embryo has advanced our understanding of the transcriptional networks driving early human embryogenesis. In addition to DUX4, loss-of-function studies have identified other key regulators of 8CLCs, including *DPPA3* and *TPRX1* (Mazid et al. 2022). However, the transcriptional networks regulating cell fate conversions in and out of the totipotent state remain poorly understood.

Eighteen Krüppel-like factors (KLFs) belonging to the zinc finger family of transcription factors have been identified in the human genome to date, including the recently annotated *KLF18* gene neighboring *KLF17*, believed to be a product of *KLF17* gene duplication (Pei and Grishin 2013). Notably, while human KLF17 is recognized as a naive pluripotency marker (Lea et al. 2021), it is also highly expressed in 8CLCs (Mazid et al. 2022; Taubenschmid-Stowers et al. 2022) and 8 cell embryos (Yan et al. 2013; Blakeley et al. 2015). KLF17-binding motifs have been shown to be enriched at regulatory regions near 8CLC identifier genes (Taubenschmid-Stowers et al. 2022). Thus, it is likely that KLF17, and perhaps the closely related *KLF18* gene, may serve as important transcriptional regulators during human embryogenesis. Expression of *KLF18* has been identified at the mRNA level following DUX4 induction after HSV-1 infection (Tan et al. 2025). However, KLF18 protein expression and its biological function remain entirely uncharacterized.

In this study, we show that DUX4 directly activates the expression of *KLF18* in muscle cells and hESCs and that loss of KLF18 results in decreased expression of a subset of DUX4-induced genes that constitute the totipotent gene signature, including *KLF17*, *CCNA1*, *KHDC1L*, and *TPRX1/2*. Chromatin immunoprecipitation showed KLF18 binding to many KLF18-dependent genes, and most of these genes required both KLF18 and DUX4 expression for transcriptional activation, indicating a feed-forward network. We identified that the *KLF18* gene is polymorphic in the human population with a variable number of 14 amino acid tandem repeats (VNTR) creating multiple KLF18 structural variants with a predicted β-solenoid structure of variable lengths.

Results

DUX4 induces KLF18, which is required for expression of a subset of the DUX4 program

The skeletal muscle cell line MB135 with a doxycycline-inducible codon-altered *DUX4* transgene (MB135iDUX4) is a validated cell model for DUX4 activity in FSHD muscle and has been shown to recapitulate ZGA gene expression mediated by DUX4 in early embryogenesis (Jagannathan et al. 2016; De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017). Induction of DUX4 in MB135iDUX4 cells robustly induced expression of both *KLF17* and *KLF18* (Fig. 1A). Previously published RNA sequencing (RNA-seq) (Jagannathan et al. 2016) and DUX4 chromatin immunoprecipitation sequencing (ChIP-seq) (Geng et al. 2012) read alignments to the *KLF18* locus revealed a DUX4-induced transcript initiating adjacent to a DUX4 binding site in an upstream mammalian-wide interspersed repeat (MIR) element. This transcript contains unannotated first and second exons positioned upstream of the annotated exons of *KLF18* (Supplemental Fig. S1A). The *KLF17* transcript is consistent with the annotated transcription start site (TSS) and exons for *KLF17* in HG38 and is not associated with an adjacent DUX4 ChIP peak. To determine the temporal dynamics of *KLF17* and *KLF18* activation downstream from DUX4, we measured mRNA expression at multiple time points following a 4 h "pulse" of doxycycline treatment using MB135iDUX4 cells to mimic the brief expression of DUX4 in development. *KLF18* was induced as early as 4 h after doxycycline treatment, followed by later expression of *KLF17* between 8 and 12 h (Fig. 1B). Doxycycline-induced MB135iDUX4 myoblasts treated with siRNAs targeting *KLF18* relative to a nontargeting control resulted in a significant loss of DUX4-induced *KLF17* expression, consistent with a temporally ordered DUX4–KLF18–KLF17 transcriptional network (Fig. 1C,D).

To further investigate the roles of KLF17 and KLF18 in modulating the DUX4 gene network, we performed RNA-seq following siRNA-mediated knockdown of each factor (Supplemental Table S1). While DUX4 has been shown to bind the promoter regions and directly regulate ~100 genes (Yao et al. 2014; Smith et al. 2023), ultimately, the expression levels of thousands of genes are altered following DUX4 expression (Jagannathan et al. 2016). To determine whether KLF18 was necessary for a portion of these genes, we first defined the universe of DUX4-regulated genes as those differentially expressed in MB135iDUX4 myoblasts transfected with control siRNAs (siCTRL) and treated with or without doxycycline for 18 h, identifying 1910 robustly upregulated and 2157 significantly downregulated genes upon DUX4 induction (basemean ≥100; $|\log_2$ fold change $|\geq 2|$; $P\text{adj} \leq 0.01$). Knockdown of KLF18 resulted in the reduced expression of 38 DUX4-induced genes, whereas KLF17 knockdown had little to no effect on DUX4-regulated gene expression

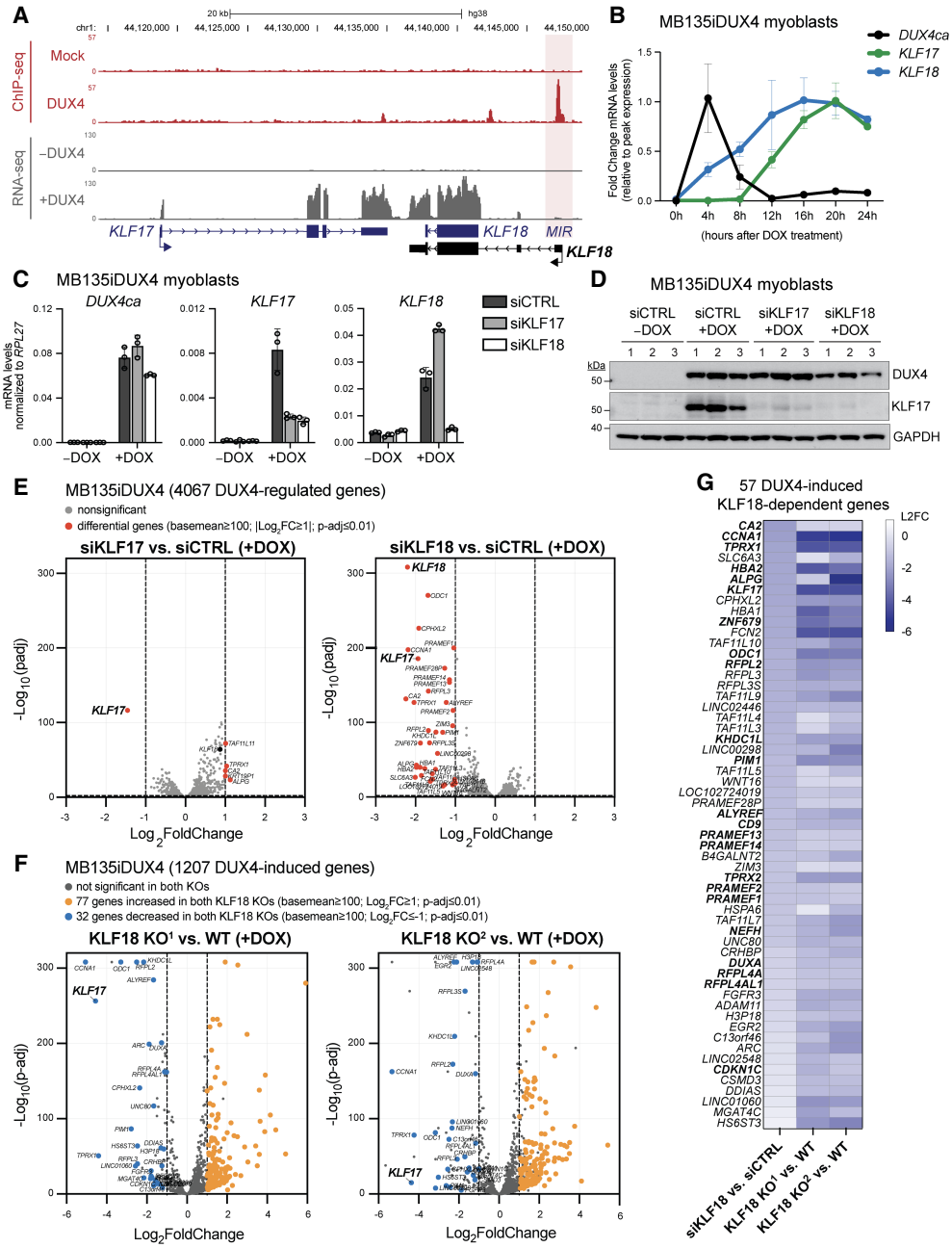


Figure 1. KLF18 regulates expression of a subset of genes downstream from DUX4. (*A*) ChIP-seq (red) and RNA-seq (gray) tracks from Geng et al. (2012) and Jagannathan et al. (2016), respectively, showing DUX4 binding at a MIR transposable element upstream of *KLF18* and expression of *KLF17* and *KLF18* mRNAs with doxycycline-induced DUX4 expression in MB135iDUX4 myoblasts. A DUX4-induced alternative upstream *KLF18* transcriptional start site creates a unique transcript (black) relative to the annotated gene (blue). (*B*) RT-qPCR analysis of MB135iDUX4 cells treated with a 4 h pulse of 1 µg/mL doxycycline (DOX), followed by a PBS wash and harvested every 4 h. Data were normalized to housekeeping gene *RPL27* and are graphed as fold change relative to the time of peak expression. Data represent mean ± SD of biological replicates; $n = 3$. (*C*,*D*) RT-qPCR (*C*) and immunoblot analysis (*D*) of MB135iDUX4 cells treated with siRNAs targeting KLF17 (siKLF17), KLF18 (siKLF18), or nontargeting controls (siCTRL) ±1 µg/mL doxycycline (DOX) for 20 h. RT-qPCR data were normalized to housekeeping gene *RPL27* and represent mean ± SD of biological replicates; $n = 3$, (DUX4ca) Codon-altered DUX4. (*E*) Volcano plot of DUX4-regulated genes (defined as basemean ≥ 100, $|\log_2$ fold change ≥ 2|, *P*adj ≤ 0.01). Genes differentially regulated with siRNA-mediated knockdown of KLF17 or KLF18 relative to control are highlighted in red (significance is defined as basemean ≥ 100, $|\log_2$ fold change ≥ 1|, $P \leq 0.01$) (see Supplemental Table S1). (*F*) Volcano plots of DUX4-induced genes (defined as basemean ≥ 100, $\log_2$ fold change ≥ 2, *P*adj ≤ 0.01) showing differential gene expression between MB135iDUX4 cells with wild-type (WT) *KLF18* alleles relative to two independent MB135iDUX4 clonal cell lines with *KLF18* mutant alleles introduced with CRISPR/Cas9-mediated indels (referred to as knockouts: KO^1 and KO^2) treated for 18 h with doxycycline (DOX). Genes that significantly decreased are highlighted in blue, and genes that significantly increased are highlighted in yellow (significance is defined as basemean ≥ 100, $|\log_2$ fold change ≥ 1|, *P*adj ≤ 0.01) (see Supplemental Table S2). (*G*) Heat map of DUX4-induced genes that were significantly downregulated with siRNA-mediated knockdown of KLF18 or CRISRP/Cas9-mediated disruption of KLF18 alleles (KLF18 KOs). ZGA/8CLC-associated genes are indicated in bold. Data are graphed as $\log_2$ fold change; $n = 3$ biological replicates per condition.

(Fig. 1E). The KLF18-dependent genes represented a subset of the DUX4-regulated ZGA/8C genes and included *KHDC1L*, *CCNA1*, and *TPRX1* but not DUX4 targets *ZSCAN4* and *H3Y* (Supplemental Fig. S2A). We further validated these findings in the context of endogenous DUX4 expression in FSHD muscle cells with siRNA-mediated knockdown of *KLF17* and *KLF18* and showed that the same subset of DUX4-induced genes depended on KLF18 in both myoblasts and differentiated myotubes (Supplemental Fig. S2B,C). These data indicate that KLF18 is necessary for DUX4-mediated activation of *KLF17* and a subset of DUX4-regulated genes associated with ZGA and the 8 cell state.

Next, we used CRISPR/Cas9-mediated editing to mutate the *KLF18* gene in MB135iDUX4 myoblasts. We used different guide RNAs (gRNAs) to generate two cell lines in order to control for potential off-target activities, and each clonal line contained frameshift mutations on both alleles (Supplemental Fig. S1B). Although complete loss of KLF18 protein cannot be directly verified due to the lack of a suitable KLF18 commercial antibody, the frameshift-inducing indels are predicted to generate loss-of-function alleles. MB135iDUX4 wild-type (WT) cells or cells with *KLF18* CRISPR indels (referred to as *KLF18* knockouts: KO^1 and KO^2) were treated with doxycycline for 18 h and harvested for RNA-seq. Correlation analysis showed strong concordance between the two *KLF18* KO cell lines ($R^2 = 0.71$) (Supplemental Fig. S1C). To ensure that changes in gene expression were due to the loss of *KLF18* and not gRNA off-target effects or variability between clonal cell lines, our analysis selected for genes significantly differentially expressed in both CRISPR knockout lines compared with the parental WT line (basemean ≥ 100; $|\log_2$ fold change $\geq 1|$; $P\text{adj} \leq 0.01$) (Supplemental Table S2).

Because siRNA-mediated knockdown indicated that KLF18 contributes to a subset of genes normally induced by DUX4 with little impact on DUX4-repressed genes, we focused our analysis on DUX4-induced genes. MB135iDUX4 myoblasts carrying WT *KLF18* were treated with or without doxycycline, resulting in robust induction of 1207 genes (basemean ≥ 100; $\log_2$ fold change ≥ 2; $P\text{adj} \leq 0.01$), of which 32 showed a significant decrease and 77 showed a significant increase in expression in both *KLF18* KO lines relative to WT ($|\log_2$ fold change $\geq 1|$; $P\text{adj} \leq 0.01$) (Fig. 1F). Twelve genes significantly downregulated in the absence of KLF18 were common to both the siRNA and KO data sets (Supplemental Fig. S1D); however, the entire set of 57 genes showed decreased expression relative to WT cells in both the siRNA and *KLF18* KO data sets (Fig 1G), suggesting that the partial overlap might reflect the use of a strict cutoff. GO analysis of these 57 genes revealed enrichment of RNAPII preinitiation complex assembly, likely due to multiple TAF11L family members and ZGA-associated and 8CLC genes (Supplemental Fig. S1E; Supplemental Table S2). The 77 DUX4-induced genes that were upregulated in the *KLF18* KO lines did not yield any statistically enriched terms ($P\text{adj} \leq 0.05$).

Although KLF18 depletion attenuated expression of a subset of DUX4-induced genes, neither siRNA-mediated knockdown nor CRISPR-mediated knockout of *KLF18* was sufficient to rescue DUX4-induced cytotoxicity in myoblasts at later time points following doxycycline induction (Supplemental Fig. S1F,G). Taken together, our results identify KLF18 as a component of the DUX4 transcriptional network, promoting activation of ZGA-associated genes and contributing to dysregulated gene expression in FSHD cell culture models.

Identification of KLF18 binding sites downstream from DUX4

Krüppel-like factors (KLFs) and specificity proteins (SPs) constitute a family of transcription factors distinguished by a highly conserved C-terminal DNA-binding motif that features three C2H2 zinc finger domains (Schuh et al. 1986; Swamynathan 2010; Pei and Grishin 2013). KLF17 has been shown to bind GC-rich and CACCC-box sequences (van Vliet et al. 2006). While the KLF18 DNA-binding motif has yet to be defined, the DNA-binding domains (DBDs) of KLF17 and KLF18 share the three conserved C2H2 zinc fingers and 54% amino acid similarity (Supplemental Fig. S3). Based on comparisons with other KLF/SP family members, it was predicted that KLF18 recognizes GC-boxes and GT-boxes (Pei and Grishin 2013), which frequently occur in promoters and enhancers.

Having demonstrated that KLF18 differentially regulates a subset of genes downstream from DUX4, we wanted to identify KLF18 binding sites throughout the genome in DUX4-expressing myoblasts. We used CRISPR/Cas9 gene editing to incorporate an N-terminal 3xFLAG tag at the endogenous *KLF18* locus in MB135iDUX4 myoblasts to best recapitulate endogenous gene expression dynamics following DUX4 induction (Supplemental Fig. S4A). Genotyping showed one functional 3xFLAG-tagged allele and one frameshift allele, mitigating concern of untagged KLF18 protein that could occupy endogenous binding sites and occlude 3xFLAG-tagged protein binding. Following doxycycline treatment, we observed DUX4-induced expression of FLAG-tagged KLF18 protein of the correct size by immunoblot analysis, as well as nuclear localization with immunofluorescence staining (Supplemental Fig. S4B,C). Furthermore, DUX4 target genes, including genes dependent on KLF18 (i.e., *KLF17*, *CCNA1*, and *KHDC1L*), were induced in our 3xFLAG-tagged engineered cell line (Supplemental Fig. S4D); thus, the N-terminal tag did not inhibit endogenous KLF18 activity.

Using a FLAG antibody, ChIP-seq was performed in MB135iDUX4 myoblasts carrying wild-type or 3xFLAG-tagged *KLF18* treated with doxycycline. We identified 958 3xFLAG-KLF18 ChIP peaks (Supplemental Table S3), a subset of which showed enrichment for DUX4 binding (Fig. 2A). Genome-wide mapping of 3xFLAG-KLF18 binding sites showed enrichment largely at gene promoters (Fig. 2B), a subset of which was associated with DUX4 binding (Fig. 2C). MEME-ChIP de novo motif discovery (Machanick and Bailey 2011) included GC-rich sequences known to be bound by SP/KLF factors, motifs bound by

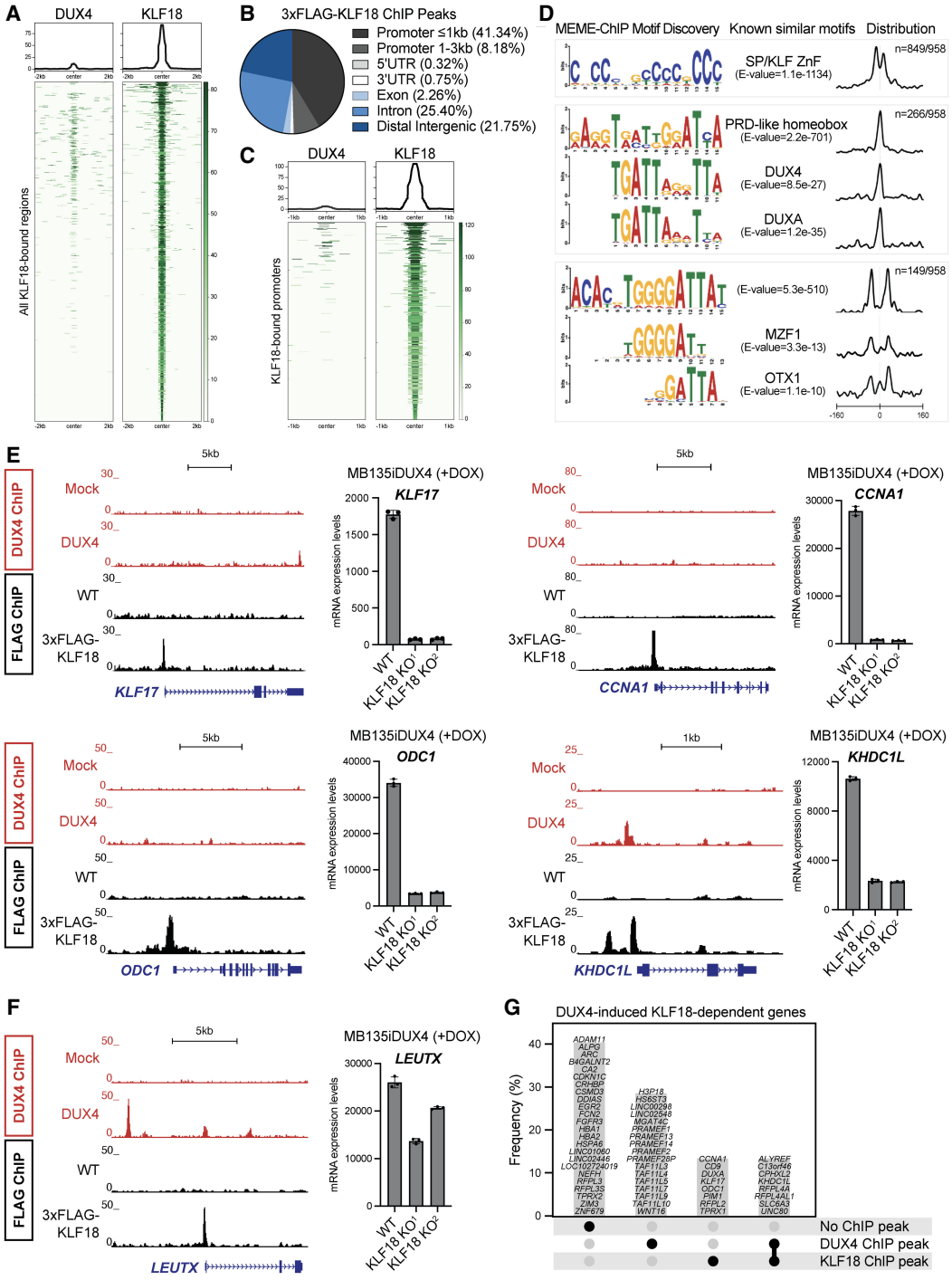


Figure 2. 3xFLAG-KLF18 ChIP-seq in MB135iDUX4 myoblasts. (*A*) Average plot (*top*) and heat map (*bottom*) of ChIP-seq reads representing DUX4 and KLF18 binding around all KLF18 ChIP peaks ($n = 958$ peaks). (*B*) Genomic distribution of KLF18 ChIP peaks. (*C*) Average plot (*top*) and heat map (*bottom*) of ChIP-seq reads representing DUX4 and KLF18 binding ±1 kb around KLF18-bound promoters ($n = 396$ promoters). (*D*) MEME-suite motif discovery of the top 958 KLF18-bound regions (q-value ≤ 0.05); sequences were derived from ChIP peak genomic coordinates. Factors listed have known binding motifs similar to the top three enriched de novo motifs reported, and *E*-values reflect the significance of the motif according to MEME or CentriMo. Histograms at the *right* show distribution (±160 bp) of the best matches to the motif in the sequences as found by CentriMo analysis; n = the number of sites contributing to the construction of the motif out of the 958 sequences. (*E*,*F*) Representative genome browser tracks of DUX4 ChIP-seq (Geng et al. 2012) and 3xFLAG-KLF18 ChIP-seq data and bar graphs showing RNA-seq mRNA expression levels of corresponding genes in MB135iDUX4 wild-type (WT) and *KLF18* knockout (KO) cell lines with 18 h of 1 µg/mL doxycycline (DOX) treatment. (*G*) Frequency of the 57 DUX4-induced KFL18-dependent genes with or without nearby DUX4 and/or 3xFLAG-KLF18 ChIP peaks (based on nearest-gene annotation) (see Supplemental Table S3).

paired-like homeobox transcription factors including DUX4 and DUXA, and a consensus motif with partial similarity to the myeloid zinc finger 1 (MZF1) and orthodenticle homeobox 1 (OTX) binding motifs (Fig. 2D).

Among the 3xFLAG-KLF18 ChIP peaks were regulatory regions of genes that required KLF18 for expression downstream from DUX4, including *KLF17*, *CCNA1*, *ODC1*, and *KHDC1L* (Fig. 2E). There were also instances of KLF18 binding at genes that did not show a significant decrease in DUX4-induced expression in both *KLF18* KOs relative to WT controls, such as *LEUTX* (Fig. 2F). Several KLF18-bound genes, including *KHDC1L* and *LEUTX*, had one or more nearby DUX4 ChIP peaks, highlighting co-occupancy at some loci. Based on nearest-neighbor relationships to 3xFLAG-KLF18 or DUX4 ChIP peaks, we asked which of the DUX4-induced KLF18-dependent genes were bound by 3xFLAG-KLF18, DUX4, or both (Fig. 2G; Supplemental Table S3). Of these 57 genes, 14% were bound by 3xFLAG-KLF18 alone, 30% were bound by DUX4 alone, and 14% were bound by both. Notably, 42% of genes that were induced by DUX4 and showed decreased expression with the loss of *KLF18* exhibited no nearby DUX4 or KLF18 ChIP peaks, including 8 cell-like cell (8CLC) genes *ALPG*, *CA2*, *CDKN1C*, *HBA1*, *HBA2*, *NEFH*, and *ZNF679*. This may be due to indirect or low-affinity binding not captured with our ChIP assay, regulation by distant or *trans*-acting enhancer elements, or activation by a KLF18-induced transcription factor.

KLF18 functions in a DUX4 feed-forward network

We overexpressed KLF17 and KLF18 proteins, as annotated by GENCODE, and asked whether these constructs were sufficient to activate transcription in the absence of DUX4. MB135 myoblasts were transiently transfected with plasmids constitutively expressing 3xFLAG-tagged codon-altered KLF17 or KLF18 or GFP as a control and harvested 48 h after transfection (Fig. 3A). Differential RNA-seq analysis showed a significant increase in 109 and 79 genes following exogenous expression of 3xFLAG-KLF17 or 3xFLAG-KLF18, respectively, relative to GFP (basemean ≥ 50; $|\log_2 \text{fold change} \geq 2|$; $P \leq 0.01$) (Fig. 3B,C; Supplemental Table S4). Of the 57 DUX4-induced KLF18-dependent genes identified via *KLF18* knockdown or knockout, only eight were significantly increased after KLF18 transfection relative to GFP (*HBA1*, *HBA2*, *SLC6A3*, *CA2*, *KHDC1L*, *CCNA1*, *ARC*, and *FGFR3*), seven of which were also induced by KLF17, while several other genes showed modest increases but remained below the significance threshold (Fig. 3D). Fifty-six genes were commonly induced by KLF17 and KLF18, and enrichment analysis of these genes showed gene expression signatures associated with nervous system function and developmental growth (Fig. 3E,F; Supplemental Table S4). Collectively, our data highlight that while KLF18 is necessary for the full activation of a subset of the DUX4 program, it is not sufficient to fully activate the expression of that subset in the absence of DUX4.

In this regard, many of the genes regulated by KLF18 in the context of DUX4 expression, like *TPRX1* and *ODC1*, did not show a statistically significant increase in expression with 3xFLAG-KLF18 overexpression in MB135 myoblasts in the absence of DUX4 (Supplemental Table S4). Our ChIP-seq analysis indicated that *TPRX1* and *ODC1* are unlikely direct DUX4 targets but rather are bound by KLF18 at their promoters. Reporter assays were used to test promoter specificity and determine whether KLF18 was sufficient to activate target genes outside their native chromatin context. We cloned an ~800 bp promoter region of *TPRX1* or *ODC1* (encompassing a KLF18 ChIP peak) upstream of a firefly *luciferase* reporter (Fig. 3G). Transient cotransfection of these luciferase reporter plasmids with an exogenous 3xFLAG-KLF18 expression vector resulted in robust *luciferase* (LUC) reporter gene activation from the *ODC1* and *TPRX1* promoters compared with their endogenous expression levels, indicating that in the absence of chromatin, KLF18 is sufficient to activate a subset of direct targets (Fig. 3H). In contrast, 3xFLAG-KLF18 failed to activate a *ZSCAN4* luciferase reporter construct, consistent with the absence of a KLF18 ChIP peak at this promoter; however, ectopic DUX4 expression alone in *KLF18*-null myoblasts (KLF18 KO[1]) robustly activated the *ZSCAN4* reporter, in line with DUX4 binding at the *ZSCAN4* promoter (Fig. 3G,H, right; Geng et al. 2012). Together, these data indicate that KLF18 alone is not sufficient to fully activate the gene program that it regulates downstream from DUX4, possibly due to the chromatin context, consistent with a component of the DUX4 feed-forward network.

KLF18 shows similar expression patterns and activity in human embryonic cells

To determine whether KLF17 and KLF18 had similar functions relative to DUX4 in embryonic stem cells, we analyzed prior RNA-seq data sets of early human embryos and induced pluripotent stem cells (iPSCs) (Hendrickson et al. 2017). *KLF17* and *KLF18* expression was enriched in the early cleavage stage embryo, mirroring transient expression of DUX4 and the ZGA transcriptional program (Fig. 4A, right). In addition, expression of DUX4 in human iPSCs induced *KLF17* and *KLF18* specifically when compared with all other Krüppel-like factors, corresponding with upregulation of other DUX4 targets (Fig. 4A, left).

Next, we transduced H9 human embryonic stem cells (hESCs) with a codon-altered doxycycline-inducible DUX4 transgene (Fig. 4B) and used CRISPR to knock out *KLF17* or *KLF18* (Fig. 4C,D; Supplemental Fig. S5). All clonal cell lines retained pluripotency, as indicated by stable expression of several stem cell markers that were unaffected by doxycycline-induced DUX4 expression (Fig. 4E). To account for the uneven levels of DUX4 induction across cell lines, DUX4-induced genes were normalized to levels of *DUX4* mRNA. Knockout of *KLF18* resulted in decreased expression of a subset of DUX4-induced genes, including *KLF17*, *KHDC1L*, *CCNA1*, *TPRX1/2*, and *ALPG*, while other well defined DUX4 targets were

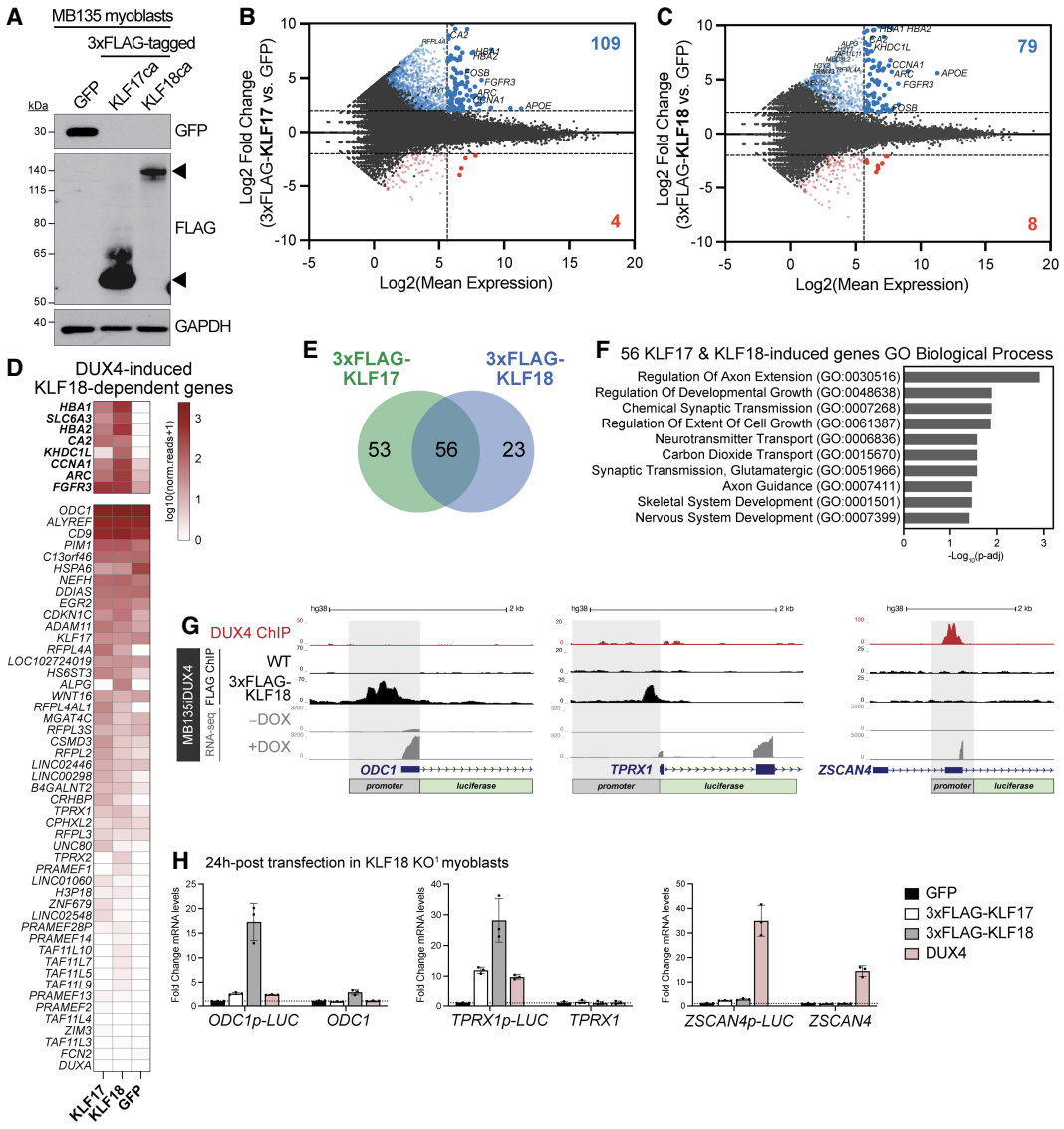


Figure 3. Overexpression of exogenous 3xFLAG-tagged KLF17 and KLF18 in myoblasts. (*A*) Immunoblot analysis of MB135 myoblasts 48 h after transfection with constructs constitutively expressing GFP or 3xFLAG-tagged codon-altered (ca) KLF17 or KLF18 GENCODE variants. (*B*,*C*) MA plot of differential RNA-seq analysis comparing 3xFLAG-KLF17ca (*B*) or 3xFLAG-KLF18ca (*C*) overexpression relative to GFP (significance is defined as basemean ≥ 50, $|\log_2$ fold change ≥ 2|, $P \leq 0.01$) (see Supplemental Table S4). (*D*) Heat map of the 57 DUX4-induced KLF18-dependent genes showing expression levels with 3xFLAG-KLF17, 3xFLAG-KLF18, or GFP transfection. Data are graphed as $\log_{10}$(normalized readcount + 1); $n = 3$ biological replicates per condition. Eight genes significantly induced by KLF18 relative to GFP are indicated in bold (significance is defined as basemean ≥ 50; $\log_2$ fold change ≥ 2; Padj ≤ 0.01). (*E*) Overlapping gene networks from RNA-seq analysis in *B* and *C*. (*F*) Gene ontology analysis of the 56 genes activated by overexpression of 3xFLAG-KLF17 or 3xFLAG-KLF18 (see Supplemental Table S4). (*G*) ChIP-seq tracks highlighting DUX4 binding (red) and KLF18 binding (black) at *TPRX1*, *ODC1*, and *ZSCAN4* promoter regions and RNA-seq tracks showing DUX4-induced expression of *ODC1*, *TPRX1*, and *ZSCAN4* in MB135iDUX4 ± doxycycline (DOX; gray). A schematic of promoter regions of interest cloned upstream of the *luciferase* reporter is shown *below*. (*H*) RT-qPCR analysis of *luciferase* (LUC) reporter expression and endogenous *TPRX1(exon3)*, *ODC1(exon5–6)*, and *ZSCAN4(exon2–3)* gene expression in MB135 24 h after cotransfection of *luciferase* reporter constructs with GFP, 3xFLAG-KLF17, 3xFLAG-KLF18, or DUX4 expression vectors. Data were normalized to housekeeping gene *RPL27* and are graphed as fold change relative to GFP (dashed line; $y = 1$). Data represent mean ± SD of biological replicates; $n = 3$.

unaffected (i.e., *H3Y*, *ZSCAN4*, and *LEUTX*) (Fig. 4F). Based on these data, we conclude that KLF18 is induced by DUX4 in hESCs and is necessary for the full expression of the totipotent gene signature, similar to our findings in DUX4-expressing muscle cells.

Comparison of human KLF17 and KLF18 with mouse orthologs and paralogs

Based on synteny and sequence similarity, human KLF17 is the proposed ortholog of the germ cell-specific mouse

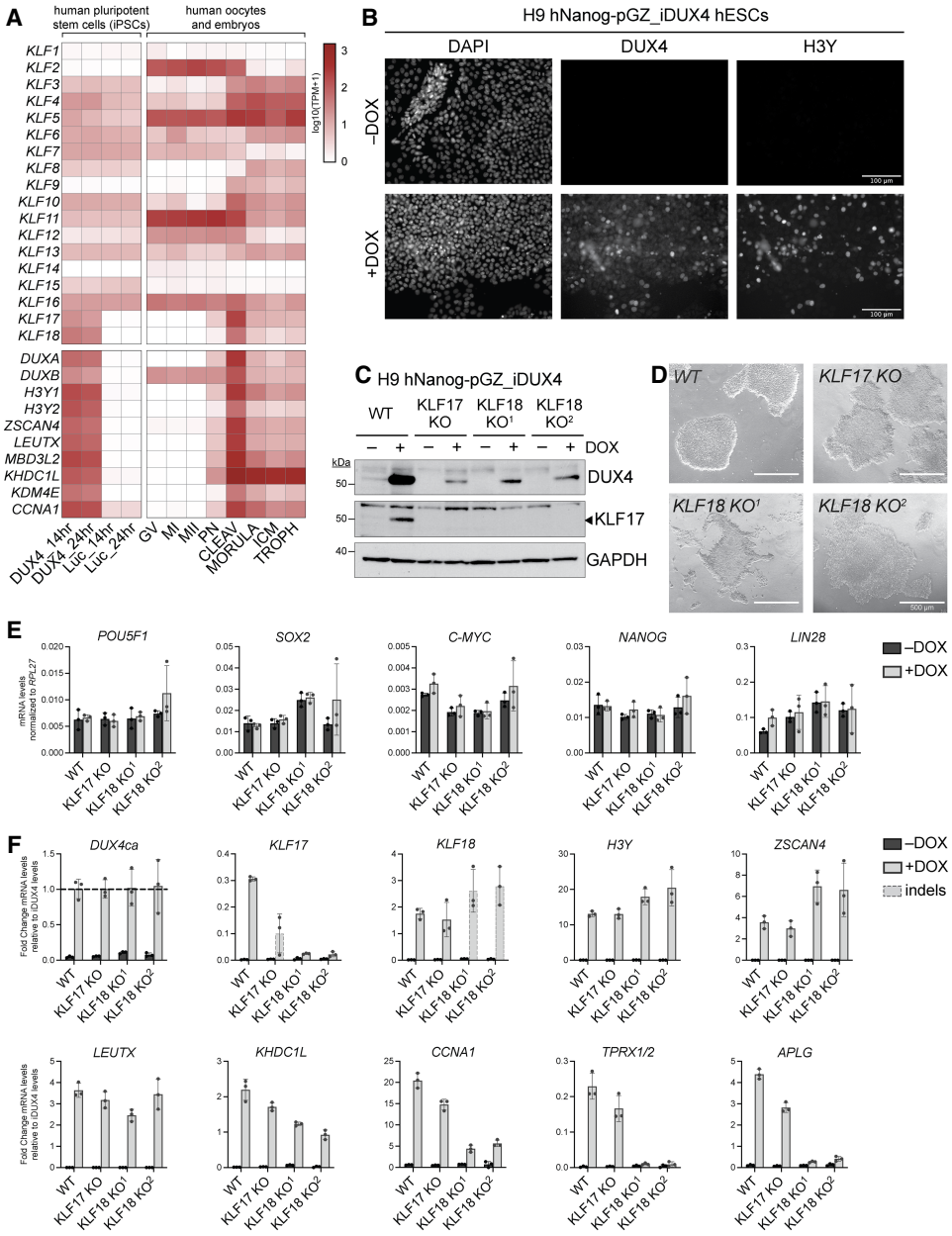


Figure 4. DUX4 activates *KLF17* and *KLF18* in human embryonic stem cells. (*A*) Heat map depicting expression of KLF genes and several DUX4-activated genes in human iPSCs after DUX4 or luciferase induction (*left*) and their expression across early embryonic development (*right*). RNA-seq were data sourced from Hendrickson et al. (2017) and are graphed as $\log_{10}$(FPKM + 1). (*B*) Immunofluorescence of DUX4 and DUX4-induced H3Y in H9 hESCs carrying codon-altered doxycycline-inducible DUX4 transgene (iDUX4) ±14 h of 1 µg/mL doxycycline (DOX) treatment. Scale bar, 100 µm. (*C*) Immunoblot analysis of DUX4 and KLF17 protein expression in H9iDUX4 cells with wild-type (WT) *KLF17* and *KLF18* alleles or with mutant alleles introduced with CRISPR/Cas9-mediated indels (knockouts [KOs]) ±16 h of 1 µg/mL doxycycline (DOX) treatment. (*D*) Phase contrast images of WT, *KLF17*, and *KLF18* CRISPR knockout (KO) monoclonal cell lines cultured without doxycycline. Scale bar, 500 µm. (*E*) RT-qPCR analysis of stem cell marker gene expression ±16 h of 1 µg/mL doxycycline (DOX) treatment. Data were normalized to housekeeping gene *RPL27*. (*F*) RT-qPCR analysis of DUX4-induced genes. Data were normalized to *RPL27* and are graphed as fold change relative to codon-altered DUX4 (*DUX4ca*) to account for doxycycline-induced transgene levels (dashed line; $y = 1$). Data represent mean ± SD of biological replicates; $n = 3$.

gene *zinc finger protein 393 (Zfp393)* (Yan et al. 2002; van Vliet et al. 2006), currently referred to as mouse *Klf17*. However, the degree of conservation between human and murine KLF17 proteins is lower than all other KLF orthologs (van Vliet et al. 2006), suggesting that KLF17 has diverged significantly in mammals. While human *KLF17* is zygotically expressed in the 8 cell embryo during ZGA, murine *Klf17* is maternally deposited and zygotically enriched at the 2 cell stage (Supplemental Fig. S6A, B). Furthermore, *Klf17* does not require *Dux* for expression

in the 2 cell mouse embryo (Supplemental Fig. S6C). It was recently reported that maternal *Klf17* knockout in mouse embryos resulted in decreased expression of 9% of ZGA genes and aberrant RNA polymerase II preconfiguration (Hu et al. 2024), indicating that maternal-deposited mouse Klf17 may serve to regulate ZGA in a different capacity than human KLF17.

Several KLF18-like genes have been identified in the mouse genome, including the gene adjacent to *Klf17*, *Gm12845*, thought to be the product of local gene duplication (Ensembl gene related to previously predicted *mCG120027*-like), as well as *Zfp352* and *Zfp353*, likely derived from retrotransposition of an ancestral *Klf18* gene (Pei and Grishin 2013). Expression of these proposed *Klf18*-like genes is enriched in the 2 cell mouse embryo and significantly reduced in *Dux* knockout embryos (Supplemental Fig. S6C). A similar dependence on Dux for *Zfp352* expression was recently demonstrated in mouse ESC populations transiently transitioning to the 2 cell-like cell (2CLC) state (Mwalilino et al. 2023). Zfp352 has been reported to regulate a subset of the 2C-like program as well as promote the dissolution of totipotency to allow for expression of later developmental programs (Li et al. 2023; Mwalilino et al. 2023). Thus, preliminary findings indicate that the rodent-specific genes *Gm12845*, *Zfp352*, and *Zfp353* may be relevant in regulating early embryonic gene expression. However, it remains to be determined whether they are functional orthologs of human KLF18.

The N-terminal domain of human KLF18 contains tandem repeats

Based on GENCODE annotation, the human KLF17 protein is 389 amino acids long and KLF18 is 1052 amino acids long (Fig. 5A; Supplemental Fig. S3). Unique among the KLF proteins, the N-terminal domain (NTD) of human KLF18 contains a series of tandem repeats that encode units of highly similar 14 amino acid repeats (Fig. 5B). This region is predicted to form a β-solenoid-like structure (Fig. 5A, right) in which repeated sequence units give rise to stacked arrays of identical residues forming a "ladder" comprised of polar residues that facilitate a network of hydrogen bonds, commonly including asparagine, serine, threonine, and glutamine (Kajava and Steven 2006; Flores-Fernández et al. 2018; Mesdaghi et al. 2023).

While KLF18 orthologs share the conserved KLF-like zinc finger DNA-binding domain, the tandem repeats are variably present in other eutherian mammals (Supplemental Fig. S6D,E). For example, despite being distant relatives on the evolutionary tree, the two-toed sloth (*Choloepus didactylus*) KLF18 ortholog has ∼40 units of a very similar 14 amino acid repeat and a predicted β-solenoid structure similar to that observed in primates. Among the Laurasiatheria clade, fruit bats (*Artibeus jamaicensis*) have diverged to have ∼30 units of a 13 amino acid repeat, whereas horses and dogs have only a few 13 and 14 amino acid repeat sequences. Mouse and rat KLF18 orthologs have a dispersed four residue motif not predicted to form a β-solenoid-like structure (Pei and Grishin 2013). However, another member of the rodent family, the golden hamster (*Mesocricetus auratus*), stands out with its extensive 27 amino acid repeat region. This suggests a dynamic evolution with either the generation or loss of repeats in the *KLF18* gene of different species.

Variable number tandem repeats create multiple KLF18 variants in the human population

The Genome Aggregation Database (gnomAD) (Chen et al. 2024) of 730,947 exome sequences and 76,215 whole-genome sequences revealed variation of sequence depth in the region of the *KLF18* repeats, suggesting that a variable number of tandem repeats (VNTR) might create structural variants in the human population, whereas *KLF17* alleles did not show significant variation (Fig. 5C). Analysis of long-read sequencing from the Human Pangenome Reference Consortium (HPRC), consisting of 90 individual component genome phases (maternal and paternal) from a cohort of genetically diverse individuals (Liao et al. 2023), identified five different *KLF18* structural variants and multiple nonsynonymous single-nucleotide polymorphisms (SNPs) (Fig. 5D; Supplemental Table S5). The longest variant, present in 4.4% of the genomes, encoded a predicted 1067 amino acid protein with 49 consecutive repeat units (48 units with 14 amino acids and one unit with 15 amino acids). The 15 amino acid repeat is deleted in the 1052 amino acid variant present in 8.9% of the genomes that matches the GENCODE KLF18. The most prevalent variant (81.1% in the HPRC data set) encodes a protein missing two adjacent 14 amino acid units to produce a 1039 amino acid protein and matches the HAPMAP annotated *KLF18* coding sequence. The CHM13 *KLF18* coding sequence is annotated as a 1025 amino acid protein missing three repeat units and occurred at a frequency of 3.3%. The smallest *KLF18* variant carries four deletions within the VNTR region coding for an 829 amino acid protein with 32 repeat units and occurring at a frequency of 2.2%.

The VNTR domain is required for activation of most KLF18-dependent genes

To determine whether the number of tandem repeats modulates KLF18 transcriptional activity in the context of DUX4, we generated MB135iDUX4 KLF18 KO cell lines with doxycycline-inducible 3xFLAG-tagged KLF18 variant transgenes integrated at the AAVS1 locus. We generated two independent clonal cell lines for the 49 unit variant (1067 amino acids; long) and two independent lines for the shortest naturally occurring 32 unit variant (829 amino acids; short) (Fig. 5E). At the protein level, the short and long variants rescued DUX4-induced KLF18-dependent expression of KLF17, KHDC1L, and TPRX1 to comparable levels. ODC1 expression was also rescued by both variants, with slightly higher induction observed following expression of the KLF18 short variant.

To assess the functional contribution of the KLF18 VNTR domain, we generated a clone of the MB135iDUX4

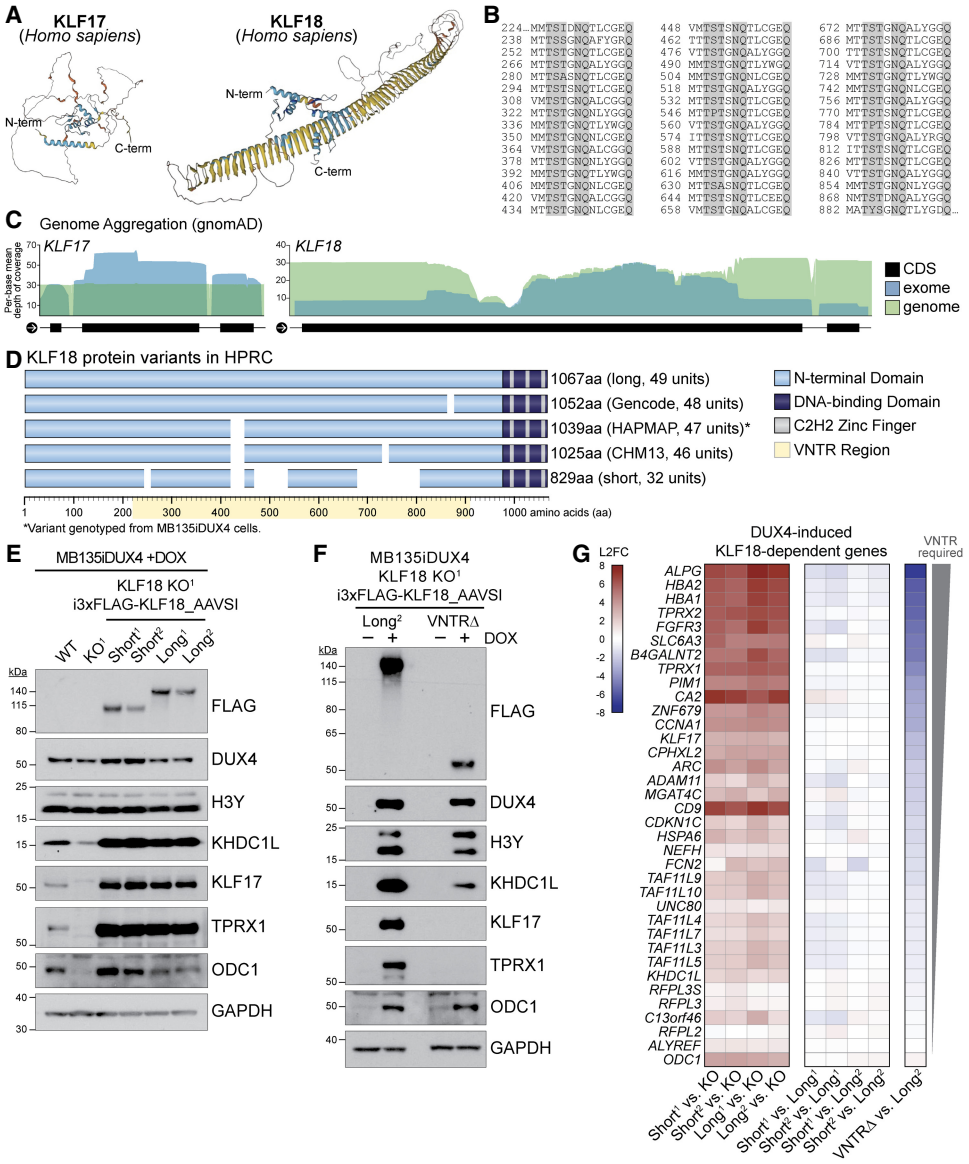


Figure 5. Structural predictions and activity of naturally occurring KLF18 variants in the human population. (*A*) AlphaFold-predicted tertiary protein structures of human KLF17 and KLF18. (*B*) Fourteen amino acid repeat sequences comprising the N-terminal region of KLF18 exhibiting a motif with recurrent polar uncharged amino acids (highlighted in gray). (*C*) Schematic from the Genome Aggregation Database profiling *KLF17* and *KLF18* exome and genome gene coverage, with transcript coding sequence (CDS) and orientation noted *below*. (*D*) Schematic of five identified KLF18 protein variants from the Human Pangenome Reference Consortium (HPRC) (see Supplemental Table S5); anticipated amino acid length based on DNA coding sequence is shown. Units = number of ~14 amino acid repeats in the VNTR. (*E*,*F*) Immunoblot analysis of MB135iDUX4 myoblasts with wild-type *KLF18* (WT) or *KLF18* knockout (KO^1) alleles and two *KLF18* KO^1-derived monoclonal cell lines expressing doxycycline-inducible 3xFLAG-tagged KLF18 short or long variants (*E*) or KLF18 long sequence without the VNTR region (Δ) from integrated transgenes at the *AAVS1* locus (*F*). Cells were treated ±0.25 µg/mL doxycycline (DOX) for 18 h. (*G*) Heat maps of RNA-seq analysis showing differential gene expression of cell lines treated with doxycycline (all conditions +0.25 µg/mL DOX for 18 h). Data are graphed as $\log_2$ fold change; $n = 3$ biological replicates per condition (see Supplemental Tables S6, S7).

KLF18 KO cells with a 3xFLAG-KLF18 long variant transgene lacking the entire ~2 kb VNTR region (VNTRΔ). Immunofluorescence confirmed that the VNTR deletion did not disrupt KLF18 nuclear localization (Supplemental Fig. S7A). However, deletion of the VNTR attenuated DUX4-mediated KLF18-dependent induction of KHDC1L expression and completely abrogated activation of KLF17 and TPRX1, while induction of ODC1 was unaffected (Fig. 5F; Supplemental Fig. S7B).

We performed RNA-seq on cells expressing each KLF18 variant and conducted pairwise short versus long comparisons across all four cell lines to control for clonal

variability (Supplemental Table S6). Of the 57 previously identified DUX4-induced KLF18-dependent genes, 36 were rescued with induction of DUX4 and 3xFLAG-KLF18 expression from the AAVS1 locus in at least one of the four clonal cell lines ($\log_2$ fold change ≥ 1; $P\text{adj} \leq 0.01$). The partial rescue may be attributed to differences in KLF18 expression relative to endogenous levels, altered temporal dynamics of induction, or consequences of the FLAG tag. Notably, none of these 36 genes showed significant differential expression in both clones when comparing the long and short variants ($|\log_2$ fold change $\geq 1|$; $P\text{adj} \leq 0.01$) (Fig. 5G). While our initial tests indicate no robust differences in activity between short and long KLF18 variants, we cannot exclude the possibility of differential activity when endogenously expressed in a physiologically relevant environment.

Comparison of the KLF18 long variant with the VNTRΔ construct showed that the VNTR region contributes to full transcriptional activation, to varying extents, of all 36 DUX4-induced KLF18-rescued genes except for *ODC1* (Fig 5G; Supplemental Table S7). These findings support a model in which the VNTR functions as a modulatory domain that mediates KLF18 transcriptional activity at most, but not all, target genes. Further studies will be required to define the molecular mechanisms by which the KLF18 VNTR contributes to target-specific transcriptional activation.

Assessment of KLF18 structural variants as modifiers of FSHD1 disease severity

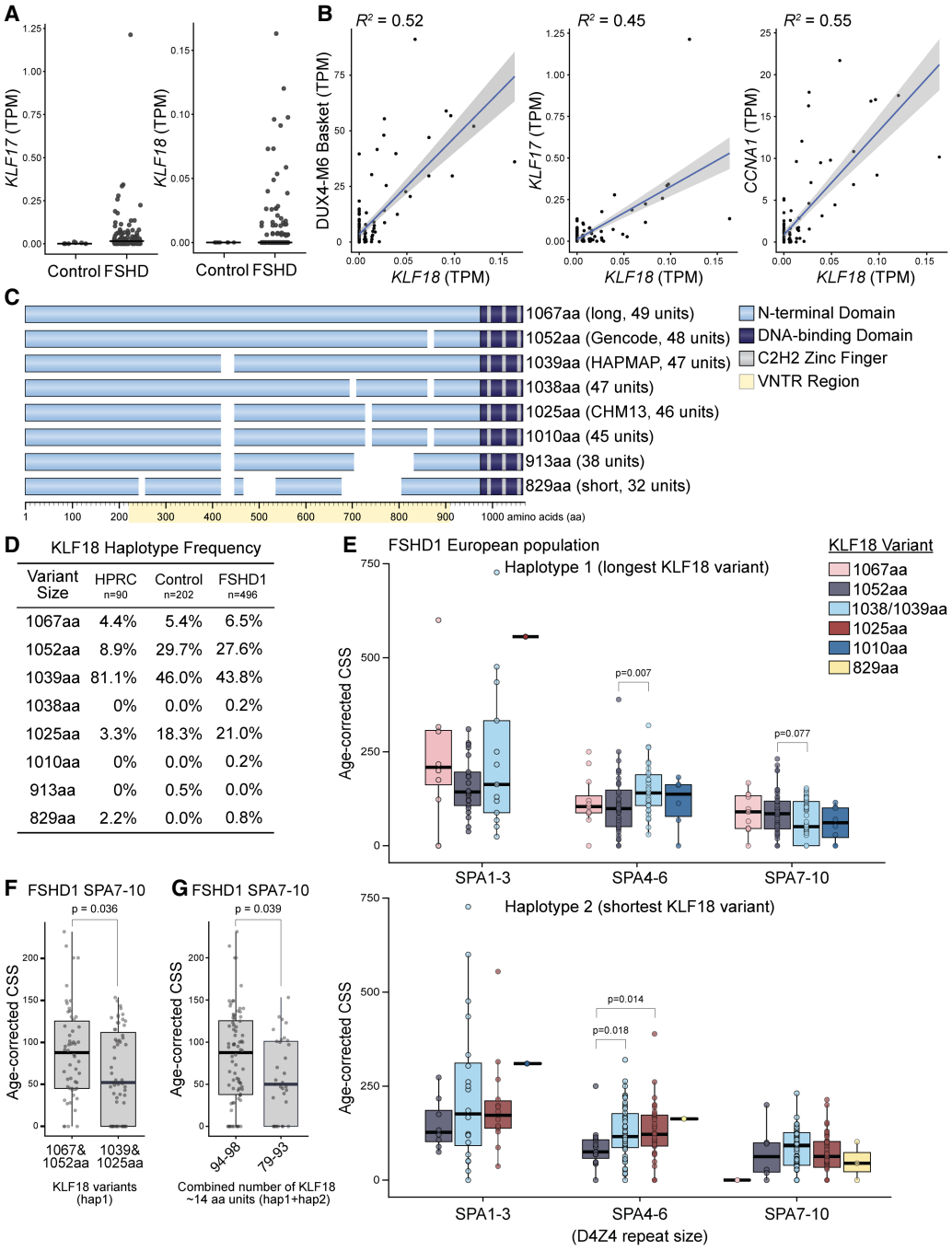
RNA-seq analysis of prior data sets from control and FSHD muscle biopsies confirmed that *KLF18* and *KLF17* mRNAs are enriched in FSHD muscle (Fig. 6A). In addition, *KLF18* expression was correlated with the expression levels of a basket of DUX4-regulated genes (*ZSCAN4*, *CCNA1*, *PRAMEF5*, *KHDC1L*, *MBD3L2*, and *H3Y*) previously used to measure FSHD disease activity (Wong et al. 2024) and with the expression of *KLF17* and *CCNA1* individually (Fig. 6B). These findings suggest that KLF18 regulates a portion of the DUX4 program in muscle from FSHD subjects and might contribute to disease severity.

To determine whether structural variants of KLF18 might modify the severity of FSHD, we identified the KLF18 structural variants in FSHD1 and control subjects that had associated D4Z4 haplotype genotyping and age-corrected clinical severity scores (ACSSs) (Ricci et al. 1999; van Overveld et al. 2005) by long-read sequencing of *KLF18* locus PCR amplicons (Supplemental Table S8). This analysis focused on samples from a predominantly European population because a similar set of non-European samples was not available. FSHD1 was genetically defined as having a short permissive allele (SPA) of 10 D4Z4 repeats or less, and control was genetically defined as having a SPA >10 repeats or no permissive allele (NP) and no clinical findings of muscular dystrophy (Ricci score = 0, ACSS = 0). Sufficient sequence depth and quality were obtained from 248 FSHD1 and 101 control subjects. The control group was enriched for family members of the FSHD1 subjects. The structural variants previously identified in the HPRC cohort were also the most prevalent in this cohort, with comparable frequencies and the addition of three new variants, each occurring once in the combined FSHD1 and control populations (Fig. 6C,D). All variants had the same translation start and stop sites and varied by the number and/or location of the VNTRs in the predicted β-solenoid.

FSHD1 severity is generally inversely correlated with the number of D4Z4 repeat units on the shortest permissive allele (SPA), with one to three D4Z4 repeats showing the earliest onset and highest severity, four to six repeats showing moderate disease onset and severity, and seven to 10 repeats showing incomplete penetrance and generally lower severity. However, within each of these groups, there is considerable variation in disease severity even within the same family, suggesting that there might be modifier loci for disease severity segregating independently from the permissive D4Z4 allele. To determine whether specific *KLF18* structural variants are associated with FSHD1 severity, we compared the ACSS for each FSHD1 subject with their *KLF18* structural variants for haplotype 1 (hap1), the longest of the two *KLF18* alleles, or haplotype 2 (hap2), the second allele of equal or shorter length (Fig. 6E). Note that this analysis did not assess the effects of individual amino acid variants or noncoding intronic variants. In the SPA 7–10 group, the group most susceptible to modifier loci, analyzing the longest *KLF18* alleles (hap1) revealed a trend where shorter structural variants were associated with lower severity scores (Fig. 6E). Within this group (hap1, SPA7–10), we compared the combined severity scores for the two longest variants (1067 and 1052 amino acids) with those of the shorter variants (1039 and 1025 amino acids) and showed a decreased average severity score associated with shorter *KLF18* variants ($P = 0.036$, Brunner–Munzel test) (Fig. 6F). Similarly, using a total KLF18 repeat score for each subject by adding the number of repeat units in both alleles showed lower severity scores in the group of subjects with 79–93 repeats compared with the group with 94–98 repeats ($P = 0.039$, Brunner–Munzel test) (Fig. 6G). Together, this suggests that in the SPA 7–10 group, the shorter structural variants are associated with milder disease. In contrast, however, the opposite is apparent in the SPA 4–6 group, where there was an association of decreased disease severity with the longer 1052 allele compared with the shorter alleles in the SPA 4–6 cohort ($P \leq 0.05$, Brunner–Munzel test) (Fig. 6E). We conclude that additional studies will be necessary to determine whether KLF18 structural variants modify disease penetrance in FSHD.

Discussion

The KLF/SP family of transcription factors is highly versatile, regulating a wide array of biological processes and having distinct roles depending on the tissue and cellular context. While human *KLF17* has been reported to be expressed in the early embryo (Yan et al. 2013; Blakeley et al. 2015), 8CLCs (Taubenschmid-Stowers et al. 2022), and



Variant Size	HPRC n=90	Control n=202	FSHD1 n=496
1067aa	4.4%	5.4%	6.5%
1052aa	8.9%	29.7%	27.6%
1039aa	81.1%	46.0%	43.8%
1038aa	0%	0.0%	0.2%
1025aa	3.3%	18.3%	21.0%
1010aa	0%	0.0%	0.2%
913aa	0%	0.5%	0.0%
829aa	2.2%	0.0%	0.8%

Figure 6. KLF18 variants in an FSHD1 population. (*A*) Gene expression levels of *KLF17* and *KLF18* in control muscle biopsies relative to FSHD muscle (transcripts per million [TPM]). (*B*) Scatter plots of gene expression levels between a basket of six DUX4 target genes (M6: *ZSCAN4*, *CCNA1*, *PRAMEF5*, *KHDC1L*, *MBD3L2*, and *H3Y*) and *KLF17* and *CCNA1* expression. Pearson correlation was calculated based on TPM. RNA-seq data were sourced from Wong et al. (2024). (*C*) Schematic of KLF18 protein variants identified in control (101 subjects) and FSHD1 (248 subjects) cohorts; anticipated amino acid length based on DNA coding sequence is shown. Units = number of ~14 amino acid repeats in the VNTR. (*D*) Frequency of KLF18 variant occurrence in the HPRC data set and in our control and FSHD1 *KLF18* long-read PacBio sequencing data set (*n* = number of haplotypes) (see Supplemental Table S8). (*E*) Box plots showing KLF18 variant distribution correlated with age-corrected clinical severity score (CSS) in FSHD1 cohorts grouped by D4Z4 repeat array size on the 4qA permissive haplotype (shortest permissive allele [SPA]). The two KLF18 alleles from each individual were graphed separately, with the longest variant (haplotype 1) shown in the *top* panel and KLF18 variants of equal size or shorter (haplotype 2) shown in the *bottom* panel. (*F*,*G*) Box plots showing the longest KLF18 variant (as depicted in *E* for haplotype 1 SPA7-10 group) classified by amino acid size (*F*), or the combined number of KLF18 tandem ~14 amino acid repeats from both KLF18 haplotypes (hap1 + hap2), correlated with age-corrected clinical severity score (CSS) in an FSHD1 European population cohort with SPA7-10 D4Z4 repeats (*G*). *P*-values are based on Brunner–Munzel test.

several tissues and cancer cells (Gumireddy et al. 2009; Dong et al. 2014; Ali et al. 2015a,b; Zhou et al. 2016; Li et al. 2021), prior to our study, the status of the previously recognized *KLF18* gene had not been experimentally investigated (Pei and Grishin 2013). We discovered that DUX4 induces expression of *KLF18*, which subsequently activates *KLF17* in muscle cells and human embryonic stem cells. Here, we characterize the transcriptional program of KLF17 and KLF18 in both the presence and absence of DUX4. While KLF17 and KLF18 were sufficient to activate a small subset of DUX4-induced genes on their own, the loss of KLF18 expression downstream from DUX4 led to much broader changes in gene expression. This implies that KLF18 acts within a feed-forward gene regulatory program driven by DUX4, potentially facilitated by a specific cellular or chromatin environment.

In agreement with KLF18 transcriptional regulation downstream from DUX4, our genome-wide ChIP-seq analysis of 3xFLAG-KLF18 binding confirmed direct binding at a subset of DUX4-induced genes. De novo motif enrichment analysis indicated that KLF18 likely binds to a GC-rich motif, similar to other members of the KLF/SP family having conserved zinc finger DNA-binding domains, and also showed enrichment for PRD-like homeobox motifs. Several DUX4-induced genes had both a DUX4 and KLF18 ChIP peak, suggesting that KLF18 and DUX4 colocalize at a subset of loci. Despite similarities in DNA motifs shared by multiple KLF family members, KLF proteins differentially regulate gene expression, likely through their divergent N-terminal domains and interaction with different cofactors, activators, or repressors (Kaczynski et al. 2003). This is supported by our finding that KLF18 differentially regulated many more genes than KLF17 did downstream from DUX4 even though their DNA-binding domains are 45% identical. Furthermore, many KLF18-bound regions did not show changes in gene expression, which is true for many other transcription factors (MacQuarrie et al. 2011); for example, DUX4, MyoD, and NF-kB (Sen and Baltimore 1986; Cao et al. 2010; Young et al. 2013; Zhang et al. 2017). This may indicate that KLF18 has context-dependent roles in ensuring proper timing of cell fate transitions, such as priming genes for future expression, organizing chromatin structure, or blocking/recruiting other factors.

In some mammalian species, previously predicted KLF18 proteins were found to contain a repeated 14 amino acid segment in their N-terminal regions (Pei and Grishin 2013). Structural modeling of the human KLF18 protein sequence predicted these repeats to form β-strands that are stabilized by hydrogen bonds and assemble into a β-solenoid-like architecture. β-Solenoids are uncommon in transcription factors and are typically found in bacterial virulence factors, viral adhesins, antifreeze proteins, and functional amyloids, where their elongated surface provides a stable platform for binding substrates and cofactors (Kajava and Steven 2006). This structural feature may act as a scaffold for recruiting epigenetic and/or transcriptional cofactors, highlighting the need for future studies on the complex protein structure of human KLF18.

While a similar repeat sequence and predicted β-solenoid structure are present in distantly related placental mammals, such as bats, two-toed sloths, and elephants, these repeats are not present in rodents and are present only to a small degree in dogs and horses. This suggests that many species have selected against the extended β-solenoid repeat structure in their KLF18 orthologs, though convergent evolution is an alternative model. In this regard, it is interesting that the number of 14 amino acid repeats is variable in the human population. Approximately 3% of the human genome consists of regions with variable numbers of tandem repeats (VNTRs; motif arrays ≥6 bp), found primarily in noncoding regions and infrequently in protein-coding genes (Ren et al. 2023). To date, VNTRs have been understudied due to challenges in detecting them with short-read sequencing. In our analysis of the HPRC long-read sequencing data, we identified VNTRs encoded within the third exon of *KLF18* that give rise to multiple structural variants that are distributed across dispersed human ancestries. This might mean that the most prevalent variants originated early in human evolution or that only a subset of variants is newly generated or maintained in the population and recurs in different ethnic backgrounds. Further population studies of *KLF18* variants with linked single-nucleotide polymorphisms (SNPs) will be necessary to understand the genetic origins of these repeat variations in the human population.

Although short and long KLF18 structural variants exhibited largely similar transcriptional activity in MB135i-DUX4 myoblasts, deletion of the entire VNTR region impaired KLF18-mediated expression of most DUX4-induced KLF18-rescued genes. We therefore examined whether variation in the *KLF18* VNTR region was associated with FSHD1 disease severity. While some subgroups showed differences in disease severity, we did not identify consistent associations of individual *KLF18* structural variants with disease severity. It should be noted that the age-adjusted severity score (ACSS) is an imperfect measure of disease severity, albeit it is the best currently available for this data set. In addition, this study was limited to studying the structural variants based on the number of repeats and did not incorporate other SNPs or indels in coding or noncoding regions. Future studies comparing disease penetrance in kindreds or using better measures of total disease burden will be necessary to better assess the association of *KLF18* haplotypes with FSHD disease severity.

The recent discovery of 8 cell-like cells (8CLCs) capturing human ZGA in naive hESCs (Mazid et al. 2022; Taubenschmid-Stowers et al. 2022) marks a major advancement since the initial identification of totipotent 2CLCs in an in vitro mouse model over a decade ago (Macfarlan et al. 2012). Despite this breakthrough, the transcriptional networks involved remain largely uncharacterized. Our study demonstrates that *KLF18* participates in a feed-forward network initiated by DUX4, wherein DUX4 induces the expression of KLF18, and many totipotency-associated genes require both DUX4 and KLF18 to be fully activated. Feed-forward networks were recognized as the most common regulatory mechanism for achieving temporal gene expression in yeast

(Lee et al. 2002; Mangan and Alon 2003), are central to the function of master regulators of cell state transitions (Penn et al. 2004), and have previously been described for other factors downstream from DUX4 (Kong et al. 2024). In contrast to a simple cascade of sequential gene activation, feed-forward networks precisely regulate both the initiation and termination of a gene network. This is critical for transient cell states, such as coordinating the transition from the totipotent to the pluripotent state in the early embryo. Brief expression of DUX4 in the cleavage stage embryo initiates a feed-forward network, activating the zygotic genome and establishing a totipotent program. Timely dissolution of this transcriptional program is essential to allow the embryo to progress and define extraembryonic and embryonic cell lineages. Transcription factors activated by DUX4—including KLF18 and its target genes, such as KLF17—may contribute to this process by exerting suppressive effects within the network, functioning as a nested incoherent feed-forward loop (Goentoro et al. 2009).

Our results collectively show that KLF17 and KLF18 are components of the DUX4 feed-forward transcriptional network, aiding in the activation of critical developmental genes and contributing to the dysregulation of gene expression in FSHD. Further research is needed to elucidate the molecular mechanisms driving their actions. A better understanding of transcriptional networks downstream from DUX4 could expose new avenues of research and possible therapeutic strategies. Thus, our study has broad implications for FSHD research and stem cell reprogramming and may provide important insights into the regulation of zygotic genome activation during human preimplantation development and the origins of placental mammals.

Materials and methods

Cell culture

Human embryonic stem cells H9 hNanog-pGZ (WiCell Research Institute) were grown in mTeSR Plus on Matrigel-coated plates. MB135 myoblasts were grown in Ham's F-10 supplemented with 10% FBS, 1% penicillin/streptomycin, 10 ng/mL rhFGF, and 1 µM dexamethasone. Differentiation of FSHD myotubes was achieved by switching myoblasts grown to confluence into DMEM, 1% penicillin/streptomycin, and 1× ITS-G for 48–72 h. Cell lines carrying puro-resistant transgenes were maintained in 2–3 µg/mL puromycin. Cells were treated with 0.25–1 µg/mL doxycycline where specified. All cell lines were cultured at 37°C in a humidified incubator supplied with 5% CO_2. Cell culture reagents are listed in Supplemental Table S9.

Cell lines

H9-hNanog-pGZ_iDUX4 cells were generated like the previously reported iDUX4 lines originally outlined by Jagannathan et al. (2016). Briefly, codon-altered *DUX4* was cloned into pCW57.1 downstream from a doxycycline-inducible promoter. Lentiviral particles were created by transfecting 293T cells with a subcloned expression vector, the psPAX2 packaging vector, and the pMD2.G envelope vector using Lipofectamine 3000 according to the manufacturer's instructions. H9 hNanog-pGZ cells were transduced and subsequently selected and maintained in puromycin.

Generation of MB135iDUX4 knockout cell lines was achieved using guide RNA (gRNA) sequences targeting *KLF18* cloned into the Cas9(BB)-2A-GFP plasmid. KLF18 gRNA_1 and gRNA_2 were used to generate KLF18 KO^1 and KO^2, respectively (see Supplemental Fig. S1B). Cells were transfected with gRNA:Cas9-GFP constructs using Lipofectamine 3000 reagent according to the manufacturer's protocol and incubated for 1.5 days. Clonal cell lines were isolated using fluorescence-activated cell sorting (FACS) for Cas9-GFP.

Generation of hESC H9-hNanog-pGZ_iDUX4 *KLF17* knockout cell lines was achieved using KLF17 gRNA_1 (see Supplemental Fig. S5) cloned into the Cas9(BB)-2A-GFP and Lipofectamine-based transfection according to the manufacturer's protocol, and cells were incubated for 1.5 days in mTeSR Plus supplemented with 5 µM ROCK inhibitor (Y-27632). Generation of H9-hNanog-pGZ_iDUX4 *KLF18* knockout cell lines was achieved using 125 pmol of Cas9-RFP protein complexed with 150 pmol of TrueGuide synthetic KLF18 gRNA_2 (see Supplemental Fig. S5) at a 1:1.2 ratio of Cas9 to gRNA in a 5 µL volume. The Cas9:gRNA RNA protein (RNP) complex was incubated for 10–20 min at room temperature. H9-hNanog-pGZ-iDUX4 were lifted with Versene, rinsed with PBS, and resuspended in 25 µL of Neon NxT resuspension genome editing buffer. The RNP complex was added to the cell suspension, electroporated using the 10 µL Neon NxT tips and electroporation system (two pulses at 1050 V for 30 msec), transferred to a Matrigel-coated plate containing prewarmed transfection culture media (DMEM/F12 supplemented with 2 mM L-glutamine, 20% KO serum replacement, 1× MEM NEAA solution, 1× BME, 4 ng/mL FGF, 2 µg/mL puro, CloneR), and incubated for 1 day. Clonal cell lines were isolated using fluorescence-activated cell sorting for either Cas9-GFP or Cas9-RFP.

Generation of 3xFLAG-tagged *KLF18* cell lines was achieved using the Alt-R CRISPR–Cas9 system and HDR donor oligos (IDT) as outlined in the manufacturer's protocol. Briefly, 125 pmol of Cas9-RFP protein was complexed with 150 pmol of TrueGuide synthetic KLF18 gRNA_2 at a 1:1.2 ratio of Cas9 to gRNA in a 5 µL volume and incubated for 10–20 min at room temperature. MB135iDUX4 cells were washed in PBS and resuspended in 20 µL of Neon NxT resuspension genome editing buffer. The 5 µL RNP complex was combined with 1.2 µL of a single-stranded DNA donor oligo (100 µM HDR donor oligo) encoding a 3xFLAG tag sequence flanked by a >50 bp homology arm with silent mutations in the gRNA seed and PAM sequence, 20 µL of the cell suspension, and 3.8 µL of PBS. The transfection mix was electroporated using the 10 µL Neon NxT tips and electroporation system (one pulse at 1400 V for 30 msec), transferred to a

plate containing prewarmed antibiotic-free culture media supplemented with 1 µM HDR enhancer, and incubated for 1 day. Clonal cell lines were isolated using fluorescence-activated cell sorting for Cas9-RFP.

MB135iDUX4 *KLF18* KO cell lines with doxycycline-inducible 3xFLAG-KLF18 (i3xFLAG-KLF18) wild-type sequence variants or VNTRΔ were generated using CRISPR/Cas9-mediated genomic integration at the AAVS1 safe harbor locus (Sadelain et al. 2011) via homology-directed repair. A multistep cloning strategy was used to generate the i3xFLAG-KLF18_BlastR_AAVS1 donor template plasmid. First, PCR and the HiFi assembly were used to insert transgenes coding for the 3xFLAG-KLF18 variants upstream of a hPGK promoter driving a blasticidin resistance marker into a vector backbone containing a CMV enhancer and promoter and homology arms to the AAVS1 safe harbor locus using the AgeI and KpnI sites to linearize the vector (derived from Addgene 158663). Second, restriction digest cloning at the NheI and AgeI sites was used to replace the CMV enhancer and promoter with a Tet-on 3G system. MB135iDUX4_KLF18 KO[1] cells were cotransfected with AAVS1-sgRNA-Cas9-expressing plasmid and the i3xFLAG-KLF18_BlastR_AAVS1 plasmids described above at a 1:2 ratio using Lipofectamine 3000 according to the manufacturer's protocol. Cells were incubated for 2 days after transfection, selected, and maintained in 10 µg/mL blasticidin. Monoclonal cell lines were isolated, and i3xFLAG-KLF18 transgene integration at the AAVS1 safe harbor locus was validated with PCR amplification from genomic DNA followed by sequencing.

Supplemental Table S9 lists all oligonucleotide sequences used for CRISPR-based gene editing and transgenes; all mutant alleles were validated with PCR and sequencing.

siRNA transfections

Transfections of siRNAs were carried out using Lipofectamine RNAiMAX according to the manufacturer's protocol. MB135iDUX4 cells were transfected with 25 pmol/mL of either target or nontargeting control siRNA for 16–20 h before each doxycycline treatment. FSHD cells were subjected to a double-transfection protocol as outlined by Resnick et al. (2019). siRNA sequences are listed in Supplemental Table S9.

Cloning and transient transfections

3xFLAG-tagged codon-altered coding sequence for KLF17 was synthesized as a custom gBlock from Integrated DNA Technologies, and coding sequences for KLF18 constructs were synthesized by GenScript (Supplemental Table S9). Constructs were cloned into expression vectors with a constitutive promoter (pCS2 or pcDNA3.1). The 10 µL Neon NxT tips and electroporation system were used to transfect plasmids according to the manufacturer's instructions (two pulses of 1400 V for 20 msec each), and cells were harvested 48 h after transfection.

For luciferase reporters, the *TPRX1* promoter and 5′ UTR sequence (812 bp upstream of translation start site) (Supplemental Table S9) was synthesized as a custom gBlock from GenScript and cloned into the pGL3-basic luciferase vector at the multiple cloning site using NheI and XhoI enzymes. The *ODC1* promoter and 5′ UTR sequence (800 bp upstream translation start site) (Supplemental Table S9) was PCR-amplified from MB135 genomic DNA using primers with compatible ends for restriction digest cloning at the multiple cloning site using KpnI and XhoI enzymes. The *ZSCAN4* promoter reporter was described previously and is available from Addgene (plasmid 69249). Uninduced MB135iDUX4_KLF18 KO[1] myoblasts were seeded in a 6 well plate and transfected with 0.25 µg of luciferase reporter plasmid, 0.05 µg of pCS2-Renilla, and 0.2 pmol of 3xFLAG-KLF17, 3xFLAG-KLF18, DUX4, or GFP pCS2-expression constructs using Lipofectamine 3000 according to the manufacturer's instructions, and cells were harvested 24 h after transfection. Plasmids and transgene sequences are listed in Supplemental Table S9.

Immunoblotting

Protein samples were harvested in RIPA buffer (150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1% SDS, 25 mM Tris-HCl at pH 7.4) supplemented with protease and phosphatase inhibitor tablets, followed by sonication for 10 min on low for 30 sec on/off using a Bioruptor standard sonicator. Lysate was cleared by centrifugation at 16,000*g* and quantified using a Pierce BCA assay. Samples were run on NuPAGE precast polyacrylamide gels and transferred to PVDF membranes. Membranes were blocked in TBS containing 0.1% Tween-20 and 5% nonfat dry milk before overnight incubation at 4°C with primary antibodies. Membranes were incubated with HRP-conjugated secondary antibodies for 1–2 h at room temperature, and chemiluminescent substrate was used for detection on film. Antibodies and reagents are listed in Supplemental Table S9.

Immunofluorescence

Cells were fixed with 2%–4% paraformaldehyde for 10 min, permeabilized 0.5% Triton X-100, incubated with primary antibodies overnight at 4°C, incubated again with fluorescently conjugated secondary antibodies for 2 h at room temperature, and counterstained with DAPI. Plates were imaged using an immersion lens, a Zeiss AxioPhot fluorescent microscope, an AxioCam MRc digital camera, and AxioVision 4.6 software.

RT-qPCR

Total RNA was isolated using the NucleoSpin RNA kit according to the manufacturer's instructions. Isolated RNA was treated with Amp-grade DNase I, heat-inactivated in the presence of EDTA, and reverse-transcribed into cDNA using a SuperScript IV first strand synthesis system following the manufacturer's protocol. Quantitative PCR was carried out on a QuantStudio 7 Flex using

iTaq SYBR Green supermix. Primers are listed in Supplemental Table S9.

RNA sequencing

Total RNA was isolated using the NucleoSpin RNA kit according to the manufacturer's instructions. RNA was submitted to the Fred Hutchinson Cancer Center Genomics Core for library preparation using the TruSeq stranded mRNA kit (Illumina) followed by size and quality analysis by TapeStation (Agilent). Libraries were sequenced on a NovaSeq 6000 using an S1 flow cell using a 150 bp paired-end reagent kit (Illumina). Read quality was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc), reads were aligned to the GRCh38 p14 primary reference human genome using Rsubread (Liao et al. 2019), gene features were called using featureCounts (Liao et al. 2014) and the GENCODE R45 annotation, and differential gene expression was analyzed using DESeq2 (Love et al. 2014).

ChIP sequencing

Cells were treated with 1 µg/mL doxycycline for 16 h. Approximately 12 million myoblasts were cross-linked for 10 min at room temperature in 1% formaldehyde, quenched by adding glycine to 125 mM final concentration, and incubated for 5 min at room temperature. Cells were washed twice with PBS and resuspended in 600 µL of ice-cold ChIP buffer (5 mM PIPES, 85 mM KCl, 0.5% NP-40 adjusted to pH 8.0 with 1 M KOH at 4°C and sterile filtered) supplemented with EDTA-free protease and phosphatase inhibitor tablets. Samples were vortexed for 15 sec, incubated for 10 min on ice, and pelleted at 9000*g* for 3 min at 4°C. The supernatant was removed, and nuclear pellets were resuspended in 600 µL of MNase digestion buffer (50 mM Tris-HCl, 5 mM CaCl2 at pH 7.9, supplemented with 100 µg/mL BSA solution, 1 mM DTT). MNase digestion was optimized to obtain 200–700 bp chromatin fragments using 600 gel units of MNase in 600 µL and incubated with mixing at 1000 rpm for exactly 10 min at 37°C. We used MNase digestion instead of sonication for reproducible chromatin fragmentation because KLF18 is a large protein and its N-terminal 3xFLAG tag could be vulnerable to degradation with sonication (Pchelintsev et al. 2016). MNase has been shown to provide high-resolution mapping of transcription factor binding sites (Henikoff et al. 2011). EDTA (50 mM at pH 8) was added to terminate MNase digestion, vortexed briefly, and incubated for 5 min on ice. Samples were centrifuged at 9000*g* for 5 min at 4°C, supernatant removed, and nuclear pellets were resuspended in 600 µL of ChIP dilution buffer (1.1% Triton X-100, 167 mM NaCl, 1.2 mM EDTA at pH 8, 16.7 mM Tris-HCl at pH 8, 0.25% SDS, 0.1% NP-40, 0.1% Na-deoxycholate) supplemented with protease and phosphatase inhibitor tablets. Nuclei were disrupted by sonication in TPX tubes for 20 min on low for 30 sec on/off using a Bioruptor standard sonicator. Samples were centrifuged at 9000*g* for 10 min at 4°C to clear insoluble material. Chromatin was diluted to bring the final concentration of SDS to <0.01%, 10 µL of rab-anti-FLAG mAb IgG (Cell Signaling Technology 14793S) was added to ~15 µg of chromatin, and samples were rotated overnight at 4°C. Fifty microliters of BSA-blocked Protein A/G magnetic beads was added to each sample and rotated for 2 h at 4°C. Beads were washed as follows: once each with low-salt wash buffer (1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl), high-salt wash buffer (1% Triton X-100, 0.1% SDS, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl), and LiCl wash buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl) and twice with PBS. Beads were resuspended in ultrapure water with 1% SDS and 0.1 M $NaHCO_3$ final concentration. Samples were treated with 20 mg of RNase A and incubated with shaking at 1000 rpm for 1 h at 37°C. To decross-link, 100 mg of Proteinase K was added, and samples were incubated with shaking at 1000 rpm overnight at 65°C. DNA was isolated with phenol-chloroform extraction and ethanol precipitation. The ThruPLEX DNA-seq kit and DNA Unique dual-index kit were used for library preparation. Libraries were size-selected with SPRI beads (1× bead:DNA ratio), quantified with qPCR, pooled, and sequenced on a NovaSeqX using a 10B flow cell and 100 bp paired-end reads (Illumina).

Raw ChIP-seq reads were initially trimmed using Trim Galore~ 0.6.7 (https://www.github.com/felixkrueger/trimgalore; Martin 2011) and then processed using ENCODE's transcription factor and histone ChIP-seq processing pipeline (Hitz et al. 2023). Peak calling was done as part of the ENCODE pipeline using SPP, and conservative IDR thresholded peaks were chosen for further analysis. MEME-ChIP v5.5.7 (Machanick and Bailey 2011) was used to analyze motifs enriched in the genomic coordinates identified for the 958 3xFLAG-KLF18 ChIP peaks (q-value < 0.05). Primary sequences were specified based on the UCSC mammal genome human hg38. The HOCOMOCO DNA H12CORE motif database was used to identify enriched motifs based on MEME and CentriMo.

Pan-genome analysis

We took the separate maternal and paternal phased genomes and annotation files from the Human Pan-Genome Consortium assemblies and extracted the annotated KLF18 mRNA sequence using gffread (Pertea and Pertea 2020). Sequences were aligned using the MUSCLE alignment algorithm in SnapGene (Edgar 2004). Individual sample metadata and KLF18 gene coordinates for a total of 90 entries are in Supplemental Table S5 and include the following: (1) GRCh38 p14/GENCODE R45 KLF18 (single entry, as GRCh38 is unphased), (2) CHM13 (single entry, as CHM13 is a haploid line), and (3) 88 maternal and paternal genomes assembled by the Human Pangenome Reference Consortium (Liao et al. 2023). The following sources were used: GitHub data sources (https://www.github.com/human-pangenomics/HPP_Year1_Assemblies), annotation and raw data source (https://s3-us-west-2.amazonaws.com/human-pangenomics/index.html), and reference genomes downloaded from

NCBI Genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_018472595.1).

AlphaFold

AlphaFold3 (Abramson et al. 2024) was used to output protein structural prediction models of human KLF17 and KLF18 orthologs as reported by NCBI.

KLF18 PacBio amplicon analysis of control and FSHD cohorts

KLF18 PCR amplicons were generated from genomic DNA of FSHD-genotyped and control individuals. Sample preparation for the PacBio Sequel II run was done by two PCR amplification steps using the Phusion PCR kit with 10% DMSO added. The first PCR step involved 35 cycles of 10 min at 98°C, 10 min at 68°C, and 5 h 72°C using tagged primers KLF18-F2 (/5AmMC6/gcagtcgaacatgtagctgactcaggtcacACTCTGTGTGTGGAGGGGACACTGG) and KLF18-R4 (/5AmMC6/tggatcacttgtgcaagcatcacatcgtagCAACCTCAATAGGGAGAATCTTTTAGCTGTACACAT) at 0.25 µM each, 0.01 U/µL enzyme, and 4 ng of gDNA template in a 10 µL reaction. The second step involved individual barcoding by amplification in 10 µL of PCR (buffer composition as in step 1) with 2 µL of 200× diluted PCR product from the first step for seven cycles of 10 min at 98°C, 10 min at 62°C, and 5 h at 72°C PCR using PacBio unique barcode primers for each sample. All samples were pooled and prepared for PacBio according to the standard procedure and run on two Sequel II chips.

CCS bam files were converted to FASTQ using PacBio BAM toolkit bam2fastq (pbtk v3.5.0), and then PacBio amplicon analysis (pbAA v1.2.0) was used to predict phased allele sequences with the forward gene-specific primer used as guide sequence. Samples with one pbAA-derived consensus sequence were assigned homozygous genotypes by default. Phased allele sequences were processed in R (R Core Team 2021) using BioStrings (https://www.bioconductor.org/packages/release/bioc/html/Biostrings.html) and ShortRead (Morgan et al. 2009) packages, and ORF prediction was done using findORFs from the ORFik package (Tjeldnes et al. 2021) after identifying and removing the intron 1 sequence from each ORF region. Plots were generated using ggplot from the tidyverse package (Wickham et al. 2019) and modified with the ggprism template (https://csdaw.github.io/ggprism).

This study was approved by the relevant Medical Ethical Committees from participating institutions. For controls, we analyzed independent European samples collected at the Leiden University Medical Center. For FSHD1, we analyzed 248 carriers of an FSHD allele for which we have a clinical severity score. Clinical evaluation was performed by an experienced neurologist after informed consent. For the clinical severity, we used the age-corrected severity score (ACSS) based on the 10-scale Ricci score: ACSS = (Ricci score/age at examination) × 1000. For all individuals, the D4Z4 repeat sizes and haplotypes have been determined previously. See Supplemental Table S8 for *KLF18* sequence data.

Data availability

RNA-seq and ChIP-seq data generated in this study have been deposited in the Gene Expression Omnibus under the following accession numbers: MB135iDUX4 siRNA knockdown RNA-seq (GSE316237), MB135iDUX4 CRISPR KO RNA-seq (GSE316239), MB135iDUX4_KLF18 KO_i3xFLAG-KLF18_AAVS1 short and long RNA-seq (GSE316240), MB135iDUX4_KLF18 KO_i3x-FLAG-KLF18_AAVS1 long and VNTRΔ RNA-seq (GSE316242), and 3xFLAG-KLF18 ChIP-seq (GSE299601).

Competing interest statement

S.J.T. and S.M.v.d.M. are board members of Renogenyx and have acted as consultants and/or members of advisory boards for several pharmaceutical companies developing therapeutics for FSHD. S.J.T., S.M.v.d.M., and R.J.L.F.L. are inventors on several FSHD-related patents and applications. A KLF18 patent application is under consideration by the Fred Hutchinson Cancer Center.

Acknowledgments

We thank the facioscapulohumeral muscular dystrophy community for their help and participation in the study. We thank Rolf Vossen for doing the PacBio run, and Davy Cats, bioinformatician, for performing the KLF18 PacBio amplicon analysis at the Leiden University Medical Center. This research was supported by the Flow Cytometry Shared Resource (RRID: SCR_022613) and the Genomics and Bioinformatics Shared Resource (RRID: SCR_022606) of the Fred Hutchinson Cancer Center/University of Washington/Seattle Children's Cancer Consortium (P30CA015704). This work was supported by grants from the National Institutes of Health (P50AR065139 and R01AR045203 to S.J.T.), the Friends of FSH Research (https://www.fshfriends.org) (S.J.T.), and the FSHD Society (S.M.v.d.M.).

Author contributions: D.C.H., A.B.O., S.R.B., P.J.v.d.V., R.J.L.F.L., S.M.v.d.M., and S.J.T. concevied the study. S.R.B. performed the computational analysis. D.C.H., A.B.O., R.J.L.F.L., and P.J.v.d.V. performed the experimentation. D.C.H., A.B.O., S.R.B., R.J.L.F.L., and P.J.v.d.V. performed the methodology. S.M.v.d.M. and S.J.T. acquired the funding. D.C.H. and S.R.B. visualized the data. D.C.H. and S.J.T. wrote the original draft of the manuscript. D.C.H. and S.J.T. reviewed and edited the manuscript. D.C.H. and S.J.T. were the project administrators. S.J.T. supervised the study.

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630:** 493–500. doi:10.1038/s41586-024-07487-w

Ali A, Bhatti MZ, Shah AS, Duong HQ, Alkreathy HM, Mohammad SF, Khan RA, Ahmad A. 2015a. Tumor-suppressive p53 signaling empowers metastatic inhibitor KLF17-dependent transcription to overcome tumorigenesis in non-small cell lung cancer. *J Biol Chem* **290:** 21336–21351. doi:10.1074/jbc.M114.635730

Ali A, Zhang P, Liangfang Y, Wenshe S, Wang H, Lin X, Dai Y, Feng XH, Moses R, Wang D, et al. 2015b. KLF17 empowers TGF-β/Smad signaling by targeting Smad3-dependent pathway to suppress tumor growth and metastasis during cancer progression. *Cell Death Dis* **6:** e1681. doi:10.1038/cddis.2015.48

Arends T, Hamm DC, van der Maarel S, Tapscott SJ. 2025. Facioscapulohumeral dystrophy: molecular basis and therapeutic opportunities. *Cold Spring Harb Perspect Biol* **17:**a041492. doi:10.1101/cshperspect.a041492

Blakeley P, Fogarty NM, del Valle I, Wamaitha SE, Hu TX, Elder K, Snell P, Christie L, Robson P, Niakan KK. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **142:** 3151–3165. doi:10.1242/dev.131235

Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, Simsek T, Garner HR, Agha-Mohammadi S, Tassin A, Coppée F, et al. 2008. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* **27:** 2766–2779. doi:10.1038/emboj.2008.201

Bosnakovski D, Gearhart MD, Toso EA, Ener ET, Choi SH, Kyba M. 2018. Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci Rep* **8:** 16957. doi:10.1038/s41598-018-35150-8

Cao Y, Yao Z, Sarkar D, Lawrence M, Sanchez GJ, Parker MH, MacQuarrie KL, Davison J, Morgan MT, Ruzzo WL, et al. 2010. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev Cell* **18:** 662–674. doi:10.1016/j.devcel.2010.02.014

Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, Alföldi J, Watts NA, Vittal C, Gauthier LD, et al. 2024. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625:** 92–100. doi:10.1038/s41586-023-06045-0

Chew G-L, Campbell AE, De Neef E, Sutliff NA, Shadle SC, Tapscott SJ, Bradley RK. 2019. DUX4 suppresses MHC class I to promote cancer immune evasion and resistance to checkpoint blockade. *Dev Cell* **50:** 658–671.e7. doi:10.1016/j.devcel.2019.06.011

Das S, Chadwick BP. 2016. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* **11:** e0160022. doi:10.1371/journal.pone.0160022

De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. 2017. DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* **49:** 941–945. doi:10.1038/ng.3858

Dong P, Kaneuchi M, Xiong Y, Cao L, Cai M, Liu X, Guo SW, Ju J, Jia N, Konno Y, et al. 2014. Identification of KLF17 as a novel epithelial to mesenchymal transition inducer via direct activation of TWIST1 in endometrioid endometrial cancer. *Carcinogenesis* **35:** 760–768. doi:10.1093/carcin/bgt369

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797. doi:10.1093/nar/gkh340

Flores-Fernández JM, Rathod V, Wille H. 2018. Comparing the folds of prions and other pathogenic amyloids. *Pathogens* **7:** 50. doi:10.3390/pathogens7020050

Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, van der Maarel SM, Ruzzo WL, Gentleman RC, Tawil R, et al. 2012. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell* **22:** 38–51. doi:10.1016/j.devcel.2011.11.013

Goentoro L, Shoval O, Kirschner MW, Alon U. 2009. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* **36:** 894–899. doi:10.1016/j.molcel.2009.11.018

Gumireddy K, Li A, Gimotty PA, Klein-Szanto AJ, Showe LC, Katsaros D, Coukos G, Zhang L, Huang Q. 2009. KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nat Cell Biol* **11:** 1297–1304. doi:10.1038/ncb1974

Hamanaka K, Šikrová D, Mitsuhashi S, Masuda H, Sekiguchi Y, Sugiyama A, Shibuya K, Lemmers R, Goossens R, Ogawa M, et al. 2020. Homozygous nonsense variant in *LRIF1* associated with facioscapulohumeral muscular dystrophy. *Neurology* **94:** e2441–e2447. doi:10.1212/WNL.0000000000009617

Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, 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. doi:10.1038/ng.3844

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci* **108:** 18318–18323. doi:10.1073/pnas.1110731108

Hitz BC, Jin-Wook L, Jolanki O, Kagda MS, Graham K, Sud P, Gabdank I, Strattan JS, Sloan CA, Dreszer T, et al. 2023. The ENCODE uniform analysis pipelines. bioRxiv doi:10.1101/2023.04.04.535623

Hu Y, Wang Y, He Y, Ye M, Yuan J, Ren C, Wang X, Wang S, Guo Y, Cao Q, et al. 2024. Maternal KLF17 controls zygotic genome activation by acting as a messenger for RNA Pol II recruitment in mouse embryos. *Dev Cell* **59:** 613–626.e6. doi:10.1016/j.devcel.2024.01.013

Jagannathan S, Shadle SC, Resnick R, Snider L, Tawil RN, van der Maarel SM, Bradley RK, Tapscott SJ. 2016. Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet* **25:** 4419–4431. doi:10.1093/hmg/ddw271

Kaczynski J, Cook T, Urrutia R. 2003. Sp1- and Krüppel-like transcription factors. *Genome Biol* **4:** 206. doi:10.1186/gb-2003-4-2-206

Kajava AV, Steven AC. 2006. β-Rolls, β-helices, and other β-solenoid proteins. *Adv Protein Chem* **73:** 55–96. doi:10.1016/S0065-3233(06)73003-0

Kong X, Nguyen NV, Li Y, Sakr JS, Williams K, Sharifi S, Chau J, Bayrakci A, Mizuno S, Takahashi S, et al. 2024. Engineered FSHD mutations results in D4Z4 heterochromatin disruption and feedforward DUX4 network activation. *iScience* **27:** 109357. doi:10.1016/j.isci.2024.109357

Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Mattéotti C, Arias C, Corona ED, Nuñez NG, Leo O, et al. 2007. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* **17:** 611–623. doi:10.1016/j.nmd.2007.04.002

Lea RA, McCarthy A, Boeing S, Fallesen T, Elder K, Snell P, Christie L, Adkins S, Shaikly V, Taranissi M, et al. 2021. KLF17 promotes human naive pluripotency but is not required for its establishment. *Development* **148:** dev199378. doi:10.1242/dev.199378

Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, et al.

2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298:** 799–804. doi:10.1126/science.1075090

Lemmers RJLF, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, Snider L, Straasheijm KR, van Ommen GJ, Padberg GW, et al. 2010. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329:** 1650–1653. doi:10.1126/science.1189044

Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, Amell AM, van der Vliet PJ, Almomani R, Straasheijm KR, 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. doi:10.1038/ng.2454

Li S, Li Y, Tan B, An Z. 2021. Effect of *KLF17* overexpression on epithelial-mesenchymal transition of gastric cancer cells. *J Int Med Res* **49:** 3000605211051581. doi:10.1177/03000605211051581

Li Z, Xu H, Li J, Xu X, Wang J, Wu D, Zhang J, Liu J, Xue Z, Zhan G, et al. 2023. Selective binding of retrotransposons by ZFP352 facilitates the timely dissolution of totipotency network. *Nat Commun* **14:** 3646. doi:10.1038/s41467-023-39344-1

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Liao Y, Smyth GK, Shi W. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47:** e47. doi:10.1093/nar/gkz114

Liao WW, Asri M, Ebler J, Doerr D, Haukness M, Hickey G, Lu S, Lucas JK, Monlong J, Abel HJ, et al. 2023. A draft human pangenome reference. *Nature* **617:** 312–324. doi:10.1038/s41586-023-05896-x

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. doi:10.1186/s13059-014-0550-8

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487:** 57–63. doi:10.1038/nature11244

Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27:** 1696–1697. doi:10.1093/bioinformatics/btr189

MacQuarrie KL, Fong AP, Morse RH, Tapscott SJ. 2011. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet* **27:** 141–148. doi:10.1016/j.tig.2011.01.001

Mangan S, Alon U. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci* **100:** 11980–11985. doi:10.1073/pnas.2133841100

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17:** 10–12. doi:10.14806/ej.17.1.200

Mazid MA, Ward C, Luo Z, Liu C, Li Y, Lai Y, Wu L, Li J, Jia W, Jiang Y, et al. 2022. Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature* **605:** 315–324. doi:10.1038/s41586-022-04625-0

Mesdaghi S, Price RM, Madine J, Rigden DJ. 2023. Deep learning-based structure modelling illuminates structure and function in uncharted regions of β-solenoid fold space. *J Struct Biol* **215:** 108010. doi:10.1016/j.jsb.2023.108010

Morgan M, Anders S, Lawrence M, Aboyoun P, Pagès H, Gentleman R. 2009. ShortRead: a Bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* **25:** 2607–2608. doi:10.1093/bioinformatics/btp450

Mwalilino L, Yamane M, Ishiguro KI, Usuki S, Endoh M, Niwa H. 2023. The role of *Zfp352* in the regulation of transient expression of 2-cell specific genes in mouse embryonic stem cells. *Genes Cells* **28:** 831–844. doi:10.1111/gtc.13070

Pchelintsev NA, Adams PD, Nelson DM. 2016. Critical parameters for efficient sonication and improved chromatin immunoprecipitation of high molecular weight proteins. *PLoS One* **11:** e0148023. doi:10.1371/journal.pone.0148023

Pei J, Grishin NV. 2013. A new family of predicted Krüppel-like factor genes and pseudogenes in placental mammals. *PLoS One* **8:** e81109. doi:10.1371/journal.pone.0081109

Penn BH, Bergstrom DA, Dilworth FJ, Bengal E, Tapscott SJ. 2004. A MyoD-generated feed-forward circuit temporally patterns gene expression during skeletal muscle differentiation. *Genes Dev* **18:** 2348–2353. doi:10.1101/gad.1234304

Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Res* **9:** I304. doi:10.12688/f1000research.23297.1

Pineda JMB, Bradley RK. 2024. DUX4 is a common driver of immune evasion and immunotherapy failure in metastatic cancers. *eLife* **12:** RP89017. doi:10.7554/eLife.89017

R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ren J, Gu B, Chaisson MJP. 2023. vamos: variable-number tandem repeats annotation using efficient motif sets. *Genome Biol* **24:** 175. doi:10.1186/s13059-023-03010-y

Resnick R, Wong C-J, Hamm DC, Bennett SR, Skene PJ, Hake SB, Henikoff S, van der Maarel SM, Tapscott SJ. 2019. DUX4-induced histone variants H3.X and H3.Y mark DUX4 target genes for expression. *Cell Rep* **29:** 1812–1820.e5. doi:10.1016/j.celrep.2019.10.025

Ricci E, Galluzzi G, Deidda G, Cacurri S, Colantoni L, Merico B, Piazzo N, Servidei S, Vigneti E, Pasceri V, et al. 1999. Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. *Ann Neurol* **45:** 751–757.<751::AID-ANA9>3.0.CO;2-M

Sadelain M, Papapetrou EP, Bushman FD. 2011. Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer* **12:** 51–58. doi:10.1038/nrc3179

Schuh R, Aicher W, Gaul U, Côte S, Preiss A, Maier D, Seifert E, Nauber U, Schröder C, Kemler R, et al. 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Krüppel, a *Drosophila* segmentation gene. *Cell* **47:** 1025–1032. doi:10.1016/0092-8674(86)90817-2

Sen R, Baltimore D. 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* **46:** 705–716. doi:10.1016/0092-8674(86)90346-6

Smith AA, Nip Y, Bennett SR, Hamm DC, Lemmers R, van der Vliet PJ, Setty M, van der Maarel SM, Tapscott SJ. 2023. DUX4 expression in cancer induces a metastable early embryonic totipotent program. *Cell Rep* **42:** 113114. doi:10.1016/j.celrep.2023.113114

Snider L, Geng LN, Lemmers RJLF, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, et al. 2010. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet* **6:** e1001181. doi:10.1371/journal.pgen.1001181

Swamynathan SK. 2010. Krüppel-like factors: three fingers in control. *Hum Genomics* **4:** 263–270. doi:10.1186/1479-7364-4-4-263

Tan J, Franke V, Neugebauer E, Pennisi S, Schächtle MA, Lagisquet J, Kuderna AK, Walter S, Welker I, Ensser A, et al. 2025. DNA

virus infections shape transposable element induction in vitro and in vivo. bioRxiv doi:10.1101/2024.04.18.589901

Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F, Nichols J, Reik W. 2022. 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell* **29:** 449–459.e6. doi:10.1016/j.stem.2022.01.014

Tawil R, van der Maarel SM, Tapscott SJ. 2014. Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet Muscle* **4:** 12. doi:10.1186/2044-5040-4-12

Tjeldnes H, Labun K, Torres Cleuren Y, Chyżyńska K, Świrski M, Valen E. 2021. ORFik: a comprehensive R toolkit for the analysis of translation. *BMC Bioinformatics* **22:** 336. doi:10.1186/s12859-021-04254-w

van den Boogaard ML, Lemmers RJLF, Balog J, Wohlgemuth M, Auranen M, Mitsuhashi S, van der Vliet PJ, Straasheijm KR, van den Akker RFP, 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. doi:10.1016/j.ajhg.2016.03.013

van den Heuvel A, Lassche S, Mul K, Greco A, San León Granado D, Heerschap A, Küsters B, Tapscott SJ, Voermans NC, van Engelen BGM, et al. 2022. Facioscapulohumeral dystrophy transcriptome signatures correlate with different stages of disease and are marked by different MRI biomarkers. *Sci Rep* **12:** 1426. doi:10.1038/s41598-022-04817-8

van Deutekom JC, Wijmenga C, van Tienhoven EA, Gruter AM, Hewitt JE, Padberg GW, van Ommen GJ, Hofker MH, Frants RR. 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. doi:10.1093/hmg/2.12.2037

van Overveld PG, Lemmers RJ, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, Padberg GW, van Ommen GJ, Frants RR, van der Maarel SM. 2003. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* **35:** 315–317. doi:10.1038/ng1262

van Overveld PG, Enthoven L, Ricci E, Rossi M, Felicetti L, Jeanpierre M, Winokur ST, Frants RR, Padberg GW, van der Maarel SM. 2005. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann Neurol* **58:** 569–576. doi:10.1002/ana.20625

van Vliet J, Crofts LA, Quinlan KG, Czolij R, Perkins AC, Crossley M. 2006. Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics* **87:** 474–482. doi:10.1016/j.ygeno.2005.12.011

Wang LH, Friedman SD, Shaw D, Snider L, Wong CJ, Budech CB, Poliachik SL, Gove NE, Lewis LM, Campbell AE, et al. 2019. MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. *Hum Mol Genet* **28:** 476–486. doi:10.1093/hmg/ddy364

Wang LH, Shaw DWW, Faino A, Budech CB, Lewis LM, Statland J, Eichinger K, Tapscott SJ, Tawil RN, Friedman SD. 2021. Longitudinal study of MRI and functional outcome measures in facioscapulohumeral muscular dystrophy. *BMC Musculoskelet Disord* **22:** 262. doi:10.1186/s12891-021-04134-7

Whiddon JL, Langford AT, Wong C-J, Zhong JW, Tapscott SJ. 2017. Conservation and innovation in the DUX4-family gene network. *Nat Genet* **49:** 935–940. doi:10.1038/ng.3846

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, et al. 2019. Welcome to the tidyverse. *J Open Source Softw* **4:** 1686. doi:10.21105/joss.01686

Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, Gruter AM, Hofker MH, Moerer P, Williamson R. 1992. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* **2:** 26–30. doi:10.1038/ng0992-26

Wong CJ, Wang LH, Friedman SD, Shaw D, Campbell AE, Budech CB, Lewis LM, Lemmers R, Statland JM, van der Maarel SM, et al. 2020. Longitudinal measures of RNA expression and disease activity in FSHD muscle biopsies. *Hum Mol Genet* **29:** 1030–1043. doi:10.1093/hmg/ddaa031

Wong CJ, Friedman SD, Snider L, Bennett SR, Jones TI, Jones PL, Shaw DWW, Blemker SS, Riem L, DuCharme O, et al. 2024. Regional and bilateral MRI and gene signatures in facioscapulohumeral dystrophy: implications for clinical trial design and mechanisms of disease progression. *Hum Mol Genet* **33:** 698–708. doi:10.1093/hmg/ddae007

Yan W, Burns KH, Ma L, Matzuk MM. 2002. Identification of Zfp393, a germ cell-specific gene encoding a novel zinc finger protein. *Mech Dev* **118:** 233–239. doi:10.1016/S0925-4773(02)00258-7

Yan L, Yang M, Guo H, Yang L, Wu J, Li R, Liu P, Lian Y, Zheng X, Yan J, et al. 2013. Single-cell RNA-seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol* **20:** 1131–1139. doi:10.1038/nsmb.2660

Yao Z, Snider L, Balog J, Lemmers RJLF, Van Der Maarel SM, Tawil R, Tapscott SJ. 2014. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet* **23:** 5342–5352. doi:10.1093/hmg/ddu251

Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, Balog J, Tawil R, van der Maarel SM, Tapscott SJ. 2013. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet* **9:** e1003947. doi:10.1371/journal.pgen.1003947

Zeng W, de Greef JC, Chen Y-Y, Chien R, Kong X, Gregson HC, Winokur ST, Pyle A, Robertson KD, Schmiesing JA, et al. 2009. Specific loss of histone H3 lysine 9 trimethylation and HP1γ/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet* **5:** e1000559. doi:10.1371/journal.pgen.1000559

Zhang Q, Lenardo MJ, Baltimore D. 2017. 30 years of NF-κB: a blossoming of relevance to human pathobiology. *Cell* **168:** 37–57. doi:10.1016/j.cell.2016.12.012

Zheng D, van den Heuvel A, Balog J, Willemsen IM, Kloet S, Tapscott SJ, Mahfouz A, van der Maarel SM. 2025. DUX4 activates common and context-specific intergenic transcripts and isoforms. *Sci Adv* **11:** eadt5356. doi:10.1126/sciadv.adt5356

Zhou S, Tang X, Tang F. 2016. Krüppel-like factor 17, a novel tumor suppressor: its low expression is involved in cancer metastasis. *Tumour Biol* **37:** 1505–1513. doi:10.1007/s13277-015-4588-3